Waggle dance distributions quantify collective foraging in honey bee colonies

# Abstract

Honeybee foraging is an extraordinary collective behaviour that is directed through the waggle dance communication system, whereby colony foraging effort is allocated through a series of feedback loops1,2. Here, we present a mathematical model that quantifies the extent to which honeybee colonies behave collectively in different ecological environments by inferring the proportion of dances performed by bees that have explored individually, and those that have been recruited to food sources. By applying this methodology to waggle dance data from twenty hives, we show our model closely fits real-world honeybee foraging patterns and demonstrate that colonies vary their use of waggle dance information across different landscapes. Our methodology provides a tool to identify the ecological conditions in which honeybee colonies rely on dance communication, opening the door to large-scale experimental exploration of the selection pressures that may have driven the evolution of this remarkable collective behaviour.

# Introduction

Within honeybee colonies, a series of simple rules that determine when and how much bees perform the celebrated waggle dance (Fig. 1) mean that choices between feeding sites occur at the level of the group rather than the individual. For example, because the number of dance circuits performed by a bee on returning from a food source reflects the net energetic benefits of the trip, more of the colony’s workforce will be recruited to the sweeter of two equidistant sources3, or the closer of two equally sweet sources4, without requiring any individual bee to compare options. This extraordinary system is a key example of how social insect colony behaviour can take on a form extending beyond that of the individual units5, functioning as a collective that makes decisions as a single entity.

Despite its intricacy and precision, a wealth of research now shows that foraging bees frequently do not use dance information to find forage sites6. Accordingly, in many situations, colonies that are prevented from communicating via dances achieve equal or even greater foraging success than their wild-type counterparts7–10. This apparent variation in the use of collective behaviour is intriguing because it may provide clues as to the selection pressures that were critical in the evolution of the dance communication system.

Identifying the circumstances in which bees use the dance is likely to be a key step towards understanding when it is likely to afford fitness benefits at the colony level. Despite the amount of research interest that has focused on the proximate mechanisms of collective decision-making via dance communication, the ultimate evolutionary drivers of its unique evolution within *Apis* -unparalleled in even a single other social insect- remain obscure12. While it is likely that the spatiotemporal distribution of forage was key, some studies find the benefits of dance communication to be realized only in challenging environments where resources are clumped or ephemeral7,8,11, while others find no effects of landscape heterogeneity13, or that dancing comes at a cost in challenging environments10 and is most beneficial when species richness is high14.

Attempts to identify the circumstances in which dance communication is important have been challenging because they require constant monitoring of real-world colony weight7,9–11,13,15 -a noisy proxy of foraging success that is influenced by many other abiotic and biotic factors, including colony health- over ecologically credible time periods16. To do this at a scale that allows inter-colony variation in foraging environments at sufficient replication is a major logistical hurdle (but see13), particularly given that multiple landscape variables may interact to determine the utility of dance communication.



Figure 1. **The honeybee waggle dance carries information about the location of a resource.** The duration of the waggle run indicates the distance to the resource and the angle of the dance relative to the vertical indicates the direction of the resource (circle in right panel), relative to the direction of the sun1. Through the observation and decoding of the waggle dance, a colony’s dance floor provides a unique opportunity to eavesdrop on the communication and decision making leading to collective foraging decisions. Overall resource “quality” -the net energetic gain of a foraging trip- is provided through the number of waggle runs performed1,17–20. Although bees that follow dances do not specifically interpret this information on an individual level3, the resulting over-representation of high-quality sites on the dance-floor means that they are more likely to encounter dances that advertise better forage2, and provides the colony with a mechanism to select the most profitable resources in their environment21.

Here, we present a method to determine the extent of waggle dance use in colony foraging within specific landscapes by identifying the proportion of foraging trips that are made by bees recruited through waggle dances. Our method examines the distribution of waggle run durations reported on the dance-floor and infers from its shape the relative contribution of dances for resources identified by individual search, “scouting” versus dances for resources found by following dances, “recruitment”2. Previously, the relative amount of scouting and recruitment trips have only been quantifiable by documenting foragers’ individual search history, which requires tracking individual bees and is time -and labour- intensive2. Our methods provide an efficient means of quantifying collective behaviour with no requirement to manipulate hive orientations, assay foraging efficiency or document individuals’ previous experience.

# Results

To reveal how the patterns of waggle dance-encoded foraging distances might differ under scenarios where colonies rely on individual search versus recruitment, we simulated honeybees foraging in a landscape where resource patches were randomly placed in the environment. Foragers could locate these under two different strategies: either acting as a scout and locating resources themselves, or following a recruit strategy and locating resources by following a random dance from the dance floor22 (Fig. 2a,b, details in Methods). As it is known in the simulation which individuals in the hive forage under what strategy, we can compare the distributions of foraging distances reported on the dance floor. Fig 2c shows that the shapes of the resource distance distributions for bees engaging in the two types of foraging trips are different. The distance distribution for the scout trips is akin close to that of an exponential distribution (Fig. 2c), which is the nearest neighbour distance distribution for foragers operating in a one-dimensional environment (see methods). The distribution of the distances reported for recruit trips (Fig. 2c) is a Rayleigh distribution which is the nearest-neighbour distribution in a two-dimensional environment23 (see Methods).

