Supporting Information

Gernat et al. 10.1073/pnas.1713568115

SI Materials and Methods

Experiments.

Animals. Bees were obtained from source colonies maintained according to standard techniques at the University of Illinois Bee Research Facility in Urbana, Illinois. Source colonies represented a mixture of European subspecies, predominantly *Apis mellifera ligustica*, and were each headed by a naturally mated queen. To obtain 1-d-old adult worker bees, we removed honeycomb frames containing sealed pupae from source colonies, placed them in a Styrofoam box, and transferred them to an incubator where they were kept in specialized emergence cages at 34 °C and 50% relative humidity.

Colony setup. Colonies were established with 1,200 1-d-old worker bees and one unrelated, naturally mated queen. Previous scientific studies successfully employed such colonies that were either of a similar size (48) or were much smaller (49). Each colony was provided with the same amount of honey and artificial "bee bread" (a mixture of 10 parts pollen, 9 parts honey, and 1 part water). The top 16 rows of honeycomb cells were provisioned with 150 g of honey, and the two rows below the honey cells were provisioned with 15 g of artificial bee bread. To help minimize food deposits on the hive window, we filled cells to only about one-half of their depth.

Workers were barcoded in two groups of ∼700 bees each. Group 1 emerged over 9 h at night before the day the colony was set up. Group 2 emerged over 9 h in the morning of the day the colony was established. Both groups emerged from the same brood frames. Workers were brushed into a plastic box with walls that were painted with Insect-A-Slip (BioQuip Products), which prevents bees from crawling out of the box. Bees were stored in an incubator until after barcoding and were fed honey ad libitum throughout the day.

To attach a barcode to a bee, we first anesthetized her on ice. She was then picked up with a pair of blunt featherweight forceps (BioQuip Products), and a small drop of Loctite Super Glue Gel Control (Henkel) was applied to her thorax. Next, a bCode was picked up with a moistened Plain Wood Applicator (Fisher Scientific) and centered on the thorax. The bCode was positioned so that its orientation vector was parallel to the anteroposterior axis of the bee. We always attached a bCode in this way to enable us to infer the bee's approximate orientation from the orientation of the bCode. Marked bees were placed individually in a Petri dish until they recovered. The entire barcoding procedure for a single bee lasted 1–2 min.

Both groups of barcoded bees were simultaneously moved into the observation hive. Bees were examined for signs of deformed wing virus infection, *Varroa destructor* mite presence, and glue residue on the bCode, and discarded if necessary. An equal number of bees from each group were individually transferred into the observation hive through a Plexiglas window with resalable openings. Next, the queen was barcoded and added to the colony. Barcoding the queen proceeded in a similar manner to barcoding a worker bee, except that she was anesthetized with CO_2. After the colony was established, we exchanged the Plexiglas window for a glass window without openings.

The hive was placed into the observation hive room and temporarily warmed to ∼37 °C to encourage the bees to spread out. Bees were allowed to acclimate until sunrise on the next day. At sunrise we started recording images. At this point workers were on average 24 h old.

Colony maintenance. Over the course of a day, the hive window acquired dirt from bees touching it with their bodies or walking on it. We therefore changed the window daily in the morning and at night to ensure a high detection rate throughout the experiment.

Forager depletion. After sundown on the second or third day of the experiment, we removed the gate that prevented bees from accessing the entrance ramp (Table S1). This allowed workers to exit the hive and to begin foraging.

On the fifth day after opening the hive entrance, we removed as many foragers as possible from the colony (Table S1). The times of the removal, originally scheduled for 8 AM until 8 PM, were adjusted according to the activity at the hive entrance. If no bees foraged, we began (ended) the removal later (earlier) (Table S1). During the removal, we alternated between collecting returning bees and passively observing the hive entrance. Collections were performed for 5 min with a portable vacuum cleaner (BioQuip Products). Bees were allowed to enter the hive during observations. If we noticed orientation flights, which are taken by bees before the onset of foraging to learn the location of their hive (50), we waited until 15 min after the cloud of orienting bees disappeared before resuming collections. No collections were performed if it rained. All collected bees were frozen.

Trials. We performed five separate trials in summer and autumn of 2013. All trials followed the same general experimental design. However, trials also differed, for example due to weather conditions and adjustments to the protocol. Most of the differences are listed in Table S1. Further differences are detailed below.

During trial 1, the hive window was changed only once per day. We experienced a camera outage of ∼30 min. In trial 2, one of the eight front lights did not work for the first 35 h of the experiment. However, since our computer vision algorithms are robust to changes in illumination, we expect this to have had a negligible effect on our results. In addition, we experienced two camera outages of ∼40 and 20 min. In trial 3, the brood frames did not yield enough bees for the second group of bees. We therefore supplemented this group with a small number of bees that were at most 3 h younger than intended. In trial 4, bees clustered on the honeycomb during the first 12 h of the experiment rather than spreading out as expected. In trial 5, the white ceiling lights were on for at most 9 h during daytime on day 1. After sunset on the second day, we turned on the ventilation system to draw a weak air current through the hive entrance into the room to discourage bees from temporarily congregating in the entrance tube, a behavior that was also observed in other trials. The forager depletion was postponed because it rained (Table S1).

Barcodes.

Design. We developed a custom matrix barcode, called "bCode," to enable a computer to automatically identify bees and track their position and orientation (Fig. S1). The bCode design is loosely based on the design of QR codes but has been tailored to suit our purposes with respect to physical size, data storage capacity, and readability. A bCode consists of small black squares that are arranged in a square lattice on a white background. The squares in the top-left, top-right, and bottom-left corner form patterns that are the same for all bCodes. Together with the white border, these fixed patterns constitute the bCode template. The finder pattern in the top-left corner is for automatically locating bCodes in a digital image. Each of the two smaller alignment patterns in the top-right and bottom-left corner marks a known point on the bCode surface. Together with the finder pattern, they allow for corrections to visual distortions resulting from a rotation of the bCode in three dimensions. They furthermore allow for determining the projected

bCode orientation, defined in terms of a vector in the plane of the image that is parallel to the left and right edge of the bCode.

Squares in the area of the bCode that are not occupied by the template represent 27 data bits. Of these, 11 bits encode an identifier, which can be used to distinguish a maximum of 2,048 bees. The remaining 16 bits are the parity symbols of a (4, 2) Reed–Solomon error-correcting code (51) with 8-bit symbols. The parity symbols are calculated from the identifier after converting it to a 16-bit word. They make it possible to automatically detect if the identifier was properly decoded and to correct certain decoding errors, ensuring that a bCode may be read correctly even if a portion is damaged, dirty, or otherwise hidden.

Manufacture. Printable bCode images had a side length of 66 dots. We anticipated that color would bleed from black squares into white squares during printing and therefore drew white squares slightly bigger than black squares. Additionally, for white squares that diagonally touch a black square, the corner dot of the black square was drawn in white. These two measures ensured appropriately sized squares with well-defined corners. bCodes were printed by Reliance Label Solutions at a resolution of 800 dpi on Food and Drug Administration-approved 70# C1S Litho paper (Avery Dennison Label and Packaging Materials) and protected with Matte Self Wound Polypropylene laminate (QSPAC). Printed bCodes were hand-cut on a wooden board using a razor blade and a steel ruler as a guide. Individual bCodes had a side length of 2.1 mm and weighed 0.6 mg, corresponding to 0.7% of the average mass of an adult worker honeybee, or 1.7% of the amount of honey a worker carries during swarming (17).

Read error rate. Pilot studies were conducted to verify that the identifier encoded on a bCode could be read reliably under experimental conditions. Verification was performed by manual visual comparison of the image of an actual bCode with the image of a computer-generated bCode, where the latter encoded the identifier our software had decoded from the actual bCode. To facilitate this comparison, the original bCode image was rotated "upright," magnified, and displayed side-by-side on a monitor with the computer-generated bCode image. If either the area encoding the identifier or the area encoding parity information of both bCodes were the same, then the original bCode was considered to have been read successfully. Obstructions of the original bCode counted as a difference if they were not mirrored by the computer-generated bCode image.

