Individuality in the hive

A Data-Driven Study of Individual and Collective Behavior

Dissertation

zur Erlangung des Grades
eines Doktors der Naturwissenschaften
(*doctor rerum naturalium*)

am Fachbereich Mathematik und Informatik
der Freien Universität Berlin

vorgelegt von
Benjamin Wild

Berlin, 2023

Erstgutachter	Prof. Dr. Tim Landgraf
Zweitgutachter	Prof. Dr. Michael L. Smith
Drittgutachter	Prof. Dr. Raúl Rojas González

Tag der Disputation	18.04.2024

Abstract

This thesis investigates the complex behavior of individuals in a group and the emergence of collective behavior through their interactions, using automatic tracking technology and novel machine learning methods. It employs honey bee colonies as a model system to address two key questions: how collective behavior arises in the absence of central organization, and how individual behaviors relate to collective behavior.

The study of emergent behavior in social animals poses a significant challenge, as it is difficult to understand how individual behaviors give rise to collective patterns. Social insect societies have a long history as model organisms in understanding these patterns, and honey bee colonies, in particular, offer a unique opportunity for such research. However, the large number of individuals in a colony, along with their continuous actions and interactions, make it impractical for human observers to study the entire colony comprehensively.

Building upon recent advancements in computer science, specifically in computer vision and machine learning, this thesis introduces novel methods for tracking and identifying all individuals in a honey bee colony over an extended period. This system uses machine learning and tracking algorithms to identify individual bees, resulting in unique and extensive datasets of honey bee trajectories. This enables the observation of large numbers of bees and their interactions over extended periods, providing foundational data for analyzing how individual honey bee behaviors relate to colony-level phenomena.

The resulting datasets are then used to explore how individual honey bee behaviors relate to collective behavior. To that end, a novel descriptor, *network age*, based on the spectral decomposition of the social interaction network, is introduced. This descriptor accurately predicts task allocation, survival, activity patterns, and future behavior of individual bees. Detailed analysis of multiple cohorts of bees reveals distinct developmental pathways and critical life changes. The thesis also analyzes variation in individual behavior and task allocation over timescales from single days to entire lifetimes. It shows how some bees consistently exhibit different movement patterns and transition to different tasks at different ages. These analyses provide insights into how individual honey bee behaviors may contribute to colony function.

Finally, the thesis builds upon these methods and generalizes them by analyzing lifetime data from multiple colonies. To that end, a novel non-negative temporal matrix factorization model is introduced. This factorization identifies the functional roles of individuals in social groups over time, regardless of when or where they lived. The method provides a quantitative framework for understanding heterogeneity in the behaviors and roles of individuals in complex social systems across time and space.

In summary, this thesis makes several key contributions. First, it demonstrates that individual behaviors and interactions can be measured on an unprecedented scale using automatic tracking technology and machine learning methods. Second, the resulting data are used to develop models that investigate individual honey bees' behavior and social interactions in the context of the collective of the colony, showing that new technologies can advance our understanding of honey bees' social organization and other complex systems. By studying social animals at the individual level, we can better understand the mechanisms underlying collective behavior.

Declaration Of Authorship

First name: Benjamin
Last name: Wild

I declare to the Freie Universität Berlin that I have completed the submitted dissertation independently and without the use of sources and aids other than those indicated. The present thesis is free of plagiarism. I have marked as such all statements that are taken literally or in content from other writings. This dissertation has not been submitted in the same or similar form in any previous doctoral procedure.

I agree to have my thesis examined by a plagiarism examination software.

Date Signature

Contents

1 Introduction

Alles ist Wechselwirkung — Everything is interconnectedness.

– Alexander von Humboldt, Travel Journal, August 1-5, 1803, Valley of Mexico.

The phenomenon of collective behavior, in which groups of individuals coordinate their actions to produce emergent phenomena at the group level, has intrigued researchers for decades. Examples of collective behavior can be seen in the synchronized movements of flocking birds and shoaling fish and in the outcomes of democratic elections in human societies. The dynamics of individual interactions play a crucial role in shaping such behavior (Farine and Whitehead, 2015; Gordon, 2010; Krause et al., 2015; Pinter-Wollman et al., 2014). Social animals, such as honey bees, are particularly adept at coordinating their actions, even in the absence of centralized control or communication (Hölldobler et al., 2009). However, understanding how individual behaviors give rise to collective patterns poses a significant challenge. Recent advances in tracking technologies and machine learning have opened up new avenues of research, providing unprecedented opportunities to study individual behavior and interactions over extended periods. This thesis, entitled *Individuality in the hive*, explores the use of automatic tracking of all individuals in a honey bee colony and modern machine learning methods to gain insight into the role of individuals and their interactions in the emergence of collective behavior.

Social insects have long been studied as model systems for investigating collective behavior, given their complex societies and intricate modes of communication. Task allocation in societies of social insects is hypothesized to arise from social interactions. However, the relationship between individual roles and social networks within a colony is poorly understood to date (Gordon, 1996; Tofts and Franks, 1992; Traniello and Rosengaus, 1997). Social interactions among individuals change depending on where and with whom they interact, and individuals can modify their behavior based on nestmate interaction (Pinter-Wollman et al., 2013, 2011; Seeley, 1992; Gordon and Mehdiabadi, 1999; Davidson and Gordon, 2017; Quevillon et al., 2015; Planckaert et al., 2019). Studies targeting specific interactions, roles, or stimuli have been conducted, but without an automated observation system, these behaviors and interactions can not be captured comprehensively, and observations are affected by human bias.

Among social insects, honey bees stand out for their richness of behavior, enabled by a range of sensory modalities and unique forms of communication, such as the famous waggle dance. This symbolic dance is an exceptional behavior of honey bees. It is unmatched in the insect world, allowing these insects to communicate the location of resources to other

colony members with impressive accuracy and precision (von Frisch, 1967). In addition to their scientific importance, honey bees play a crucial role in agriculture and are of great economic significance. The phenomenon of Colony Collapse Disorder, which has led to the decline of honey bee populations, underscores the importance of understanding the behavior and functioning of honey bee colonies (Oldroyd, 2007). Beekeeping, a practice that dates back thousands of years, has provided us with valuable knowledge and tools to study honey bees and their intriguing behaviors (Crane, 1999). For instance, standard observation frames commonly used by beekeepers can be repurposed for scientific observation, and markers used to identify queens in hives can serve as unique identifiers of all individuals in a colony (Wario et al., 2015). By studying honey bees in their natural colony environment, we can better understand their complex communication and social behavior. As honey bee colonies offer a model system to observe many individuals and their actions and interactions over a prolonged period, they provide a unique opportunity to study collective behavior.

Understanding the mechanisms underlying collective behavior is of fundamental importance, with implications for fields such as ecology, robotics, and the social sciences (Gordon, 2014; Rossi et al., 2021; Duan et al., 2023; Blumer, 1971) . This work posits that to understand emergent group behaviors, we must first understand individuals and their rich set of behaviors and social interactions within their society. By building models that explain individual behaviors, we can study how these behaviors give rise to collective patterns. Honey bee colonies are composed of thousands of individuals that engage in a wide range of behaviors and modes of interaction. To build models of individual behavior and interactions over long periods, we must first track and measure the behavior of each individual in the colony. This presents many technical challenges, such as the ability to distinguish and automatically track individual bees in the colony, the rapid movements of bees, and the many interactions that occur. Additionally, bees can disappear within the comb or leave the hive to forage, making long-term tracking challenging. Furthermore, the colony should not be exposed to light in the visible spectrum of honey bees, making tracking even more difficult. Once individual bees are tracked, the question arises of how to use this data to model individual behavior within the context of colony-level behavior. As learning to predict is closely related to understanding (Clark, 2013; Brown et al., 2020), the success of such a model can be measured by its ability to predict future behavior based on an individual's actions and interactions within the colony. For instance, can we predict an individual bee's future task allocation based on its current behavior? There are technical challenges in building such models, using large amounts of data, and evaluating and interpreting these models. Furthermore, such models should be able to generalize across individuals living at different times and from different colonies. This requires developing methods that generate consistent descriptors that capture the variability of individual behaviors, regardless of the origin of the individual.

Previous work on tracking social animals has relied heavily on manual observation, which is time-consuming and limited in scope. Researchers have manually followed indi-

vidual bees in order to study food exchange interactions (Naug, 2008), colony proximity networks (Baracchi and Cini, 2014), foraging behaviors (Biesmeijer and Seeley, 2005), and waggle dances (Couvillon et al., 2014). However, manual tracking is limited by the number of individuals that can be observed and the time required to analyze the data. In recent years, computer vision software has become a popular tool for automatically identifying and tracking animals (Krause et al., 2013; Dell et al., 2014), with tracking the position of an animal often being sufficient to infer its behavior (Kabra et al., 2013; Eyjolfsdottir et al., 2017; Blut et al., 2017). Tracking bees within a colony is challenging due to their similar appearance and frequent occlusions. Therefore, marker-based insect tracking systems have been proposed, which use a binary code for error correction, but these systems require expensive equipment and may result in reduced accuracy and sample rate if detections cannot be decoded (Mersch et al., 2013; Crall et al., 2015; Gernat et al., 2018). This thesis builds upon the BeesBook tracking system (Wario et al., 2015), which uses less expensive recording equipment than other solutions and achieves high recall and precision in localizing tags and decoding IDs without relying on error-correcting codes by using a unique tag design and machine learning to identify individuals (Wild et al., 2018). However, linking detections only based on matching IDs can lead to errors, so a more elaborate tracking algorithm is required to follow individuals robustly.

To this end, the thesis aims to go beyond the state of the art by developing and applying novel methods to track and analyze honey bee behavior over long periods. One of the key contributions of this work is the development of a reliable system for automatically tracking individual honey bees in a colony, using machine learning methods to identify and distinguish between individuals. This system is described in Chapter 3 - Tracking All Members of a Honey Bee Colony Over Their Lifetime Using Learned Models of Correspondence. This system makes possible the observation of large numbers of individuals, their actions, and their interactions over extended periods, generating a unique and extensive dataset that has not been possible before.

After establishing a reliable tracking system and the technical infrastructure for storing and processing large amounts of tracking data, the focus of the subsequent works in this thesis was on analyzing the generated datasets. Previous studies examining the behavior of individuals in social insect societies and how it relates to the collective have primarily relied on human observation, assigning behaviors to individuals using ethograms (Lindauer, 1952; Seeley, 1982; Seeley and Kolmes, 1991; Johnson, 2003; Siegel et al., 2013; Smith et al., 2017; Perez and Johnson, 2019). Network analysis of social interaction networks has shown promise as a technique for studying social animals. This is because interaction networks can be relatively easily constructed from tracking data, and network analysis represents each individual as embedded within the broader social structure of the group (Krause et al., 2015; Pinter-Wollman et al., 2014; Brask et al., 2021). However, researchers have only studied the interaction networks from a colony-wide or temporal perspective to a limited extent (Gernat et al., 2018; Hasenjager et al., 2020), and tracking has often been restricted to specific individuals or short intervals (Naug, 2008; Blut et al., 2017;

Siegel et al., 2013; Bozek et al., 2021). Previous research has shown that individual bees spend most of their time in specific nest regions during each life stage, interacting with their nestmates (Seeley, 1982; Johnson, 2010). The patterns of interaction with others may depend on factors beyond location, such as previous interactions or genetic diversity within the colony (Farina, 2000; Girard et al., 2011). While social interactions enable the exchange of information and can have long-term effects on an individual's behavior, identifying an individual bee's role in a colony based on its patterns of interaction is still challenging, especially with numerous individuals and various interaction modes. Recent advances in automated tracking such as the system described in this thesis now provide the opportunity to obtain behavioral metrics that were previously beyond the scope of human observation, allowing for a more comprehensive, data-driven analysis of individual variability and inter-individual differences across different time scales (Wario et al., 2015; Mersch et al., 2013; Crall et al., 2015; Gernat et al., 2018; Wild et al., 2018; Bozek et al., 2021; Crall et al., 2018; Jones et al., 2020; Richardson et al., 2021).

To that end, this thesis presents novel methods of social network analysis to analyze individual behaviors and interactions within the colony, allowing for a deeper understanding of how individual behavior relates to the colony's behavior as a whole. Chapter 4 - Social networks predict the life and death of honey bees introduces a succinct descriptor of an individual's social network, called *network age*, which can accurately predict task allocation, survival, activity patterns, and future behavior. This descriptor is used to analyze the developmental trajectories of multiple cohorts of individuals in a natural setting and identify distinct developmental pathways and critical life changes.

However, it is important to note that there is significant behavioral variability over longer periods, among individuals, and particularly between individuals from different colonies. Honey bee task allocation is influenced by multiple factors, including age, genetics, and social interactions (Huang et al., 1994; Beshers and Fewell, 2001; Robinson, 2002; Grozinger et al., 2007; Johnson, 2008a; Cook and Breed, 2013; Cook et al., 2019). While young bees tend to work inside the nest, and older bees tend to forage (Seeley, 1982), individuals may switch between or perform multiple tasks, suggesting that task repertoires are a more useful categorization than age alone for describing variation in worker behavior (Seeley, 1982; Johnson, 2010; Beshers and Fewell, 2001; Jeanson and Weidenmüller, 2014). Chapter 5 - Behavioral variation across the days and lives of honey bees introduces a quantitative framework to describe individual behavioral variation within a colony from a single day to an entire lifetime. This analysis shows that bees differ in their spatial use, detection, and movement, and that inter-individual differences exist in the movement characteristics of individuals across their entire lives. Furthermore, this method reveals that bees differ in how quickly they transition through behavioral space to ultimately become foragers, with fast-transitioning bees living the shortest lives.

Notably, the previous analyses were conducted on a single colony, highlighting the need for further studies to explore behavioral and social interaction dynamics across multiple colonies. Furthermore, methods like *network age* are not suited to derive behavioral

descriptors from long-term datasets. To address these limitations, the final work in this thesis investigates whether consistent semantic embeddings that allow for the comparison of individuals from multiple colonies and long-term recordings can be automatically learned. In recent years, semantic embeddings have become popular in machine learning to solve related problems by learning interpretable representations of entities based on high-dimensional data. These methods map entities into a learned vector space, allowing for downstream tasks such as identifying genres of movies or related words in natural language models (Frome et al., 2013; Asgari and Mofrad, 2015; Camacho-Collados and Pilehvar, 2018; Nelson et al., 2019; Koren et al., 2009; Mikolov et al., 2013). In the context of social interactions, embeddings can be learned from raw data and used to understand how environmental conditions affect behavioral changes within a group (Richardson et al., 2021). One of the challenges of learning semantic descriptors in dynamic social systems is that individuals' behavior changes over time, and their interactions with others are not consistent. While methods like *network age*, as introduced in Chapter 4, can also be interpreted as a way to extract semantic embeddings from temporal snapshots of social networks of animal societies, this approach does not capture the changes in behavior and interactions that occur over long periods and cannot produce behavioral descriptors that are consistent across several colonies.

This limitation is addressed in the final major contribution of this thesis. Chapter 6 - Learning to embed lifetime social behavior from interaction dynamics demonstrates that common factors that partially determine individuals' roles, such as age, can be used to learn semantically consistent functional descriptors of individual behaviors and interactions over time in highly dynamic social systems. This work also shows that embeddings learned from social interaction data without additional supervision can be used to reveal how environmental conditions affect behavioral changes within the group. Non-negative matrix factorization (NMF) is presented as a principled and scalable method to learn embeddings from data that can be represented in matrix form, such as interaction networks, allowing for soft clustering and the derivation of embeddings from social interaction networks. The resulting embeddings are biologically relevant and consistent over time, allowing for the comparison of individuals' functional roles regardless of when or in which colony they lived.

