Ecology dictates the value of memory for foraging bees

Christopher D. Pull[1,3], Irina Petkova[1], Cecylia Watrobska[1], Grégoire Pasquier[1], Marta Perez Fernandez[2] and Ellouise Leadbeater[1]

[1]Department of Biological Sciences, Royal Holloway, University of London, Egham, Surrey, TW200EX, UK

[2]Department of Geography, Royal Holloway, University of London, Egham, Surrey, TW200EX, UK

[3]Current address: Department of Zoology, University of Oxford, Oxford, OX1 3SZ, UK

Summary

"Ecological intelligence" hypotheses posit that the benefits of cognitive investment vary with foraging ecology, and provide a key framework for understanding the evolution of animal learning and memory[1–4]. However, although certain ecological selection pressures have been found to correlate with brain or neural region size[5–8], empirical evidence to show that any specific cognitive trait is useful in certain environments but not others is currently lacking. Here, we assay the short-term memory of bumblebee (*Bombus terrestris audax*) workers from 25 identically reared colonies, before allowing each colony to forage in a landscape where forage availability varies seasonally. Through analysis of the bees' lifetime foraging careers, comprising >1700 foraging trips over two years, we show that performance on a task designed to test short-term memory predicts individual foraging efficiency – a fitness proxy that is key to colony reproductive output – in plentiful spring foraging conditions. However, this relationship is reversed during the summer floral dearth, when the costs of cognitive investment may outweigh the benefits. Our results provide evidence that the value of a cognitive trait depends upon the prevailing ecological conditions and suggest that temporal changes in that environment could place contrasting selection pressures on memory within a single species.

The evolution of cognitive traits is thought to be constrained by significant constitutive and induced costs[9–11], such that net benefits will only emerge where these traits can markedly enhance performance in fitness-critical contexts[12]. However, ecological selection pressures are unlikely to be uniform either within species nor even individual lifetimes, with consequences for cognitive evolutionary trajectories [13,14]. For example, if a particular cognitive trait enhances fitness only when food is plentiful, the costs of maintaining it during harsher periods may constrain its evolution[15]. Studying within-species variation in the relationship between cognitive traits and fitness-determining attributes could, therefore, identify potential ecological factors that both drive and constrain the evolution of cognitive traits. Yet, direct evidence linking within-species variation in learning and memory to fitness proxies is still rare[16–19] due to the considerable challenges of standardizing confounding noncognitive factors (e.g. previous experience, parasite load or motivation) that can influence both cognitive assay

performance and fitness, when working with wild animals[20–24]. As a result, it has not yet been possible to link specific ecological drivers to variation in the relationship between cognitive performance and fitness proxies.

Here, we capitalize upon the unique features of a social insect system to examine how the benefits afforded by a specific cognitive trait may vary with ecological circumstances. *Bombus* are a temperate group, and in most species, workers begin to emerge in the early spring and colony foraging continues until late summer[25]. The availability of floral forage varies considerably across this time, typically reaching a peak in the spring that recedes to a trough in late summer, with local variations[26,27]. Foraging bees can visit hundreds of individual flowers across repeated foraging bouts every day and cognitive traits such as learning and memory are thought to be fundamental in maximising colony foraging success[28]. In laboratory settings, abundant evidence has documented the role of medium and long-term memory formation in the development of preferences for particular flower types or patches[29]. Other work has shown that short-term memory (STM), a physiologically distinct memory phase that requires neither transcription nor translation and forms after a single trial, contributes to within-patch decisions[30,31]. However, attempts to link memory to real-world foraging efficiency have focussed on short summer foraging windows[19,32], and so do not capture any variation in resource availability. Here, we investigate the importance of STM for bumblebee foraging success across a highly variable foraging season.

We reared 25 young, commercially supplied colonies (mean initial number of workers on delivery ± SD = 39 ± 16.8) under identical laboratory conditions following a staggered design from April to September over two successive years (Fig. 1a). For each colony, a mean of nine (range 7-13) worker bees of known age underwent cognitive testing in a four-arm Radial Arm Maze (RAM; Fig. 1b). Originally developed for rodent toxicology studies, the RAM is a win-shift paradigm where performance is assayed as the number of revisits made to arms previously depleted of food within a foraging bout, and has been used to effectively test STM in bees[31] (for additional validation see Supplementary Fig. 1). In each case, bees were trained to asymptotic performance prior to testing[31] (Extended Data Figure 1), such that cognitive assays required two weeks per colony. After cognitive testing was complete, tested workers were tagged with an RFID chip and screened for gut parasites – which may compromise cognition[33] – but no infections were found. Each colony was then provided through-the-wall access to the external environment (Fig. 1c), comprising broadleaved mixed woodland and parkland surrounded by suburban housing and gardens (Fig. 1d). Of the 230 bees tested, 144 (63%) foraged in the wild; we recorded nectar and pollen foraging efficiency by individually weighing bees on entry and exit, until death, sampling ~ 47% of their total foraging career (n = 1735 nectar and pollen foraging trips).

