



OPEN

## Automated monitoring of honey bees with barcodes and artificial intelligence reveals two distinct social networks from a single affiliative behavior

Tim Gernat<sup>1,2</sup>, Tobias Jagla<sup>2</sup>, Beryl M. Jones<sup>1,5,6</sup>, Martin Middendorf<sup>2</sup> & Gene E. Robinson<sup>1,3,4</sup>✉

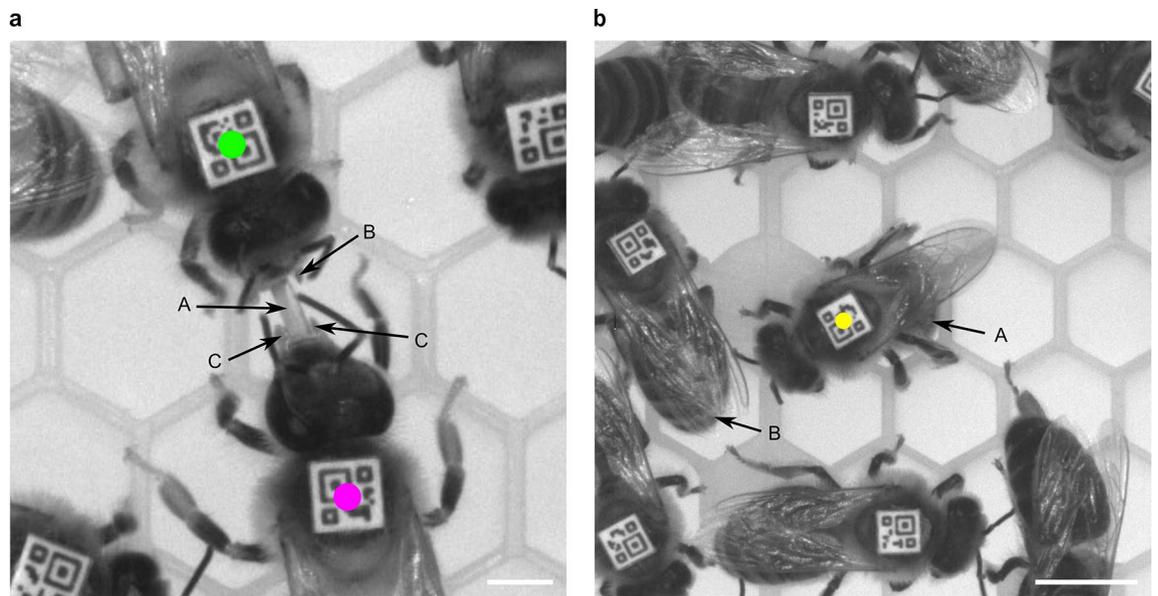
Barcode-based tracking of individuals is revolutionizing animal behavior studies, but further progress hinges on whether in addition to determining an individual's location, specific behaviors can be identified and monitored. We achieve this goal using information from the barcodes to identify tightly bounded image regions that potentially show the behavior of interest. These image regions are then analyzed with convolutional neural networks to verify that the behavior occurred. When applied to a challenging test case, detecting social liquid transfer (trophallaxis) in the honey bee hive, this approach yielded a 67% higher sensitivity and an 11% lower error rate than the best detector for honey bee trophallaxis so far. We were furthermore able to automatically detect whether a bee donates or receives liquid, which previously required manual observations. By applying our trophallaxis detector to recordings from three honey bee colonies and performing simulations, we discovered that liquid exchanges among bees generate two distinct social networks with different transmission capabilities. Finally, we demonstrate that our approach generalizes to detecting other specific behaviors. We envision that its broad application will enable automatic, high-resolution behavioral studies that address a broad range of previously intractable questions in evolutionary biology, ethology, neuroscience, and molecular biology.

Barcode-based tracking makes it possible to automatically distinguish hundreds of individuals in digital videos, and to record their location and heading direction over long time periods at a high spatiotemporal resolution<sup>1–4</sup>. This is otherwise not possible, and has already generated a wealth of individualized data that is transforming how ethologists study the behavior of animals, especially for species that naturally interact in large collectives, such as ants and honey bees<sup>4–8</sup>.

In addition to knowing where individuals are located, it is often necessary to also know what they are doing, to understand both individual and group-level behavior and its neural and molecular underpinnings. However, with the exception of locomotion, barcodes are unable to automatically generate behavioral information directly. When studying other behaviors, researchers therefore tend to resort to proxies that infer coarse-grained behavioral states from changes in the location and orientation of an individual's barcode, and social interactions from the relative position of individuals to each other<sup>2–7,9</sup>. Such proxies have a limited capacity for distinguishing specific behaviors<sup>10</sup> and may thus result in high error rates.

Convolutional neural networks (CNNs) are a promising technology for developing detectors for specific behaviors. They can be trained to accurately identify digital images that show a particular object, and learn

<sup>1</sup>Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, 1206 West Gregory Drive, Urbana, IL 61801, USA. <sup>2</sup>Swarm Intelligence and Complex Systems Group, Department of Computer Science, Leipzig University, Augustusplatz 10, 04109 Leipzig, Germany. <sup>3</sup>Neuroscience Program, University of Illinois at Urbana-Champaign, 505 South Goodwin Avenue, Urbana, IL 61801, USA. <sup>4</sup>Department of Entomology, University of Illinois at Urbana-Champaign, 320 Morrill Hall, Urbana, IL 61801, USA. <sup>5</sup>Present address: Department of Ecology and Evolutionary Biology, Princeton University, 106A Guyot Lane, Princeton, NJ 08544, USA. <sup>6</sup>Present address: Lewis-Sigler Institute for Integrative Genomics, Princeton University, South Drive, Princeton, NJ 08544, USA. ✉email: generobi@illinois.edu