We performed this verification on 5,000 randomly selected bCode images from a 1-wk-long test trial. The setup used in this trial differed from the setup used for experiments in that it did not include a backlight, four deep-red L300-660W linear lights (Smart Vision Lights) illuminated the hive, and there was no filter on the lens. bCodes differed in that a white grid separated squares to ensure well-defined square edges and corners. We expect that these differences had a marginal effect on our error rate estimate for the bCodes used in the actual experiment. An equal number of bCode images were randomly assigned to two human evaluators. They determined that 98.58 ± 0.2% of the bCodes were read correctly by the computer program.

Detection rate. To obtain an estimate of the detection rate, we manually counted all bees that were not automatically identified in each of 79 images. These images were sampled every 3 h from trial 3. Of the bees with a bCode that was at least partially visible, 87 ± 2% were automatically detected. Out of all bees in an image, including bees with a bCode that was not visible, we automatically detected 54 ± 9%. Most of the bees with an invisible bCode were either inside a honeycomb cell (75 ± 9%) or walked on the hive window (23 ± 9%) with their bCode facing the comb.

Effect on bees. To test whether the barcoding procedure or the bCode itself affected survival, we set up five colonies for a pilot study. Each colony contained 999 1-d-old worker bees and one queen. Equal proportions of the worker bees were unmarked, marked with a glue dot on their thorax, or marked with a bCode. Honey was provided ad libitum, as well as varying amounts of either pollen or artificial bee bread. The colony entrance was kept closed for 2–3 d to ensure that the bees did not attempt to fly before being mature enough to do so (52). At the end of day 7, we killed the bees and counted the number of survivors in each group and the number of bees that lost their glue dot or bCode.

On average, survival of bees without a bCode, bees with a glue dot, and barcoded bees was 83.2%, 81%, and 78.2%, respectively. The difference in survival between the three groups was not statistically significant (Friedman test, $X^2 = 5.2$, $n = 5$, $P > 0.07$). For barcoded workers, the average daily mortality during the time the hive entrance was open was 4.5%. Mortality before opening the entrance was negligible for all three groups.

Casual observations of barcoded workers and workers without a bCode suggest that barcoded bees behaved normally. The overall activity level of barcoded bees and bees without a bCode appeared to be similar, and barcoded individuals performed all tasks, including but not limited to cell cleaning, brood care, attending the queen, undertaking, nectar processing, capping cells, guarding, and foraging. In casual observations, we noticed no apparent differences with respect to the frequency of antennal contacts and trophallaxis within or between barcoded workers and workers without a bCode.

Durability. By the evening of day 7 during the pilot studies with the five colonies, an average of 1.6% of the bees had lost their glue dot or bCode. Thus, at most 3.2% of the bCodes had fallen off in each test colony. bCodes that were still attached to bees showed no signs of wear. In an additional long-term experiment with a colony that initially contained 1,200 barcoded workers and 1 barcoded queen, scanning of some of the bCodes revealed no obvious deterioration of bCode material after 4 wk, except for a thin patina on the laminate of some bCodes that had no effect on bCode readability.

Tracking.

Image preprocessing. To increase the probability of detecting and correctly decoding bCodes, we preprocessed each camera image. Images were resized to 140% of their original resolution. Each resized image was sharpened by digital unsharp masking (53) with a radius of four pixels, a sigma of 2, an amount of 400%, and a threshold of 0. Parameters for image preprocessing were empirically determined.

Barcode detection. We detected potential bCode locations by first converting the preprocessed camera image to a binary representation. This was accomplished by setting all pixels below a threshold b to black and all other pixels to white. Pixel rows and columns of the binary image were then scanned for the characteristic finder pattern of bCodes—an alternating sequence of black-and-white pixels with a length ratio of 1:1:2:1:1. Deviations from this ratio of up to 50% were allowed. A pixel that was part of both a horizontal and a vertical finder pattern sequence marked a potential bCode location.

Next, the immediate image region around a potential bCode location was searched for alignment patterns. We detected alignment patterns using the same method as employed for the finder pattern. However, to find alignment patterns, we searched for a sequence of black, white, and black pixels with a length ratio of 1:2:1. If more than one alignment pattern was found, we used the finder pattern and each possible combination of two alignment patterns to determine the image area of candidate bCodes. If only one alignment pattern was detected, we used its location and the location of the finder pattern to estimate the two possible positions of the missing alignment pattern before determining candidate bCode regions.

Candidate bCode images were extracted from the binary image representation and information about the actual position and the expected position of the three fixed patterns was used to correct perspective distortion. We then determined the color of

the squares that make up a bCode. The color of the invariant squares was compared against the bCode template. If less than 85% of the template was preserved, we discarded the candidate bCode image. The white bCode border was excluded from this comparison as it was often badly conserved. If the bCode image was well conserved, we decoded and verified the identifier. Identifiers that neither passed error detection nor error correction were retained, but flagged as being misread.

To make bCode detection robust to inhomogeneous illumination, we repeated the procedure described above using empirically determined thresholds $b \in \{10\,g \mid 3 \leq g \leq 11\}$. bCodes that were detected at multiple thresholds in the same image were considered to belong to the same bee if the distance between the bCode centers was smaller than one-half the bCode side length. The images of these bCodes were grouped and the average location of the grouped bCodes and their most common identifier were recorded.

Tracking data postprocessing. Tracking data were postprocessed to correct tracking errors and to facilitate subsequent behavioral analyses. First, we deleted records of bCodes from which the identifier could not be decoded correctly. Second, the records of bCodes that were too big or too small to be real were removed. Third, we deleted the records of all bCodes with an identifier that occurred twice per time point. Fourth, records of bCodes that moved faster than 5 cm/s between successive time points were filtered out, as such fast movement exceeds the mean walking speed of a bee [3 cm/s (54)] and likely resulted from incorrectly decoded identifiers.

To further facilitate behavioral analysis, we performed the two additional postprocessing steps described below.

First, we removed all records of a bee after her time of death. This was necessary because dead bees may fall to the hive floor with their barcode facing the camera. We defined a bee's time of death as the last time when she was observed for at least 4 min during a 5-min window and was on average located above the third row of honeycomb cells counted from the floor of the hive.

Manual verification of the automatically determined times of death was carried out for all 2,047 bees detected in trial 3 (note that due to bCode read errors, the number of bees detected was greater than the actual colony size). A bee was considered alive if (*i*) her mean movement speed fluctuated and was low but not zero, (*ii*) her mean height above the hive floor fluctuated at a high frequency and she was mostly located above the third row of cells, and (*iii*) her mean detection rate fluctuated at a high frequency and went regularly beyond the 2,000 detections per h mark.

A comparison of automatically and manually predicted times of death showed that both times closely agreed with each other; only 0.6% of the bees were assigned a considerably different time of death. Note that incorrectly predicted times of death also include the failure to reject spurious "bees" that did not actually exist.

Second, we deleted all records of bees with a bCode that was detected much better or worse than the majority of the bCodes. A bCode may, for example, perform differently if it is permanently dirty. To identify such barcodes, we calculated for each bCode the fraction of interdetection durations lasting 1 s out of the interdetection durations lasting at most 5 s. We then used the interquartile range rule to determine outliers, which were deleted from the tracking data. This resulted in the removal of 43 ± 14.61 individuals per trial. The queen was excluded from this postprocessing step. Together, all postprocessing steps resulted in the removal of 5.9 ± 0.91% of the tracking records per trial.

Experimental Setup.