Overall, this thesis investigates the mechanisms underlying collective behavior in complex animal societies without central organization. Specifically, it examines individual behaviors in the context of global collective behavior. To achieve this goal, the thesis addresses the following two overarching research questions:

1. How does collective behavior emerge in societies without central organization?

2. What is the relationship between individual behaviors and the global behavior of the collective?

To answer these questions, the thesis studies honey bee colonies as a model system and pursues the following three specific research aims:

1. Develop a computer vision and machine learning-based tracking system to collect unprecedentedly detailed data on individual honey bee movement and interactions.
2. Develop analytical methods to explore individual honey bee behaviors and interactions and connect them to collective colony-level behavior.
3. Generalize the findings and develop methods to analyze large datasets comprising multiple colonies and extended periods, including non-overlapping cohorts of individuals living in different conditions.

By studying individual honey bee behaviors, this thesis investigates how novel technologies can enhance our understanding of complex systems and demonstrates the importance of measuring individual-level behaviors to understand emergent collective phenomena. The thesis focuses on developing methods to analyze complex systems and argues that computational approaches, illustrated through the example of honey bee colonies, can significantly advance our understanding of collective behavior. Its contributions are supported by the release of large-scale datasets, with all methods and software made publicly accesible.

2 Contributions To Individual Publications

In this section, the contributions to the individual publications according to § 7 Abs. 3 of the doctoral regulations from 1st December 2021 are summarized:

Tracking all members of a honey bee colony over their lifetime using learned models of correspondence (Chapter 3)
Franziska Boenisch, Benjamin Rosemann, Benjamin Wild, David Dormagen, Fernando Wario, and Tim Landgraf.

BR and TL: Conceptualization; FB, BR, BW, DD, and TL: Methodology; FB and BR: Software; TL: Resources, supervision, and project administration; FB, BR, and BW: Data curation; FB, BW, DD, and TL: Writing–original draft; FB, BW, FW, DD, and TL: Writing–review and editing and visualization.

In the development of the *BeesBook* data processing pipeline, I was instrumental in the first two stages: the detection and decoding of binary honey bee markers. These stages formed the foundation for the third stage presented in this paper. I contributed to the experimental setup of the long-term video data processing, the marking of bees, and development of software for recording, storing, cleaning, validating, and processing this data. Furthermore, I was responsible for integrating the results of the detection and decoding into a database, which was crucial for the tracking stage. I contributed to the design and development of the tracking methodolgy and implemented several evaluation metrics and was deeply involved in the writing process. My contributions extended to creating datasets by applying the tracking method presented in this paper to the experimental data we recorded from 2016 to 2019.

Social networks predict the life and death of honey bees (Chapter 4)
Benjamin Wild, David M. Dormagen, Adrian Zachariae, Michael L. Smith, Kirsten S. Traynor, Dirk Brockmann, Iain D. Couzin, and Tim Landgraf.

Conceptualization: BW, DMD, AZ, TL, MLS, KST; Methodology: BW, DMD, AZ, TL; Software: BW, DMD; Resources, supervision: TL, DB, IDC; Project administration:

TL; Data curation: BW, DMD, MLS; Writing: BW, DMD, TL, MLS, KST, IDC; Visualization: BW, DMD.

I developed the concept of network age and played a key role in the development, analysis, visualization and interpretation of the method. I was heavily involved in the design and implementation of software crucial for data acquisition and analysis. I contributed to the data acquisition and to the field experiments by training bees to use artificial feeding stations. I significantly contributed to writing and revising the manuscript, ensuring that all aspects of the study were clearly presented and scientifically robust.

Behavioral variation across the days and lives of honey bees. (Chapter 5)
Michael L. Smith, Jacob D. Davidson, Benjamin Wild, David M. Dormagen, Tim Landgraf, and Iain D. Couzin.

Methodology: MLS, JDD, BW, DD, TL, IDC; Investigation: MLS, JDD, TL, IDC; Fieldwork: MLS; Software: BW, DMD, TL; Visualization: JDD; Supervision: TL, IDC; Writing-original draft: MLS, JDD; Writing-review and editing: MLS, JDD, BW, DMD, TL, IDC.

I played a key role in replicating the *BeesBook* tracking system in Konstanz, enabling and assisiting in extensive longitudinal observations in two recording seasons over two years. I contributed to the data acquisition, processing, storage, analysis, and to the conceptualization of the experimental setup and paper. Additionally, I contributed to the writing and editing of the paper.

Learning to embed lifetime social behavior from interaction dynamics. (Chapter 6)
Benjamin Wild, David M. Dormagen, Michael L. Smith, and Tim Landgraf.

Conceptualization: BW, TL; Methodology: BW; Software: BW; Resources, supervision: TL; Project administration: TL; Data curation: BW, DMD; Writing: BW, DMD, MLS, TL; Visualization: BW.

I was primarily responsible for the conceptualization, methodology, software development, data curation, writing, and visualization aspects of the project. Specifically, I conceived the methodological framework, developed and implemented the method, curated and processed the data, and wrote the initial draft of the manuscript which was then revised and edited by all authors.

3 Tracking All Members of a Honey Bee Colony Over Their Lifetime Using Learned Models of Correspondence

Franziska Boenisch [1], Benjamin Rosemann [1], Benjamin Wild [1], Fernando Wario [1], David Dormagen [1], Tim Landgraf [1,*]

[1] Dahlem Center of Machine Learning and Robotics, Dept. Mathematics and Computer Science, Freie Universität Berlin, Berlin, Germany

[*] Corresponding author: tim.landgraf@fu-berlin.de

3.1 Preface

The paper *Tracking All Members of a Honey Bee Colony Over Their Lifetime Using Learned Models of Correspondence* presents a new approach to automatically track individual honey bees within a colony, allowing for the observation of large numbers of individuals and their interactions over extended periods. Previous studies of social insect behavior have predominantly relied on manual observation, but this approach is limited by the number of individuals that can be observed and the time required to analyze the data. The proposed system builds upon the BeesBook tracking system (Wario et al., 2015; Wild et al., 2018) but employs a more elaborate tracking algorithm that uses classifiers learned from annotated data to predict the correspondence of two consecutive detections, resulting in improved marker decoding accuracy. In addition to the proposed used of machine learning methods to identify and distinguish between individuals, a multi-step algorithm to produce motion paths was introduced. The system was tested on approximately 2,000 marked honey bees over 10 weeks, and the resulting trajectories represent a unique and extensive resource for investigating honey bee collective behavior. The paper highlights the challenges of tracking bees within a colony, but also demonstrates the potential of computer vision and machine learning techniques for addressing these challenges. The methods described in this paper are foundational for the works described in the following chapters, as they enabled the

creation of the first ever large scale honey bee trajectory dataset, containing tracks for all bees in a colony with high temporal and spatial resolution. These data were used in the following works to automatically quantify space use and movement characteristics of the individuals, and to detect interaction behaviors and build temporal social networks of all individuals in a honey bee colony.

This chapter was previously published as:
Boenisch et al. (2018a) — Franziska Boenisch, Benjamin Rosemann, Benjamin Wild, David Dormagen, Fernando Wario, and Tim Landgraf. Tracking all members of a honey bee colony over their lifetime using learned models of correspondence. *Frontiers in Robotics and AI*, 5, 2018a. URL https://doi.org/10.3389/frobt.2018.00035

3.2 Author contributions

BR and TL: Conceptualization; FB, BR, BW, DD, and TL: Methodology; FB and BR: Software; TL: Resources, supervision, and project administration; FB, BR, and BW: Data curation; FB, BW, DD, and TL: Writing–original draft; FB, BW, FW, DD, and TL: Writing–review and editing and visualization.

3.3 Abstract

Computational approaches to the analysis of collective behavior in social insects increasingly rely on motion paths as an intermediate data layer from which one can infer individual behaviors or social interactions. Honey bees are a popular model for learning and memory. Previous experience has been shown to affect and modulate future social interactions. So far, no lifetime history observations have been reported for all bees of a colony. In a previous work we introduced a recording setup customized to track up to 4000 marked bees over several weeks. Due to detection and decoding errors of the bee markers, linking the correct correspondences through time is non-trivial. In this contribution we present an in-depth description of the underlying multi-step algorithm which produces motion paths, and also improves the marker decoding accuracy significantly. The proposed solution employs two classifiers to predict the correspondence of two consecutive detections in the first step, and two tracklets in the second. We automatically tracked ~2000 marked honey bees over 10 weeks with inexpensive recording hardware using markers without any error correction bits. We found that the proposed two-step tracking reduced incorrect ID decodings from initially ~13% to around 2% post-tracking. Alongside this paper, we publish the first trajectory dataset for all bees in a colony, extracted from ~3 million images covering three days. We invite researchers to join the collective scientific effort to investigate this intriguing animal system. All components of our system are open-source.

3.4 Introduction

Social insect colonies are popular model organisms for self-organization and collective decision making. Devoid of central control, it often appears miraculous how orderly

termites build their nests or ant colonies organize their labor. Honey bees are a particularly popular example - they stand out due to a rich repertoire of communication behaviors (von Frisch, 1965; Seeley, 2010) and their highly flexible division of labor (Johnson, 2010; Robinson, 1992). A honey bee colony robustly adapts to changing conditions, whether it may be a hole in the hive that needs to be repaired, intruders that need to be fended off, brood that needs to be reared, or food that needs to be found and processed. The colony behavior emerges from interactions of many thousand individuals. The complexity that results from the vast number of individuals is increased by the fact that bees are excellent learners: empirical evidence indicates that personal experience can modulate communication behavior (Grüter et al., 2006; Balbuena et al., 2012; Grüter and Farina, 2009; Grüter and Ratnieks, 2011; Goyret and Farina, 2005; De Marco and Farina, 2001; Richter and Waddington, 1993). Especially among foragers, personal experience may be very variable. The various locations a forager visits might be dispersed over large distances (up to several kilometers around the hive) and each site might offer different qualities of food, or even pose threats. Thus, no two individuals share the same history and experiences. Evaluating how personal experience shapes the emergence of collective behavior and how individual information is communicated to and processed by the colony requires robust identification of individual bees over long time periods.

However, insects are particularly hard to distinguish by a human observer. Tracking a bee manually is therefore difficult to realize without marking these animals individually. Furthermore, following more than one individual simultaneously is almost impossible for the human eye. Thus, the video recording must be watched once per individual, which, in the case of a bee hive, might be several hundred or thousand times. Processing long time spans or the observation of many bees is therefore highly infeasible, or is limited to only a small group of animals. Most studies furthermore focused on one focal property, such as certain behaviors or the position of the animal. Over the last decades, various aspects of the social interactions in honey bee colonies have been investigated with remarkable efforts in data collection: Naug (Naug, 2008) manually followed around 1000 marked bees in a one hour long video to analyze food exchange interactions. Barachi and Cini (Baracchi and Cini, 2014) manually extracted the positions of 211 bees once per minute for 10 hours of video data to analyze the colony's proximity network. Biesmeijer and Seeley (Biesmeijer and Seeley, 2005) observed foraging related behaviours of a total of 120 marked bees over 20 days. Couvillon and coworkers manually decoded over 5000 waggle dances from video (Couvillon et al., 2014). Research questions requiring multiple properties, many individuals, or long time frames are limited by the costs of manual labor.

In recent years, computer vision software for the automatic identification and tracking of animals has evolved into a popular tool for quantifying behavior (Krause et al., 2013; Dell et al., 2014). Although some focal behaviors might be extracted from the video feed directly (Berman et al., 2014; Wiltschko et al., 2015; Wario et al., 2017), tracking the position of an animal often suffices to infer its behavioral state (Kabra et al., 2013; Eyjolfsdottir et al., 2017; Blut et al., 2017). Tracking bees within a colony is a particularly

challenging task due to dense populations, similar target appearance, frequent occlusions and a significant portion of the colony frequently leaving the hive. The exploration flights of foragers might take several hours, guard bees might stay outside the entire day to inspect incoming individuals. The observation of individual activity over many weeks, hence, requires robust means for unique identification.

For a system that robustly decodes the identity of a given detection, the tracking task reduces to simply connecting matching IDs. Recently, three marker-based insect tracking systems (Mersch et al., 2013; Crall et al., 2015; Gernat et al., 2018) have been proposed that use a binary code with up to 26 bits for error correction (Thompson, 1983). The decoding process can reliably detect and correct errors, or, reject a detection that can not be decoded. There are two disadvantages to this approach. First, error correction requires relatively expensive recording equipment (most systems use at least a 20 MP sensor with a high quality lens). Second, detections that could not be decoded can usually not be integrated into the trajectory, effectively reducing the detection accuracy and sample rate.

In contrast to these solutions, we have developed a system called BeesBook that uses much less expensive recording equipment (Wario et al., 2015). Figure 3.1 shows our recording setup, Figure 3.2 visualizes the processing steps performed after the recording. Our system localizes tags with a recall of 98% at 99% precision and decodes 86% IDs correctly without relying on error correcting codes (Wild et al., 2018). See Figure 3.3 for the tag design. Linking detections only based on matching IDs would quickly accumulate errors, long-term trajectories would exhibit gaps or jumps between individuals. Following individuals robustly, thus, requires a more elaborate tracking algorithm.

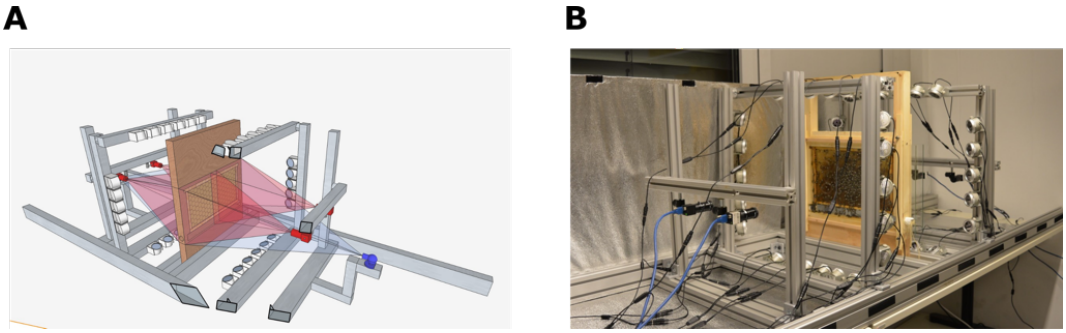

Figure 3.1: (**A**) Schematic representation of the setup. Each side of the comb is recorded by two 12 MP PointGrey Flea3 cameras. The pictures have an overlap of several centimeters on each side. (**B**) The recording-setup used in summer 2015. The comb, cameras and the infrared lights are depicted, the tube that can be used by the bees to leave the setup is not visible. During recording, the setup is covered. Figures adapted from (Wario et al., 2015).

The field of multiple object tracking has produced numerous solutions to various use-cases such as pedestrian and vehicle tracking (for reviews see Cox (1993); Wu et al.