**Figure 1.** Examples of trophallaxis and egg-laying, illustrating the stark differences between these two behaviors. **(a)** In honey bees, trophallaxis is a social behavior during which two individuals orally exchange a liquid containing nutrients, hormones, and signaling molecules. One of the two bees (green) acts as the donor while the other bee (magenta) acts as the recipient. These roles are maintained for the duration of the behavior. Our automatic trophallaxis detector identifies this behavior by recognizing that the recipient has placed her proboscis (A) between the open mandibles (B) of the donor. In this image, only one of the donor's mandibles is clearly visible. The other one is partially hidden by her right antenna. To distinguish donor and recipient, our detector may take advantage of the fact that the mandibles (C) of the recipient are closely aligned with her proboscis while those of the donor are wide open. In addition, the head of the recipient appears to be bigger and shaped differently, because she has tilted it back to be able to extend her proboscis. Note that the proboscis and a cell wall of the plastic honeycomb look very similar; whether the mandibles are visible depends on how a bee holds her head and antennae. Scale bar, 2 mm. **(b)** Egg-laying is a solitary behavior during which an individual bee (yellow) places an egg inside a honeycomb cell. To position the egg, the bee inserts her abdomen into a cell and remains stationary while releasing the egg. Our automatic detector identifies this behavior by distinguishing bees whose abdomens are hidden (A) from those whose abdomens are visible (B). Note that under more crowded conditions than in this image it will be more challenging to make this distinction, because the space behind an egg-layer might be (partially) occupied by the abdomen of another bee. Scale bar, 5 mm.

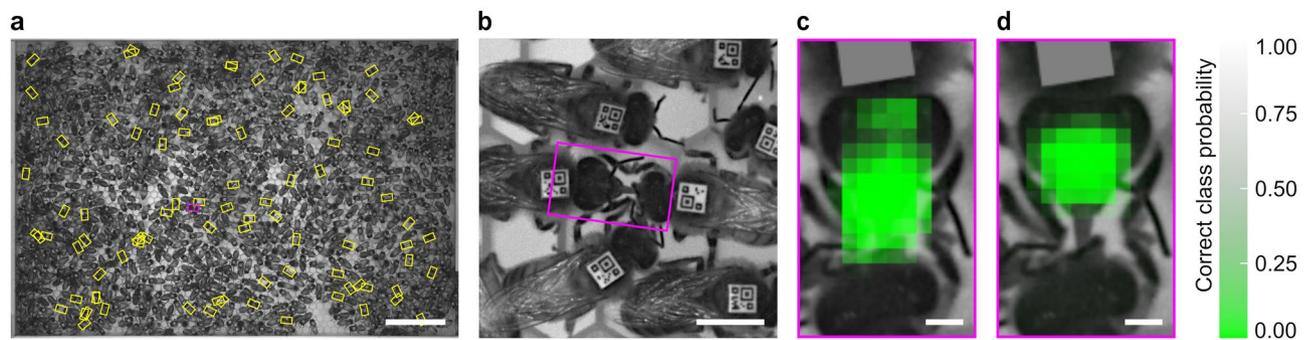
independently which features of the object are most diagnostic<sup>11</sup>. CNNs have been used for tracking animals<sup>12–14</sup>, estimating their pose<sup>15–18</sup>, and detecting behaviors performed in isolation<sup>19,20</sup> or in small groups<sup>20,21</sup>. However, they have not yet been widely employed to identify behavior in videos showing hundreds of closely interacting individuals. Obstacles to more widespread adoption of CNNs include issues that arise when detecting small objects in large scenes<sup>22</sup>, detecting dense or partially occluded objects<sup>22</sup>, and assigning detected behaviors to the correct individuals.

We present a method that combines CNNs with barcode-based tracking to accurately identify specific behaviors in large animal collectives. Complementing the innovations in ref.<sup>23,24</sup>, our key contribution lies in combining tracking information, such as an animal's location and orientation, with domain knowledge about the behavior of interest to perform a precise yet computationally inexpensive region proposal that acts as an attention mechanism for the CNN. In addition, we leverage information from the barcodes to simplify the behavior classification task by rotating the proposed image regions to correspond to a predefined reference frame. Using this approach, we designed a proof-of-principle detector for mouth-to-mouth liquid transfer (trophallaxis) and then used it to gain new insights into this still poorly understood behavior. We also designed a detector for egg-laying to show that our methodology can be applied to very different behaviors.

## Results and discussion

Trophallaxis is an important social behavior during which two adult worker honey bees touch each other with their antennae while orally transferring liquid containing food<sup>25</sup> and signaling molecules<sup>26</sup> (Fig. 1a). This behavior is challenging to detect automatically because honey bee colonies contain hundreds to tens of thousands of individuals densely crowded together (Fig. 2a). Moreover, owing to the small size of the honey bee, a detector needs to focus on millimeter-sized body parts, such as the mouthparts, to distinguish trophallaxis from visually similar behaviors, such as antennation.

To identify trophallaxis, our automatic detector first uses location and heading direction, obtained from each bee's barcode, to select pairs of bees that are in the proper position relative to each other to perform the behavior (Supplementary Fig. 1). For each pair, the detector then estimates a rectangular image region that shows only



**Figure 2.** Automatic trophallaxis detection and recipient identification. **(a)** Typical image captured by the honey bee tracking rig, showing barcoded bees inside the observation hive. Rectangles outline image regions identified by our barcode-based region proposal procedure for trophallaxis. Scale bar, 5 cm. **(b)** Zoom-in on the magenta region proposal in **(a)**, which shows the head, forelegs and part of the thorax of two bees that engage in trophallaxis. Scale bar, 5 mm. **(c)** CNN input image created from the region proposal in **(b)** superimposed with a map of the output of the CNN for detecting the occurrence of trophallaxis, as a function of the position of a grey square that occludes some pixels (see “Methods”). A false negative from the input image is more likely if pixels corresponding to the mouthparts and proboscis are obscured, indicating that the CNN is able to distinguish salient features of trophallaxis from the background. Scale bar, 1 mm. **(d)** The same input image as in **(c)** but superimposed with a map of the correct class probability for recipient identification. This map suggests that the CNN for identifying the recipient relies on visual cues obtained from the mouthparts of the recipient, which, unlike those of the donor, are always visible and closely aligned with the proboscis. Scale bar, 1 mm.