Observation hive. Bees were housed in a one-sided glass-walled observation hive with one honeycomb (Fig. S3). The hive frame was made from four wooden pieces. These pieces fit tightly around a white plastic honeycomb (Honey Super Cell) that was cut to a size of 348 × 232 mm to match the dimension of the camera sensor. We chose a plastic honeycomb rather than one made from natural beeswax because its white color provides a better contrast to the body of honeybees and therefore simplifies image processing. Additionally, the plastic honeycomb is perfectly straight and cannot be altered by the bees. This enabled us to precisely control the maximum distance between the hive window and the comb.

The honeycomb was covered by a glass window. The distance between the window and the honeycomb was 8 mm. This gap was wide enough for bees to perform all behaviors normally, yet narrow enough to ensure that bees could not crawl over each other and thus hide their barcodes from the camera. The window was secured with a set of clamps and seated in rails that were attached to the top and bottom bar of the frame, respectively. The rails allowed us to regularly exchange a dirty hive window for a clean one with minimal disturbance to the bees. Condensation on the window was prevented by ventilation holes between the rails and the hive frame.

The lower-right corner of the hive frame was cut out to allow bees to move from the honeycomb onto a small ramp that led to a 52 × 8-mm entrance tunnel that passed through the hive frame. The tunnel exit on the front of the hive frame was covered with a wire mesh to permit wind blowing into the tunnel to exit into the room instead of into the hive. The tunnel exit on the back of the hive frame was fitted with a round plastic connector that allowed us to attach the hive to a tube with an inner diameter of 52 mm. The inside of the connector contained two ramps that provided a gentle transition from the tube into the entrance tunnel.

Lights. Bees were illuminated with 850-nm infrared light, invisible to honeybees (32), from eight L300-850W linear lights (Smart Vision Lights). Two of these lights were attached on each side of a 550 × 430-mm aluminum frame, which was mounted on a tripod and positioned at a distance of 0.8 m in front of the hive. To further increase the contrast between the honeycomb and the bees, we also illuminated the back of the honeycomb with 850-nm infrared light from one SOBL-CBO-350x300–850 backlight (Smart Vision Lights).

In the first three trials, at least some lights were on throughout the trial (Table S1). During these trials the backlight was cooled by six IXP-34-16 Case Fans (SilenX Corporation) to reduce heat stress on the bees. Fans were not in contact with the hive. During the last two trials, all lights were controlled via a CTL-IO-4 I/O Module (Smart Vision Lights), which enabled the camera to briefly turn on the lights when it captured an image. During these trials, the backlight was directly attached to the back of the observation hive (Fig. S3), since it no longer required cooling.

Camera. Images were recorded with a monochrome Prosilica GX6600 machine vision camera (Allied Vision). The camera was fitted with a Nikkor AF 135 mm *f*/2 D DC prime lens (Nikon). The lens was covered by a B+W 72-mm IR Dark Red (092M) filter (Schneider Optics). This filter blocks visible light and thus ensured that images were not overexposed when the deep-red overhead light was turned on during manual observations. The camera was mounted on a tripod and placed 1.3 m in front of the hive so that the entire honeycomb was visible. At the camera's resolution of 6,576 × 4,384 pixels, bCodes had a side length of ∼38 pixels in the captured images.

Computer. The camera was connected to the Ethernet port of a standard personal computer with a 3.3-GHz Intel i3 processor and 4-GB RAM. This computer captured 1 image/s and saved it to a redundant array of independent disks with 16-TB storage space. At a frame rate of 1 image/s, this disk array provided enough storage space to continuously record for 26 d. Images were saved as JPEGs with a quality setting of 90, such that total file sizes for each 8- to 11-d trial were 4–6 TB.

Observation hive room. Experimental colonies were maintained in a dedicated room in the University of Illinois Bee Research Facility in Urbana, Illinois (Fig. 1*A*). In the west wall of this room was a 305-mm-long tube to which we connected the observation hive.

This enabled the bees to leave the hive to forage normally outdoors. Ants and other insects were prevented from entering the hive by isolating it with Tangle-Trap Adhesive (Tanglefoot) and a water barrier.

The room was completely dark during experiments, except for the duration of manual observations when we reflected the light of a deep-red L300-660W linear light (Smart Vision Lights), which bees cannot see well, off the wall opposite the observation hive. The door of the observation hive room was covered with an opaque black cloth to reduce light exposure when an observer entered the room.

To control hive temperature, we isolated the room from the air conditioning system of the building and used a space heater to maintain an average room temperature between 31.7 and 32.8 °C (Table S1). This resulted in an average hive temperature between 33.5 and 35.1 °C (Table S1). Hive temperature was measured with a U12-006 data logger (Onset Computer Corporation) through the center of the honeycomb. Room temperature was measured with a U12-011 data logger (Onset Computer Corporation) that was located next to the observation hive.

Trophallaxis Detection.

General procedure. To detect instances of trophallaxis in a single image, we first identified pairs of bees that were close and faced each other as potential trophallaxis partners. Each pair was then further examined to confirm that the two bees were indeed engaged in trophallaxis, using custom computer-vision algorithms. These algorithms first attempted to determine the exact shape, size, position, and orientation of the head of both bees. They then tested whether the heads were connected by a thin, dark shape that resembles a proboscis or antenna. If such a shape was found, we recorded a trophallaxis event. To reduce the number of false positives, trophallaxis predictions were integrated over multiple images during postprocessing. Details of our trophallaxis detection procedure are outlined below.

Potential partners. We defined potential trophallaxis partners geometrically as pairs of bees (i, j) that are within reach and face each other (Figs. 2 and 3*A*). To identify bees that are within reach, we calculated the distance $d_{i,j}$ between the most anterior point P on each bee's anteroposterior axis that is still on her body, assuming that the anteroposterior axis passes through the bCode center and is parallel to the bCode orientation vector o. Bees were classified as within reach if $D_{\min} \leq d_{i,j} \leq D_{\max}$ holds for threshold parameters $D_{\min}$ and $D_{\max}$. For pairs of bees (i, j) that passed this criterion, we calculated for bee i the angle γ_i between her bCode orientation vector o_i and the vector from point P_i to point P_j. For bee j, the angle γ_j between o_j and the vector from P_j to P_i was calculated. If $\gamma_i + \gamma_j \leq \Gamma_{\max}$ holds for threshold parameter $\Gamma_{\max}$, we classified the two bees as facing each other and considered them to be potential trophallaxis partners.

Head detection. As a first step to automatically confirming that potential trophallaxis partners were indeed engaged in trophallaxis, we determined the shape, size, position, and orientation of the head of both bees. Heads were modeled as a closed path composed of four quadratic Bézier curves $L(\delta)$, $R(\delta)$, $F(\delta)$, and $B(\delta)$, which represent the left eye, right eye, front of the head, and back of the head, respectively (Fig. S2*A*). The parameter δ was restricted to the range $0 \leq \delta \leq 1$, with the extremes denoting the end points of a curve. We randomly generated 40 candidate head models for each bee, defined by parameters describing the end points and curvature of the four Bézier curves, and used particle swarm optimization (55) to iteratively improve the match of these models to the image region estimated to contain a bee's head. After 50 iterations, the model that best matched the image was selected for each bee (Fig. S2*B*).

To quantify how well a head model matches the image, the particle swarm optimizer considered the three criteria described below.

First, we favored head models with dark pixels in the area of the model that includes the eyes and the vertex (top of the head). This area A was enclosed by curves $L(\delta)$, $B(\delta)$, $R(\delta)$, and line segments connecting points $F(0)$ and $F(1)$ to the head center P' (Fig. S2*A*). Point P' was located at the average position of the end points of $L(\delta)$ and $R(\delta)$. We thus excluded the area of the head model that is most variable between bees due to hair color differences. We then averaged the deviation over all pixel intensities a inside A from the expected pixel intensity G as $h = \text{mean}(\max(|a - G|, G))$. A head model is more likely to fit an actual head if h is closer to 0.