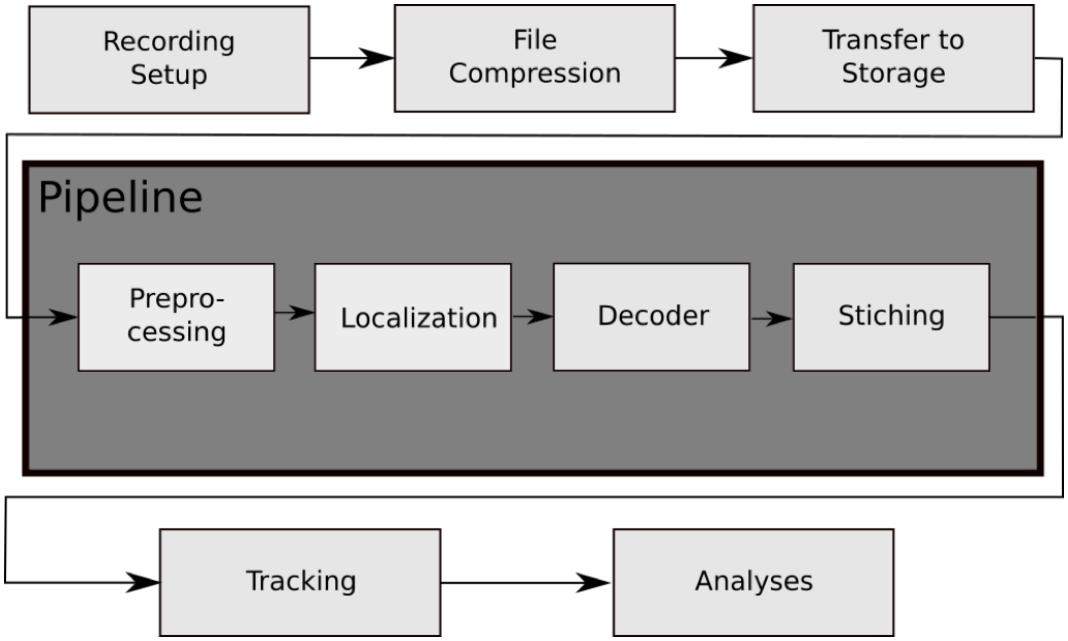

Figure 3.2: The data processing steps of the BeesBook project. The images captured by the recording setup are compressed on-the-fly to videos containing 1024 frames each. The video data is then transferred to a large storage from where it can be accessed by the pipeline for processing. Preprocessing: histogram equalization and subsampling for the localizer. Localization: bee markers are localized using a convolutional neural network. Decoding: a second network decodes the IDs and rotation angles. Stitching: the image coordinates of the tags are transformed to hive coordinates and duplicate data in regions where images overlap are removed.

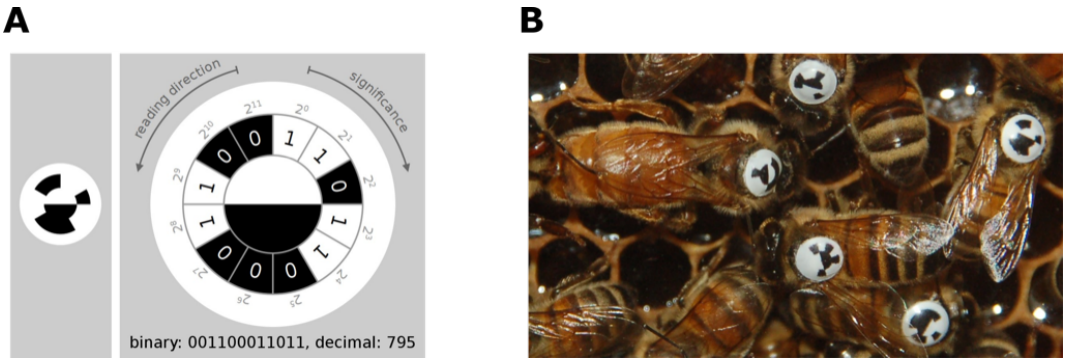

Figure 3.3: (**A**) The tag-design in the BeesBook project uses 12 coding segments arranged in an arc around two semi-circles that encode the orientation of the bee. The tag is glued onto the thorax such that the white semi-circle is rotated towards the bee's head. Figure adapted from (Wario, 2017). (**B**) Several tagged honey bees on a comb. The round and curved tags are designed to endure heavy duty activities such as cell inspections and foraging trips.

(2013); Luo et al. (2021); Betke and Wu (2016)). Animals, especially insects, are harder to distinguish and solutions for tracking multiple animals over long time frames are far less numerous (see Dell et al. (2014) for a review on animal tracking). Since our target subjects may leave the area under observation at any time, the animal's identity cannot be preserved by tracking alone. We require some means of identification for a new detection, whether it be paint marks or number tags on the animals, or identity-preserving descriptors extracted from the detection.

While color codes are infeasible with monochromatic imaging, using image statistics to "fingerprint" sequences of visible animals (Pérez-Escudero et al., 2014; Kühl and Burghardt, 2013; Wang and Yeung, 2013) may work even with unstructured paint markers. Merging tracklets after occlusions can then be done by matching fingerprints. However, it remains untested whether these approaches can resolve the numerous ambiguities in long-term observations of many hundreds or thousands of bees that may leave the hive for several hours.

In the following, we describe the features that we used to train machine learning classifiers to link individual detections and short tracklets in a crowded bee hive. We evaluate our results with respect to path and ID correctness. We conclude that long-term tracking can be performed without marker-based error correction codes. Tracking can, thus, be conducted without expensive high-resolution, low-noise camera equipment. Instead, decoding errors in simple markers can be mitigated by the proposed tracking solution, leading to a higher final accuracy of the assigned IDs compared to other marker-based systems that do not employ a tracking step.

3.5 Description of Methods

3.5.1 Problem statement and overview of tracking approach

The tracking problem is defined as follows: Given a set of detections (timestamp, location, orientation and ID information), find correct correspondences among detections over time (tracks) and assign the correct ID to each track. The ID information of the detections can contain errors. Additionally, correct correspondences between detections of consecutive frames might not exist due to missing detections caused by occluded markers. In our dataset, the ID information consists of a number in the range of 0 to 4095, represented by 12 bits. Each bit is given as a value between 0.0 and 1.0 which corresponds to the probability that the bit is set.

To solve the described tracking problem, we propose an iterative tracking approach, similar to previous works (for reviews, see Luo et al. (2021); Betke and Wu (2016)). We use two steps: 1. Consecutive detections are combined into short but reliable tracklets (Rosemann., 2017). 2. These tracklets are connected over longer gaps (Boenisch, 2017). Previous work employing machine learning mostly scored different distance measures separately to combine them into one thresholded value for the first tracking step (Wu and Nevatia, 2007; Huang et al., 2008; Wang et al., 2014; Fasciano et al., 2013). For merging longer tracks, boosting models to predict a ranking between candidate tracklets have been proposed (Huang et al., 2008; Fasciano et al., 2013). We use machine learning models in both steps to learn the probability that two detections, or tracklets, correspond. We train the models on a manually labeled dataset of ground truth tracklets. The features that are used to predict correspondence can differ between detection level and tracklet level, so we treat these two stages as separate learning problems. Both of our tracking steps use the Hungarian algorithm (Kuhn, 1955) to assign likely matches between detections in subsequent time steps based on the predicted probability of correspondence. In the following, we describe which features are suitable for each step and how we used various regression models to create accurate trajectories. We also explain how we integrate the ID decodings of the markers along a trajectory to predict the most likely ID for this animal, which can then be used to extract long-term tracks covering the whole lifespan of an individual. See Figure 3.4 for an overview of our approach.

3.5.2 Step 1: Linking consecutive detections

The first tracking step considers detections in successive frames. To reduce the number of candidates, we consider only sufficiently close detections (we use approximately 200 pixels, or 12mm).

From these candidate pairs we extract three features:

1. Euclidean distance between the first detection and its potential successor.

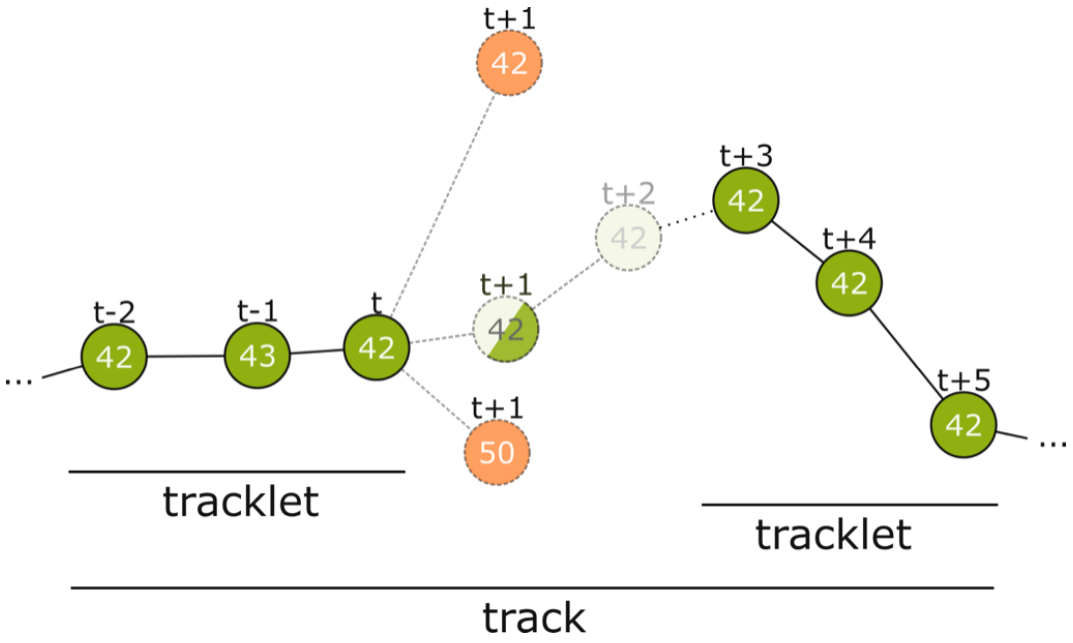

Figure 3.4: Overview of the tracking process. The first step connects detections from successive frames to tracklets without gaps. At time step t only detections within a certain distance are considered. Even if a candidate has the same ID (top-most candidate with ID 42) it can be disregarded. The correct candidate may be detected with an erroneous ID (see t-1) or may even not be detected at all by the computer vision process. There may be close incorrect candidates that have to be rejected (candidate with ID 43 at t+1). The model assigns a correspondence probability to all the candidates. If none of them receive a sufficient score the tracklet is closed. In time step t+3 a new detection with ID 42 occurs again and is extended into a second tracklet. In tracking step 2, these tracklets are combined to a larger tracklet or track.

2. Angular difference of both detections' orientations on the comb plane.

3. Manhattan distance between both detections' ID probabilities.

We use our manually labeled training data to create samples with these features that include both correct and incorrect examples of correspondence. A support vector machine (SVM) with a linear kernel (Cortes and Vapnik, 1995) is then trained on these samples. We also evaluated the performance of a random forest classifier (Ho, 1995) with comparable results. We use the SVM implemented in the scikit-learn library (Pedregosa et al., 2011). Their implementation of the probability estimate uses Platt's method (Platt, 1999). This SVM can then be used get the probability of correspondence for pairs of detections that were not included in the training data. To create short tracks (tracklets), we iterate through the recorded data frame by frame and keep a list of open tracklets. Initially, we have one open tracklet for each detection of the first frame. For every time step, we use the SVM to score all new candidates against the last detection of each open tracklet. The Hungarian algorithm is then used to assign the candidate detections to the open tracklets. Tracklets are closed and not further expanded if their best candidate has a probability lower than 0.5. Detections that could not be assigned to an existing open tracklet are used to begin a new open tracklet that can be expanded in the next time step.

3.5.3 Step 2: Merging tracklets

The first step yields a set of short tracklets that do not contain gaps and that could be connected with a high confidence. The second tracking step merges these tracklets into longer tracks that can contain gaps of variable duration (for distributions of tracklet and gap length in our data see Section 3.6). Note that a tracklet could consist of a single detection or that its corresponding consecutive tracklet could still begin in the next time step without a gap. To reduce computational complexity we define a maximum gap length of 14 time steps ($\sim$ 4s in our recordings).

Similar to the first tracking step, we use the ground truth dataset to create training samples for a machine learning classifier. We create positive samples (i.e. fragments that should be classified as belonging together) by splitting each manually labeled track once at each time step. Negative samples are generated from each pair of tracks with different IDs which overlapped in time with a maximum gap size of 14. These are also split at all possible time steps. To include both more positive samples and more short track fragments in the training data, we additionally use every correct sub-track of length 3 or less and again split it at all possible locations. This way we generated 1.021.848 training pairs, 7.4% of which were positive samples.

In preliminary tests, we found that for the given task of finding correct correspondences between tracklets, a random forest classifier performed best among a selection of classifiers available in scikit-learn (Boenisch, 2017).

Tracklets with two or more detections allow for more complex and discriminative features compared to those used in the first step. For example, matching tracklets separated by longer gaps may require features that reflect a long-term trend (e.g. the direction of motion).

We implemented 31 different features extractable from tracklet pairs. We then used four different feature selection methods from the scikit-learn library to find the features with the highest predictive power. This evaluation was done by splitting the training data further into a smaller training set and validation set. The methods used were Select-K-Best, Recursive Feature Elimination, Recursive Feature Elimination with Cross-Validation and the Random Forest Feature Importance for all possible feature subset sizes as provided by scikit-learn (Pedregosa et al., 2011). In all these methods, the same four features (number 1 to 4 in the listing below) performed best according to the ROC AUC score (Spackman, 1989) that proved to be a suitable metric to measure tracking results. Therefore, we chose them as an initial subset.

We then tried to improve the feature subset manually according to more tracking-specific metrics. The metrics we used were the number of tracks in the ground truth validation set that were reconstructed entirely and correctly, and the number of insertions and deletes in the tracks (for further explanation of the metrics see Section 3.6). We added the features that lead to the highest improvements in these metrics on our validation set. This way, we first added feature 5 and then 6. After adding feature 6, the expansion of the subset with any other feature only lead to a performance decrease in form of more insertions and less complete tracks. We therefore kept the following six features. Visualizations of features 2 to 5 can be found in Figure 3.5.

1. Manhattan distance of both tracklets' bitwise averaged IDs.

2. Euclidean distance of last detection of tracklet 1 to first detection of tracklet 2.

3. Forward error: Euclidean distance of linear extrapolation of last motion in first tracklet to first detection in second tracklet.

4. Backward error: Euclidean distance of linear extrapolation of first motion in second tracklet to last detection in first tracklet.

5. Angular difference of tag orientation between the last detection of the first tracklet and the first detection of the second tracklet.

6. Difference of confidence: all IDs in both tracklets are averaged with a bitwise median, we select the bit that is closest to 0.5 for each tracklet, calculate the absolute difference to 0.5 (the confidence) and compute the absolute difference of these two confidences.

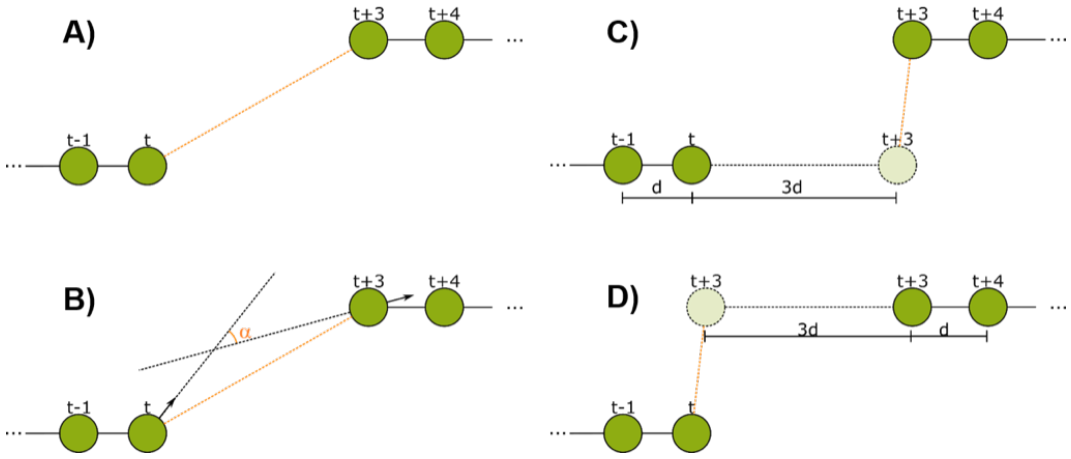

Figure 3.5: The spatial features used in the second tracking step. A) Euclidean distance between the last detection of tracklet 1 and the first detection of tracklet 2. B) Forward error: Euclidean distance of the extrapolation of the last movement vector in tracklet 1 to the first detection in tracklet 2. C) Angular difference between the tag orientations of the last detection in tracklet 1 and the first detection in tracklet 2. D) Backward error: Euclidean distance between the reverse extrapolation of the first movement vector of tracklet 2 to the last detection of tracklet 1.