Detector	F <sub>1</sub> score	Sensitivity	Specificity	PPV	NPV	Speed (MP/s)
Trophallaxis	0.89	0.89	1.00	0.90	1.00	55.38
Egg-laying	0.75	0.63	1.00	0.93	1.00	20.87

**Table 1.** Detailed detection and runtime performance estimates. Note that the runtime estimate for trophallaxis detection includes detecting the occurrence of trophallaxis as well as identifying the recipient. *PPV* positive predictive value, *NPV* negative predictive value.

the bees’ mouthparts and head (Fig. 2a,b), which will be examined to verify trophallaxis. This region proposal excludes  $97.9 \pm 1.0\%$  (mean  $\pm$  standard deviation,  $n = 300$ ) of a video frame from further consideration, thus increasing the detector’s computational efficiency and reducing its false positive rate. Image regions are then independently preprocessed and scored with a CNN that was trained to estimate the probability of a region showing trophallaxis (Supplementary Table 1 and Fig. 2c). During post-processing, these probabilities are integrated across successive video frames to determine when trophallaxis begins and ends.

Performance measurements (Table 1) showed that our trophallaxis detector outperforms the state of the art<sup>8,24</sup>. Its Matthews correlation coefficient (MCC) is 0.89, 0.28 higher than that of the best automatic honey bee trophallaxis detector so far<sup>8</sup>. Most of this improvement is due to a 67% higher sensitivity and an 11% higher positive predictive value, which means that our detector identifies more trophallaxis interactions and generates fewer false positives. Unlike other honey bee trophallaxis detectors<sup>8,9</sup>, it also accurately scores interactions between adult worker bees and the queen. Moreover, at 55 megapixel/s, it processes videos 16 times faster than the reference detector<sup>8</sup>, and thus enables larger-scale experiments.

When studying social behavior, it is often important to know the role played by each individual, i.e., who is the “donor” and who is the “recipient” of a particular behavioral interaction. For honey bee trophallaxis, this involves distinguishing the individual that has only opened her mouthparts (donor) from the one that has also extended her proboscis (recipient) (Fig. 1). Previously described automatic trophallaxis detectors are either unable to determine the direction of liquid transfer<sup>8,9,23</sup> or require fluorescence-labeled liquids and a more complex tracking system to do so<sup>24</sup>. To establish the liquid transfer direction directly from videos, we trained a second CNN (Supplementary Table 1 and Fig. 2d) that identifies the liquid recipient, and operated it in parallel with the CNN that detects the occurrence of trophallaxis. When applied to automatically identified trophallaxis partners, this CNN has a MCC of 0.97, which means that in almost all cases it correctly determined a bee’s trophallactic role.

To verify that our trophallaxis detector generates plausible results, we used it to monitor three honey bee colonies, each consisting of  $\sim 1,000$  barcoded individuals, at 1 frame/s for five consecutive days. We then employed a temporally explicit epidemiological model to perform *undirected* spreading simulations on the bees’ trophallaxis networks (Table 2). Because these simulations ignore the direction of liquid transfers, they can be used to study information and disease transmission via physical contacts that take place during trophallaxis, such as antennation or touch, but not how the exchanged liquid flows through the colony. Consistent with previous results<sup>8</sup>, we found that undirected spreading through the observed networks of physical contacts was faster than through temporally randomized counterparts (Fig. 3a, Supplementary Figs. 2a and 3a) until most individuals

Trial	V	E	I	Undirected			Directed		
				$s_{min}$	$s_{max}$	$p_{s=0}$	$s_{min}$	$s_{max}$	$p_{s=0}$
1	1050	115,269	150,859	-0.00	6.81	0.97	-0.19	1.50	0.29
2	902	97,591	132,591	-0.01	4.99	0.95	-0.23	1.00	0.26
3	815	60,540	84,164	-0.04	3.40	0.84	-0.13	1.18	0.36

**Table 2.** Trophallaxis network properties. For each trial we show the number of nodes (V), edges (E), and trophallaxis interactions (I), the smallest spreading speedup ( $s_{min}$ ), the biggest spreading speedup ( $s_{max}$ ), and the mean prevalence when the speedup of spreading is zero ( $p_{s=0}$ ) for the undirected and the directed spreading simulations. In the case of directed spreading,  $s_{min}$ ,  $s_{max}$ , and  $p_{s=0}$  were computed with respect to the temporally and directionally randomized reference networks.

were “infected”, confirming that the temporal structure of these contacts can accelerate the transmission of information or disease<sup>8</sup>.

Knowledge of the direction of liquid transfers allowed us for the first time to also perform *directed* spreading simulations, which model the flow of liquid through the colony (Table 2). These simulations showed that liquid flow was much slower than undirected spreading via physical contacts (Fig. 3, Supplementary Figs. 2, 3), most likely because in directed simulations a recipient cannot “infect” a donor. Moreover, relative to temporally and directionally randomized counterparts, spreading through the observed liquid exchange network was accelerated only during the early phase of the spreading process (Fig. 3b, Supplementary Figs. 2b, 3b); once  $0.30 \pm 0.05$  (mean  $\pm$  standard deviation) individuals were “infected”, liquid flow was inhibited. By randomizing only either the directionality or the temporal structure of trophallaxis we showed that the temporal structure of trophallaxis is the predominant driver of this dynamic, while its directional nature acts to always inhibit the flow of liquid (Fig. 3b, Supplementary Figs. 2b, 3b). This finding sheds light on the forces that shape the dynamics of this important social behavior in honey bee colonies.

To demonstrate that our approach to automatic behavior detection generalizes to other behaviors, we also developed a detector for egg-laying (Fig. 1b), a solitary behavior performed by worker honey bees when the colony has lost its queen and is unable to replace her<sup>27</sup>. Because, unlike for the above case, we included no proxy for identifying bees that are likely to lay an egg in the region proposal procedure for this detector, it is applied to all bees in an image and therefore generates more false positives. We addressed this problem by adding to the detector a second CNN that was trained to identify and filter out false positives generated by the first CNN. This approach increased the detector’s MCC from 0.64 to 0.76 (see Table 1 for additional performance values). These results demonstrate that our method can be applied to automatically detect very different behaviors.