We found that for nectar-collecting trips, the relationship between RAM score and foraging efficiency reversed in direction across the foraging season (Fig. 2 and Supplementary Table 1; linear mixed effects model [LMER] containing RAM score x week interaction: ΔAIC to next best model = 4.7; interaction estimate ± 95% confidence interval = 0.05 ± 0.02 to 0.09). Specifically, bees with better RAM scores collected more nectar in spring but less nectar in summer, compared to bees with poorer RAM scores. Bees exhibited an increase in nectar foraging efficiency at the start, and a decrease towards the end, of their foraging careers, presumably as they gained experience and then underwent senescence, in accordance with previous results[25] (Extended Data Fig. 2a; estimate ± 95% CI = –0.03 ± –0.04 to –0.02). Larger bees were also more efficient foragers (Extended Data Fig. 3b; estimate ± 95% CI = 0.34 ± 0.15 to 0.52) and all bees foraged more efficiently on cooler, more humid days (Extended Data Fig. 2c; estimate ± 95% CI = 0.01 ± 0.01 to 0.02). Neither year of the experiment (Extended Data Fig. 2d; estimate ± 95% CI = –0.28 ± –0.58 to 0.04) nor bee age at initial release (Extended Data Fig. 2e; estimate ± 95% CI = –0.03 ± -0.06 to 0.001) had significant impacts. Pollen foraging efficiency is not expected to vary as drastically with STM because flowers typically contain more pollen than can be extracted in a single visit and pollen collection is somewhat passive[34,35]. Avoidance of visited flowers within patches is thus less relevant and, accordingly, STM did not predict pollen foraging efficiency (Supplementary Table 2; generalised linear mixed effects model [GLMER]: ΔAIC between intercept-only null model and next best alternative = 2.66).

Since our bees were RFID-tagged, we could also relate RAM performance to survival and total lifetime foraging effort (total number of trips), again across the whole season. However, both variables were solely predicted by age, with bees that were older on release dying sooner (cox proportional hazard model: estimate ± 95% CI = 0.07 ± 0.01 to 0.12) and conducting fewer trips (GLMER: estimate ± 95% CI = –0.35 ± –0.55 to –0.15). Adding RAM score to the models did not sufficiently improve fit in either case (Supplementary Tables 3 and 4). Additionally, comparisons of tested and non-tested control bees revealed that cognition testing itself had no measurable impact on either the foraging performance or survival of bees (Supplementary Analysis and Results; Supplementary Tables 5-7).

Our results show that the value of STM use for nectar collection changes over the course of the foraging season. It is unlikely that this finding is the result of differences arising from confounding variables typically thought to influence cognition studies[20–24]: all colonies were raised under identical conditions, bees were free from gut parasites during cognitive testing, there were no differences in the prior experience or knowledge of bees, bees were of known age and size, and hunger- or condition-driven motivational effects should be minimal due to the fact the bees forage to meet the demands of the colony. Instead, we believe that our findings are due to shifting patterns of floral resources, which are typically abundant in spring but limited in summer[26], including at suburban sites in London similar

to our own[36]. To track floral resource phenology, we combined land classification techniques with transect and quadrat sampling to measure floral generic diversity within a 500 m radius around our bee colonies, spanning mixed broadleaved woodland and parkland, local public green spaces and residential gardens. Although bumblebees can forage further, average foraging distances are typically shorter; for example, one study found that 40% of *B. terrestris* foraging occurred within 100 m of the nest[37]. Weekly surveys of floral resources revealed an overall decline in the number of genera in flower towards the end of the season for all land types except grassland, which did not change (Fig. 3a; all statistics given in Supplementary Tables 8 and 9). Floral resources in wooded areas, which is the dominant land type in our survey area (woodland = 48%, open woodland = 11%), showed the sharpest linear decline from a peak in spring, whereas others, such as private gardens (26%), increased towards a peak in early summer before declining by the end of summer. This pattern is also mirrored in the gross foraging trends found for all bees across the season: nectar and pollen foraging efficiency peaked in spring and was at its lowest at the end of summer, albeit with a brief revival in early autumn (Fig. 3b-c). Bout duration (Fig. 3d), which correlates with resource availability in bumblebees[38], followed the same pattern. Finally, identification of pollen loads collected by bees in the second year of our experiment showed a gradual decline in floral generic diversity from a spring peak of fifteen genera to a late summer trough of one genus (Extended Data Fig. 3).