Track ID assignment

After the second tracking step, we determine the ID of the tracked bee by calculating the median of the bitwise ID probabilities of all detections in the track. The final ID is then determined by binarizing the resulting probabilities for each bit with probability threshold 0.5.

Parallelization

Tracks with a length of several minutes already display a very accurate ID decoding (see Section 3.6). To calculate longer tracks of up to several days and weeks, we execute the tracking step 1 and step 2 for intervals of one hour and then merge the results to longer tracks based on the assigned ID. This allows us to effectively parallelize the tracking calculation and track the entire season of ten weeks of data in less than a week on a small cluster with less than 100 CPU cores.

3.6 Results and evaluation

We marked an entire colony of 1953 bees in a two days session and continuously added marked young bees that were bred in an incubation chamber. In total, 2775 bees were

marked. The BeesBook system was used to record 10 weeks of continuous image data (3 Hz sample rate) of a one-frame observation hive. The image recordings were stored and processed after the recording season. The computer vision pipeline was executed on a Cray XC30 supercomputer. In total, 3,614,742,669 detections were extracted from 67,972,617 single frames, corresponding to 16,993,154 snapshots of the four cameras. Please note that the data could also be processed in real-time using consumer hardware (Wild et al., 2018).

Two ground truth datasets for the training and evaluation of our method were created manually. A custom program was used to mark the positions of an animal and to define its ID (Mischek, 2016). Details on each dataset can be found in Table 3.1. To avoid overfitting to specific colony states, the datasets were chosen to contain both high activity (around noon) and low activity (in the early morning hours) periods, different cameras and, therefore, different comb areas. Dataset *2015.1* was used to train and validate classifiers and dataset *2015.2* was used to test their performance.

Dataset	2015.1	2015.2
Date	18.09.2015	22.09.2015
Times	11:36; 04:51	13:36
Frames	201 (3 fps)	200 (3 fps)
Detections	18085	10945
False positives	222 (1.23%)	82 (0.75%)
Individuals	144	98

Table 3.1: Dataset *2015.1* was used for training and dataset *2015.2* for testing. The number of detections is the number of tags localized and decoded by the deep learning approach over all frames in the dataset. The number of false positives shows how many times the deep learning pipeline detects a detection when there is none. The number of individuals indicates how many different bees are present in the dataset.

Dataset *2015.1* contains 18085 detections from which we extracted 36045 sample pairs (i.e. all pairs with a distance of less than 200 pixels in consecutive frames). These samples were used to train the SVM which is used to link consecutive detections together (tracking step 1). Hyperparameters were determined manually using cross-validation on this dataset. The final model was evaluated on dataset *2015.2*.

Tracklets for the training and evaluation of a random forest classifier (tracking step 2) were extracted from datasets *2015.1* respectively *2015.2* (see Section 3.5 for details). Hyperparameters were optimized with hyperopt-sklearn (Komer et al., 2014) on dataset *2015.1* and the optimized model was then tested on dataset *2015.2*.

To validate the success of the tracking, we analyzed its impact on several metrics in the tracks, namely:

1. ID Improvement

2. Proportion of complete tracks
3. Correctness of resulting tracklets
4. Length of resulting tracklets

To be able to evaluate the improvement through the presented iterative tracking approach, we compare the results of the two tracking steps to the naive approach of linking the original detections over time based on their initial decoded ID only, in the following referred to as "baseline". For an overview on the improvements achieved by the different tracking steps see Table 3.2.

	baseline	after step 1	after step 2	perfect tracking
incorrect detection IDs	13.3%	3.9%	1.9%	0.6%
incorrect track IDs	63.5	27.2%	18.2%	8.2%
complete tracks	10.2%	26.5%	70.4%	77.6%
detections missing from their track (deletions)	32.2%	1.38%	2.37%	0%
tracks with at least one deletion	94.6%	26.7%	18.25%	0%

Table 3.2: Different metrics were used to compare the two tracking steps to both a naive baseline based on the detection IDs and to manually created tracks without errors (perfect tracking). In all cases, the baseline performs worst and the two tracking steps successively improve the performance.

ID IMPROVEMENT An important goal of the tracking is to correct IDs of detections which could not be decoded correctly by the computer vision system. Without the tracking algorithm described above, all further behavioral analyses would have to consider this substantial proportion of erroneous decodings. In our dataset, 13.3% of all detections have an incorrectly decoded ID (Wild et al., 2018).

In the ground truth dataset we manually assigned detections that correspond to the same animal to one trajectory. The ground truth data can therefore be considered as the "perfect tracking". Even on these perfect tracks the median ID assignment algorithm described above provides incorrect IDs for 0.6% of all detections, due to partial occlusions, motion blur and image noise. This represents the lower error bound for the tracking system. As shown in Figure 3.6, the first tracking step reduces the fraction of incorrect IDs from 13.3% to 3.9% of all detections. The second step further improves this result to only 1.9% incorrect IDs.

Most errors occur in short tracklets (see Figure 3.7). Therefore, the 1.9% erroneous ID assignments correspond to 18.2% of the resulting tracklets being assigned an incorrect

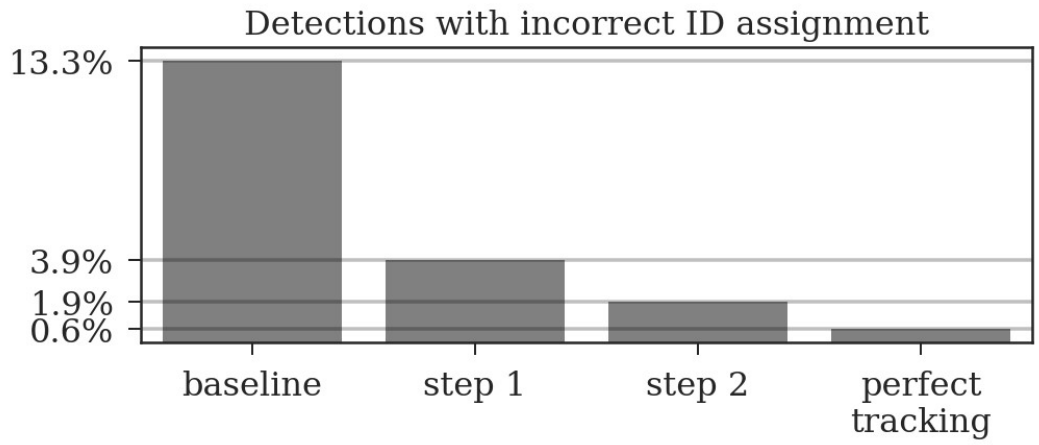

Figure 3.6: Around 13% of the raw detections are incorrectly decoded. The first tracking step already reduces this error to around 4% and the second step further reduces it to around 2%. Even a perfect tracking (defined by the human ground truth) would still result in 0.6% incorrect IDs when using the proposed ID assignment method.

median ID. This is an improvement over the naive baseline and the first tracking step with 63.5% and 27.2% respectively. A perfect tracking could reduce this to 8.2% (see Figure 3.8).

Proportion of complete tracks Almost all gaps between detections in our ground truth tracks are no longer than 14 frames (99.76%, see Figure 3.9). Even though large gaps between detections are rare, long tracks are likely to contain at least one such gap: Only around one third (34.7%) of the ground truth tracks contain no gaps and 77.6% contain only gaps shorter than 14 frames. As displayed in Figure 3.10, the baseline tracking finds only 10.2% complete tracks without errors (i.e. 30% of all tracks with no gaps). Step 1 is able to correctly assemble 26.5% complete tracks (i.e. around 76.5% of all tracks containing no gaps). Step 2 correctly assembles 70.4% complete tracks (about 90.4% of all tracks with a maximum gap size of less than 14 frames).

Correctness of resulting tracklets To characterize the type of errors in our tracking results, we define a number of additional metrics. We counted detections that were incorrectly introduced into a track as *insertions*. Both tracking steps and the baseline inserted only one incorrect detection into another tracklet. Thus less than 1% of both detections and tracklets were affected.

We counted detections that were missing from a tracklet (and were replaced by a gap) as *deletions*. In the baseline, 32.2% of all detections were missing from their corresponding track (94.6% of all tracks had at least one deletion). After the first step, 1.38% of detections

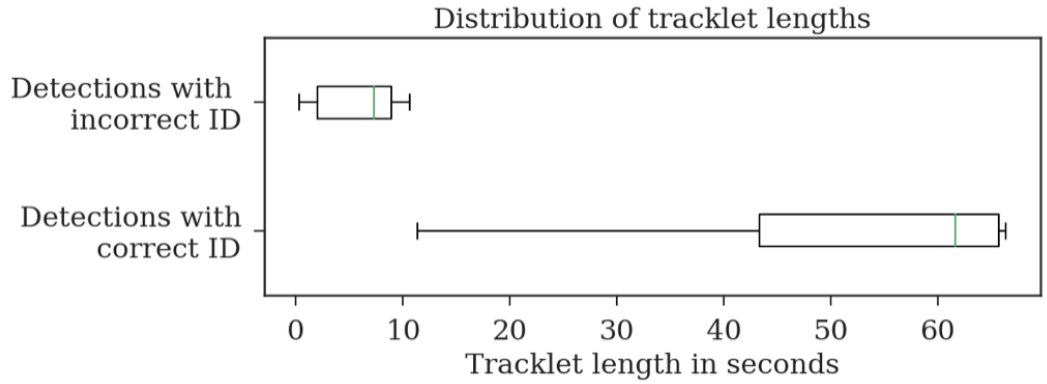

Figure 3.7: Evaluation of the tracklet lengths of incorrectly assigned detection IDs after the second tracking step reveals that all errors in the test dataset *2015.2* happen in very short tracklets. Note that this dataset covers a duration of around one minute.

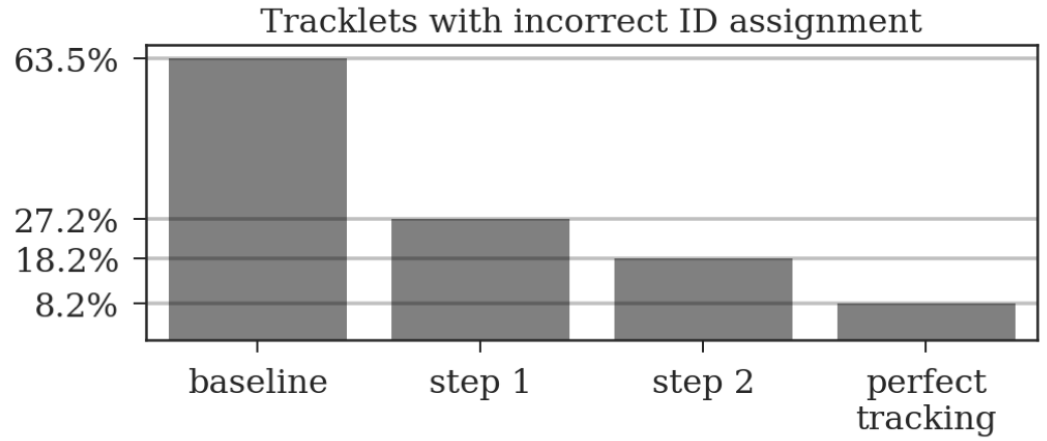

Figure 3.8: A naive tracking approach using only the detection IDs would result in around 64% of all tracks being assigned an incorrect ID. Our two-step tracking approach reduces this to around 27% and 18% respectively. Due to the short length of most incorrect tracklets, these 18.2% account for only 1.9% of the detections. Using our ID assignment method without any tracking errors would reduce the error to 8.2%.

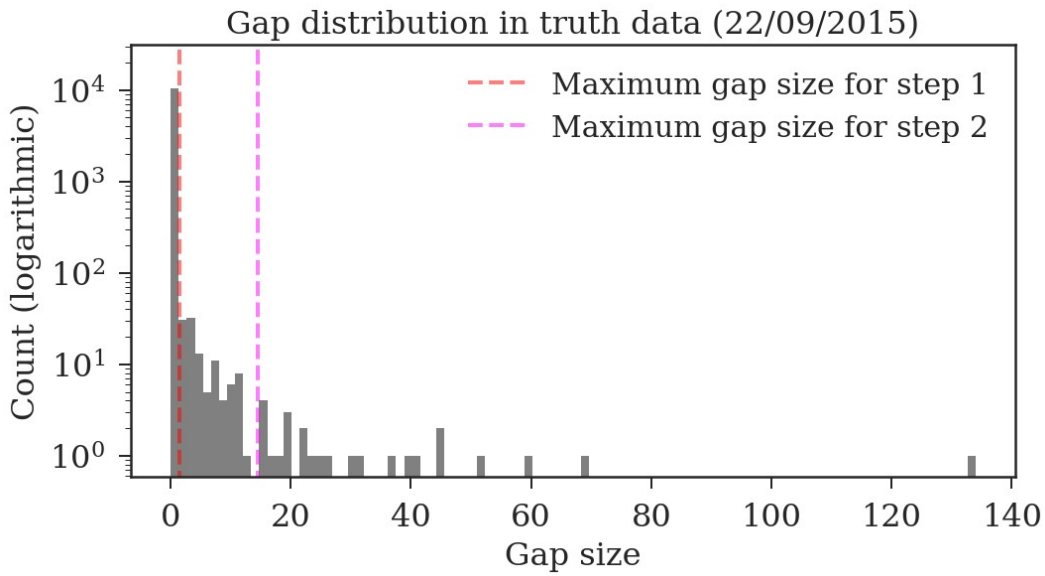

Figure 3.9: Distribution of the gap sizes in the ground truth dataset *2015.2*. Most corresponding detections (i.e. 97.9%) have no gaps and can be therefore be matched by the first tracking step. The resulting tracklets are then merged in the second step. The maximum gap size of 14 covers 99.76% of the gaps.

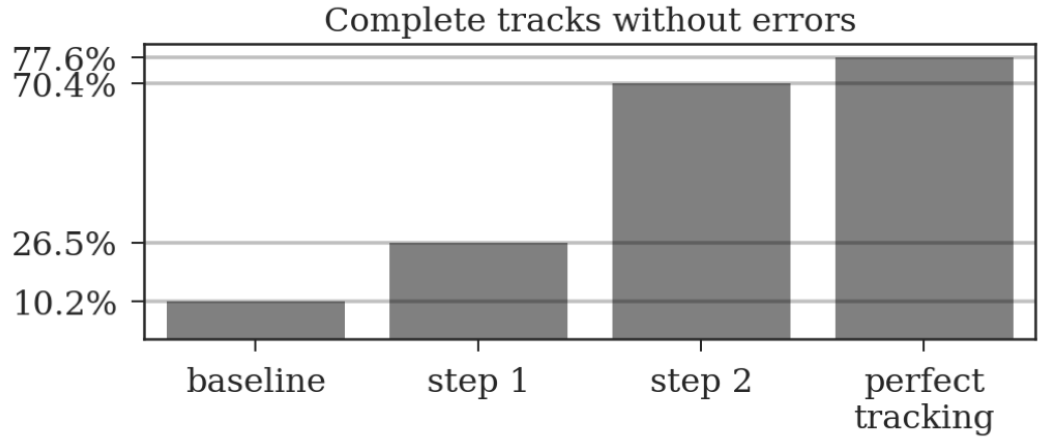

Figure 3.10: A *complete track* perfectly reconstructs a track in our ground truth data without any missing or incorrect detections. Even a perfect tracking that is limited to a maximum gap size of 14 frames could only reconstruct around 78% of these tracks. The naive baseline based only on the detection IDs would assemble 10% without errors while our two tracking steps achieve 26.5% and 70.4% respectively.

were missing from their track, affecting 26.7% of all tracks. After the second step, 2.37% of all detections and 18.25% of all tracks were still affected.