## Conclusion

We demonstrated that barcode-and-CNN-based detectors accurately identify specific behaviors even under the challenging condition of a dense and populous society. We achieved this by combining information obtained from an individual’s barcode with domain knowledge about the behavior of interest, and showed that our approach can be used to detect very different behaviors.

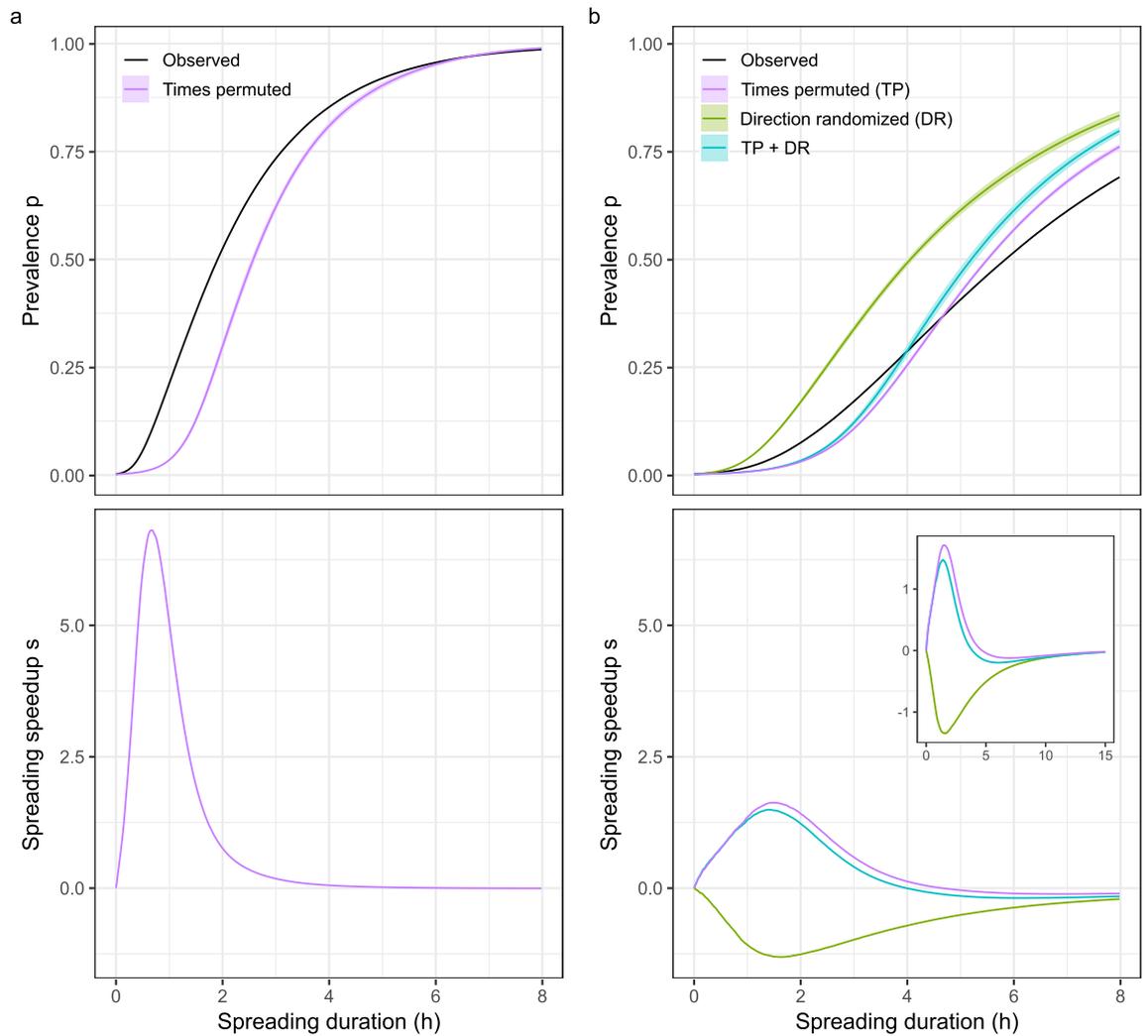
Behavioral data obtained with one of our new detectors revealed that trophallaxis results in two distinct social networks, each with different transmission dynamics. One network represents physical contacts during trophallaxis and the other network represents liquid transfers among bees. Trophallaxis is a complex behavior that involves multiple sub-behaviors, such as antennation and liquid transfer, and each of these behaviors could serve as a separate transmission channel. Moreover, trophallaxis is known to be involved in different biological processes, such as social learning<sup>28</sup>, the regulation of division of labor<sup>29</sup>, and disease transmission<sup>30</sup>. It is thus not surprising that it results in multiple social networks. However, our finding that the two trophallaxis-related networks exhibit different transmission dynamics was unexpected, and sheds light on how trophallaxis might mediate multiple biological processes simultaneously.

Our approach to automatically detecting behavior can be applied to all animals that perform visually recognizable behaviors, and whose position can be tracked. We therefore envision that this method will make it possible to perform quantitative analysis of behavior in a variety of species, including those that live in large collectives, such as ants, bees, and birds. This will enable automatic, high-resolution behavioral studies that address a broad range of previously intractable questions in evolutionary biology, ethology, neuroscience, and molecular biology.

## Methods

**Honey bee tracking.** Colonies were established as in ref.<sup>8</sup>. Briefly, up to 1,370 one-day-old adult honey bee workers (*Apis mellifera*) were individually outfitted with a barcode by chilling them until they stopped moving and then gluing a bCode barcode to their thorax. For recording trophallaxis interactions, we also barcoded an unrelated, naturally mated queen.