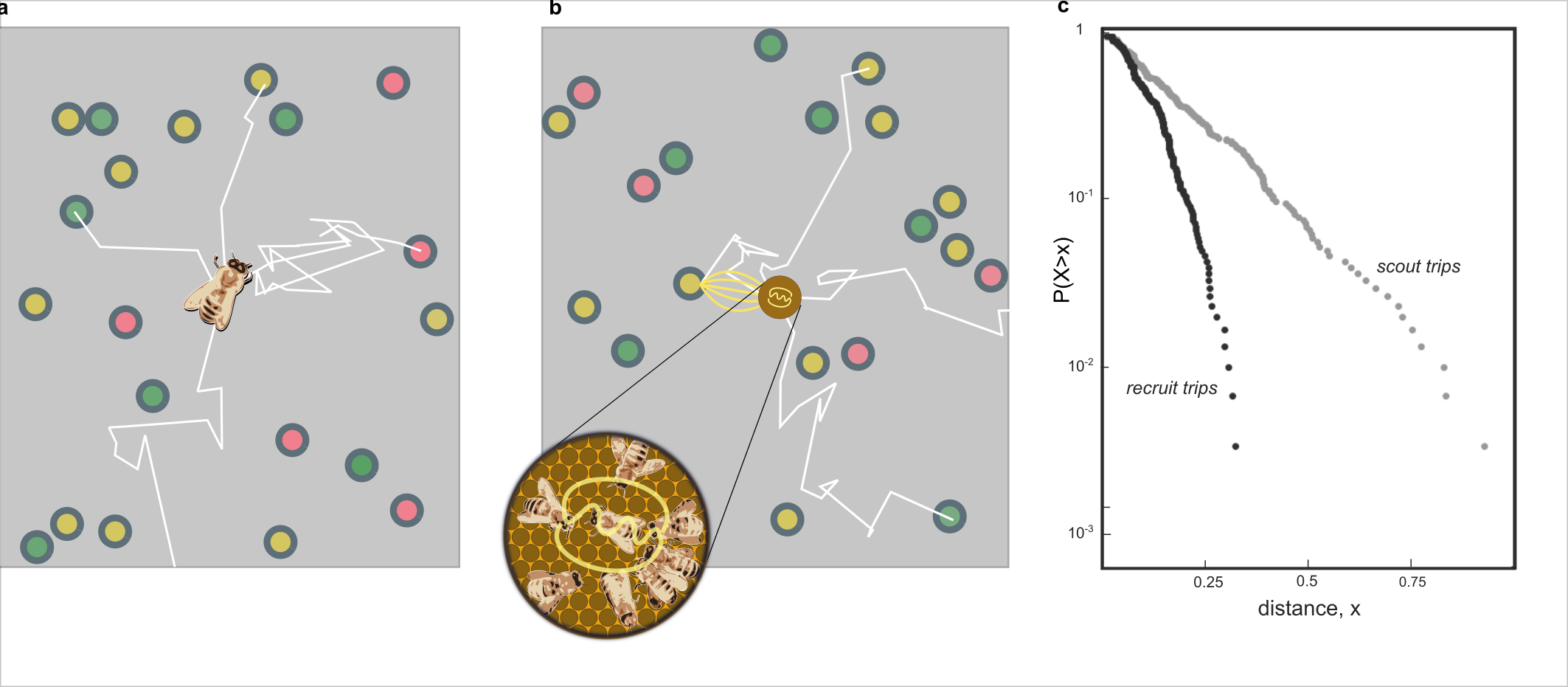


Figure 2. **Simulating honey bee foraging**. In our simulation model with scouting only (a), foragers leave the hive on a search path (white lines) and continue until they encounter a resource (circles, colours indicate different resource quality). When foraging with recruitment (b) foragers continue to identify resources in scouting trips (white lines) and convey this information on the dance floor (brown disc) where foragers can sample dances reporting on scouting and recruiting trips and follow these directions (yellow lines). (c) Complementary cumulative frequencies of foraging distances reported from scouting and recruit trips. Note the difference in the shape of the distributions

On the dance-floor, the number of waggle runs performed for a resource depends on its profitability19. Honeybees achieve this by measuring the energetic efficiency of a foraging trip through the ratio of energetic gain to energetic cost19,21. By combining this profitability bias with the distributions identified in our simulations, we can accurately describe the distribution of waggle runs reported on the dance-floors of real honeybee colonies as a superposition of scout and recruit distributions (Fig 3, see Methods for details)

This description intrinsically captures honeybee foraging as collective decision making, whereby the foraging sites represented on the dance floor derive from a mixture of individual search and waggle dance information, modified by the profitability rule that biases recruitment towards closer or sweeter patches. The extent to which collective decision making is used is expressed in the proportion of scout dances, , and the proportion of recruit dances, . Scouting and recruiting are not fixed behavioural categories, because individual bees can engage in both over the course of their foraging lifetime, and foragers can dance on return from any successful trip irrespective of whether they were recruited to the forage site or found it individually24.

In a hypothetical colony that relies only on individual search to find foraging sites, all trips are “scout” trips, but as recruitment becomes more important, the proportion of “recruit” trips will increase. By setting , we obtain a model based on the sole use of individual search (scouts only) and by allowing the proportion of scout trips, , to take on any value between 0 and 1, we can derive the extent to which foraging occurs collectively. (Fig 2). In fitting these two models to distribution of waggle run durations decoded from real honeybee colonies we are thus able, using model selection25, to infer if and, by estimating the parameter , to what extent honeybee colonies use waggle dance information when foraging.