Second, we expect to see a light intensity transition between the honeycomb (white) and the eyes of a bee (black). To test whether the curves $L(\delta)$ and $R(\delta)$ were positioned over such a transition, we first approximated the image gradient by convolving the original image with a Sobel operator (56) that used 5×5 kernels (Fig. S2*C*). We then calculated the average gradient magnitude $0 \leq m \leq 1$ over all pixels on the curves $L(\delta)$ and $R(\delta)$. A head model is more likely to fit an actual head if m is big. This heuristic works even if the image background is dark, because the hairs between the individual lenses of a bee's compound eye reflect light, which creates a dim halo in front of the dark background.

Third, the image region around a bee's head typically contains multiple light intensity transitions. In addition to the intensity transition at the head, intensity transitions also surround, for example, the feet or antennae of the bee. To discourage the head model from being attracted to the gradients surrounding other body parts, we also considered the image gradient direction. We did so by drawing the curves $L(\delta)$ and $R(\delta)$ onto an otherwise empty image and estimated the image gradient of that image as before. Next, we calculated the absolute difference of this gradient's direction and the gradient direction of the original image over all pixels on the curves $L(\delta)$ and $R(\delta)$. A head model is more likely to fit an actual head if the mean $0 \leq \mu \leq \pi$ of these differences is small.

The overall match quality q of a head model to an image was calculated as $q = (h/G + m + \mu/\pi)/3$. Thus, a model fits the image well if the area containing the eyes and vertex is dark, if the image intensity changes strongly at the edge of the eyes, and if the direction of change of the image intensity agrees with the curves that model the eyes.

Refinement of the potential trophallaxis pair list. Head detection provides new information about the position and orientation of a bee's head. This information is more accurate than the estimates derived from the position and orientation of her bCode. We used this new information to refine the list of potential trophallaxis partners (i, j), employing the same procedure that was used to create this list. However, this time we used the head centers P_i' and P_j' instead of the most anterior points P_i and P_j, a vector from point $B(0.5)$ to point $F(0.5)$ (head orientation vector) instead of the bCode orientation vector when calculating γ, and thresholds $D'_{\min}$, $D'_{\max}$, and $\Gamma'_{\max}$ instead of $D_{\min}$, $D_{\max}$, and $\Gamma_{\max}$, respectively.

Touch detection. For each of the potential trophallaxis pairs that passed the refinement procedure, we checked whether their heads are connected by a dark, thin shape that resembles an antenna or proboscis. To simplify this task, we removed details in the bright areas of the image by setting all pixel intensities (an integer between 0 and 255) that were above the empirically determined threshold of 140 to that threshold (Fig. 3*B*). We then applied Bernsen's thresholding method (57) with circular windows, a radius of five pixels, and a contrast threshold of 30 to separate dark foreground pixels from bright background pixels (parameters were empirically determined) (Fig. 3*C*). Next, noise was reduced by promoting background pixels to foreground pixels if their von Neumann neighborhood consisted exclusively of foreground pixels, applying a median filter with a radius of 0.5 pixels, and repeating the background pixel promotion procedure.

To check whether two heads were connected by a thin shape, we first calculated for each foreground pixel the local thickness

(58), which corresponds to the diameter of the largest circle that contains the foreground pixel but no background pixels (Fig. 3*D*). The local thickness was used later to determine the thickness of connections between the two heads.

Next, we limited our search for a thin connection to an area between the heads of the potential interaction partners. This area was defined as the intersection of two half-disks around the head center points P_i' and P_j'. The radius of each half-disk was set to the maximum proboscis length Q reported in Waddington and Herbst (59). The diameter segment of each half-disk was oriented orthogonal to the head orientation vector. Foreground pixels outside the search area were discarded.

The remaining foreground pixels were skeletonized to obtain a thin representation of the image inside the search area (Fig. 3*E*). The skeleton was then simplified by demoting foreground pixels to background pixels if they had more than two foreground pixels in their von Neumann neighborhood. Skeleton pixels on the curves $L(\delta)$, $B(\delta)$, $R(\delta)$ of the head of one trophallaxis partner were designated as source pixels while skeleton pixels on the same curves of the head of the other potential trophallaxis partner were designated as target pixels. We then traversed the skeleton to find a nonintersecting path between a source pixel and a target pixel.

If a path was found, we checked its length, as well as the local thickness profile of the pixels on the path, to determine whether it could represent a proboscis or antenna. Paths needed to be shorter than the maximum proboscis length Q. The expected thickness profile allowed an arbitrary thickness for the first and last R pixels of a path. This accommodates a limitation of the head model, which does not fit the clypeus (part of the "face") or mandibles and labrum (mouthparts) well, if a bee tilts her head all of the way back. The middle section of the path was required to have a maximum thickness of S pixels. If a path fulfilled these criteria, we recorded a trophallaxis contact (Fig. 3*F*). If the path did not fulfill the criteria, we continued to traverse the image skeleton until a path either did fulfill them, or all possible paths had been examined.

Postprocessing. Trophallaxis contacts between the same two bees that were detected in successive images (i.e., 1 s apart) were combined into one trophallaxis interaction. Prior observations in beehives indicate that trophallaxis can range in duration from as little as 1 s to more than 2 1/2 min (60–62), but trophallaxis shorter than 3 s may not always result in the transfer of liquid food (60). We therefore discarded all trophallaxis interactions that were observed for less than 3 s to focus on the subset of detections where fluid was likely exchanged. Next, interactions were merged if it was likely that they were part of the same trophallaxis event. Interactions were merged if they involved the same two bees, were separated by less than 60 s, and if at each point in time between the two interactions to be merged at least one of the bees was invisible or both bees passed the first step of the trophallaxis detector (i.e., they are close and face each other). The latter conditions prevent interactions from being joined if there is evidence that the bees did not interact. It is possible that merging interactions in this way joins separate interactions. However, since it is unlikely that the same bees interact repeatedly within a short period of time, this is a suitable approach for dealing with short trophallaxis detection gaps. Merged interactions lasting longer than 3 min were discarded, because such long trophallaxis interactions are rare and their detection likely results from situations where, for example, bees are crowded in or sleep next to each other with antennae or legs crossed.

Parameterization. The procedure for geometrically identifying pairs of bees that might be engaged in trophallaxis has four parameters: the distance of the point P from the bCode center, and three thresholds for determining whether bees are close and face each other. We estimated the distance of P from the bCode center by manually measuring the distance between the bCode center and the point halfway between the bases of the antennae for multiple randomly selected bees from all trials. We then calculated the mean of these measurements and rounded it to the nearest pixel. The three remaining parameters were determined together with the parameters for refining the list of potential trophallaxis partners.

Head detection has one free parameter, the expected pixel intensity G. To estimate G, we manually sampled the pixel intensity (an integer between 0 and 255) of a random head pixel in multiple images from all trials, calculated the average intensity, and rounded it to the nearest integer.

The remaining parameters for creating the initial list of potential trophallaxis partners, as well as the parameters for refining this list after head detection, were determined through an optimization procedure that operated on a trophallaxis image library *L1*. To create this library, we randomly sampled from trial 3 images of pairs of bees from the set of all bees that are close enough to engage in trophallaxis, that is, $d_{i,j} < Q$. A single person then manually annotated each image as to whether the two focal bees performed trophallaxis. In total, the library *L1* contains 39,863 images. In 1,045 of these images, the focal bees engaged in trophallaxis.