The benefits afforded by cognitive abilities such as learning and memory are thought to become most apparent in demanding foraging conditions, such as high-elevation winters[6], but here we found that strong RAM performance was associated with higher foraging efficiency in plentiful spring conditions rather than harsh summers. Our findings are consistent with the nature of the within-patch foraging task that the RAM seeks to mirror, if individuals that can remember and avoid visited flowers move through patches more efficiently. In spring, when resource patches are plentiful near to the colony, such within-patch efficiency savings may have greater impact than in summer, when reaching and moving between floral patches requires long travel times that thus dwarf any within-patch savings, and long-term memory of patch locations may be more important. Accordingly, previous work in the same species has shown that performance in a task requiring longer-term memory predicted foraging efficiency at the height of summer (but see[32]). The RAM requires formation of memories based on a single trial; in *Drosophila*, the ability to form single-trial memories (or multiple trials with very short inter-trial intervals; termed anaesthesia-resistant memory), trades off against long-term memory[39], but this possibility has not been explored in bees. Alternatively, although we found no significant effect of STM on individual survival, reduced foraging efficiencies in individuals with good STM in the summer dearth could reflect sublethal costs of cognitive investment that emerge when flight times are longer, manifested through flight efficiency or the need to consume more of the nectar they collect whilst foraging[12]. Future research should investigate the link between

such metabolic costs and investment in memory, because sublethal stressors can critically compromise colony reproductive success[40].

Although it has been repeatedly proposed that variation in the benefits afforded by cognitive traits may vary with environmental conditions[8,13,41], even within individual lifetimes, direct demonstrations of such fluctuations in selective value have not been performed. Here, we provide evidence that the benefit of a cognitive trait can indeed vary considerably, even within a relatively short single foraging season, becoming apparent in plentiful, species-rich floral environments but reversing in the summer dearth. In animals that will experience variability within a single lifetime, such patterns might be expected to compromise selection for certain cognitive traits; in bumblebees, the unit of selection is the colony, and future studies could evaluate the possibility that bee colonies may benefit from worker phenotypic diversity in cognitive abilities, which may even vary with the stage in the foraging season at which an individual worker emerges. Consequently, our study finds strong support for the ecological intelligence framework[2–4] and points to the major role that ecological differences can potentially play in generating and maintaining the intraspecific cognitive variation found across the animal kingdom[20–23].

Acknowledgments

We are grateful to Rosie Wright, Matthew Rawlings, Monika Yordanova, Harry Siviter, Matthew Hasenjager and Fabio Manfredini for help with data collection, the Egham and Englefield Green residents who provided access to their gardens for floral surveys, and Imperial College London, Silwood Park for providing access to weather station data. This research was supported by a Leverhulme Trust Research Project Grant to E.L. (RPG-2016-444) and C.W. was funded through the IC-RHUL BBSRC DTP (grant no: BB/M011178/1).

Author contributions

E.L. conceived the experiment and C.P. contributed to the design of the study. C.P. and I.P. ran the main experiment, G.P. ran and analysed RAM validation and C.W. and M.P.E analysed pollen samples. C.P. performed the statistical analysis with input from E.L, and C.P. and E.L. wrote the manuscript with all authors providing final comments.

Competing interests

The authors declare no competing interests.

Materials & Correspondence

Should be addressed to C.P.

Methods

Overview

From April to October in 2018 and 2019, following a staggered design (Fig. 1a), we performed laboratory-based cognitive testing on 230 worker bees from 25 commercially sourced colonies (Fig. 1b; n = 12 in 2018 and 13 in 2019; one colony removed in April 2018 due to colony death pre-release). After a two-week testing period, tested bees were screened for gut parasites and RFID-tagged, and colonies were connected to a through-the-wall external access hatch (Fig. 1c). Nectar and pollen foraging efficiency was monitored for two weeks (approx. 6h per day, 4-5 days per week). At the same time, testing began for the next colony in the cycle (Fig. 1a). After foraging efficiency recording had ceased, colonies were permitted to continue foraging for another two weeks, during which time survival and foraging activity (but not foraging efficiency) was recorded through automated RFID readers. Throughout, we conducted weekly surveys of the land surrounding the foraging colonies, to record changes in the abundance and diversity of floral resources over the seasons (Fig. 1d).

Bee colonies

We used commercial colonies of UK-native *Bombus terrestris audax* (BioBest, Waterloo, Belgium), housed in two-chamber plastic nestboxes (28 [l] x 16 [w] x10.5 [h] cm). Commercially supplied colonies are reared from domestic *Bombus* lines and queens had emerged as gynes, hibernated, and founded colonies in identical controlled indoor conditions. All colonies were young on arrival (mean initial number of workers ± SD = 39 ± 16.8), and only bees that emerged post-arrival (identified through tagging all bees with numbered discs on arrival and new bees after emergence [Abelo, UK]) were tested in our cognitive assay. Prior to cognitive testing, colonies were fed an *ad libitum* supply of inverted sugar syrup via an in-nest feeder (45% [w/v]; Thorne, Windsor, U.K.), with one 1.5g pollen ball (2:1 honeybee-collected pollen: 45% inverted sugar) added daily (two on Fridays; none on weekends). During cognitive testing, pollen feeding continued, but syrup feeders were removed and between 10-15 ml (depending on colony size and stores) was pipetted into nectar pots per weekday evening. Colonies were not fed once given outside access.