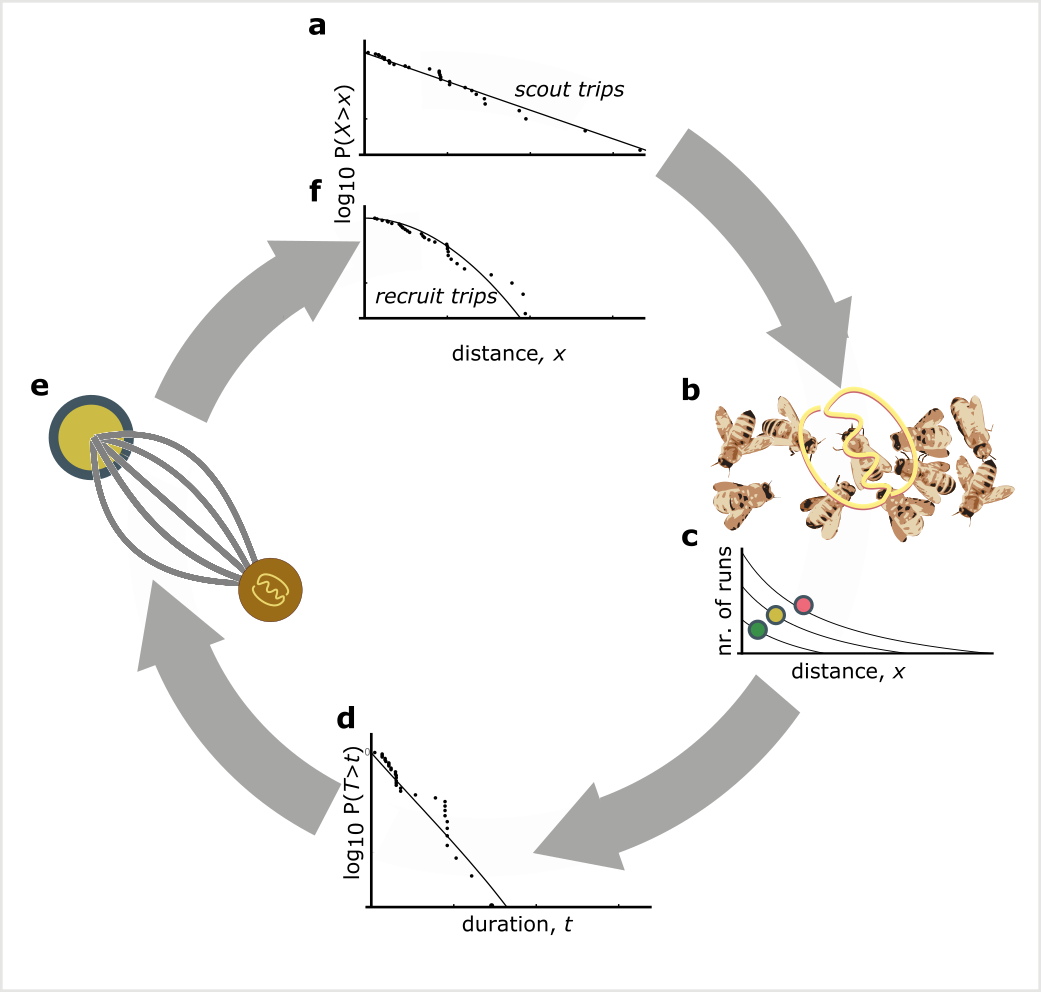


Figure 3. **The rationale of the foraging model**. The distances of resources encountered by scouts are distributed exponentially (a). These dances are advertised on the dance floor (b). Dances for resources that are closer or higher in quality are repeated more often (c). As a consequence, dances for more profitable resource are over-represented and sampling foragers are biased to the more profitable resources (d). After successfully visiting advertised resources, recruits also dance for them leading to further amplification of this bias towards the most profitable resource in the vicinity of the hive (e). The distances of recruiting trips are than distributed through a Rayleigh distribution (f). The distances reported on the dance floor are a mixture of the scout and recruiting trips and can be calculated from the distance distributions of the scouting and recruiting trips, taking the reporting bias into account (see Methods for detail).

To evaluate the use of waggle dance information and individual search in honeybee colonies foraging in ‘natural’ landscapes, we analysed a pre-existing dataset of 2827 waggle dance observations from 20 observation hives, recorded between April-September 2017, (previously described in26). Hives that contributed to this dataset had been situated at different locations in South East England (see Methods, figure 4A) and visited every two weeks for a period of 24 weeks. On each visit, two hours of continuous waggle dance data was recorded by training a camcorder onto the dance floor. The footage of the dances was decoded manually26–28 to extract waggle run durations. Using this data, for each site we fit both the collective and individual models and used model selection to determine which provided the better explanation of the data, and (if the collective model provided a better fit) to quantify the relative use of social information through estimating the parameter . In each case, We calculated the goodness-of-fit using a Kolmorgorov-Smirnov (KS) test to ascertain if the model provided a plausible explanation of the data29,30.

For 16 out of 20 study hives, a model of collective foraging provided a better description of the data than one of individual foraging (Fig. 4a). In the other 4 sites, despite the collective model having the higher maximum likelihood, the individual model had a higher AIC value and so is more parsimonious (Supplementary Table 1). In all but one site, the collective model had a good fit (using a Kolmogorov-Smirnov statistic of , see methods) to the empirical waggle run durations (Fig. 4b), whereas the scout model was significantly different to the observed data in 8 sites (Kolmogorov-Smirnov statistic , Fig. 4b). The sites shown in Figs 4c-d are representative examples showing the model fits where the individual (Fig. 4c) and collective (Fig. 4d) models fit best. Note the closeness of the fit to the data, illustrating the overall quality of the model description.