To fix the parameters involved in creating and refining the list of potential trophallaxis partners, we first defined a truncated trophallaxis detector that operated on this list, but did not check whether the heads of potential interaction partners are actually connected by a dark, thin shape. We applied this truncated detector to the trophallaxis image library and estimated its quality as the product of its sensitivity and its positive predictive value. This process was repeated for different parameter combinations $\langle D_{\min}, D_{\max}, \Gamma_{\max}, D'_{\min}, D'_{\max}, \Gamma'_{\max}\rangle$, which we sampled uniformly from the extreme regions of the parameter space $\{d_{i,j}\colon (i,j) \in L1\} \times \{d_{i,j}\colon (i,j) \in L1\} \times \{\gamma_{i,j}\colon (i,j) \in L1\} \times \{d_{i,j}\colon (i,j) \in L1\} \times \{d_{i,j}\colon (i,j) \in L1\} \times \{\gamma_{i,j}\colon (i,j) \in L1\}$. The parameter combination that resulted in the highest quality score was used for the full trophallaxis detector that also includes touch detection.

Touch detection has two parameters, the maximum terminal thick segment length R and the maximum path thickness S. To estimate R, we manually measured the shortest distance from the border of a food recipient's head model to the tip of her mandibles. We repeated this procedure for multiple randomly selected recipients, calculated the mean, and rounded it to the nearest pixel. The parameter S was set to the maximum proboscis width measured in a random sample of recipients. In both cases, images of food recipients were obtained from the trophallaxis library *L1*.

Performance. To evaluate the performance of the trophallaxis detector, we created a second image library *L2*. This library was made by extracting images of all pairs of bees that are in the proper position for trophallaxis (i.e., $d_{i,j} < Q$ and $\gamma_i + \gamma_j \leq \Gamma_{L2}$) in 100 randomly chosen triples of successive whole-hive images. The additional condition $\gamma_i + \gamma_j \leq \Gamma_{L2}$ was introduced to reduce the number of pairs that needed to be annotated, and its parameter Γ_{L2} was set to capture 95% of the trophallaxis events in *L1*. Using triples of successive images enabled us to include in our performance estimates the effect of postprocessing, which aggregates trophallaxis detections over 3 s. To reduce the effect of potential biases, images for *L2* were annotated by a different person than the images for *L1*. In addition, *L2* images were sampled from a different trial (trial 2) than images for *L1*.

Evaluating our detector on *L2* demonstrates that it has a very high specificity and negative predictive value (Table S3). This result shows that most instances of bees not engaged in trophallaxis are correctly identified as such and that the detector generates only a very small number of false-negative predictions. The sensitivity indicates that our detector correctly identifies ~50% of the trophallaxis interactions (Table S3). Out of all trophallaxis detections, only 19% are false positives (Table S3).

Note that these estimates are conservative for trophallaxis interactions that are longer than 3 s, because these interactions consist of multiple 3-s-long segments.

The current state of the art for automatically detecting a head-to-head interaction between insects is to use information about their relative position and orientation. However, it is well known that spatiotemporal information alone may not be sufficient to distinguish a true social interaction from a colocalization in space and time (63, 64). We demonstrated that this is also the case for honeybee trophallaxis, using a truncated detector that only employs spatiotemporal information. Essentially, this truncated detector corresponds to the first stage of the full trophallaxis detector that identifies potential trophallaxis partners. As expected, the truncated detector did not predict trophallaxis reliably, mostly because it had a very low positive predictive value (Table S3). Our approach for detecting trophallaxis thus represents a substantial improvement over the current state of the art.

SI Text

Effect of Short Waiting Times on Spreading Dynamics. Deterministic SI spreading progresses along temporal paths that connect an infected node to an uninfected node and take the least time to traverse. This suggests that temporal correlations on adjacent edges, that is, short waiting times, could explain fast spreading. Since the probability of short waiting times is much higher in a trophallaxis network than in its temporally randomized counterparts (Fig. 4*A* and Fig. S4*A*), we hypothesized that spreading dynamics in the trophallaxis networks are driven by interactions that result in short waiting times.

To test this hypothesis, we removed all interactions from the trophallaxis networks of trial 1 that resulted in waiting times shorter than $W = 168$ s and repeated our spreading simulations. The value of the waiting time threshold W was set to the center of the last bin of the waiting time distribution $P(\tau)$ that had a higher probability in the empirical network than in the temporally randomized reference networks. To compensate for the longer average waiting time due to the removal of interactions, we required that the time $\hat{t}_k$ was greater than 10 min when calculating the prevalence $\check{p}$ at the end of the accelerated spreading period.

We find that, despite the in silico removal of 55% of the interactions, spreading continues to be accelerated in the empirical networks. The speedup of spreading was $s = 14.2\%$ and $\check{p} = 95.2\%$ of the bees were infected at the end of the period of accelerated spreading. This indicates that, while interactions resulting in short waiting times contribute to fast spreading, they are not the sole driver.

Effect of Mean Waiting Time on Spreading Dynamics. If a temporal network violates the ongoing link picture (65), applying the original RP null model (5) may result in a temporally randomized network with a higher mean waiting time than the original network. In principle, this increase could slow down simulated spreading on a temporally randomized network. Our modified version of the RP null model mitigates this issue by assigning individuals to interactions that occurred before their time of death. However, applying the modified RP null model may still result in an artificial increase of the mean waiting time if a randomized interaction involving individual i occurs before (after) the first (last) interaction of that individual.

For our datasets, the average difference of the mean waiting time in the temporally randomized reference networks and that same quantity in the trophallaxis network is 9.07 ± 4.27 s ($n = 5$). While this difference is small, it does, on average, apply to each waiting time of all individuals and could explain why, especially during the early spreading phase, the randomized network is slower.

To quantify how much the waiting time differences affect our results, we repeated our spreading simulations, using a version of the RP null model that preserves the mean waiting time for each bee. This was achieved by fixing the time $\dot{t}_i$ ($\underline{t}_i$) of the first (last) interaction of each individual i. In addition, interaction times $\theta_{i,j}$ and $\theta_{i',j'}$ of pairs of individuals (i, j) and (i', j') were only permuted if the condition $\dot{t}_{i'} \le \theta_{i,j} \le \underline{t}_{i'} \wedge \dot{t}_{j'} \le \theta_{i,j} \le \underline{t}_{j'} \wedge \dot{t}_i \le \theta_{i',j'} \le \underline{t}_i \wedge \dot{t}_j \le \theta_{i,j'} \le \underline{t}_j$ was satisfied.

We find that neither the speedup s and nor the grand mean prevalence at the end of the accelerated spreading period $\check{p}$ changes much if the mean waiting time is preserved during temporal randomization (Table S2). This result excludes the possibility that accelerated spreading is due to a lower mean waiting time in the temporally randomized networks.

Effect of False Positives and False Negatives in Trophallaxis Detection. To understand how the false-negative rate of the trophallaxis detector affects our conclusions about spreading on the trophallaxis network, we subsampled the interactions of trial 4 to create thinned versions of the original trophallaxis network, and repeated some of our analyses on these thinned networks.

Trophallaxis interactions were subsampled at sampling rates $\rho \in \{0.2, 0.4, 0.6, 0.8\}$. In each case, a fraction ρ of the original interactions was retained, which lowered the average interaction rate from λ in the original trophallaxis network to $\lambda_\rho = \rho\lambda$ in the subsampled networks. We then repeated the SI spreading simulations on each subsampled network and on the two ensembles of randomized reference networks. In contrast to our original simulations, spreading was simulated for 20 h rather than 10 h, because we expected spreading dynamics on the subsampled networks to be slower than on the original trophallaxis network due to the lower average interaction rate λ_ρ.

Fig. S6 shows that, even in the subsampled networks, simulated spreading is enhanced relative to temporally randomized networks and relative to temporally and topologically randomized networks. As anticipated, spreading dynamics in the subsampled networks are slower than in the original network. In particular, the time $t^{20\%}$ varies inversely with the sampling rate over all values ρ (Fig. S7*A*). Thus, the average interaction rate λ_ρ is the only quantity that sets the observed timescale for spreading on the networks.