We also evaluated whether incorrect detections were contained in a track in situations where the correct detection would have been available (instead of a gap) as *mismatches*, but no resulting tracks contained such mismatches.

Length of resulting tracklets The ground truth datasets contain only short tracks with a maximum length of one minute. To evaluate the average length of the tracks, we also tracked one hour of data for which no ground truth data was available. The first tracking step yields shorter fragments with an expected length of 2:23 minutes, the second tracking step merges these fragments to tracklets with an expected duration of 6:48 minutes (refer to Figure 3.11 for tracklet length distributions).

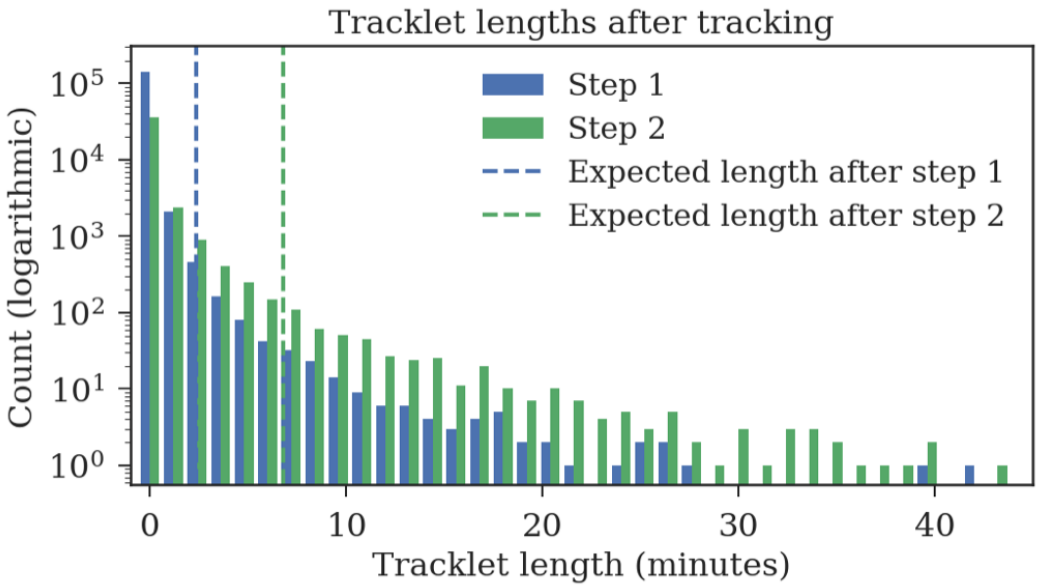

Figure 3.11: Track lengths after tracking one hour of video data at three frames per second. The expected length of a track is 2:23 minutes after the first step and 6:48 minutes after the second step.

3.7 Discussion

We have presented a multi-step tracking algorithm for fragmentary and partially erroneous detections of honey bee markers. We have applied the proposed algorithm to produce long-term trajectories of all honey bees in a colony of approximately 2000 animals. Our dataset

comprises 71 days of continuous positional data at a recording rate of 3 Hz. The presented dataset is by far the most detailed reflection of individual activities of the members of a honey bee colony. The dataset covers the entire lifespan of many hundreds of animals from the day they emerge from their brood cell until the day they die. Honey bees rely on a flexible but generally age-dependent division of labor. Hence, our dataset reflects all essential aspects of a self-sustaining colony, from an egg-laying queen and brood rearing young workers, to food collection, and colony defense. We have released a three days sample dataset for the interested reader (Boenisch et al., 2018b). Our implementation of the proposed tracking algorithm is available online[1].

The tracking framework presented in the previous sections is an essential part of the BeesBook system. It provides a computationally efficient approach to determine the correct IDs for more than 98% of the individuals in the honey bee hive without using extra bits for error correction.

Although it is possible to use error correction with 12 bit markers, this would reduce the number of coding bits and therefore the number of observable animals. While others chose to increase the number of bits on the marker, we solved the problem in the tracking stage. With the proposed system, we were able to reduce hardware costs for cameras and storage. When applied to the raw output of the image decoding step, the accuracy of other systems that use error-correction (for example Mersch et al. (2013)) may even be improved further.

Our system provides highly accurate movement paths of bees. Given a long-term observation of several weeks, these paths, however, can still be considered short fragments. Since the IDs of these tracklets are very accurate, they can now be linked by matching IDs only.

Still, some aspects of the system can be improved. To train our classifiers, we need a sufficiently large, manually labeled dataset. Rice et al. (2015) proposed a method to create a similar dataset interactively, reducing the required manual work. Also, the circular coding scheme of our markers causes some bit configurations to appear similar under certain object poses. This knowledge could be integrated into our ID determination algorithm. The IDs along a trajectory might not provide an equal amount of information. Some might be recorded under fast motion and are therefore less reliable. Other detections could have been recorded from a still bee whose tag was partially occluded. Considering similar readings as less informative might improve the ID accuracy of our method. Still, with the proposed method there are only 1.9% detections incorrectly decoded, mostly in very short tracklets.

The resulting trajectories can now be used for further analyses of individual honey bee behavior or interactions in the social network. In addition to the three day dataset published alongside this paper, we plan to publish two more datasets covering more than 60 days of recordings, each. With this data we can investigate how bees acquire information

[1] https://github.com/BioroboticsLab/bb_tracking

in the colony and how that experience modulates future behavior and interactions. We hope that through this work we can interest researchers to join the collective effort of investigating the individual and collective intelligence of the honey bee, a model organism that bears a vast number of fascinating research questions.

Conflict of Interest Statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Funding

FW received funding from the German Academic Exchange Service (DAAD). DD received funding from the Andrea von Braun Foundation. This work was in part funded by the Klaus Tschira Foundation. We also acknowledge the support by the Open Access Publication Initiative of the Freie Universität Berlin.

Acknowledgments

We are indebted to the help of Jakob Mischek for his preliminary work and his help with creating the ground truth data.

4 Social networks predict the life and death of honey bees

Benjamin Wild[1*]†, David M Dormagen[1]†, Adrian Zachariae[2]†, Michael L Smith[3,4,5], Kirsten S Traynor[1,6], Dirk Brockmann[2,7], Iain D Couzin[3,4,5], Tim Landgraf[1*]

[1] Department of Mathematics and Computer Science, Freie Universität Berlin, Berlin, Germany.
[2] Robert Koch Institute, Berlin, Germany.
[3] Department of Collective Behaviour, Max Planck Institute of Animal Behavior, Konstanz, Germany.
[4] Centre for the Advanced Study of Collective Behaviour, University of Konstanz, Konstanz, Germany.
[5] Department of Biology, University of Konstanz, Konstanz, Germany.
[6] Global Biosocial Complexity Initiative, Arizona State University, Tempe, USA.
[7] Institute for Theoretical Biology, Humboldt University Berlin, Berlin, Germany.

* Corresponding authors: b.w@fu-berlin.de, tim.landgraf@fu-berlin.de
† Equal contribution

4.1 Preface

Building upon the tracking system introduced in Chapter 3, the paper *Social networks predict the life and death of honey bees* proposes a method to assess the social network of honeybees and its effects on their development and behavior, without interfering with the colony. Social networks can be used to understand individual roles within a colony, but the relationship between individual roles and the social network is not well understood. This work introduces a succinct descriptor of an individual's social network, called *network age,* which can accurately predict task allocation, survival, activity patterns, and future behavior. A detailed analysis of the developmental trajectories of multiple cohorts of individuals in a natural setting using *network age* identifies distinct developmental pathways and critical life changes. Along with this work, the first ever large scale honey bee tracking dataset was released (Wild et al., 2021a), opening up a broad range of future studies. This dataset contains interaction matrices and metadata of all individuals in a single honey bee colony, including data on their interactions, age, location, and velocity, as well as subsampled

position information for each bee during the recording period. In summary, this work investigates the relationship between an individual's social network and its lifetime role within a complex honey bee society, and introduces a low-dimensional descriptor, *network age*, that allows for the prediction of task allocation, mortality, and behavioral patterns.

This chapter was previously published as:
(Wild et al., 2021b) — Benjamin Wild, David M. Dormagen, Adrian Zachariae, Michael L. Smith, Kirsten S. Traynor, Dirk Brockmann, Iain D. Couzin, and Tim Landgraf. Social networks predict the life and death of honey bees. *Nature Communications*, 12(1):1110, 2021b. URL https://doi.org/10.1038/s41467-021-21212-5

4.2 Author contributions

Conceptualization: BW, DMD, AZ, TL, MLS, KST; Methodology: BW, DMD, AZ, TL; Software: BW, DMD; Resources, supervision: TL, DB, IDC; Project administration: TL; Data curation: BW, DMD, MLS; Writing: BW, DMD, TL, MLS, KST, IDC; Visualization: BW, DMD.

4.3 Abstract

In complex societies, individuals' roles are reflected by interactions with other conspecifics. Honey bees (*Apis mellifera*) generally change tasks as they age, but developmental trajectories of individuals can vary drastically due to physiological and environmental factors. We introduce a succinct descriptor of an individual's social network that can be obtained without interfering with the colony. This 'network age' accurately predicts task allocation, survival, activity patterns, and future behavior. We analyze developmental trajectories of multiple cohorts of individuals in a natural setting and identify distinct developmental pathways and critical life changes. Our findings suggest a high stability in task allocation on an individual level. We show that our method is versatile and can extract different properties from social networks, opening up a broad range of future studies. Our approach highlights the relationship of social interactions and individual traits, and provides a scalable technique for understanding how complex social systems function.

4.4 Introduction

In complex systems, intricate global behaviors emerge from the dynamics of interacting parts. Within animal groups, studying interactions helps to elucidate the individuals' functions (Farine and Whitehead, 2015; Gordon, 2010; Krause et al., 2015; Pinter-Wollman et al., 2014). Descriptors of individuals derived from social interaction networks have been used to investigate e.g. pair-bonding (Psorakis et al., 2012), inter-group brokering (Lusseau and Newman, 2004), offspring survival (Cheney et al., 2016), cultural spread (Aplin et al., 2015; Claidière et al., 2013), policing behavior (Flack et al., 2006), leadership (Mehra et al., 2006; Sueur and Petit, 2008; Strandburg-Peshkin et al., 2018), organization of food retrieval (Planckaert et al., 2019), the ability to affect behavioral change (Rosenthal et al., 2015), and behavior during famine events (Sendova-Franks et al., 2010). As our ability to collect detailed social network data increases, so too does our need to develop tools for understanding the significance and functional consequences of these networks (Gomez-Marin et al., 2014).

Social insects are an ideal model system to study the relationship between social interactions and individual roles, because task allocation has long been hypothesized to arise from interactions (Gordon, 1996; Tofts and Franks, 1992; Traniello and Rosengaus, 1997). The relationship of individual roles within the colony and the social network, however, is not well understood. Individuals, for example, can modify their behavior based on nestmate interaction (Pinter-Wollman et al., 2013, 2011; Seeley, 1992; Gordon and Mehdiabadi, 1999), and interactions change depending on where and with whom individuals interact (Pinter-Wollman et al., 2011; Davidson and Gordon, 2017; Quevillon et al., 2015; Planckaert et al., 2019). These studies typically target specific types of interactions (e.g. food-exchange), specific roles within task allocation (e.g. foraging), or specific stimuli within the nest (e.g. brood), but an automatic observation system could capture behaviors and interactions within a colony more comprehensively and without human bias. Measuring the multitude of social interactions and their effect on behavior, and the social networks over the lifetime of individuals without interfering with the system (e.g., by removing individuals) is an open problem.

In honey bees, task allocation is characterized by temporal polyethism (Naug, 2008; Baracchi and Cini, 2014; Schneider and Lewis, 2004), where workers gradually change tasks as they age: young bees care for brood in the center of the nest, while old bees forage outside (Seeley, 1982; Huang and Robinson, 1996). Previous works often used few same-aged cohorts resulting in an unnatural age distribution (Naug, 2008; Baracchi and Cini, 2014; Gernat et al., 2018; Huang and Robinson, 1996). The developmental trajectory of individuals can, however, vary drastically due to internal factors (i.e. genetics, ovary size, sucrose responsiveness (Amdam and Omholt, 2003; Ihle et al., 2010; Pankiw and Page Jr., 1999; Scheiner, 2012; Wang et al., 2012, 2010; Münch et al., 2008)), nest state (i.e. amount of brood, brood age, food stores (Dreller et al., 1999; Seeley, 1989a; Traynor et al., 2015)), and the external environment (i.e. season, resource availability, forage success (Ament et al., 2010; Huang and Robinson, 1995; Toth et al., 2005; Wang et al., 2016)). These myriad influences on maturation rate are difficult to disentangle, but all drive the individual's behavior and task allocation. Due to the spatial organization of honey bee colonies, task changes also result in a change of location, with further implications on the cues that workers encounter (Seeley, 1982). How and when bees change their allocated tasks in a natural setting has typically been

assessed through destructive sampling (e.g. for measuring hormone titers of selected individuals), but understanding how all these factors combine would ideally be done in an undisturbed system.

With the advent of automated tracking, there has been renewed interest in how interactions change within colonies (Blut et al., 2017; Mersch et al., 2013), how spatial position predicts task allocation (Crall et al., 2018), and how spreading dynamics occur in social networks (Gernat et al., 2018). Despite extensive work on the social physiology of honey bee colonies (Seeley, 1995), few works have studied interaction networks from a colony-wide or temporal perspective (Gernat et al., 2018; Hasenjager et al., 2020). While there is considerable variance in task allocation, even among bees of the same age, it is unknown to what extent this variation is reflected in the social networks. In large social groups, like honey bee colonies, typically only a subset of individuals are tracked, or tracking is limited to short time intervals (Naug, 2008; Blut et al., 2017; Siegel et al., 2013; Bozek et al., 2021).

Tracking an entire colony over a long time would allow one to investigate the stability of task allocation. Prior research has shown that during each life stage, an individual spends most of its time in a specific nest region (Seeley, 1982; Johnson, 2010), interacting with nestmates, but with whom they interact may depend on more than location alone (e.g. previous interactions, or the genetic diversity within the colony (Farina, 2000; Girard et al., 2011)). Social interactions permit an exchange of information and can have long-term effects on an individual's behavior (Cholé et al., 2019). While honey bees are well-known for their elaborate social signals (e.g. waggle dance, shaking signal, stop signal (von Frisch, 1967; Nieh, 2010; Seeley et al., 2012)), they also exchange information through food exchange, antennation, or simple colocalization (Balbuena et al., 2012; Goyret and Farina, 2003). However, identifying an individual bee's role in a colony based on its characteristic patterns of interaction remains challenging, particularly with large numbers of individuals and multiple modes of interaction.

In this work we investigate the relationship between an individual's social network and its lifetime role within a complex society. We developed a tracking method for unbiased long-term assessment of a multitude of interaction types among thousands of individuals of an entire honey bee colony with a natural age distribution. We introduce a low-dimensional descriptor, network age, that allows us to compress the social network of all individuals in the colony into a single number per bee per day. Network age, and therefore the social network of a bee, captures the individual's behavior and social role in the colony and allows us to predict task allocation, mortality, and behavioral patterns such as velocity and circadian rhythms. Following the developmental trajectories of individual honey bees and cohorts that emerged on the same day reveals clusters of different developmental paths, and critical transition points. In contrast to these distinct clusters of long-term trajectories, we find that transitions in task allocation are fluid on an individual level. We show that the task allocation of individuals in a natural setting is stable over long periods, allowing us to predict a worker's task better than biological age up to one week into the future.