All barcoded bees passing quality control checks of barcode placement and readability were moved into a glass-walled observation hive, which was placed in a dark, climate-controlled room and connected to the outside environment via a plastic tube to enable normal foraging. The observation hive held a 348 mm  $\times$  232 mm white plastic honeycomb, one side of which was inaccessible to the bees. The other side was provisioned with enough honey and pollen for the duration of the experiment and was covered by an exchangeable glass window. The



**Figure 3.** Simulated spreading through honey bee trophallaxis networks. Panels show data from Trial 1; see Supplementary Figs. 2, 3 for Trial 2 and 3, respectively, which yielded similar results. **(a)** Undirected spreading results, modeling transmission via physical contacts during trophallaxis. Top panel: Average fraction of “infected” bees (prevalence) as a function of spreading duration. Black line, prevalence in the observed trophallaxis network; magenta line, mean prevalence, averaged across 5 temporally randomized reference networks; magenta band, point-wise 95% confidence interval. Bottom panel: Spreading speedup, corresponding to the normalized difference between the prevalence in the observed network and the mean prevalence across the 5 temporally randomized reference networks, as a function of spreading duration. The spreading speedup is positive if the prevalence in the observed trophallaxis network is higher than in the randomized reference networks, zero if there is no difference, and negative if it is lower. Simulation results indicate that the temporal pattern of trophallaxis accelerates undirected spreading. **(b)** Directed spreading results, modeling liquid flow. Panels as in **a**. Green, prevalence and spreading speedup in directionally randomized reference networks; turquoise, prevalence and spreading speedup in temporally and directionally randomized reference networks. Simulation results indicate that the temporal pattern of trophallaxis accelerates directed spreading during the early phase of the spreading process, but inhibits it later. The directional nature of trophallaxis inhibits directed spreading throughout. Inset: Spreading speedup as a function of spreading duration until, like for undirected spreading, almost all bees are “infected”, showing that all three spreading speedups remain negative for longer spreading durations (corresponding prevalence curves in Supplementary Fig. 4a).

distance between this window and the honeycomb was short enough to ensure that the bees formed a monolayer, which does not affect their behavior<sup>8</sup> and prevented them from obscuring their barcodes.

Trophallaxis interactions and egg-laying events were monitored with the tracking systems described in ref.<sup>30</sup> (Supplementary Fig. 5) and ref.<sup>8</sup>, respectively. Video of the honeycomb was captured for up to 7 consecutive days at a frequency of 1 frame/s with a computer-controlled 29 megapixel monochrome machine vision camera under infrared light, invisible to the bees, and stored on hard drives for later processing. Barcode detection and decoding was performed on a compute cluster, using the software and procedures described in ref.<sup>8</sup>.

**Ground truth.** To create a ground truth for the trophallaxis detector, we manually annotated images of pairs of bees that were in the proper position to perform this behavior. For each image we recorded whether the two bees were engaged in trophallaxis and, if so, which bee was the recipient.

The first set of images consisted of the image library *L1* described in ref.<sup>8</sup>. These images show random bee pairs that were selected solely based on whether they were within reach. To determine if two bees are within reach, the coordinate of their mouthparts was estimated by translating the barcode center of each bee by a fixed distance in the direction of her barcode orientation vector, which was assumed to be parallel to her anteroposterior axis (Supplementary Fig. 1). If the distance between these coordinates was shorter than 7 mm, which is the maximum proboscis length of a honey bee<sup>31</sup>, the image was annotated. An image was considered to show trophallaxis if the proboscis of one bee touched the head of the other bee close to her mouthparts (Fig. 2b). Otherwise, it was annotated as not showing trophallaxis.

Examination of the images in library *L1* revealed that only approximately 1 in 40 images showed trophallaxis. Because CNNs learn to classify images better if the number of positive and negative examples is more balanced<sup>32</sup>, and manual annotation is labor intensive, additional images were selected by also requiring that bees needed to face each other, which is necessary for trophallaxis. Whether two bees faced each other was established similar to ref.<sup>8</sup>. We calculated the angle between the barcode orientation vector of each bee and a line through the mouthpart location of both bees (Supplementary Fig. 1). If the sum of the two resulting angles was less than 104 degrees, the image was annotated. This criterion captures 95% of the trophallaxis contacts in *L1*. Using this filter as a proxy for identifying trophallaxis, we annotated an additional 6045 random bee pair images.

To evaluate the performance of the trophallaxis detector when its predictions are integrated over time, we furthermore used the trophallaxis proxy to annotate all bee pairs in 600 random triples of successively recorded images of the entire observation hive. Each bee pair was annotated by three annotators to be able to reduce annotation errors. Consensus among the three annotators was reached through majority voting. A subset of the resulting annotations was previously published as image library *L2*<sup>8</sup>.

The ground truth for the egg-laying detector was created by manually annotating 1323 random images of the entire hive. Bees that either had inserted their abdomen into a honeycomb cell (Supplementary Fig. 6) or appeared to be in the process of doing so were called egg-layers. All other bees were annotated as having laid no egg. In addition, we annotated each egg layer in up to two images that were recorded before and after these hive images.

The final trophallaxis and egg-laying ground truths consisted of 142,182 images of bee pairs and 723,995 images of individual bees, respectively. Both ground truths were split into disjunct training, calibration, and test data sets as shown in Supplementary Table 4.

**Region proposal.** To identify image regions that are likely to show bees engaged in trophallaxis, we leveraged the trophallaxis proxy described earlier (Supplementary Fig. 1). For each pair of potential trophallaxis partners identified by this proxy, we extracted the region showing the heads of both bees by translating the barcode center of each bee by a fixed distance in the direction of her barcode orientation vector. The midpoint of the line segment defined by these two coordinates was used as the center point of a 96 px × 160 px region of which the longer sides were parallel to the aforementioned line segment (Fig. 2b). The top edge of this rectangle was defined to be the short edge closest to the head of the bee with the bigger ID.

For detecting egg-layers, the image region focusing on the abdomen of a bee was extracted by translating the center of a bee's barcode by a fixed distance in the opposite direction of her barcode orientation vector. The resulting coordinate was used as the midpoint of one edge of a 130 px × 130 px region of which the top edge was parallel and closest to the lower edge of the bee's barcode (Supplementary Fig. 6). The image region showing the entire bee as well as her immediate surroundings was obtained by first translating the center of her barcode by a fixed distance in the direction of the barcode orientation vector. The resulting coordinate was then used as the top-right corner of a 256 px × 256 px region of which the diagonal between the top-right corner and the bottom-left corner passed through the barcode center (Supplementary Fig. 6).