If the mean prevalence is plotted as a function of dimensionless time $\lambda_\rho t$, which normalizes for the effect of the sampling rate ρ on spreading speed, then the prevalence curves for the empirical networks subsampled at sampling rates $\rho \ge 0.6$ collapse on top of each other (Fig. S7*B*). This observation suggests that we would obtain qualitatively similar results if the sensitivity of the trophallaxis predictor were lower than it actually is. For sampling rates $\rho \le 0.4$, the prevalence curves of the subsampled empirical networks are more similar to the curves of the randomized reference networks (Fig. S7*B*) and the speedup of spreading s decreases (Fig. S7*C*), likely because at these low subsampling rates the network structures responsible for accelerated spreading are destroyed.

Taken together, these results suggest that the results of our spreading simulations are robust to the false-negative rate of trophallaxis detection. Furthermore, it appears that the enhancement of spreading can be captured at false-negative rates that are even higher than those of the trophallaxis predictor.

False positives are spurious trophallaxis detections that may be generated when two bees are close and face each other (Fig. 2) for at least 3 s. The precise spatiotemporal pattern of spurious detections is unknown. However, as a first approximation, we can assume that they are generated at random times between individuals that fulfill the aforementioned criteria. Such random events can be expected to make an observed network more similar to its randomized counterpart. Because neither burstiness nor accelerated spreading manifests in the randomized reference networks, we expect that our quantitative estimate of these two network features represents a lower bound.

Interactions with the Queen. As the single female reproductive bee in a colony, the queen receives special attention from her many more or less sterile daughters. In addition to feeding her, worker bees frequently form a circle around the queen, examine her intensely with their antennae and forelegs, and lick her with their tongues (17). Our trophallaxis detector cannot distinguish these behaviors, which occur only briefly and relatively infrequently among worker bees, from trophallaxis and therefore generates many false-positive detections that involve the queen, resulting in a positive predictive value of only 0.11. We therefore excluded interactions with her from our analyses. However, to test whether our results change if these interactions are included, we repeated our analyses on networks that include all predicted interactions with the queen.

We find that our main results do not depend strongly on the presence of interactions with the queen. Trophallaxis continues to be bursty ($B = 0.33 \pm 0.04$, $n = 5$), and spreading dynamics are faster in the bee networks than in temporally randomized reference networks until most bees ($\check{p} = 80.6 \pm 6.5\%$, $n = 5$) are infected. The speedup of spreading s is $49.5 \pm 14.2\%$ ($n = 5$), and the difference between the trophallaxis network and its temporally randomized counterparts is statistically significant (conditional uniform graph test, $n = 100$, $P < 0.01$ for all trials).

Like in networks without the queen, network topology has a small but statistically significant effect on spreading dynamics (Mann–Whitney U test, $U \geq 89$, $n = 100$, $P < 0.04$ for all trials). However, in three trials, randomizing network topology leads to a slowdown ($-1.3 \pm 0.5\%$), and in two trials it leads to a speedup ($0.4 \pm 0.3\%$). This inconsistency is likely rooted in the fact that the majority of trophallaxis detections with the queen are false positives.

Finally, accelerated spreading continues to be resilient to a strong demographic perturbation. Simulated spreading through the empirical postremoval networks is faster than through temporally and topologically randomized reference networks (conditional uniform graph test, $n = 100$, $P < 0.01$ for all trials).

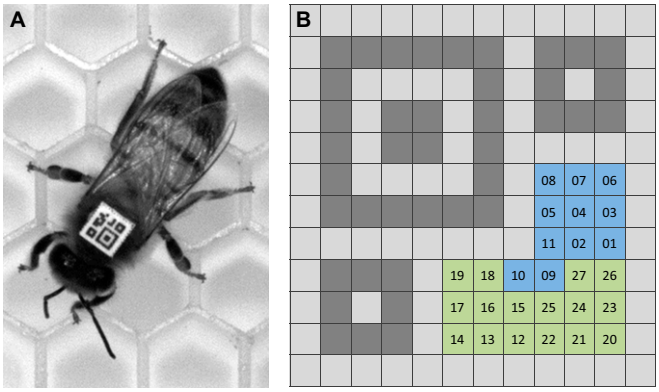

Fig. S1. Matrix barcode (bCode) for automatically identifying insects. (*A*) A barcoded honeybee. The bCode on this bee encodes ID 1405. (*B*) Schema of a bCode. Gray squares are the bCode template. Background squares are shown in light gray. The finder pattern and the two alignment patterns are shown in dark gray. Blue squares encode the identifier, and green squares encode parity symbols. Identifier and parity symbols are encoded as binary numbers, and squares labeled with bigger numbers represent higher-order bits of the identifier and the parity symbols. If a labeled square is black in an actual bCode, it represents a 1; if it is white, it represents a 0.

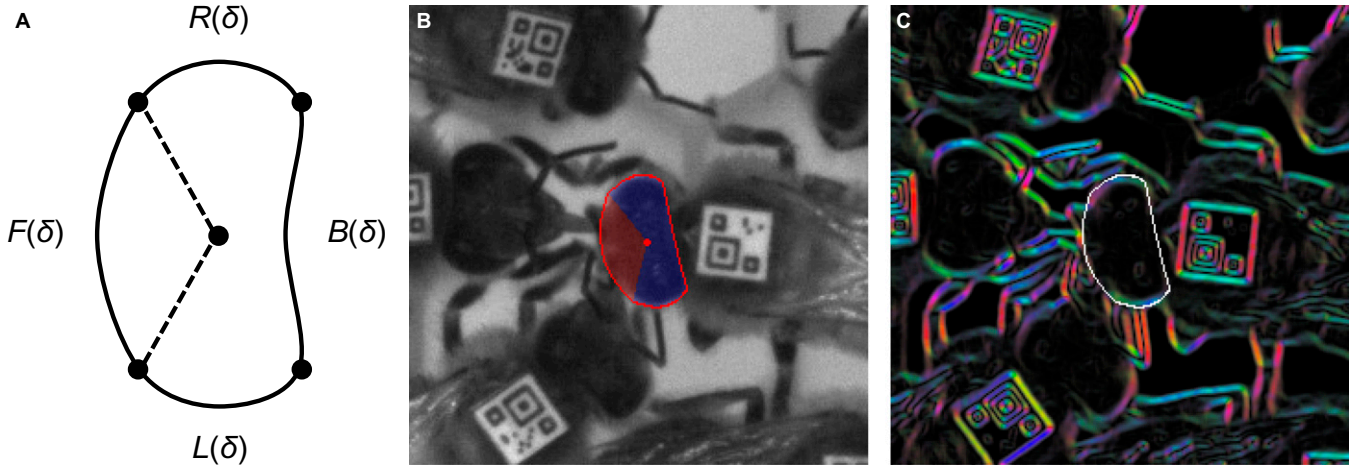

Fig. S2. Honeybee head detection. (*A*) Schema of the model used by the particle swarm optimizer to detect bee heads in a digital image. The model is composed of four quadratic Bézier curves, $L(\delta)$, $R(\delta)$, $F(\delta)$, and $B(\delta)$, that delineate the left eye, right eye, front of the head, and back of the head, respectively. The end points of these curves and the center of the head model are shown as black disks. $F(\delta)$ and the two dashed lines circumscribe the area expected to show the most differences between bees due to variation in hair color. (*B*) Result of applying the head detector to the image region estimated to contain the head of the center bee. (*C*) Image gradient of the image shown in *B*. The intensity and color of a pixel correspond to the strength and angle of the image gradient, respectively. The fitted head model is outlined in white.