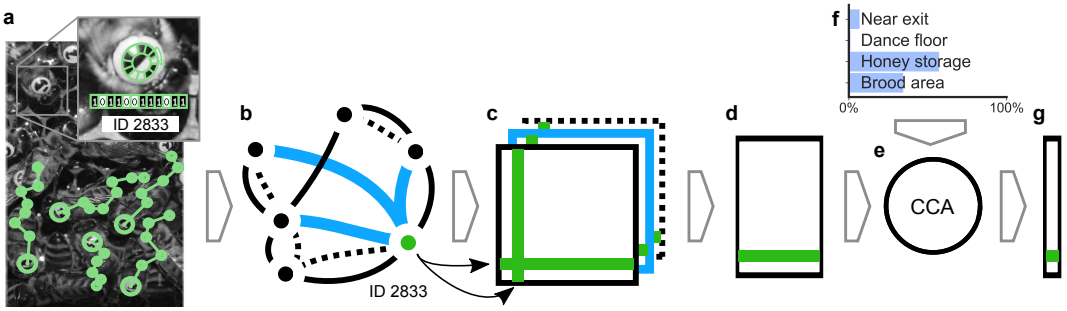

Figure 4.1: **Network age, a one-dimensional descriptor of an individual's role within the colony, based on an individual's interaction pattern.** Using the BeesBook automated tracking system, we obtain lifetime tracking data for individuals (**a**). These tracks are used to construct multiple weighted social interaction networks (**b**). We aggregate daily networks (**c**) to then extract embeddings that group bees together with similar interaction patterns, using spectral decomposition (**d**). Finally, we use a linear transformation (**e**; CCA = canonical-correlation analysis) that maximizes correlation with the fraction of time spent in different nest areas (**f**) to compress them into a single number per day called 'network age' (**g**).

4.5 Results

4.5.1 What is network age?

To obtain the social network structure over the lifetime of thousands of bees, we require methods that will track the tasks and social interactions of many individuals over consecutive days. We video recorded a full colony of individually marked honey bees (*A. mellifera*) at 3 Hz for 25 days (from 2016-08-01 to 2016-08-25) and obtained continuous trajectories for all individuals in the hive (Wario et al., 2015; Boenisch et al., 2018a). We used a two-sided single-frame observation hive with a tagged queen and started introducing individually tagged bees into the colony approximately one month before the beginning of the focal period (see Methods: Recording setup, data extraction, preprocessing for details). To ensure that no unmarked individuals emerged inside the hive, we replaced the nest substrate regularly (approx. every 21 days). In total, we recorded 1920 individuals aged from 0 days to 8 weeks.

A worker's task and the proportion of time she spends in specific nest areas are tightly coupled in honey bees (Seeley, 1982). We annotated nest areas associated with specific tasks (e.g. brood area or honey storage) for each day separately (see Methods: Nest area mapping and task descriptor), as they can vary in size and location over time (Smith et al., 2016). We then use the proportion of time an individual spends in these areas throughout a day as an estimate of her current tasks.

We calculated daily aggregated interaction networks from contact frequency, food exchange (trophallaxis), distance, and changes in movement speed after contacts (see Methods: Social networks). These networks contain the pairwise interactions between individuals over time. For each day and interaction type, we extract a compact representation that groups bees together with similar interaction patterns, using spectral decomposition (Belkin and Niyogi, 2003; von Luxburg,

2007). We then combine each bees' daily representations of all interaction types and map them to a scalar value (network age) that best reflects the fraction of time spent in the task-associated areas using CCA (canonical-correlation analysis; (Hotelling, 1936; Knapp, 1978)). Note that network age is solely a representation of the social network and not of location; the fraction of time spent in the task-associated areas is only used to select which information to extract from the social networks (e.g., by assigning higher importance to proximity contacts, see Section 4.13.1). Network age can still represent an individual's location, but only if this information is inherently present in the social networks. Network age thus compresses millions of data points per individual and day (1919 potential interaction partners, each detected $127\,501 \pm 50\,340$ times on average per day, with four different interaction types) into a single number that represents each bee's daily position in the multimodal temporal interaction network. Since CCA is applied over the 25 days of the focal period, network age can only represent interaction patterns that are consistent over time. See Figure 4.1 for an overview and Methods: Network age - CCA for a detailed description of the methods.

Network age is a unitless descriptor. We scale it such that 90% of the values are between 0 and 40 to make it intuitively comparable to a typical lifespan of a worker bee during summer, and because biological age is commonly associated with task allocation in honey bees. This scaling can be omitted for systems where behavior is not coupled with biological age.

4.5.2 Network age correctly identifies task allocation

Because of the inherent coupling of tasks and locations in a honey bee colony, we expect a meaningful measure of social interaction patterns to be correlated with the individual's spatial preferences. We quantify to what extent network age captures this correlation by using multinomial regression to predict the fraction of time each bee spends in the annotated nest areas (see Methods: Task prediction models and bootstrapping). Note that while we also used these spatial preferences to select which information to extract from the interaction networks, it is not certain whether the spatial information is contained in the social network in the first place, and how well a single dimension can capture it. Furthermore, the social network structure could vary over many days with changing environmental influences, preventing the extraction of a stable descriptor. The regression analysis allows us to compare different variants of network age to biological age as a reference.

To evaluate the regression fit, we use McFadden's pseudo R^2 scores R^2_{McF} (McFadden et al., 1973). Network age is twice as good as biological age at predicting the individuals' location preferences, and therefore their tasks (network age: median $R^2_{McF} = 0.682$, 95% CI [0.678, 0.687]; biological age: median $R^2_{McF} = 0.342$, 95% CI [0.335, 0.349]; 95% CI of effect size [0.332, 0.348], N=128; likelihood-ratio χ^2 test $p \ll 0.001$, N=26 403, Table 4.2, see Methods: Statistical comparison of models for details). Network age provides a better separability of time spent in task-associated nest areas than biological age (Figure 4.2a, example cohort in Figure 4.2c, Section 4.13.1 for all cohorts). Network age correlates with location because of the inherent coupling between tasks and nest areas. Still, it is not a direct measure of location: bees with the same network age can exhibit different spatial distributions and need not directly interact (see Section 4.13.2).

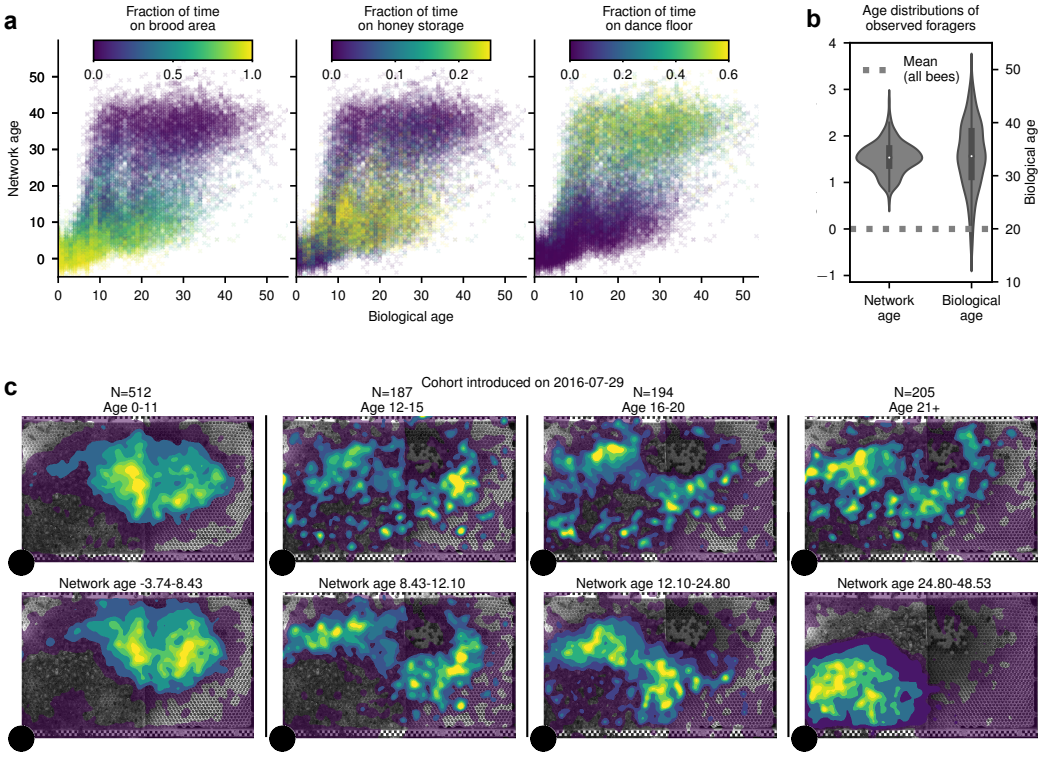

Figure 4.2: **Network age is an accurate descriptor of task allocation. a** The proportion of time spent on task-associated locations in relation to biological age and network age with each cross representing one individual on one day of her life. For a given value on the y-axis (network age) colors are more consistent than for a given value on the x-axis (biological age). **b** Z-transformed age distributions for known foragers visiting a feeder (N=40 observed individuals). The variance in biological age is greater than the variance in network age (boxes: center dot, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range). Corresponding biological ages are also shown on the right y-axis (original biological age: 34.2 $\pm$ 7.9, original network age: 38.3 $\pm$ 4.6; mean $\pm$ standard deviation). **c** Spatial distributions of an example cohort over time (bees emerged on 2016-07-29, 64 individuals over 25 days), grouped by biological age (top row) versus network age (bottom row). Note how network age more clearly delineates groups of bees than biological age, with bees transitioning from the brood nest (center of the comb), to the surrounding area, to the dance floor (lower left area). The shaded areas depict density percentiles (brightest to darkest: 99%, 97.5%, 95%, 80%, 70%, 20%).

While we can improve the predictive power of network age by extracting a multi-dimensional descriptor instead of a single value (see Methods: Network age - CCA, Methods: Task prediction models and bootstrapping for details), the improvements for additional dimensions are marginal

compared to the difference in predictiveness between the first dimension of network age and biological age (see Table 4.2). This implies that a one-dimensional descriptor captures most of the information from the social networks that is relevant to the individuals' location preferences and therefore their tasks.

We experimentally demonstrated that network age robustly captures an individual's task by setting up sucrose feeders and identifying workers that foraged at the feeders (known foragers, N=40, methods in Methods: Forager groups experiment). We then compared the biological ages of these known nectar foragers to their network ages. We made these two quantities comparable by z-transforming them because they do not have the same unit of measure. As expected, foragers exhibited a high biological age and a high network age, whereas biological age exhibited significantly larger variance than network age (Figure 4.2b; Levene's test (Levene, 1960), performed on z-transformed values: $p \ll 0.001$, N=200). Indeed, while we observed a forager as young as 12 days old, that individual had a network age of 25.5, demonstrating that network age more accurately reflected her task than her biological age (z-transformed values: biological age -0.46; network age 0.61).

Tagging an entire honey bee colony is laborious. However, by sampling subsets of bees, we find that network age is still a viable metric, even when only a small proportion of individuals are tagged and tracked. With only 1% of the bees tracked, network age is still a good predictor of task (median $R^2_{McF} = 0.516$, 95% CI [0.135, 0.705], N=128) while increasing the number of tracked individuals to 5% of the colony results in a R^2_{McF} value comparable to the fully tracked colony (5% of colony tracked: median $R^2_{McF} = 0.650$, 95% CI [0.578, 0.705], N=128; whole colony tracked: median $R^2_{McF} = 0.682$, 95% CI [0.678, 0.687], N=128; see Section 4.13.2). Similarly, we find that an approximation of network age can be calculated without annotated nest areas: Network age can be extracted in an unsupervised manner using PCA on the spectral embeddings of the different interaction type matrices (median $R^2_{McF} = 0.646$, 95% CI [0.641, 0.650], N=128, see Methods: Network age - PCA).

4.5.3 Developmental changes over the life of a bee

Network age reveals differences in interaction patterns and task allocation among same-aged bees (Figure 4.3a). After around six days of biological age, the network age distribution becomes bimodal (see Methods: Quantifying when bees first split into distinct network age modes). Bees in the functionally old group (high network age) spend the majority of their time on the dance floor, whereas same-aged bees in the functionally young group (low network age) are found predominantly in the honey storage area (Figure 4.3a). Transitions from high to low network age are a rare occurrence in our colony (see Section 4.13.2).

We attribute the split on the population level to distinct patterns of individual development. Clustering the time series of network ages over the lives of bees identifies distinct developmental paths within same-aged cohorts. We set the number of clusters to three as this is the minimum number of clusters that separates an early and a late transition from low to high network age in all tested cohorts (see Methods: Network age transition clustering for further details). In the cohort that emerged on 2016-08-01, the first developmental cluster (blue, Figure 4.3b) rapidly transitions to a high network age (likely corresponding to foraging behavior) after only 11 days.

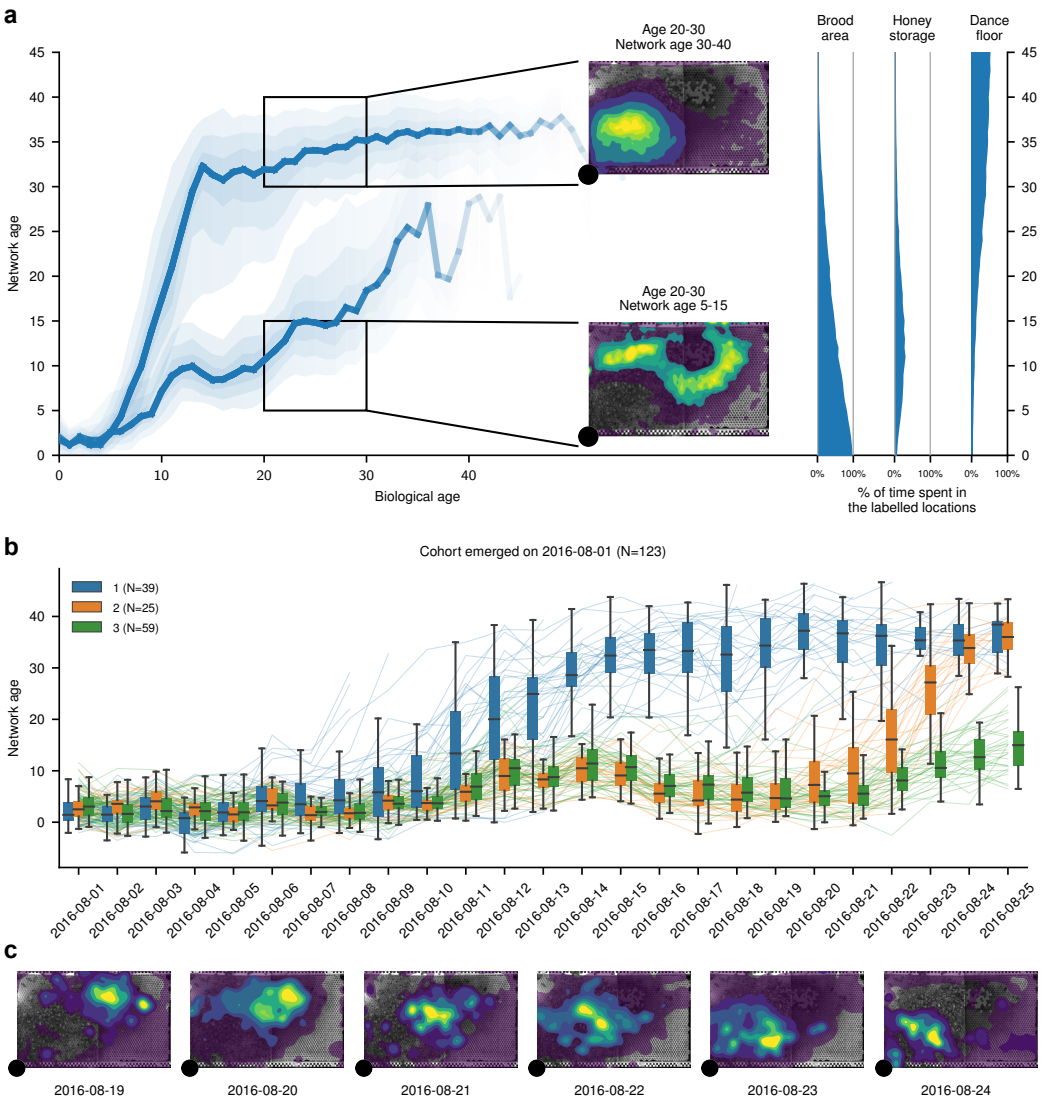

Figure 4.3: **Network age reveals distinct developmental paths. a** Left: The median of network ages over biological ages for all individuals that lived more than 11 days split by a threshold on the age of 11 (T=23.07, calculated using Otsu's method (Otsu, 1979).The upper line contains all bees that fall above this threshold (N=832), the lower contains all bees below that threshold (N=563). The shaded areas depict 20%, 40%, 60% data intervals. We observe a split in network age corresponding to different tasks: The upper heatmap (network age 30-40, biological age 20-30, 577 bees, 857 283 data points) corresponds to the dance floor, while the lower heatmap (network age 5-15, biological age 20-30, 381 bees, 742 622 data points) borders between dance floor and brood nest. Right: The mean fraction of time a bee with a given network age spends on our annotated regions throughout a day. **b** Lines depict the network age of individual bees of a same-aged cohort with the colors indicating clusters of their network age over time. Boxes summarize bees belonging to each cluster for a given day (center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range). **c** Heatmaps showing the spatial distribution of bees in the developmental cluster 2 (orange) from 2016-08-19 to 2016-08-24. The smooth transition in network age (orange in line plot, **b**) from one mode to another corresponds to a smooth transition in spatial location (heatmaps, **c**). The shaded areas depict density percentiles (brightest to darkest: 99%, 97.5%, 95%, 80%, 70%, 20%).