Radial arm maze

We used a radial arm maze (Fig. 1b) to assess the STM of individual bees. The RAM is a win-shift paradigm in which all arms are initially baited with food rewards that are not replaced upon removal, within one bout. Revisits to depleted arms constitute “errors” and the number of errors within a single foraging bout is the measure of STM[42]. The RAM is therefore an ecologically relevant task that mimics natural nectar foraging within a flower patch. Our RAM consisted of an octagonal four-arm array in which differentiation of the arms was possible through a laminated black and white panoramic image of the laboratory at the ends of each arm. Rewards were accessed via removable

blue, rectangular Perspex platforms – henceforth "flowers" – at the end of each arm (colour number = 744; 7.5 [l] x 3 [w] x 0.5 [h] cm) that slotted through holes in the maze wall (Fig. 1b). Bees retrieved a sucrose reward (see "Cognitive testing") by alighting on flowers and inserting their proboscises through small holes in the RAM wall (Fig. 1b inset).

Cognitive testing

Cognitive testing began 12 days after arrival in the laboratory (to allow bees of known age to emerge) and commenced with a group training period (~1h) whereby bees were allowed to forage freely in the RAM and all arms were continually rewarding (an *ad libitum* supply of 2M sucrose solution), and to enter and leave the arena at will. Motivated foragers were then tested in the RAM alone for 12 foraging bouts, during which all arms were baited with 20 µl of 2M sucrose solution that was not replaced within a bout (except the last non-depleted flower when bees were fed to repletion). After each landing, platforms were replaced with identical clean, unrewarded replacements to preclude the use of scent marks. Access to the nest during testing was prevented via sliding shutters, unless a bee spent >30 s trying to return or attempted to return more than twice within a bout (to maintain foraging motivation). Consequently, in some bouts not all arms were visited. All bouts were filmed for later video analysis.

Parasite screening and RFID chips

We screened faecal samples from all tested bees for the presence of gut parasites (*Apicystis bombi*, *Nosema bombi* and *Crithidia bombi*) at the end of the two weeks of cognition testing (Nikon e50i) and measured intra-tegula distance as a proxy for overall size, as in other studies (e.g.[19,31]). An RFID chip (Microsensys GmbH) encoding a unique 16-digit identification number was also superglued (Loctite) to the thorax. We performed the same procedure for a cohort of non-tested bees from each colony that were observed foraging during group training (mean of four bees per colony, total n = 99), in order to confirm that our testing regime had no effect on subsequent foraging efficiency.

Measuring foraging performance

We measured the foraging success of individual bees on exiting and re-entering the colony using weight-averaging scales for moving subjects (mean of three repeat measurements with 2s averaging and accuracy of ± 2 mg; Advanced portable balance Scout STX123 120g; OHAUS Corporation) and their lifetime foraging activity and survival using an RFID system (MicroSensys GmBH, Fig. 1c). Per trip nectar values were calculated by subtracting the bee's outbound weight from her inbound weight, minus the weight of any pollen. We removed pollen from one leg of the bee – via a trap door in the tube – and weighed and froze (-20) the pollen for later identification. The weight of the pollen was doubled and subtracted from the bees' weight. On entry and exit, bees passed through two RFID readers (Microsensys GmbH) that recorded identity and travel direction. We collected foraging

performance data for ~ 6 hours per day, four-five days a week, for two weeks per colony. In total, 6616 were trips recorded across all colonies and bees. For analysis, nectar trips (n = 1202 trips by 134 bees) were defined as those where < 3 mg of pollen was collected, and pollen trips > 3 mg (n = 526 trips by 91 bees), based on histograms of foraging data.

Colonies then foraged for a further two weeks with only RFID data collection to assay survival, in which time >99% of RFID-chipped bees eventually failed to return to the colony and were presumed dead. We affixed brightly coloured plastic cones onto the outdoor entrance of nests and replaced pollen that we removed daily for microscopic pollen identification using a local reference collection with equal quantities of honeybee collected pollen (Koppert, UK). Colonies were euthanized at the end of four weeks of foraging.

Floral resource surveys and weather data

We used QGIS to (*i*) classify the 500 m radius surrounding our colonies into broad land use types likely to contain forage: woodland, grassy woodland, grassland and landscaped and (*ii*) select sampling sites within these categories. Additionally, we arranged access to private suburban gardens (total sites per land use type = 12, except grass = 24 and gardens = 11). Surveys were performed for one day each week, whereby each site was visited once every three weeks on a rotational basis. Methods were customised to each land-use type: in grass and landscaped areas, we sampled 0.25m^2 quadrats; in woodland and open woodland we surveyed 30 m transects; in gardens we counted every genus of plant in flower. In year one, we utilised a different approach to record transect and garden data that was discarded after the 4th survey; hence we only include transect and garden data from year two in survey blocks 1-4 of Fig. 3a. A three-week average was calculated using generic richness data from all three sets of sites for the same survey period in both years for graphical representation in Fig. 3A. Local meteorological data was collected from a weather station located ~ 8 km away at Imperial College London, Silwood Park. Hourly weather data (temperature, humidity and wind speed) was averaged to produce a daily mean for analysis.