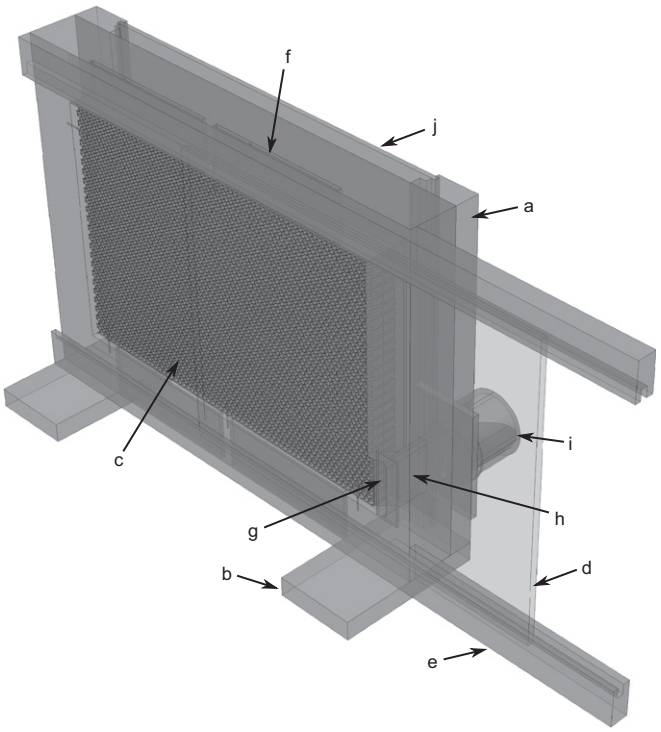

Fig. S3. Semitransparent view of the observation hive with honeycomb and attached backlight. Parts occurring multiple times were labeled only once for visual clarity. The observation hive consisted of a wooden frame (*a*), stabilized by two feet (*b*). The frame held a single plastic honeycomb (*c*). The honeycomb was covered by a glass window (*d*, partially slid to the right) that was supported by two rails (*e*). Holes (*f*) at the top and bottom of the hive frame provided ventilation. A short ramp (*g*) allowed bees to access the entrance tunnel (*h*) that opened out into a plastic connector with two ramps (*i*). An infrared LED backlight (*j*) increased the contrast between the plastic comb and bees.

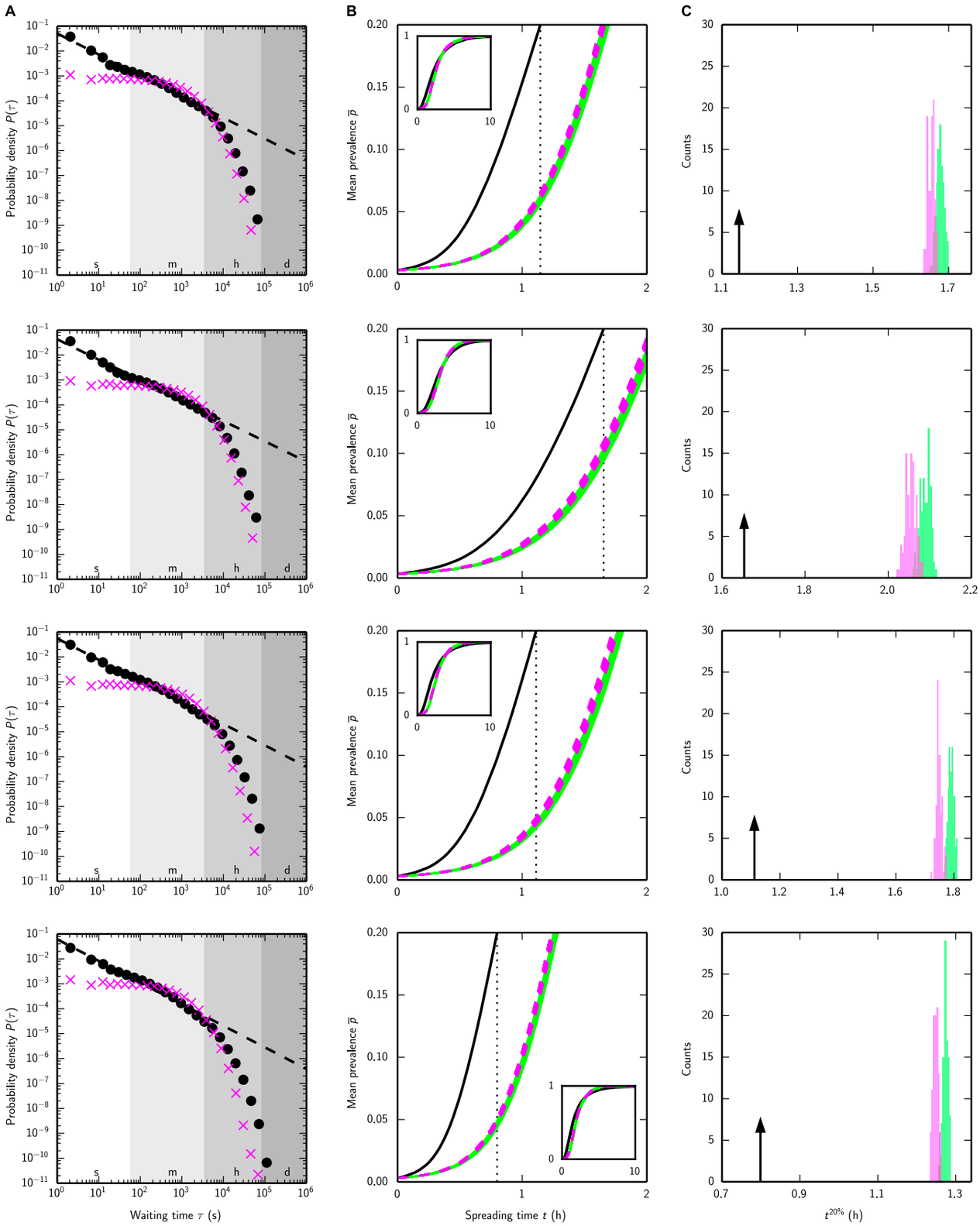

Fig. S4. Simulated spreading in honeybee trophallaxis networks is faster than in randomized reference networks, despite bursty interaction patterns. Rows correspond to trials 2–5. (*A*) Distribution of log-binned waiting times between interactions for the empirical network (black circles) and 100 temporally randomized reference networks (magenta crosses). Dashed line: power law fit to the empirical waiting times (see Table 1 for exponents of the fit). Lanes labeled s, m, h, and d denote seconds, minutes, hours, and days, respectively. (*B*) Mean fraction of bees infected via deterministic SI spreading (mean prevalence, controlled for mortality), averaged over 1,000 simulation runs, as a function of spreading time. Solid black line: empirical trophallaxis network; magenta dashed lines: 100 temporally randomized reference networks; green lines: 100 temporally and topologically randomized reference networks; dotted black line: time when the mean prevalence reaches 20% in the empirical network. (*Inset*) Mean prevalence as a function of spreading time until almost all bees have been infected. (*C*) Histogram of the mean time required to reach 20% prevalence ($t^{20\%}$) for the 100 temporally randomized reference networks (magenta) and the 100 temporally and topologically randomized reference networks (green). Arrow indicates when the prevalence reaches 20% in the empirical network.

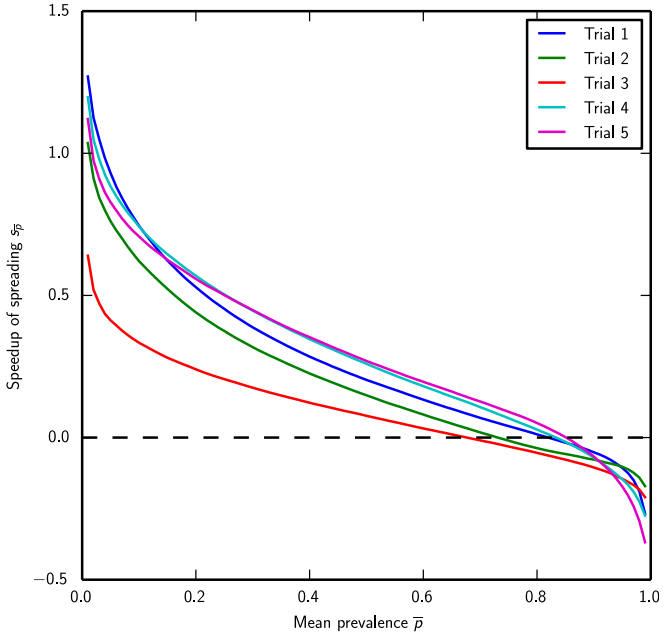

Fig. S5. Speedup of spreading $s_{\bar{p}}$ as a function of the mean prevalence $\bar{p}$, measured in the empirical network. The dashed black line highlights the speedup $s = 0$.