**Image preprocessing.** CNN input images were created by extracting the proposed image region and rotating it “upright”, so its top edge was on the x-axis of the image coordinate system. Pixel intensities were then clamped at a value of 200 to remove bright details in the background that we did not expect to provide information about the behavior of interest, such as the honeycomb structure and specular reflections in the honeycomb cell contents (Fig. 2c). For trophallaxis detection, we furthermore filled the bounding box of the focal bees' barcode with a uniform color to prevent the CNN from associating parts of the barcode pattern with a behavior (Fig. 2c). Finally, pixel intensities were mean-centered and scaled to [-1, 1] to provide a consistent input range for the CNNs.

**CNN architecture and training.** Images of potential trophallaxis partners and egg-layers were classified with two CNNs each. While the exact details of these CNNs varied, they had a similar architecture, consisting of 2 or 3 convolutional layers, max-pooling layers, and 2 fully connected layers (Supplementary Tables 1, 2, and 3). The output of convolutional layers was standardized with batch normalization before being passed to a rectified linear unit activation function, and the output of the final activation function was transformed by a softmax function, so it can be interpreted as the probability of the input image showing the behavior of interest.

CNN training consisted of initializing the weights of the network to values drawn from a normal distribution with a mean of 0 and a standard deviation of 1, truncated at 2 standard deviations. Network weights were then optimized for 10,000 iterations, using the Adam algorithm, which was configured as recommended by its authors (alpha = 0.001, beta1 = 0.9, beta2 = 0.999, epsilon = 0.0000001)<sup>33</sup>, with a cross-entropy loss function. We

used batches of 256 images, which were augmented as shown in Supplementary Table 5 and normalized to mean zero and unit variance. To avoid overfitting, we applied a L2 weight decay of 0.005 and a dropout of 0.5 in the first fully connected layer.

The CNNs for detecting the occurrence of trophallaxis and whether a bee's abdomen was inserted into a honeycomb cell were trained on all positive examples and a matching number of random negative examples from the trophallaxis and egg-laying training data set, respectively. For training the CNN that identifies the recipient, we used all positive examples of the trophallaxis training data set and a corresponding number of random negative examples. The CNN that uses an image of the entire bee to identify false positives was trained on the 4044 false positives generated by the CNN that checks for the visual absence of a bee's abdomen and a matching number of random positive examples from the egg-laying training data set.

**Behavior detection.** To detect trophallaxis or egg-laying in individual images, we first performed the corresponding region proposal to obtain image regions that potentially show the behavior of interest. Each of these regions was extracted, preprocessed as described above, and independently scored by the two CNNs that had been trained to detect the behavior. This approach is computationally less efficient than a strictly serial operation where, for example, the CNN for identifying the recipient processes the input image only if the score for the occurrence of trophallaxis exceeds a specified threshold. It was nevertheless chosen because it makes it possible to store both CNN scores in a file, which can later be post-processed to yield behavior predictions of varying stringency without having to process the video again.

Successive per-image trophallaxis and egg-laying detections were thresholded and linked together to yield behavior predictions that are easier to analyze. Linked predictions were filtered to improve their quality. For trophallaxis, this procedure is identical to the post-processing steps described in ref.<sup>8</sup>. Briefly, detections between the same two individuals were concatenated if they occurred in successive video frames. Concatenated detections shorter than 3 s were discarded, because such short trophallaxis interactions may not always result in the transfer of liquid<sup>34</sup>. The remaining interactions were merged if they were less than 60 s apart, and discarded if their total duration exceeded 180 s. The former threshold is based on the personal observation that trophallaxis partners typically do not engage in trophallaxis again during this time period, while the latter threshold was chosen to be slightly longer than the longest trophallaxis interaction reported in ref.<sup>35</sup>.

For egg-laying, linking and filtering consisted of concatenating the thresholded detections from successive video frames into egg-laying events. Concatenated egg-laying events shorter than 3 s were discarded. The remaining events were merged if they involved the same bee, occurred within 10 s of each other, and the distance of the average position of the events was shorter than 11.2 mm (the width of two honeycomb cells). These conditions ensured that egg-laying predictions were only merged if they appeared to belong to the same real event.

**Detector calibration.** The trophallaxis detector and egg-laying detectors each have free parameters. For the trophallaxis detector, these parameters are the minimum and maximum distance between a pair of bees, the maximum sum of the angle between the bees' orientation vector and a line through their estimated mouthparts location, a threshold for the output of the CNN that determines which scores correspond to the occurrence of trophallaxis, and a threshold for the output of the CNN that identifies the recipient. For the egg-laying detector, free parameters are the CNN thresholds that determine which output scores identify a (potential) egg layer.

To fix the free parameters of each behavior detector, we applied it to the calibration data set of its ground truth. We then performed a grid search on the parameter space of the detector, and chose the parameter combination that maximized the product of the detector's sensitivity and positive predictive value.

**Detector evaluation.** We estimated the fraction of an image that the trophallaxis detector's region proposal procedure excludes from automatic visual inspection by applying it to the "center" image of all 300 observation hive image triples from which the detector's test data set was created. We then added the number of pixels inside the proposed image regions and divided the total by the number of pixels per image and by the number of observation hive images.

To test if the CNNs for detecting the occurrence of trophallaxis and for identifying the recipient are sensitive to salient features of the behavior, we performed an occlusion sensitivity analysis<sup>36</sup>. We systematically moved a 41 px × 41 px big grey square (occluder) over the CNN input image, restricting the occluder center to pixels inside the image. For each occluder position, we processed the altered input image with the respective CNN and recorded the CNN's output. CNN outputs were then spatially coarse-grained into 12 × 20 square bins by averaging over all outputs inside a bin. Bins with a low mean score represent occluder positions that lead to a misclassification of the altered input image.

Detector performance was evaluated by applying each detector to its respective test data set, using the parameters we had obtained during detector calibration. This results in an estimate of a detector's performance on images of the entire hive, since the prevalence of trophallaxis and egg-laying in the respective test data sets is the same as in hive images. These performance estimates are conservative, because the test data sets consist of 3 s long segments of behavior occurrences that likely lasted longer. Longer behavior occurrences consist of multiple such segments and are therefore more likely to be detected. Moreover, due to the high positive predictive value of both detectors on individual images, the probability of a spurious detection decreases sharply as the duration of the detected behavior increases.