Statistical analysis

All statistical analyses were conducted using R version 4.1.0 and an information-theoretic approach. For foraging and survival models, we built a candidate model that contained every hypothesised covariate (see below and Supplementary Tables 1-10 for model descriptions) and compared this to (*i*) the same model, but with STM score as a predictor, (*ii*) the same model but including an interaction between STM score and week (*iii*) a null model containing only the intercept and random factors. Model selection was based on ΔAIC or ΔAICc (depending on sample size), where a cut-off of >2 was used to identify the best model. If nested models had comparable fit the simplest model was selected.

We modelled nectar foraging efficiency using a linear mixed effects regression (LMER), and pollen foraging efficiency with a general linear mixed effect regression (GLMER), which had log-link Gamma errors. Both included foraging efficiency as the response (mg/min; ordered quantile normalisation transformation using the BestNormalize package due to heteroscedasticity) and week-of-year, STM score (log(n+1) transformed to reduce influence of outliers) and their interaction as our predictors of interest, with initial bee age (at time of release), bee size, year of experiment, foraging experience (number of days since foraging began), and a composite "weather" score as covariates. The composite weather score was produced via principal component analysis to reduce temperature, humidity, and wind speed into a single component (~ 84% variation captured). To account for collinearity between weather scores and week, weather scores in the main analysis are the residuals of general additive model (GAM) predicting weather based on week; thus, any reported effects of "weather" reflect those that occurred in addition to effects of week. We included foraging experience as a quadratic polynomial to account for its non-linear effect, which improved nectar model fit. For each bee, we fitted random intercepts that interacted with random slopes for experience (because improvement in performance could vary between bees and is likely affected by starting performance). To model pollen foraging efficiency, we used an identical procedure except that foraging experience was not fit as a polynomial and uncorrelated random slopes and intercepts were included, to improve fit.

To analyse survival data, we used a Cox proportional hazards model with week of year, STM score and their interaction as our predictors of interest, and bee age (at time of release), bee size, composite weather and year of experiment as covariates, including colony as a random effect using the shared frailty function. We used a GLMER with Poisson error distribution to model lifetime foraging effort, including number of RFID-recorded trips as the response and an observation-level random effect to account for overdispersion[43]; all other parameters were identical to the nectar foraging model and colony was included as a random intercept.

To analyse generic floral richness, we built linear regressions that included weekly generic richness as the response, season as a linear predictor or quadratic polynomial (depending on fit), and year of experiment as a covariate. For gardens and woodland data, we applied a log(n) and ORQ transformation respectively to the response to achieve normality in model residuals.

For mixed effects modelling, we used the 'lme4' package and checked all model assumptions via residual plots and functions in 'influence.ME' and 'performance' packages. For the survival model, we plotted residuals, tested for non-proportional hazards and assessed the model for influential observations.

Data Availability

Data will be deposited in Dryad upon acceptance.

References

1. Gibson, K. R. Cognition, brain size and extraction of embedded foods. in *Primate ontogeny and social behaviour* (eds. Else, J. & Lee, P.) 93–105 (Cambridge University Press, 1986).
2. Clutton Brock, T. H. & Harvey, P. H. Primates, brains and ecology. *J. Zool.* **190**, 309–323 (1980).
3. González-Forero, M. & Gardner, A. Inference of ecological and social drivers of human brain-size evolution. *Nature* **557**, 554–557 (2018).
4. Rosati, A. G. Foraging cognition: reviving the ecological intelligence hypothesis. *Trends Cogn. Sci.* **21**, 691–702 (2017).
5. Sol, D., Duncan, R. P., Blackburn, T. M., Cassey, P. & Lefebvre, L. Big brains, enhanced cognition, and response of birds to novel environments. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 5460–5465 (2005).
6. Pravosudov, V. V. & Clayton, N. S. A test of the adaptive specialization hypothesis: population differences in caching, memory, and the hippocampus in black-capped chickadees (*Poecile atricapilla*). *Behav. Neurosci.* **116**, 515–522 (2002).
7. DeCasien, A. R., Williams, S. A. & Higham, J. P. Primate brain size is predicted by diet but not sociality. *Nat. Ecol. Evol.* **1**, 1–7 (2017).
8. Sayol, F. *et al.* Environmental variation and the evolution of large brains in birds. *Nat. Commun.* **7**, 13971 (2016).
9. Aiello, L. C. & Wheeler, P. The expensive-tissue hypothesis: the brain and the digestive system in human and primate evolution. *Curr. Anthropol.* **36**, 199–221 (1995).
10. Mery, F. & Kawecki, T. J. The effect of learning on experimental evolution of resource preference in Drosophila melanogaster. *Evolution* **58**, 757–767 (2004).
11. Mery, F. A cost of long-term memory in *Drosophila*. *Science* **308**, 1148–1148 (2005).
12. Burns, J. G., Foucaud, J. & Mery, F. Costs of memory: lessons from ‘mini’ brains. *Proc. R. Soc. B Biol. Sci.* **278**, 923–929 (2011).
13. Allman, J. M., McLaughlin, T. & Hakeem, A. Brain structures and life-span in primate species. *Proc. Natl. Acad. Sci.* **90**, 3559–3563 (1993).
14. Van Woerden, J. T., Van Schaik, C. P. & Isler, K. Effects of seasonality on brain size evolution: evidence from strepsirrhine primates. *Am. Nat.* **176**, 758–767 (2010).
15. van Woerden, J. T., Willems, E. P., van Schaik, C. P. & Isler, K. Large brains buffer energetic effects of seasonal habitats in catarrhine primates. *Evolution* **66**, 191–199 (2012).