The second cluster (orange) transitions at around 21 days of biological age, while bees in the third cluster (green) remain at a lower network age throughout the focal period. We see similar splits in developmental trajectories for all cohorts, although the timing of these transitions varies (see Methods: Network age transition clustering for additional cohorts). Such divergence in task allocation has been previously shown in bees; factors that accelerate a precocious transition to foraging include hormone titers (Robinson and Ratnieks, 1987), genotype (Pankiw and Page Jr., 1999), physiology, especially the number of ovarioles (Amdam et al., 2004), and sucrose response threshold (Scheiner et al., 2004).

The transition from low to high network age over multiple days is characterized by a gradual shift in the spatial distribution (see example in Figure 4.3c), highlighting that an individual's task changes gradually. The network age of most bees is highly repeatable (median $R = 0.612$ 95% CI=[0.199, 0.982], see Methods: Repeatability for details), indicating task stability over multiple days. Both findings (gradual change over a few days and high repeatability) are consistent with the dynamics of the underlying physiological processes, such as vitellogenin and juvenile hormone, that influence task allocation and the transition to foraging (Amdam and Page Jr, 2010).

4.5.4 Network age predicts an individual's behavior and future role in the colony

Network age predicts task allocation (i.e. in what part of the nest individuals will be) up to ten days into the future. Knowing the network age of a bee today allows a better prediction of the task performed by that individual next week than her biological age informs about her current tasks (Figure 4.4c, binomial test, $p \ll 0.001$, N=55 390, 95% CI of effect size [0.055, 0.090], N=128, see Methods: Future predictability for details). We confirm that this is only partially due to network age being repeatable (see Methods: Future predictability). We do note, however, that our ability to predict the future tasks of a young bee is limited, especially before cohorts split into high and low network age groups (Figure 4.3a). This limitation hints at a critical developmental transition point in their lives, an attractive area for future study.

We explicitly optimized network age to be a good predictor of task-associated locations. However, we find that network age predicts other behaviors better than biological age, including an individual's impending death (network age: median $R^2 = 0.165$, 95% CI [0.158, 0.172], versus biological age: median $R^2 = 0.064$, 95% CI [0.059, 0.068]; 95% CI of effect size [0.037, 0.039], N=128, likelihood-ratio χ^2 test $p \ll 0.001$, N=26 403). Biologically-young but network-old bees have a significantly higher probability of dying within a week (80.6% N=139) than do biologically-old but network-young bees (42.1% N=390; χ^2 test of independence $p \ll 0.001$ N=529; see Section 4.13.3 for details). This is likely because a biologically-young bee with a high network age, i.e. a bee that starts to forage earlier in life and faces more perils imposed by the outside world, is more likely to die than a bee of the same age with a low network age. This finding is consistent with previous work showing increased mortality with precocious foraging (Perry et al., 2015; Rueppell et al., 2008).

We measure movement patterns of individual bees such as daily and nightly average speed, the circadian rhythm, and the time of an individual's peak activity. While these properties are related to task allocation due to the diurnal nature of foraging, they are not direct measures of an

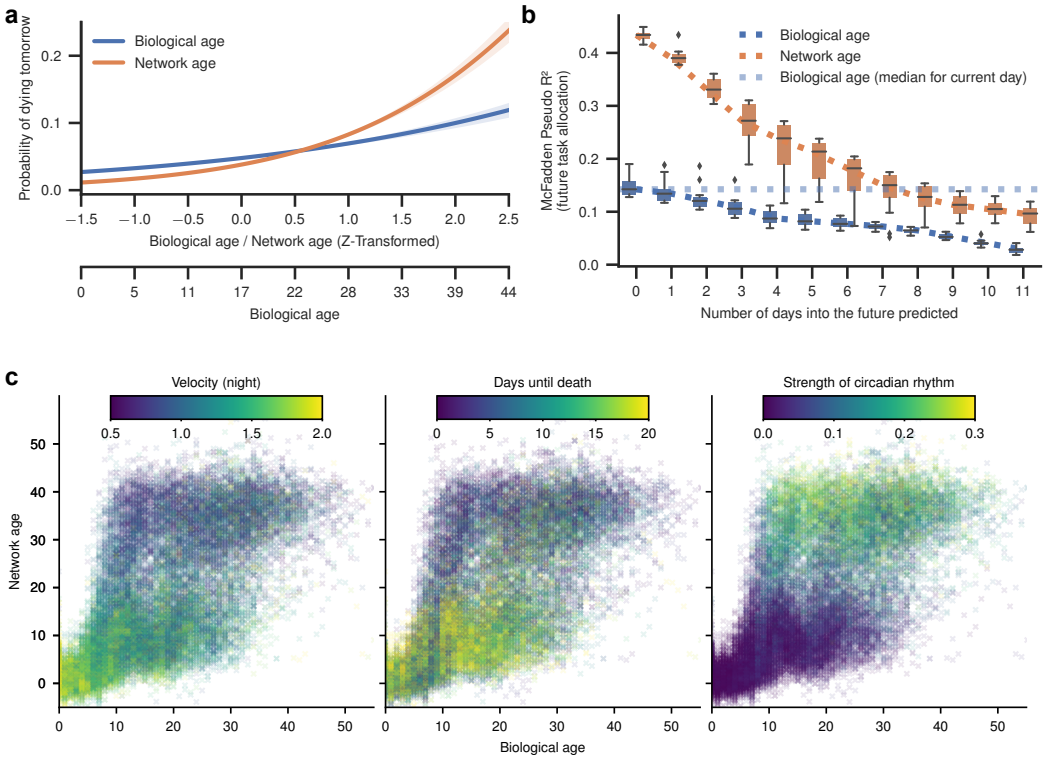

Figure 4.4: **Network age can be used to predict other properties, such as mortality and circadian rhythms. It also predicts an individual's future task allocation. a** An individual's mortality on the next day based on her age (x-axes show original and z-transformed biological age and z-transformed network age). Bees with a low network age have lower mortality than biologically young bees; bees with a high network age have higher mortality than biologically old bees (shaded areas: 95% bootstrap confidence intervals for the regression estimates). **b** Network age can be used to predict task allocation and future behaviors. Network age predicts the task of an individual 7 days into the future better than biological age predicts the individual's task the same day (blue dotted line). Each box comprises N=12 scores from models with N=12 days of training data (center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range; points, outliers). **c** Selected properties mapped for network age over biological age with each cross representing one individual on one day of her life. Note that for a given value on the y-axis (network age) colors are more consistent than for a given value on the x-axis (biological age).

individual's location. Network age also captures these movement patterns better than biological age (likelihood-ratio χ^2 test $p \ll 0.001$, N=26 403, see Table 4.1b for 95% CI of effect sizes).

To investigate whether network age is a good predictor of future task allocation and behavior only because it captures the spatial information contained in the social network, we repeat the analyses above using the time spent in task-associated locations as independent variables. We find that network age, even though it was extracted using this spatial information as a guide, is still a better predictor of an individuals' behavior (for all dependent variables likelihood-ratio χ^2 test $p \ll 0.001$, N=26 403, see Location (1D) in Section 4.13.3 and Table 4.1d for 95% CI of effect sizes). This difference in predictive power suggests that the multimodal interaction network contains more information about an individual than spatial information alone.

While we focus on predicting tasks from network age, we can control the information we extract from the observed social networks and derive variants of network age better suited for other research questions. By replacing the 'task associated location preferences' in the final step of our method with 'days until death', we extracted a descriptor that captures social interaction patterns related to mortality. This descriptor improves the prediction of the individuals' death dates by 31% compared to network age (median increase in $R^2 = 0.05$; 95% CI [0.04, 0.06] N=128, see Methods: Targeted embedding using CCA and Table 4.1c), opening up novel social network perspectives for studies such as the risk factors of disease transmission. Similarly, we extracted descriptors optimized to predict the movement patterns introduced in the last paragraph (for all except 'Time of peak activity' likelihood-ratio χ^2 test $p \ll 0.001$, N=26 403, see Table 4.1c for 95% CI of effect sizes). These targeted embeddings provide precise control over the type of information we extract from the social networks and extend the network age method to address other important research questions in honey bees and other complex animal societies.

4.6 Discussion

Combining automated tracking, social networks, and spatial mapping of the nest, we provide a low-dimensional representation of the multimodal interaction network of an entire honey bee colony. While many internal and external factors drive an individuals' behavior, network age represents an accurate way to measure the resulting behavior of all individuals in a colony non-invasively over extended periods.

We use annotated location labels to select which information to extract from the social network, but stress that network age can only contain information inherent in the social network. Therefore, the predictive power of network age demonstrates that the social interaction network by itself comprehensively captures an individual's behavior. We show that network age does not only separate bees into task groups, such as foragers and nurses, but also allows us to follow maturing individuals as they develop. A recent work derived a social maturity index in colonies of the social ant *C. fellah* (Richardson et al., in press), highlighting a strong separation of nurses and foragers in the social network and high variability in transition timing. Similarly, network age is a fluid measure and the age at which individuals change between the task groups is highly variable. However, we find distinct clusters of developmental trajectories at the colony level, with some groups entering critical developmental transitions earlier in life than others. Further investigating

the precise combination of internal and external factors that drive those transitions is a promising direction for future research.

These transition points are also reflected in changes in nest location, because spatial preferences, task allocation, and interactions are inherently coupled in honey bees. However, we show that network age is more than just a representation of location: Bees with the same network age do not necessarily share a location in the nest, and the time spent in task-associated locations is less predictive of an individual's current and future behavior than network age. Additionally, we calculate a variant of network age that is not guided by auxiliary spatial information but instead extracts the information with the highest variance from the social networks (Network age PCA). The PCA variant is still predictive of task allocation, suggesting that location is the predominant signal in the social network. However, the higher predictive power of the CCA network age variant and the targeted embeddings indicates that there is more information in the social network that our method can extract.

In this study, we extract network age from daily aggregated interaction networks, and thereby disregard potentially relevant intraday information. Furthermore, honey bees have a rich repertoire of interaction behaviors, of which we only capture a subset. The inclusion of intraday data or additional interaction types could reveal further differences between individuals (e.g., the temporal aspects of intraday interaction networks can disentangle the contribution of different modes of interactions (Hasenjager et al., 2020)). While we study one colony in this work, we observe thousands of individuals and many overlapping cohorts. Our findings, in particular the existence of distinct developmental trajectories, and the fluidity and long-term stability in task allocation on an individual level, are consistent in all cohorts in our study. While some details, e.g. the timing of developmental transitions, might depend on environmental circumstances, we believe that these results transfer to other colonies. There is no straight-forward extension of the method to extract a common embedding of social networks that do not share individuals (e.g., over different experimental treatments or repetitions). Still, specific hypotheses can be tested using network age as long as a treatment group is compared to a control group within each trial. For example, while the meaning of specific values of network age can differ slightly between colonies, a group treated with pesticides could show differences in development, as measured by network age, relative to a control group from the same colony. Analyzing how network age changes within a day, over other datasets with possibly other types of interactions, or how network age shifts in response to disease pressure or experimental manipulation of age demography would be potentially fruitful areas for future investigation, as previous work has shown that there is a relationship between pathogens and interaction behaviors (Naug, 2008; Smith, 2012; Lecocq et al., 2016; Geffre et al., 2020).

Network age can be repurposed and extended for other research questions: We show that (1) variants of network age capture different aspects from the social networks related to mortality, velocity, or circadian rhythms, and (2) with a subsample of only 5% of the bees in the colony, we can extract a good representation of the social network. This makes the method applicable to systems with far more individuals or with much less required experimental effort for a comparable number of individuals. Network age could be calculated in real-time, opening up a wide range of possibilities for future research: For example, it would be possible to selectively remove bees that have just begun a developmental change to determine their influence on colony-wide task allocation. Sequencing individual bees could determine how known internal drivers of behavioral

transition, like the double-repressor co-regulation of vitellogenin and juvenile hormone (Amdam and Omholt, 2003), are reflected in the social network. Our perspective captures both internal and external influences that impact social interactions and is thus applicable to all complex systems with observable multimodal interaction networks. Network age can be adapted to questions that explore social interaction patterns independent of age and division of labor, making it broadly applicable to any social system. As such, our method will permit future research to analyze how complex social animal groups use and modify interaction patterns to adapt and react to biotic and abiotic pressures.

4.7 Methods

4.7.1 Recording setup, data extraction, preprocessing

We set up our observation hive on 2016-06-24, with a queen and approximately 2,000 young bees (*Apis mellifera*) sourced from a local host colony. To obtain newly emerged bees, we incubated brood from the host colony, and later from the observation colony in an incubator at 34°C. Freshly emerged bees were marked every weekday. All bees were removed from the brood comb each day before marking, so the maximum age in each batch of bees was 24 hours. After removing the hair from the bees' thoraxes with a wet toothpick, we applied shellac onto the thorax and attached a curved, circular tag. The number of bees marked per batch varied, but never exceeded 156. Marked bees were introduced to the colony through a backdoor entrance. After the initial marking period (27 days, starting 2016-06-28), the video recording was started 2016-07-24 and stopped 2016-09-19. Marking newborn bees continued approximately twice a week with the latest introduction on 2016-08-23. A total of 3,166 individuals were marked. We recorded 1920 individuals from 30 cohorts during the focal period. See Figure 4.13 for the number of bees that were alive on each day. Bees had free access to the outside environment via a tube connected to the observation hive. We use the BeesBook (Wario et al., 2015) recording setup (scaffold, cameras, lighting, storage, marking procedure), however, for the experiments described here, we used custom-built IR flash circuit boards triggered by an Arduino controller, synchronized with the high-res cameras (Mönck, 2021). Combs were imaged at 3 Hz, alternating between sides of the observation hive, to avoid low contrast due to backlighting.