Image processing time was measured by averaging detector runtime across the 900 (545) hive images from which the trophallaxis (egg-laying) test data set was created. These measurements were performed on a cluster with 2.9 GHz Intel Core i9-7920X CPUs and a RAID 6 storage array consisting of 6 HGST

H3IKNAS800012872SWW hard drives. During runtime measurements, both detectors were restricted to a single hardware thread and had access to 2 GB RAM.

**Trophallaxis networks.** We constructed one temporal network from the trophallaxis detections of each colony. Nodes in these networks represent one bee. Pairs of distinct nodes were connected with a directed edge if the corresponding bees exchanged liquid at least once. A list of elapsed times, counting from the beginning of the experiment, was assigned to each edge to specify when the liquid transfers were initiated. These contact initiation times had a resolution of 1 s and enabled our spreading simulations to maintain the precise time-order of transfer events.

**Spreading simulations.** To simulate the transmission of information, pathogens, and liquid, we employed a temporally explicit version of the deterministic susceptible-infected model<sup>37</sup>. This model assumes that individuals are in one of two states, “susceptible” or “infected.” Simulations begin by setting all bees to susceptible, choosing a trophallaxis interaction uniformly at random, and “infecting” the two bees involved in this interaction, independent of how long they had interacted. Spreading dynamics were then simulated over an 8 h time window. During this time, an infected donor infects a susceptible recipient with a probability of 1 when they engage in trophallaxis (directed transmission). For models of undirected transmission, the trophallactic role of individuals was ignored, which means recipients were also able to infect donors.

For each simulation, we recorded the fraction of infected individuals  $f(t) = i(t)/S(t)$ , where  $i(t)$  is the number of infected bees alive at time  $t$  after the first infection, and  $S(t)$  is the colony size at time  $t$ . To obtain a more robust estimate of the fraction of infected individuals, we averaged it over  $R = 1000$  simulation runs and calculated the prevalence  $p(t) = \frac{\sum_{r=1}^R f_r(t)}{R}$ .

To establish whether the prevalence for an observed interaction sequence,  $p(t)$ , is greater than expected by chance, we compared it to the prevalence for  $N = 5$  randomized interaction sequences. For undirected spreading simulations, randomized interaction sequences were created with the PTN null model<sup>38</sup>, which shuffles the contact times among the observed interactions while ensuring that no individual gets assigned an interaction after its time of death. For directed spreading simulations, interaction sequences were randomized by reversing the direction of trophallaxis with a probability of 0.5 before the PTN null model was applied. To separately study the effect of the directional nature and of the temporal structure of trophallaxis on directed spreading, we performed only one of these randomizations, respectively. Spreading dynamics were characterized by calculating the spreading speedup  $s(t) = (\hat{p}(t) - \bar{p}(t)) / \min(\hat{p}(t), \bar{p}(t))$ , where  $\bar{p}(t) = \sum_{n=1}^N \frac{p_n(t)}{N}$ , is the mean prevalence across the  $N$  randomized interaction sequences.

### Data availability

The datasets generated and analyzed during the current study are available in Zenodo, <https://doi.org/10.5281/zenodo.5582779>.

### Code availability

Computer code for detecting trophallaxis and egg-laying is available in Zenodo, <https://doi.org/10.5281/zenodo.5584131>.

Received: 27 September 2022; Accepted: 20 December 2022

Published online: 27 January 2023

### References

1. Wario, F., Wild, B., Couvillon, M. J., Rojas, R. & Landgraf, T. Automatic methods for long-term tracking and the detection and decoding of communication dances in honeybees. *Front. Ecol. Evol.* **3**, 1–14 (2015).
2. Crall, J. D., Gravish, N., Mountcastle, A. M. & Combes, S. A. BEEtag: A low-cost, image-based tracking system for the study of animal behavior and locomotion. *PLoS ONE* **10**, 1–13 (2015).
3. Alarcón-Nieto, G. *et al.* An automated barcode tracking system for behavioural studies in birds. *Methods Ecol. Evol.* **9**, 1536–1547 (2018).
4. Mersch, D. P., Crespi, A. & Keller, L. Tracking individuals shows spatial fidelity is a key regulator of ant social organization. *Science* **340**, 1090–1093 (2013).
5. Crall, J. D. *et al.* Neonicotinoid exposure disrupts bumblebee nest behavior, social networks, and thermoregulation. *Science* **362**, 683–686 (2018).
6. Stroeymeyt, N. *et al.* Social network plasticity decreases disease transmission in a eusocial insect. *Science* **362**, 941–945 (2018).
7. Crall, J. D. *et al.* Spatial fidelity of workers predicts collective response to disturbance in a social insect. *Nat. Commun.* **9**, 1201 (2018).
8. Gernat, T. *et al.* Automated monitoring of behavior reveals bursty interaction patterns and rapid spreading dynamics in honeybee social networks. *Proc. Natl. Acad. Sci.* **115**, 1433–1438 (2018).
9. Blut, C. *et al.* Automated computer-based detection of encounter behaviours in groups of honeybees. *Sci. Rep.* **7**, 17663 (2017).
10. Krause, J. *et al.* Reality mining of animal social systems. *Trends Ecol. Evol.* **28**, 541–551 (2013).
11. LeCun, Y., Bengio, Y. & Hinton, G. E. Deep learning. *Nature* **521**, 436–444 (2015).
12. Romero-Ferrero, F., Bergomi, M. G., Hinz, R. C., Heras, F. J. H. & de Polavieja, G. G. Idtracker.Ai: Tracking all individuals in small or large collectives of unmarked animals. *Nat. Methods* **16**, 179–182 (2019).
13. Bozek, K., Hebert, L., Mikheyev, A. S. & Stephens, G. J. Towards dense object tracking in a 2D honeybee hive. In *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* 4185–4193 (2018) doi: <https://doi.org/10.1109/CVPR.2018.00440>.
14. Gal, A., Saragosti, J. & Kronauer, D. J. C. anTraX, a software package for high-throughput video tracking of color-tagged insects. *Elife* **9**, 1–32 (2020).