16. Sonnenberg, B. R., Branch, C. L., Pitera, A. M., Bridge, E. & Pravosudov, V. V. Natural selection and spatial cognition in wild food-caching mountain chickadees. *Curr. Biol.* **29**, 670-676.e3 (2019).
17. Cole, E. F., Morand-Ferron, J., Hinks, A. E. & Quinn, J. L. Cognitive ability influences reproductive life history variation in the wild. *Curr. Biol.* **22**, 1808–1812 (2012).
18. Keagy, J., Savard, J.-F. & Borgia, G. Male satin bowerbird problem-solving ability predicts mating success. *Anim. Behav.* **78**, 809–817 (2009).
19. Raine, N. E. & Chittka, L. The correlation of learning speed and natural foraging success in bumble-bees. *Proc. R. Soc. B Biol. Sci.* **275**, 803–808 (2008).
20. Thornton, A., Isden, J. & Madden, J. R. Toward wild psychometrics: linking individual cognitive differences to fitness. *Behav. Ecol.* **25**, 1299–1301 (2014).
21. Rowe, C. & Healy, S. D. Measuring variation in cognition. *Behav. Ecol.* **25**, 1287–1292 (2014).
22. Morand-Ferron, J., Cole, E. F. & Quinn, J. L. Studying the evolutionary ecology of cognition in the wild: a review of practical and conceptual challenges. *Biol. Rev.* **91**, 367–389 (2016).
23. Boogert, N. J., Madden, J. R., Morand-Ferron, J. & Thornton, A. Measuring and understanding individual differences in cognition. *Philos. Trans. R. Soc. B Biol. Sci.* **373**, (2018).
24. Pritchard, D. J., Hurly, T. A., Tello-Ramos, M. C. & Healy, S. D. Why study cognition in the wild (and how to test it)? *J. Exp. Anal. Behav.* **105**, 41–55 (2016).
25. Goulson, D. *Bumblebees: their behaviour and ecology*. (Oxford University Press, UK, 2004).
26. Leong, M. & Roderick, G. K. Remote sensing captures varying temporal patterns of vegetation between human-altered and natural landscapes. *PeerJ* **3**, e1141 (2015).
27. Timberlake, T. P., Vaughan, I. P. & Memmott, J. Phenology of farmland floral resources reveals seasonal gaps in nectar availability for bumblebees. *J. Appl. Ecol.* **56**, 1585–1596 (2019).
28. Klein, S., Cabirol, A., Devaud, J.-M., Barron, A. B. & Lihoreau, M. Why bees are so vulnerable to environmental stressors. *Trends Ecol. Evol.* **32**, 268–278 (2017).
29. Chittka, L. & Thomson, J. D (eds.). *Cognitive ecology of pollination*. (Cambridge University Press, 2002). doi:10.1046/j.1439-0310.2002.00774.x.
30. Brown, M. F. & Demas, G. E. Evidence for spatial working memory in honeybees (*Apis mellifera*). *Journal of Comparative Psychology* vol. 108 344–352 (1994).
31. Samuelson, E. E. W., Chen-Wishart, Z. P., Gill, R. J. & Leadbeater, E. Effect of acute pesticide exposure on bee spatial working memory using an analogue of the radial-arm maze. *Sci. Rep.* **6**, 38957 (2016).
32. Evans, L. J., Smith, K. E. & Raine, N. E. Fast learning in free-foraging bumble bees is negatively correlated with lifetime resource collection. *Sci. Rep.* **7**, 496 (2017).
33. Gomez-moracho, T., Heeb, P. & Lihoreau, M. Effects of parasites and pathogens on bee

cognition. **42**, 2–15 (2017).