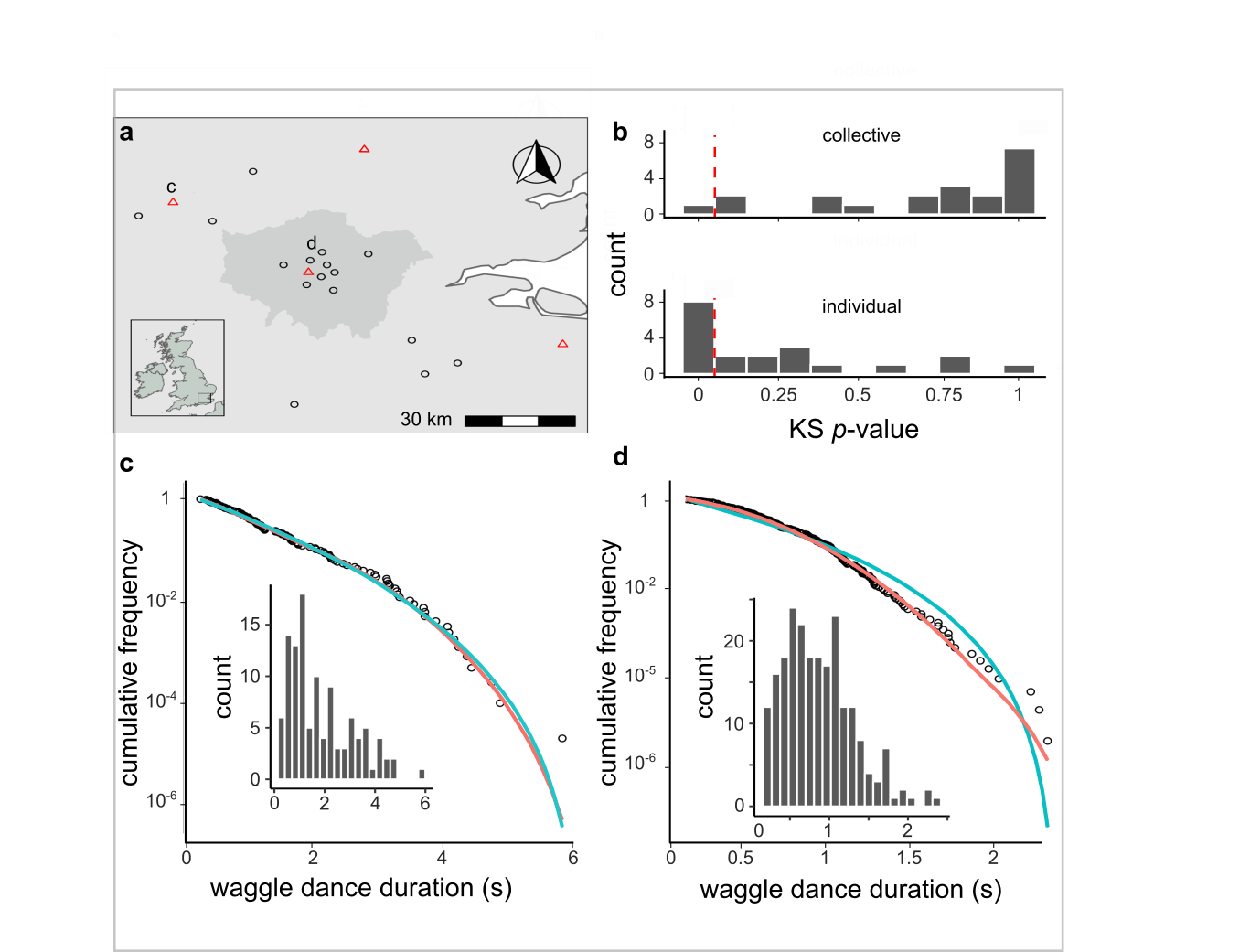


Figure 4. **The honey bee foraging model fitted to data from 20 hives**. (a) Location of study hives in Southern England, shaded area in the main plot indicates Greater London. For 16 hives for the collective foraging model provided best explanation (black circles) or for 4 hives the individual search model provided the best explanation(red triangles) as indicated by lowest AIC score. (b). Distribution of goodness of fit confidence values for each model fit to waggle run durations from each site. The p-value is derived from a bootstrapped two-sided KS test comparing the fitted model predictions to the empirical data, the red dashed line marks the significance threshold of 0.05. For values exceeding the threshold there is no statistically significant difference between the model and the data, indicating the model provides a good fit. For the hive in (c) the individual model (blue line) provided a better fit than the collective foraging model (red line). For the hive in (d) the collective foraging model (red line) provided a better fit than the individual model (blue line). Panels show the compliment cumulative frequencies with binned frequency distributions as inset.

Our results show that, whilst colony-level foraging is mostly comprised of a mixture of scout and recruit foraging trips, in some circumstances, colony foraging can be better described by individual foraging alone. Thus, in some environments, the majority of foraging trips involve scouting to find new food sites rather than recruitment through dances. Note that this does not imply that these bees do not engage in dance following, because bees regularly follow dances but choose not to visit the advertised site31, although it has also been shown that bees may cease dance following if it is proving unproductive10,31. Evaluating how these individual decisions influence the collective, however, has historically been a challenge as it is effectively impossible to track an individual’s foraging behaviour over a landscape. As our results show that individual foraging accurately describes colony foraging in 11 different sites, and more parsimoniously than a model of collective foraging in 4 sites, these findings support the idea that individual decision making can dominate colony foraging and demonstrates further evidence for flexible waggle dance use by honey bee colonies.

Further quantification of the use of waggle-dance recruitment within all colonies, as a proportion of all foraging trips, can be achieved by extracting the estimated proportion of scout trips, , for each site. Since our sites varied in land-use characteristics and potentially thus forage distributions (although in this case, not by design for this study), we investigated whether these estimates might correlate with land-use. We first classified the different land-use types of the area surrounding each site32 to obtain a standardised land-use profile for the urban and agri-rural environments separately as many land-use types present in urban areas do not occur in agri-rural environments and vice versa (see methods). We then performed a Partial Least Squares (PLS) analysis33 (see methods) to determine the principal components that represent combinations of land-use types which explained the most variation in the proportion of scout dances within agri-rural and urban environments. As for one of our sites neither model provided a plausible description, this site was removed from the PLS analysis. Due to our small sample size (10 urban and 9 agri-rural sites), we used jackknife resampling to evaluate the robustness of our results to influential points (see Methods, Supplementary Material).

In the agri-rural environments the first principal component is a combination of land use types that explained ~73% of the variation in the proportion of scouts (beta regression: , = 4.9, p < 0.05, Fig. 5a). This principal component correlates positively with arable land (29% of land coverage; Table 1) and negatively with built-up areas (17% of land coverage); note that it also correlates negatively with non-agricultural unmanaged green space and water, but together these represent less than 3% of land-use (Table 1; Fig. 5b). These land-use types maintain a significant correlation with the first principal component over the jackknifed PLS (Fig. 5b), with the exception of non-agricultural unmanaged green space which sits on the boarder, indicating the results are robust. As arable land increases whilst built-up areas decrease, the proportion of trips that are driven by individual search increases. Arable land in the UK is typically considered nutritionally poor for bees (note that oilseed rape fields were not included within this category), while there is evidence to suggest that the residential areas that were captured within the “built-up” category are forage-rich hotspots, typically supporting relatively high bee diversity and abundance within gardens34.

**Table** **1**: Percentage area covered for each land-use type in the agri-rural and urban environments in the sites studied.

| **Environment** | **Land-use** | **% coverage** |
| --- | --- | --- |
| Agri-rural | Arable | 28.300 |
| Agri-rural | Pasture | 23.300 |
| Agri-rural | Woodland | 21.100 |
| Agri-rural | Built Up Area | 15.000 |
| Agri-rural | Fruit | 3.080 |
| Agri-rural | Oilseed Rape | 2.900 |
| Agri-rural | Non Agricultural Unmanaged Green Space | 2.810 |
| Agri-rural | Non Agricultural Managed Green Space | 1.760 |
| Agri-rural | Other Agricultural | 1.500 |
| Agri-rural | Water | 0.178 |
| Urban | Sparse Residential | 34.800 |
| Urban | Continuous Central | 24.300 |
| Urban | Dense Residential | 21.800 |
| Urban | Parks Allotments Cemeteries | 7.880 |
| Urban | Woodland | 4.220 |
| Urban | Water | 3.360 |
| Urban | Amenity Grassland | 2.620 |
| Urban | Railway | 1.040 |