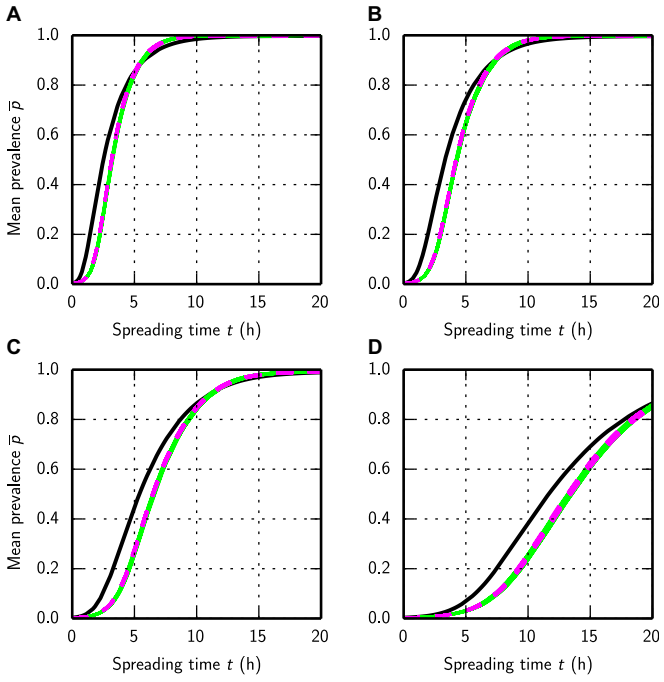

Fig. S6. Mean prevalence for subsampled empirical networks (black solid line), 100 temporally randomized networks (magenta dashed line), and 100 temporally and topologically randomized networks (green solid line). Simulated spreading dynamics in the randomized networks are similar enough that the prevalence curves for these networks appear to fall on top of each other. (*A*) Sampling rate $\rho = 0.8$. (*B*) $\rho = 0.6$. (*C*) $\rho = 0.4$. (*D*) $\rho = 0.2$.

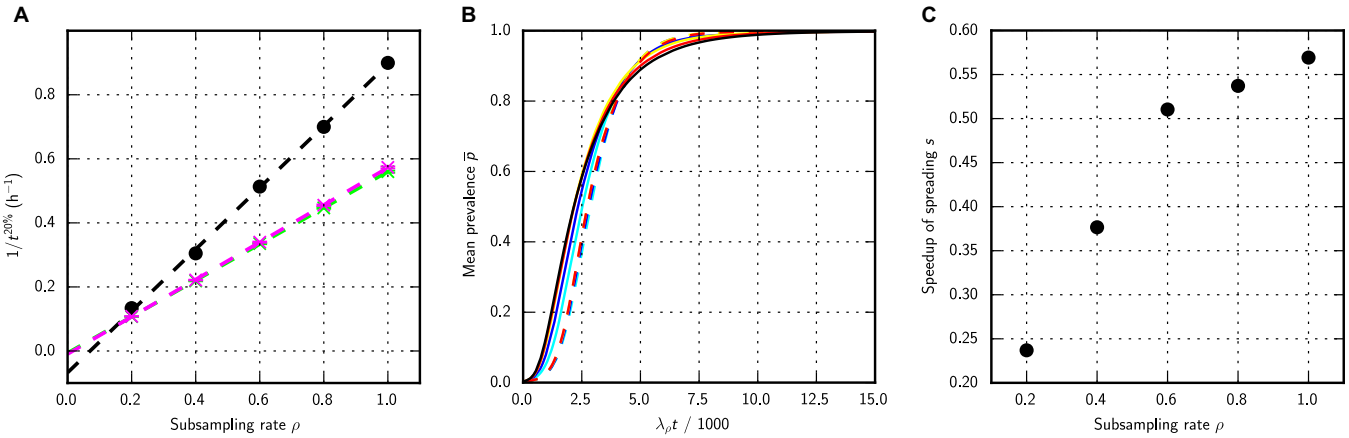

Fig. S7. Comparison of spreading dynamics in the original trophallaxis network and networks subsampled at sampling rates ρ. (*A*) Variation of the reciprocal time $1/t^{20\%}$ with the rate of subsampling ρ, for the empirical networks (black circles), the temporally randomized networks (magenta crosses), and the temporally and topologically randomized networks (green crosses). The dashed lines are to guide the eye. (*B*) Mean prevalence in empirical (solid) and temporally randomized (dashed) networks, at sampling rates $\rho = 1$ (black), $\rho = 0.8$ (red), $\rho = 0.6$ (yellow), $\rho = 0.4$ (blue), and $\rho = 0.2$ (turquoise) as a function of dimensionless time, namely, the spreading time t multiplied by the interaction rate λ_ρ for that subsampled network. The prevalence curves for the randomized networks approximately collapse on top of one another. The curves for empirical networks at sampling rates $\rho \geq 0.6$ fall on top of each other, while the curves for empirical networks at increasingly lower sampling rates become more similar to randomized networks. (*C*) The speedup of spreading s for different sampling rates ρ.

Table S1. Overview of trials

				Forager removal					Temperature, °C		
Trial	Begin date	End date	EO	Day	Begin time	End time	CS	BR, %	Hive	Room	LED lights always on
1	07/03	07/13	3	8	08:25	20:00	948	29	34.4 ± 0.2	31.7 ± 0.2	Yes
2	07/16	07/23	2	7	10:05	19:50	991	46	35.1 ± 0.4	32.1 ± 0.7	Yes
3	08/06	08/14	2	7	08:31	20:00	973	35	34.5 ± 0.4	31.8 ± 0.2	Backlight
4	09/16	09/25	2	7	12:01	18:18	995	13	33.5 ± 0.7	32.6 ± 0.6	No
5	09/30	10/09	2	9	11:56	18:39	892	21	33.5 ± 0.6	32.8 ± 0.5	No

BR, percent bees removed during the forager removal; CS, colony size at the time when the forager removal began; EO, number of days after which the hive entrance was opened.

Table S2. Preserving the mean waiting time in temporally randomized trophallaxis networks has a negligible effect on simulated spreading dynamics

Trial	$(s - \hat{s})/s$	$(\check{\bar{p}} - \hat{\bar{p}})/\check{\bar{p}}$
1	0.014	0.002
2	0.028	0.001
3	0.012	0.012
4	0.010	0.006
5	0.004	0.005

The normalized difference of the speedup of spreading with ($\hat{s}$) and without (s) preserving the mean waiting time and the normalized difference of the grand mean prevalence at the end of the period of accelerated spreading with ($\hat{\bar{p}}$) and without ($\check{\bar{p}}$) preserving the mean waiting time are shown for each trial.

Table S3. Trophallaxis detection performance for the full detector with computer vision and the truncated detector without computer vision

Performance measure	Full detector	Truncated detector
Sensitivity	0.53	0.73
Specificity	0.98	0.71
Positive predictive value	0.81	0.28
Negative predictive value	0.93	0.94

Detection performance was evaluated after integrating detections over three successive images.

Other Supporting Information Files

Dataset S1 (TXT)