34. Haynes, J. & Mesler, M. Pollen foraging by bumblebees: foraging patterns and efficiency on *Lupinus polyphyllus*. *Oecologia* **61**, 249–253 (1984).
35. Thorp, R. W. The collection of pollen by bees. *Plant Syst. Evol.* **222**, 211–223 (2000).
36. Samuelson, A. E., Schürch, R. & Leadbeater, E. Dancing bees evaluate agricultural forage resources as inferior to central urban land. *J. Appl. Ecol.* (In Press).
37. Wolf, S. & Moritz, R. F. A. Foraging distance in *Bombus terrestris* L. (Hymenoptera: Apidae). *Apidologie* **39**, 419–427 (2008).
38. Redhead, J. W. *et al.* Effects of habitat composition and landscape structure on worker foraging distances of five bumble bee species. *Ecol. Appl.* **26**, 726–739 (2016).
39. Lagasse, F., Moreno, C., Preat, T. & Mery, F. Functional and evolutionary trade-offs co-occur between two consolidated memory phases in *Drosophila melanogaster*. *Proc. R. Soc. B Biol. Sci.* **279**, 4015–4023 (2012).
40. Bryden, J., Gill, R. J., Mitton, R. A. A., Raine, N. E. & Jansen, V. A. A. Chronic sublethal stress causes bee colony failure. *Ecol. Lett.* **16**, 1463–1469 (2013).
41. Sol, D. The cognitive-buffer hypothesis for the evolution of large brains. in *Cognitive Ecology II* (eds. Reuven, D. & Ratcliffe, J. M.) 111–134 (University of Chicago Press, 2009). doi:10.7208/chicago/9780226169378.003.0007.
42. Foreman, N. & Ermakova, I. The radial arm maze: twenty years on. in *A Handbook of Spatial Research Paradigms and Methodologies* (eds. Foreman, N. & Raphael, G.) 87–143 (1998).
43. Harrison, X. A. Using observation-level random effects to model overdispersion in count data in ecology and evolution. *PeerJ* **2**, e616 (2014).

Figures

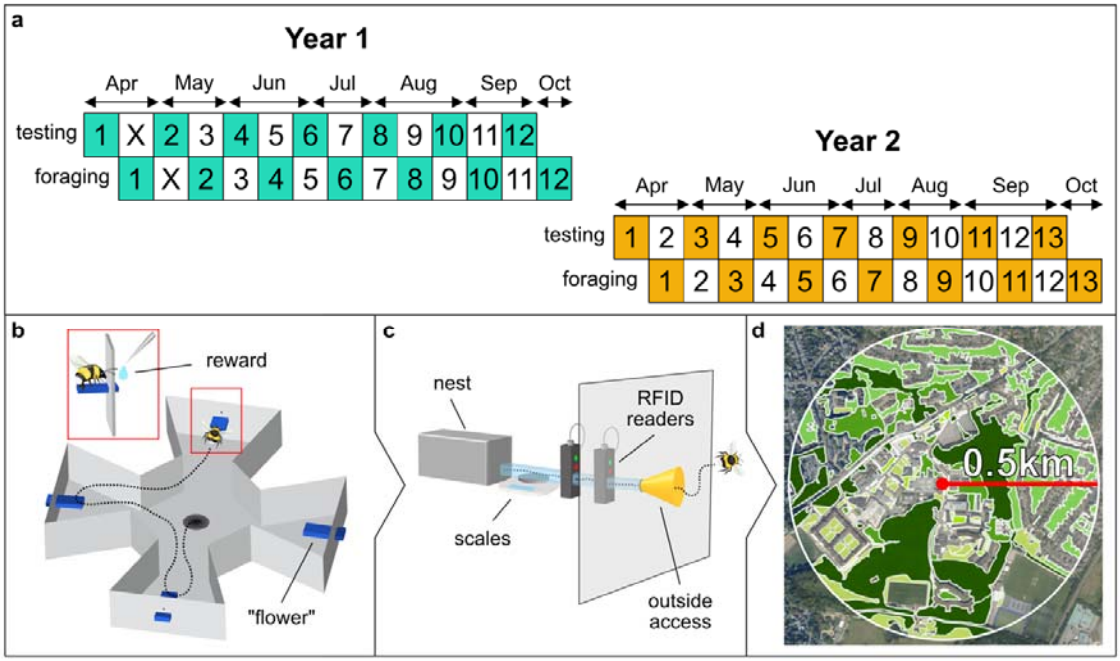

Fig 1. Lab-to-field testing of bee cognition and foraging efficiency. a, Experimental timeline showing staggered design: numbers represent different colonies (*n* = 25 colonies) that underwent cognitive testing in the laboratory ('testing') that were then provided outside access to forage ('foraging'); whilst one colony foraged the next was being tested. This design was replicated over two years and an approximate monthly timeline is displayed per year; X denotes one colony that died during testing. **b,** Four-arm radial arm maze (RAM) used to assay the short-term memory (STM) of individual bees; inset shows bee extracting a sucrose reward. **c,** Following RAM testing, colonies were provided with outside access and foraging efficiency was measured individually by weighing bees on exiting and returning to the nest, passing through RFID readers **d,** Bee colonies were surrounded by a mix of university parkland, suburban greenspace, and private gardens. The most dominant land type was mixed woodland (dark green). Non-coloured areas represent unsampled areas unlikely to contain floral resources.