15. Mathis, A. *et al.* DeepLabCut: Markerless pose estimation of user-defined body parts with deep learning. *Nat. Neurosci.* **21**, 1281–1289 (2018).
16. Pereira, T. D. *et al.* Fast animal pose estimation using deep neural networks. *Nat. Methods* **16**, 117–125 (2019).
17. Pereira, T. D. *et al.* SLEAP: Multi-animal pose tracking. Preprint at *bioRxiv* 276246 (2020).
18. Graving, J. M. *et al.* Deepposekit, a software toolkit for fast and robust animal pose estimation using deep learning. *Elife* **8**, 1–42 (2019).
19. Bohnslav, J. P. *et al.* DeepEthogram: A machine learning pipeline for supervised behavior classification from raw pixels. Preprint at *bioRxiv* 312504 (2020).
20. Carreira, J. & Zisserman, A. Quo vadis, action recognition? A New model and the kinetics dataset. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (eds Carreira, J. & Zisserman, A.) 4724–4733 (IEEE, 2017). <https://doi.org/10.1109/CVPR.2017.502>.
21. Norouzzadeh, M. S. *et al.* Automatically identifying, counting and describing wild animals in camera-trap images with deep learning. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E5716–E5725 (2018).
22. Zou, Z., Shi, Z., Guo, Y. & Ye, J. Object detection in 20 Years: A survey 1–39 (2019).
23. Wild, B. *et al.* Social networks predict the life and death of honey bees. *Nat. Commun.* **12**, 1–12 (2021).
24. Baltiansky, L., Sarafian-Tamam, E., Greenwald, E. & Feinerman, O. Dual-fluorescence imaging and automated trophallaxis detection for studying multi-nutrient regulation in superorganisms. *Methods Ecol. Evol.* **12**, 1441–1457 (2021).
25. Winston, M. L. *The Biology of the Honey Bee* (Harvard University Press, 1991).
26. LeBoeuf, A. C. *et al.* Oral transfer of chemical cues, growth proteins and hormones in social insects. *Elife* **5**, e20375 (2016).
27. Jones, B. M. *et al.* Individual differences in honey bee behavior enabled by plasticity in brain gene regulatory networks. *Elife* **9**, e62850 (2020).
28. Farina, W. M., Grüter, C., Acosta, L. & Mc Cabe, S. Honeybees learn floral odors while receiving nectar from foragers within the hive. *Naturwissenschaften* **94**, 55–60 (2007).
29. Seeley, T. D. Social foraging by honeybees: How colonies allocate foragers among patches of flowers. *Behav. Ecol. Sociobiol.* **19**, 343–354 (1986).
30. Geffre, A. C. *et al.* Honey bee virus causes context-dependent changes in host social behavior. *Proc. Natl. Acad. Sci.* **117**, 10406–10413 (2020).
31. Waddington, K. D. & Herbst, L. H. Body size and the functional length of the proboscis of honey bees. *Fla. Entomol.* **70**, 124–128 (1987).
32. Buda, M., Maki, A. & Mazurowski, M. A. A systematic study of the class imbalance problem in convolutional neural networks. *Neural Netw.* **106**, 249–259 (2018).
33. Kingma, D. P. & Ba, J. L. Adam: A method for stochastic optimization. *3rd Int. Conf. Learn. Represent. ICLR 2015—Conf. Track Proc.* 1–15 (2015).
34. Farina, W. M. & Wainselboim, A. J. Thermographic recordings show that honeybees may receive nectar from foragers even during short trophallactic contacts. *Insectes Soc.* **48**, 360–362 (2001).
35. Allen, M. D. The honeybee queen and her attendants. *Anim. Behav.* **8**, 201–208 (1960).
36. Zeiler, M. D. & Fergus, R. Visualizing and understanding convolutional networks. In *Computer Vision—ECCV 2014. Lecture Notes in Computer Science* Vol. 8689 (eds Fleet, D. *et al.*) 818–833 (Springer, 2014).
37. Anderson, R. M. & May, R. M. Infectious diseases of humans: Dynamics and control. In *Epidemiology and Infection* (eds Anderson, R. M. & May, R. M.) (Oxford University Press, 1992).
38. Li, M., Rao, V. D., Gernat, T. & Dankowicz, H. Lifetime-preserving reference models for characterizing spreading dynamics on temporal networks. *Sci. Rep.* **8**, 709 (2018).

## Acknowledgements

We thank the University of Illinois School of Life Sciences Machine Shop for constructing bee tracking equipment; Reliance Label Solutions for printing bCodes; and the Carl R. Woese Institute for Genomic Biology Computer Network Resource Group for computational support. We are grateful to J. Peng for providing computational resources. We thank J. Cullum, S. Bransley, and A. Ray for manual image annotation; T. L. Harrison and A. L. Sankey for bee management; and A. R. Hamilton, A. Ray, S. Bransley, J. Cullum, K. Wilk, J. Falk, A. Zhang, L. Block, and V. Bagchi for assistance with field work. We also thank members of the M.M. and G.E.R. laboratories for discussions; and members of the M.M. and G.E.R. laboratories, and Jan Aerts for comments that improved the manuscript. This material is based on work supported by the National Academies Keck Futures Initiative Grant NAKFI CB4 (to T.G.), a Christopher Family Foundation Grant (to G.E.R.), National Institutes of Health Grant R01GM117467 (to G.E.R. and N. Goldenfeld), and Defense Advanced Research Projects Agency Grant HR0011-16-2-0019 (to G.E.R. and H. Zhao).

## Author contributions

T.G., T.J., B.M.J., and G.E.R. conceived the idea. T.G., T.J., and B.M.J. created ground truths. T.G. and T.J. designed and trained CNNs. T.G., T.J., and M.M. designed behavior detectors. T.J. implemented the software. M.M. provided computational resources. T.G., B.M.J., and G.E.R. designed the experiments. T.G. and B.M.J. performed experiments. T.G. analyzed experimental data. T.G. and G.E.R. wrote the manuscript, with input from all coauthors. M.M. and G.E.R. supervised the project.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-26825-4>.

**Correspondence** and requests for materials should be addressed to G.E.R.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply 2023