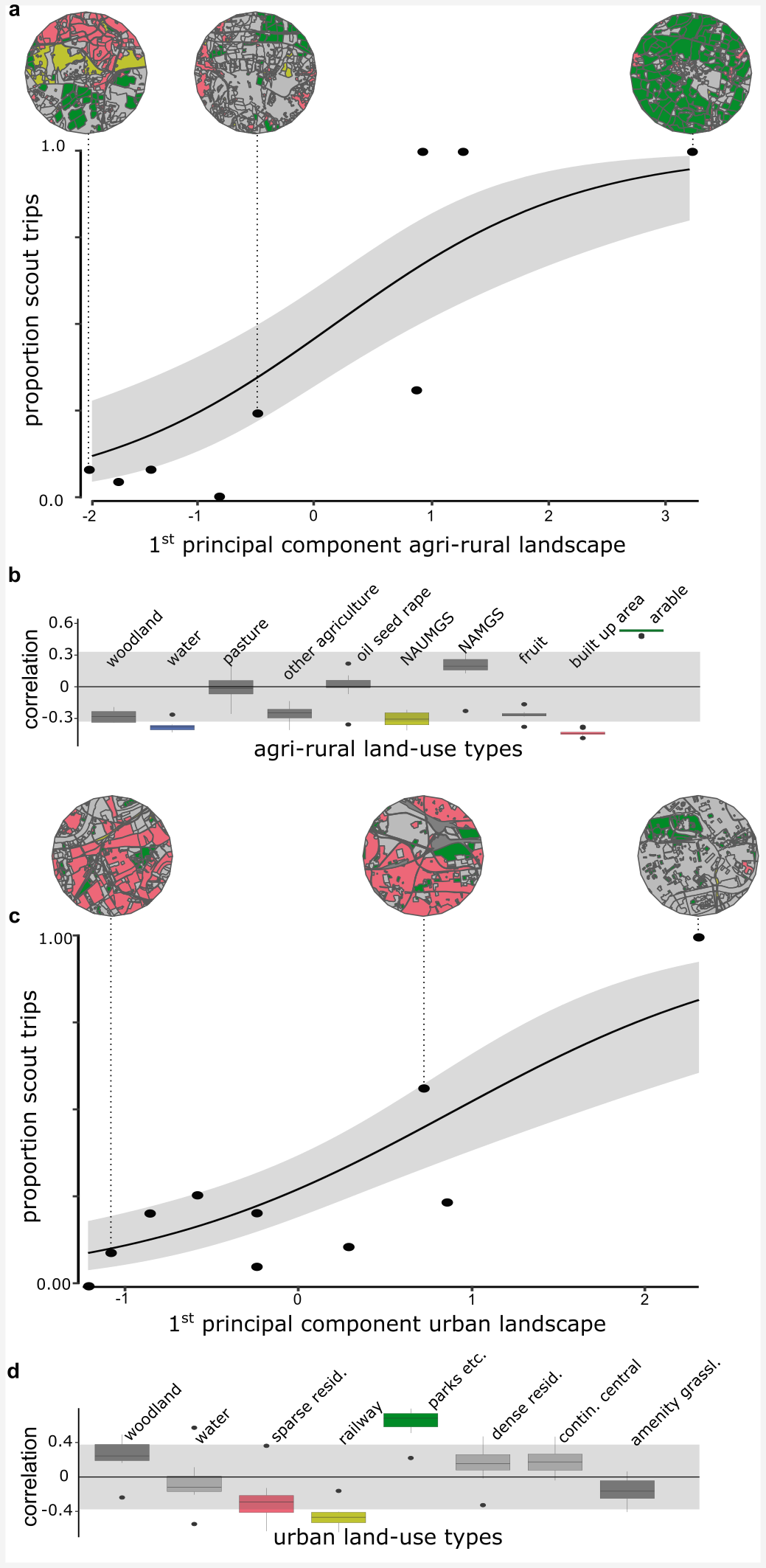


Figure 5. **Collective foraging correlates with land-use**. The proportion of scouts for each site against the first principal component derived from a Partial Least Squares analysis of land-use type. Beta regression shows the relationship (black line) between first principal component and the proportion of scouts, with 95% CI shown by the grey shaded area for agri-rural (a) and urban landscapes (c). The correlations between first principal component and each land-use type are shown for agri-rural (b) and urban landscapes (d). Correlations outside the shaded area significantly contribute to the first principal component. Colours correspond to the land use as shown in maps (circular insets) for selected sites. NAUMGS (resp. NAMGS) stands for non-agricultural unmanaged (resp. managed) green space.

In the urban environment, the first principal component explained ~73% of the variance in the proportion of scouts (beta regression: = 0.73, = 10.4, p < 0.05, Fig. 5c). This component correlates positively with parks, allotments and cemeteries (8% of land cover; Table 1) and negatively with railways, however this accounts for ~1% of land cover (Table 1). The dominant land-use type by coverage, sparse residential (land coverage ~35%; Table 1), has a significant negative correlation with the first principal component for some sites removed over the jackknife sampling (Supplementary Fig. 2.), however, it is overall not a significant contributor to the variance in the loadings (Fig. 5c.). Parks, allotments and cemeteries are typically forage-rich for honeybees34, a result which was not replicated here. However, note that the amount of land cover for this component is limited and the jackknifed PLS identified substantial variation in the loadings identified for the first principal components (Supplementary Fig. 3.).

# Discussion

Here, we have presented a model to quantify the use of waggle-dance communication in collective decision-making by honeybee colonies. Whilst recruitment is well known to occur through other mechanisms besides the waggle dance, such as through olfactory cues35, the exceptionally close fit of our model to waggle run durations underlines the importance of the waggle dance in honey bee foraging. The fitting of our model to waggle dance observations provides a time and labour efficient methodology to quantify collective behaviour. Further, by identifying the land-use combinations which most influence variation in the proportion of scouts, our model provides a tool to map the environment along the major axis of honeybee information use and visualise how land-use influences the use of the waggle dance.