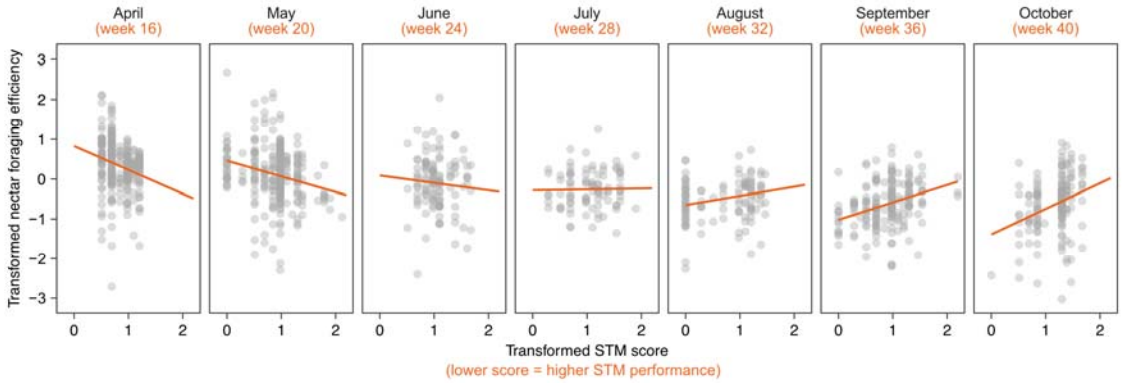

Fig 2. Seasonal reversal in bee STM and foraging efficiency relationship. Partial residual plots from a linear mixed effect model with an interaction between bee STM score and week of year (*n* = 1202 nectar foraging trips by 134 bees). Fitted lines indicate the relationship between STM and nectar foraging efficiency at four-week intervals across the entire foraging season, while holding the effect of other numeric predictors constant at their median and by setting year to "two" (most common value). Nectar foraging efficiency and STM score are presented, as analysed, on transformed scales (ORQ normalisation and log(n+1), respectively); for reference, untransformed nectar values range from -6.75–14.85 mg/min and STM scores 0–8 RAM errors. Negative nectar values result from bees leaving with more nectar than they return with[25].

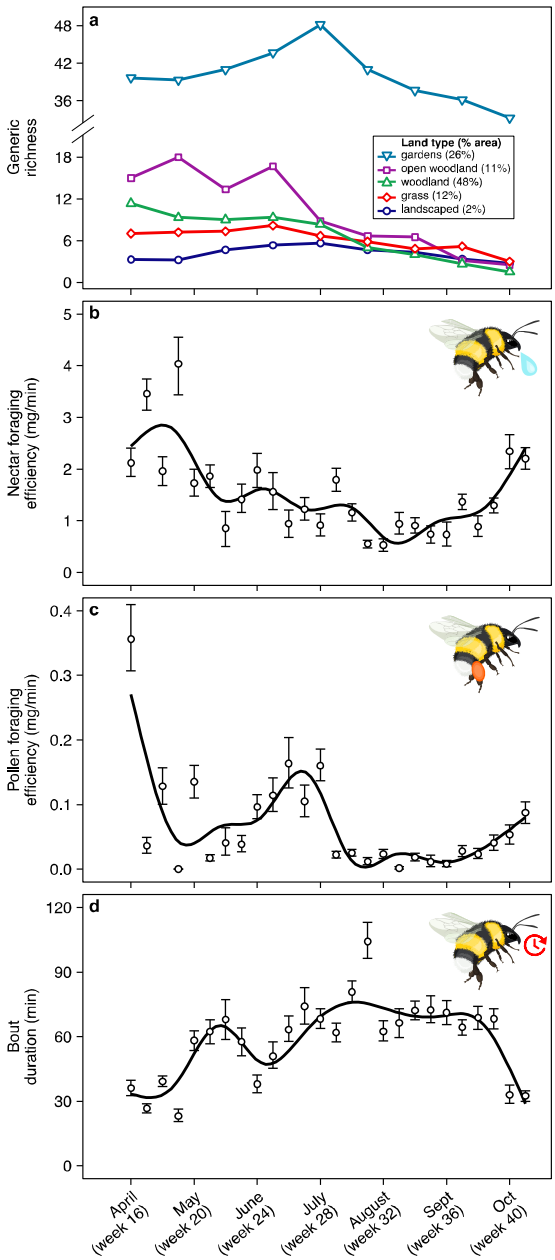

Fig 3. Seasonal variation in floral resources and bumblebee foraging. a, Shifts in floral generic richness across the foraging season, for five land-use types surveyed within a 500 m radius around foraging bee colonies. Each point represents a three-week average (or two weeks for the final survey) based on both years of the study (except surveys 1-4 for *gardens*, *open woodland* and *woodland*, which comprise only year 2 data; see methods) and the total coverage of each land type within the survey area is shown as a percentage in the legend. **b-c,** Nectar foraging efficiency, pollen foraging efficiency and bout duration of whole colonies (n = 6616 foraging trips) across both years of the foraging season. Points ± errors bars show weekly mean ± 95% CI and trend lines were fitted using a generalized additive model (GAM) smoother function.