Whilst our results show a decrease in reliance on waggle dance recruitment as resources become harder to find and foraging trips become longer26, they do not yet provide conclusive evidence for such a pattern. Our study was designed as a proof-of-principle based on an existing dataset, we did not systematically choose sites based on forage availability, and instead inferred forage availability based on land-use at each site. The dance recordings in our dataset were collected over an extended period of five months, over which time forage availability in the landscape likely changed considerably and non-uniformly across sites. We thus cannot rule out that longer term effects not captured in the data used in this study, such as resource stability16, may also have contributed to our estimates of waggle dance use.

Our results within the agri-rural sites are in agreement with other studies which evaluated foraging performance with and without the dance10,11. These results suggest that, even though collective foraging is not always beneficial, in environments where high quality resources are present, but relatively scarce, collective foraging is being carried out by colonies, suggesting that under these conditions exchanging social information through the waggle dance confers benefits. In both agri-rural and urban landscapes we have shown that the proportion of scouting trips change with land-use (illustrated in the change in land-use shown in the maps from left to right in Fig 5). We arrived at this conclusion through the analysis of waggle dance data, using a mathematical model to fit the waggle dance durations against. This method provides a time and labour efficient methodology to quantify collective behaviour. We analysed an existing data set and thus provide a proof of concept for how this new toolkit can be used to evaluate factors influencing waggle dance recruitment.

The analysis of waggle dance data can make an important contribution to our understanding of social information use and provides a methodology to further evaluate how honey bees use their unique dance language. Recently, technological advances have emerged which enable colony metrics to be collected faster, more accurately and over greater time spans than could be gathered by hand, allowing individuals to be tracked within colonies and theories of individual behaviour to be evaluated in more depth than could have been done previously36. By piecing together the behavioural response of individuals and combining these with landscape analyses, we have found an accurate mathematical description of colony foraging which extends our ability to quantify collective behaviour across environments. With the advances in the decoding of the waggle dance through automated methods37, we face the prospect of waggle dance data becoming “big data”. Our methodology thus provides a means of analysing such large data sets to inform the debate about the importance of collective decision making, as well as providing useful colony metrics of foraging activity.

# References

1. Von Frisch, K. *The Dance Language and Orientation of Bees*. (Harvard University Press, 1967). doi:[10.4159/harvard.9780674418776](https://doi.org/10.4159/harvard.9780674418776).

2. Seeley, T. *The Wisdom of the Hive*. 317 (1995). doi:[10.1016/j.desal.2010.03.003](https://doi.org/10.1016/j.desal.2010.03.003).

3. Seeley, T. D., Camazine, S. & Sneyd, J. Collective decision-making in honey bees: how colonies choose among nectar sources. *Behavioral Ecology and Sociobiology* **28**, 277–290 (1991).

4. Hasenjager, M. J., Hoppitt, W. & Leadbeater, E. Do honey bees modulate dance following according to foraging distance? *Animal Behaviour* **184**, 89–97 (2022).

5. Sumpter, D. J. The principles of collective animal behaviour. *Philosophical Transactions of the Royal Society B: Biological Sciences* **361**, 5–22 (2006).

6. Grüter, C. & Farina, W. M. The honeybee waggle dance: can we follow the steps? *Trends in ecology & evolution* **24**, 242–247 (2009).

7. Sherman, G. & Visscher, P. K. Honeybee colonies achieve fitness through dancing. *Nature* **419**, 920–922 (2002).

8. Dornhaus, A. & Chittka, L. Why do honey bees dance? *Behavioral Ecology and Sociobiology* **55**, 395–401 (2004).

9. Grüter, C., Segers, F. H. & Ratnieks, F. L. Social learning strategies in honeybee foragers: Do the costs of using private information affect the use of social information? *Animal Behaviour* **85**, 1443–1449 (2013).

10. I’Anson Price, R., Dulex, N., Vial, N., Vincent, C. & Grüter, C. Honeybees forage more successfully without the ‘dance language’ in challenging environments. *Science Advances* **5**, (2019).

11. Dornhaus, A., Klügl, F., Oechslein, C., Puppe, F. & Chittka, L. Benefits of recruitment in honey bees: effects of ecology and colony size in an individual-based model. *Behavioral Ecology* **17**, 336–344 (2006).

12. I’Anson Price, R. & Grüter, C. Why, when and where did honey bee dance communication evolve? *Frontiers in Ecology and Evolution* **3**, 125 (2015).

13. Nürnberger, F., Steffan-Dewenter, I. & Härtel, S. Combined effects of waggle dance communication and landscape heterogeneity on nectar and pollen uptake in honey bee colonies. *PeerJ* **2017**, e3441 (2017).

14. Donaldson-Matasci, M. C. & Dornhaus, A. How habitat affects the benefits of communication in collectively foraging honey bees. *Behavioral Ecology and Sociobiology* **66**, 583–592 (2012).

15. Kirchner, W. H. & Grasser, A. The significance of odor cues and dance language information for the food search behavior of honeybees (Hymenoptera: Apidae). *Journal of Insect Behavior* **11**, 169–178 (1998).

16. Schürch, R. & Gruẗer, C. Dancing Bees Improve Colony Foraging Success as Long-Term Benefits Outweigh Short-Term Costs. *PLOS ONE* **9**, e104660 (2014).

17. Boch, R. Die Tänze der Bienen bei nahen und fernen Trachtquellen. *Zeitschrift für Vergleichende Physiologie* **38**, 136–167 (1956).

18. Esch, H. Über die Schallerzeugung beim Werbetanz der Honigbiene. *Zeitschrift für Vergleichende Physiologie* **45**, 1–11 (1961).

19. Seeley, T. D. & Tovey, C. A. Why search time to find a food-storer bee accurately indicates the relative rates of nectar collecting and nectar processing in honey bee colonies. *Animal Behaviour* **47**, 311–316 (1994).

20. Seeley, T. D., Mikheyev, A. S. & Pagano, G. J. Dancing bees tune both duration and rate of waggle-run production in relation to nectar-source profitability. *Journal of Comparative Physiology - A Sensory, Neural, and Behavioral Physiology* **186**, 813–819 (2000).

21. Seeley, T. D. Honey bee foragers as sensory units of their colonies. *Behavioral Ecology and Sociobiology* **34**, 51–62 (1994).

22. Seeley, T. D. & Towne, W. F. Tactics of dance choice in honey bees: do foragers compare dances? *Behavioral Ecology and Sociobiology* **30**, 59–69 (1992).

23. Pyke, G. H. Optimal foraging in bumblebees and coevolution with their plants. *Oecologia 1978 36:3* **36**, 281–293 (1978).

24. Beekman, M., Gilchrist, A. L., Duncan, M. & Sumpter, D. J. What makes a honeybee scout? *Behavioral Ecology and Sociobiology* **61**, 985–995 (2007).

25. Burnham, K. & Anderson, D. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach.* (Springer New York, 2002). doi:[10.1007/b97636](https://doi.org/10.1007/b97636).

26. Samuelson, A. E., Schürch, R. & Leadbeater, E. Dancing bees evaluate central urban forage resources as superior to agricultural land. *Journal of Applied Ecology* (2021) doi:[10.1111/1365-2664.14011](https://doi.org/10.1111/1365-2664.14011).

27. Couvillon, M. J. *et al.* Intra-dance variation among waggle runs and the design of efficient protocols for honey bee dance decoding. *Biology Open* **1**, 467–472 (2012).

28. Schürch, R. *et al.* Dismantling Babel: creation of a universal calibration for honey bee waggle dance decoding. *Animal Behaviour* **150**, 139–145 (2019).

29. Goldstein, M. L., Morris, S. A. & Yen, G. G. Problems with fitting to the power-law distribution. *The European Physical Journal B - Condensed Matter and Complex Systems 2004 41:2* **41**, 255–258 (2004).

30. Clauset, A., Rohilla Shalizi, C. & J Newman, M. E. POWER-LAW DISTRIBUTIONS IN EMPIRICAL DATA. (2009).

31. Grüter, C. & Ratnieks, F. L. Honeybee foragers increase the use of waggle dance information when private information becomes unrewarding. *Animal Behaviour* **81**, 949–954 (2011).

32. Samuelson, A. E. & Leadbeater, E. A land classification protocol for pollinator ecology research: An urbanization case study. *Ecology and Evolution* **8**, 5598–5610 (2018).

33. Carrascal, L. M., Galván, I. & Gordo, O. Partial least squares regression as an alternative to current regression methods used in ecology. *Oikos* **118**, 681–690 (2009).

34. Baldock, K. C. *et al.* A systems approach reveals urban pollinator hotspots and conservation opportunities. *Nature Ecology and Evolution* **3**, 363–373 (2019).

35. Arenas, A., Fernández, V. M. & Farina, W. M. Floral odor learning within the hive affects honeybees’ foraging decisions. *Naturwissenschaften* **94**, 218–222 (2007).

36. Wild, B. *et al.* Social networks predict the life and death of honey bees. *Nature Communications 2021 12:1* **12**, 1–12 (2021).

37. Wario, F., Wild, B., Rojas, R. & Landgraf, T. Automatic detection and decoding of honey bee waggle dances. *PLOS ONE* **12**, e0188626 (2017).

Forage site distributions quantify collective foraging in honey bee colonies: Methods

# Acknowledgments

This work was supported by the Biotechnology and Biological Sciences Research Council (BBSRC) through grant BB/M011178/1.

# Conflict of interest

The authors declare no conflict of interest.

# Code and data availability

All code for analysis is available on GitHub at (insert link to finalised public github repo or archive). All data is available via the Dryad Digital Repository <https://doi.org/10.5061/dryad.c2fqz618f>1.

# Data collection

Methods for how the data used for this study was collected can be found in full in Materials and Methods sections 2.1, 2.2 and 2.3 of1.

## Land-use preference analysis

Methods for how land-use types were classified can be found in Materials and Methods sections 2.6 of1.

## Waggle dance decoding

Methods for waggle dance decoding are fully described in Materials and Methods section 2.4 in1.

# Simulation

All simulation code was written in Python version 3.9 and uses the Pandas2 and Scipy3 packages.

A circular environment is first created with radius . The number of resources in the environment is generated as a random Poisson variable with rate equal to 5000 multiplied by the area of the environment. These are placed on polar coordinates with a uniformly selected angle, , between 0 and 2 and a radial value, , between 0 and , determined from the square root of the uniform position values multiplied by . These polar coordinates are converted to Cartesian coordinates. Each location is then assigned to an instance of a resource object along with a random quality of between 0 and 10. This quality is combined with the distance of the resource to the centrally located hive to form a measure of how profitable the resource is (see model, equation?).

One-hundred honeybee objects are created, 20 of which 20 are on scouting trips and the rest recruited to follow scout dances. Scouts leave the environment following a random path through the environment generated by sampling a uniform random step length and angle. The number of paths the scout draws when searching is also determined as a uniform random number. Each straight line in the random path is converted to a rectangle with length equal to the path section length and a constant width of ~0.01 to represent an area the scout searches along that path. Of all the resources contained in the boxes drawn from the scout’s path, the one closest to the colony is selected as the resource patch that the scout will report and will communicate its location if the quality of the resource resource exceeds a minimum threshold. Communication is simulated by pooling together all the resource patches found. If no resources are contained in the scout’s path, they will not add any resources to the scout pool and draw a new path in the next foraging iteration.

Recruits represent honeybee objects which do not perform the searches the scouts do. Instead, they sample from the pool of resource objects reported by scouts. This sampling is done by selecting resources with a probability which is skewed towards the profitability of the resource, meaning more profitable resources have a greater chance of being selected by recruits. Recruits will then visit these resources and in the next iteration will add their resource to the pool of scout dances. Consequently, the pool of dances represents resources discovered by scouts and resources exploited by recruits. When a resource is depleted, it is removed from the environment and so any foragers that were foraging on it would select a different resource from the dance floor.

The simulation was run 100 times and every 5 time steps all distances reported by scouts and recruits were recorded and combined. This was done to reflect the way foraging data is collected in real honeybee studies. We fit an exponential and minimum of an exponential distribution to both the distribution of foraging distances reported by the scout and recruit objects. Fitting was done by deriving the maximum likelihood estimate for each model fit on each data source through their analytical solutions: , minimum of the exponential with a minimum foraging distance: . As the exponential assumes distributions start from 0 the data was transformed to start from 0 by subtracting the minimum foraging distance from all foraging distances () before fitting.

# Model

To describe the distribution of dance durations on the dance floor we formulated a generic model the duration of waggle dances. In the model resource patches are assumed to be randomly placed in the environment. Foragers scout for these patches. The rationale of the model is illustrated in Fig. 3: upon visiting a resource patch, foragers translate the profitability of a resource into the number of repeats of the dance. The number of repeats of the dance is a function of quality an distance. Recruits sample random dances and report the location of successful visits to resource patches on the dance floor. Through the feedback and over-representation of profitable resources on the dance floor recruits will converge to visiting the most profitable resource in vicinity of the hive. The distribution of dance durations is the superposition of scouting and recruiting trips.

As the resources patches are randomly placed in the environment, the distance after which the first resource is discovered approximately follows an exponential distribution (given by ). Through the feedback mechanism that the dance floor provides, the colony can, collectively, locate the most profitable resource in its environment. For randomly placed resources in a two dimensional environment the distance to the nearest point is distributed according to a Rayleigh distribution (given by ) (Pyke 1978). Our simulation model shows that this describes the distances at which recruits visit resources well. Knowing the distance distributions of scout and recruit trips we then assume that the a proportion of all trips are scout trips. With this information we can specify the distributions of distances on the dance floor (see Supplementary Material for details).

We implemented this in a full model which describes the distance distribution of the environment has different resource types (See Supplementary Material). However, in the full model the number of parameters increases with . Even if the number of resources is low, it turned out to be cumbersome to estimate the parameters and the model tends to overfit. To facilitate estimation of the parameter we therefore used a simplified model to estimate the fraction of scout trips, where represents the lowest duration considered, and here a minimum waggle run duration in the data set.

For the simplified the model we assumed that the number of dances depends weakly on distance and there is a sizable quality differences between resources of a decent size and that there is a sizable intensity of the high quality resource. Foragers on scouting trips are more likely to report larger distances than foragers on recruiting trips. By linearising the function that translates the profitability into the number of waggle dance run for the largest dance duration and normalising, we arrive at simplified distribution for dance durations for scouting trips:

where we used the shorthand The parameter is the maximum dance duration by scouts, is the intensity of resources found by scouts and is factor that normalises the distribution.

Recruit trips will be predominantly to high quality resources. Only if the nearest high quality patch is very far away will there be a more profitable patch of lesser quality available, and this happens only rarely if the intensity of the best quality resource is sizable. After linearising the function that translates the profitability into the number of waggle dance runs for short dance durations and normalising the distribution of dance durations reported from recruit trips in the simplified model is:

where

is the normalisation factor, the parameter is the rate with which dances repeats depends on distance for recruit trips and is the intensity of high quality resources reported by recruited foragers.

The simplified distribution function is

which we used for parameter estimation and model fittings.

# Statistical analysis

All analysis code is written in R4.

## Model fitting

All models are fit using Maximum likelihood estimation5 by summation of the log of the simplified distribution function outlined in the methods section: model. The numerical optimisation routine is written in c++ and uses the Nelder-Mead simplex algorithm6 implemented in the ‘NLopt’ library7 and interfaced to R4 using ‘Rcpp’8.

The most parsimonious model is assessed using Akaike information criterion (AIC)5,9 and Akaike weights. The model with the lowest AIC score is deemed to be the most parsimonious.

Goodness of fit is assessed using the two-sample Kolmorgorov-Smirnov (KS) test10 and implemented in R using the ks.boot function of the package ‘Matching’ in R11.

## Partial Least Squares analysis

Prior to conducing the PLS we removed any sites in which the models fit was significantly different to the waggle dance durations for that site. This resulted in one agri-rural site (ROT) and no urban sites being removed from the analysis.

As our estimated proportion of scouts is continuos on the interval we used the R package plsRbeta12 to conduct the partial least squares analysis and performed a beta regression on the results using the R package betareg13. As the betareg package only works on the open interval the data, , was transformed using the following equation: $, as outlined in the betareg package documentation. After analysis the data was back transformed to the original values for the plots in Fig 5.

For the jackknifed resampling we iterated through the each site and removed it from the pool of data and then ran the PLS as described above, recoding the loadings for each iteration (see Supplementary Material for loadings with each site removed). The PLS loadings for each land-use type are plotted as a box plot in Fig 5. to show the spread of these variable types. A loading was determined to be significantly correlating with the first principal component if contributed more than its expected variance.

# References

1. Samuelson, A. E., Schürch, R. & Leadbeater, E. Dancing bees evaluate central urban forage resources as superior to agricultural land. *Journal of Applied Ecology* (2021) doi:[10.1111/1365-2664.14011](https://doi.org/10.1111/1365-2664.14011).

2. McKinney, W. pandas: a Foundational Python Library for Data Analysis and Statistics. *Proceedings of the 9th Python in Science Conference* (2011).

3. Jones, E., Oliphant, T. & Peterson, P. SciPy: Open source scientific tools for Python. (2001).

4. R Core Team. *R: A Language and Environment for Statistical Computing*. (R Foundation for Statistical Computing, 2020).

5. Burnham, K. & Anderson, D. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach.* (Springer New York, 2002). doi:[10.1007/b97636](https://doi.org/10.1007/b97636).

6. Nelder, J. A. & Mead, R. A Simplex Method for Function Minimization. *The Computer Journal* **7**, 308–313 (1965).

7. Johnson, S. G. The NLopt nonlinear-optimization package. (2020).

8. Eddelbuettel, D. & François, R. Rcpp: Seamless R and C++ integration. *Journal of Statistical Software* **40**, 1–18 (2011).

9. Aho, K., Derryberry, D. & Peterson, T. Model selection for ecologists: the worldviews of AIC and BIC. *Ecology* **95**, 631–636 (2014).

10. Goldstein, M. L., Morris, S. A. & Yen, G. G. Problems with fitting to the power-law distribution. *The European Physical Journal B - Condensed Matter and Complex Systems 2004 41:2* **41**, 255–258 (2004).

11. Sekhon, J. S. Multivariate and Propensity Score Matching Software with Automated Balance Optimization: The Matching package for R. *Journal of Statistical Software* **42**, 1–52 (2011).

12. Bertrand, F. *et al.* Régression Bêta PLS. *Journal de la Société Française de Statistique* **154**, 143–159 (2013).

13. Cribari-Neto, F. & Zeileis, A. Beta Regression in R. *Journal of Statistical Software* **34**, 1–24 (2010).