A total of 46 TB of video data was recorded and continuously moved to a network storage unit at the North-German Supercomputing Alliance (HLRN). After the recording season, the data was processed to detect and decode the bee markers (Wild et al., 2018) and track these detections through time (Boenisch et al., 2018a). The output data consists of timestamps, planar positions, three-dimensional rotations, IDs, and confidence scores, for the decoded IDs.

Two cameras were used for each side of the nest with both viewports overlapping partially. Six reference points were marked on each comb side such that four points were visible in each of the cameras' recordings with the center two points visible in both cameras. Reference points were identified and their image coordinates were extracted manually. The coordinates were then used to calculate the homography between the comb and the image plane. The homography was then used to rectify the tracking data which translated image coordinates to a metric reference frame, i.e. the nest surface. See Figure 4.14 for a schematic of the setup.

Resulting tracking data was post-processed before entering the analysis. We discarded detections with low decoding confidence, i.e. detections that the machine vision pipeline could not reliably decode. Remaining implausible detections (e.g. of IDs that had not been tagged yet) were removed in an additional filter step. The distribution of the number of detections of all IDs is strongly bimodal, the larger mode representing those bees which actually are in the observation hive, and the smaller mode representing erroneous decodings of tags which are not present on the given day. We use Otsu's method (Otsu, 1979) to automatically determine the threshold which best separates those two modes and filter out all potentially incorrect IDs. The tracking data, therefore, contains gaps due to falsely filtering out correct detections, but also due to occlusions, such as when bees inspect cells or depart the nest on foraging trips.

4.7.2 Forager groups experiment

Between 2016-07-28 and 2016-08-22, foragers were trained to a feeder (see Figure 4.15 for a photo of the feeder) offering unscented sucrose solution by gradually moving it from the colony (52.457 032, 13.296 635) to a sequence of locations 'F1' to 'F4' (see Table 4.3). For days over which the feeder was moved, high sugar concentration was used and iteratively changed to control the number of new foragers. Once the final locations were reached, the feeder offered the highest concentration for 1-2 hours per day. After a minimum of three days, training to the next location in the list was resumed. We photographed all bees landing at the location and manually transcribed the identity and time of arrival. A list of foragers visiting the feeder is given in Table 4.4. The network age values around the day we first observed each bee at the feeding side is given in Figure 4.16.

4.7.3 Bayesian lifetime model

The death date of an individual could ideally be computed as the first date she was not detected in the hive. Unfortunately, this does not work in practice for two reasons. First, tags are sometimes incorrectly decoded, and because of the number of detections we have for each day, this means that most IDs will be detected at least a couple of times per day. Second, some bees were not visible at all on some days, even though they are not dead yet (see Figure 4.17 for an example).

We, therefore, use a Bayesian changepoint model to robustly estimate the death dates of all individuals. An individual is defined to be alive on all days since she emerged and was introduced into the colony (day e) up to the change point $d = e + l$, where l is the number of days she was alive. We use a weakly informative prior $N(35, 50)$ for the number of alive days l. We model the probability that a bee is detected at least as often as a threshold t while she is alive and less often than t when she is not alive, using a Bernoulli distribution. We use a $Beta(5, 1)$ prior for this probability because we know that, typically, an alive bee will have many detections. For the threshold t we use an informative $Beta(25, 1)$ prior because we know that a dead bee will have very few detections, if any. Note that we normalize the detection counts to $[0, 1]$ when fitting the model, i.e. for each bee we divide the counts of daily detections by the maximum count of detections of that tag over the entire recording period. We sample this model using pymc3 and the NUTS sampler (Salvatier et al., 2016). For each bee, we compute 2,000 tuning samples and 1,000 samples. The date of death is determined using the mean of those last 1,000 Monte Carlo samples.

4.7.4 Social networks

Proximity interaction network Two bees were defined to be in proximity if their tags were less than 2 cm from each other (~1.4 body lengths) over at least 0.9 seconds (three frames with our recording frame rate). We construct affinity networks based on the counts of these proximity interactions without taking the duration of each contact into account to reduce the effect of bees resting next to each other.

Euclidean proximity networks Euclidean proximities were determined for each pair of bees when both bees were visible. The daily average distance d between two individuals was then transformed to two affinity matrices, the first derived by applying a Gaussian similarity function $d' = (-d^2/2\gamma^2)$ with $\gamma = max(D)/4$ and the second by subtracting from the maximum distance ($d' = max(D) - d$). D is the matrix of all daily average distances on the same day.

Trophallaxis networks We constructed an interaction network representing trophallaxis interactions (food exchange). To filter our data to detect trophallaxis events, we use a two-step approach. We first use a fast logistic regression with low precision to discard most of the non-trophallaxis encounters. We then use a slower convolutional neural network to further refine the results with higher recall.

To train the two classification models, we manually labeled bee interactions in our dataset by observing video sequences. To increase the fraction of positive events, we queried our data for bees that are close to each other, and approximately facing each other. Note that we did not distinguish the directionality or different types of trophallactic interactions.

This ground truth data contains 140 trophallaxis events out of the distinct 2,651 events in total. For some events, we annotated a begin and end timestamp and could, therefore, use multiple frames for the training. In total, we had 25,835 training samples, each consisting of a pair of bee IDs, a timestamp, and a label (trophallaxis / not trophallaxis).

Because the prefiltering of the training data can introduce a sampling bias we created another test set by labeling all possible interactions in 33 randomly sampled frames, containing a total of 15 trophallaxis events and 39,051 negative events (we use every pair of bees with a thorax distance of less or equal than 3 cm as a possible candidate). This test set represents our data distribution without any bias.

In the classification step, we look at all pairs of bees with their thoraxes at a distance between 0.731 cm and 1.204 cm (i.e. the 99th percentile of the positive events in the training data) together in a frame. For a pair of bees (i, j) we have the locations of the thorax on the hive in millimeters (xy_i, xy_j) and their orientations (α_i, α_j). We calculate the approximate head position h_i as $xy_i + d * [cos(\alpha_i), sin(\alpha_i)]$ where d=3.19 mm. We calculate their relative orientation as $[cos(\alpha_i), sin(\alpha_i)] \cdot [cos(\alpha_j), sin(\alpha_j)]^T$. We then perform logistic regression on the euclidean distance of their thorax locations, the euclidean distance of their head locations, and their relative orientation.

The logistic regression was trained on the manually labeled samples, setting the threshold to get a recall of 85% at a precision of 21% (on a 20% validation test). This regression discards 62% of the true negative samples (i.e. the specificity). For the remaining data points, those that were classified as possible trophallaxis by the first classifier, we extract trajectories of both bees for around

5 seconds (15 frames) around the possible trophallaxis events. We then use a convolutional neural network, again trained on the manually labeled data.

Evaluated on the test set, the two combined filters yield a recall of 60% at a precision of 47%, discarding 99.97% of negative samples (i.e. the specificity).

Interaction effect networks For each proximity interaction with a duration no longer than 60 seconds (to exclude bees resting next to each other) and with a minimum gap of at least 5 seconds since the last interaction of the same two individuals, we compute the difference in mean velocities within 30 second time windows before and after the interaction. This is done for both partners, and so we derive four networks based on the mean and cumulative changes, each split into negative and positive values. We use separate networks for the positive and negative values because this allows us to define affinity matrices which can only have positive edge weights.

Temporal aggregation and post-processing of networks Time-aggregated networks were constructed by defining the weighted edge strength as the number of times two individuals were in proximity or engaged in trophallaxis. The networks were aggregated over 24 hours without overlap. Edges in both networks are undirected. For subsequent analyses, all networks were represented as a square adjacency matrix with each element i, j representing the affinity of bee i with bee j on this day, given by the interaction mode (e.g. for the network of trophallaxis counts, a high value represents many trophallaxis interactions between the two individuals). Each matrix is then preprocessed using a rank transform and normalized such that 0 represents the lowest affinity and 1 the highest affinity. Ties are resolved by assigning the same rank to identical affinities.

4.7.5 Nest area mapping and task descriptor

We manually outlined the capped brood area and visible honey storage cells for every day in background images from 2016-07-30 until 2016-09-05. To obtain the open brood area, we calculated the area of the comb that would become capped within 8 days. We extracted the background images by extracting the first frame from every video we recorded over a specific day (approximately one image every 5.6 minutes), and then applying a rolling median filter with window size 10 to these images. We then calculated the modal pixel value, for every pixel, over all the median images. For each side of the comb, we stitched together the background images from the two cameras on that side.

To get the approximate location of the dance floor, we used the detected waggle runs of our waggle dance detection system (Wario et al., 2017) that had high confidence (>=0.9) (see Figure 4.18). As nearly all waggle detections happened on one side of the comb, we exclusively labelled this area as the dance floor. We then fitted an ellipse to the detections using scikit-image (van der Walt et al., 2014), which we scaled manually to not intersect with the exit area. The dance floor area was consistent throughout the experiment, so in cases where we did not have waggle dance data for a given day, we interpolated the dance floor area over the adjacent days. Finally, we used a kaiser window applied over the consecutive days (window size=5, beta=5) to smooth the annotations. We considered the region 7.5 cm around the exit tube as the nest region close to the exit.

To generate a task descriptor for every bee, we fetched one high confidence detection (>0.9) per bee for every minute of a day. We then counted how many of these detections per bee fell into the annotated regions. Then we normalized these counts per bee to 1 by dividing through the sum. This descriptor, therefore, contains the fraction of time each individual spends in each of the annotated regions. Data points outside of the annotated areas are ignored for this descriptor but are used in all other parts of this work. For all evaluations, we consider the brood area region to be the sum of the annotated open and closed brood cell regions. See Figure 4.18 for an example of the annotations.

4.7.6 Network age - From networks to spectral embeddings to CCA

Network age is derived from the raw interaction matrices using spectral decomposition and canonical correlation analysis (CCA). For each day and interaction mode, the graph of interactions between all bees that were alive (see Methods: Bayesian lifetime model for the definition of alive bees) on that day is retrieved as an adjacency matrix as described in Methods: Social networks.

For each preprocessed affinity matrix, spectral embeddings (Belkin and Niyogi, 2003) are calculated using the python package scikit-network. We compute the first eight embedding dimensions (see Section 4.13.2 for an evaluation of the performance of different numbers of embeddings).

For the non-symmetric interaction effect matrices, we use bispectral decomposition (Abdi, 2007) to obtain one set of embeddings each for the rows and for the columns, to represent the two directions of an interaction.

For different days the eigenvectors of the embeddings and therefore the embedding values themselves can have an inverted sign. To correct this, we flip the sign of the values if the Spearman correlation between consecutive days is negative.

For every day, we now have a high-dimensional embedding per bee. We reduce the dimensionality further by applying canonical-correlation analysis (CCA). We use CCA to find a linear transformation of the network embeddings to a three-dimensional vector that maximizes the correlation to a projection of the bees' task descriptors (as introduced in Methods: Nest area mapping and task descriptor). We use the CCA implementation in scikit-learn (Pedregosa et al., 2011). We use the first dimension of this vector as 'network age' throughout this paper, but also evaluate multidimensional variants (Network age 2D, Network age 3D) in Methods: Task prediction models and bootstrapping, Methods: Statistical comparison of models, and Methods: Prediction of other behavior-related measures.

For every dimension and day, we use robust scaling based on the 5th and 95th percentile of the network age distribution. A network age of zero corresponds to the 5th percentile and 40 corresponds to the 95th percentile. This stabilizes the distribution over time and also maps the values to a range comparable with the biological age of honey bees. We note that this scaling slightly improves the prediction of task allocation, but that the method also works without it. We enforce that the 5th percentile of network age corresponds to bees with a lower biological age than the 95th percentile such that biological age and network age have the same directionality.

4.7.7 Network age - Unsupervised variant using Principal Component Analysis

We also calculate a variation of network age that does not require the annotated location descriptors. Instead of applying CCA to the concatenated spectral embeddings (see Methods: Network age - CCA), we instead use Principal Component Analysis (PCA) to reduce the dimensionality. This unsupervised network age still predicts task allocation better than biological age (see Section 4.13.2, Methods: Task prediction models and bootstrapping for details).

4.7.8 Task prediction models and bootstrapping

To evaluate how well biological age and the different variants of network age represent an individual's task allocation, we use these measures as features to predict the proportion of time individuals spend in the brood area, dance floor, honey storage and near the exit (see Methods: Nest area mapping and task descriptor for details on the nest area mapping). We evaluate the areas individually and in combined models. We evaluate different complexities of models (linear vs. nonlinear) and different independent variables (e.g. network age and biological age).

To test different complexities of the relationships, we evaluate both a generalized linear model (GLM, the default model) and a small neural network consisting of two fully connected layers (listed as 'nonlinear' in Table 4.2) for each of the combinations of independent and dependent variables. The hidden layer of the neural network has a dimensionality of 8 and uses tangens hyperbolicus as its nonlinearity.

To evaluate the performance of the models for each area separately we select a sigmoid as the link function of the GLM and the activation function of the neural network's last layer. We then optimize and calculate the likelihood of the data assuming a binomial distribution.

We also evaluate both models to simultaneously predict all four values of an individual's task allocation distribution. To this end, we choose a softmax function as the link function of the GLM and the neural network's final activation function. We then optimize and calculate the likelihood of the data assuming a multinomial distribution.

For all the combinations of independent and dependent variables, we repeat the described procedure for 128 bootstrap samples. For each model, we retrieve the final likelihood of the data L_1. We use PyTorch (Paszke et al., 2019) and the L-BFGS optimizer to obtain maximum likelihood estimates of the models. We also always fit a null-model only consisting of the intercepts and retrieve its likelihood L_0. For each model and bootstrap iteration, we calculate McFadden's pseudo R^2 (McFadden et al., 1973) as $R^2_{McF} = 1 - ((L_1)/(L_0))$. We then calculate the median and 95% confidence intervals from these bootstrap samples. See Table 4.2 for an overview of the results for all evaluated models. We test the significance of these results separately with the tests described in Methods: Statistical comparison of models.

4.7.9 Statistical comparison of models

We use bootstrapped confidence intervals of the effect strength to investigate whether a model based on one feature (e.g. network age) explains the dependent variables (e.g. task allocation distributions) significantly better than the same model based on a different feature (e.g. biological

age). Additionally, we use a likelihood ratio χ^2 test to answer whether one feature (e.g. network age) provides additional information over biological age in a combined model.

Bootstrapped confidence intervals of the effect strength We draw 128 bootstrap samples of the combined daily bee data. For each sample, we calculate either the McFadden's pseudo R^2 in the case of the task allocation models (see Methods: Task prediction models and bootstrapping, Methods: Future predictability) or the R^2 in the case of the other measures (see Methods: Prediction of other behavior-related measures, Section 4.13.3) for both a model based on biological age and the independent variable we want to compare with (e.g. network age). For each of these paired samples, we calculate the difference in scores of the two models. From these 128 differences, we calculate a two-sided 95% confidence interval of the effect strength. If the null hypothesis (difference in scores is zero or less) is not contained in the confidence interval, we can reject the null hypothesis at an alpha level of 2.5%.

Likelihood ratio test As the likelihood ratio test requires a nested model for testing, we compare a model based solely on biological age with a model based on a combination of biological age and the independent variable we want to compare with (e.g. network age).

We fit each model to the data and calculate the likelihoods of the data under the fitted models (L_1 for the combined model and L_0 for the model based on biological age). The likelihood ratio is given by $LR = -2 * ln(L_0/L_1)$. If the null hypothesis that the models are equal were true, LR would approximately follow a chi-squared distribution with k degrees of freedom (with k=4 in the case of the task allocation model from Methods: Task prediction models and bootstrapping and k=1 in case of the general regression model for Methods: Prediction of other behavior-related measures, Section 4.13.3). We use the cumulative density function of the chi-squared distribution to calculate the p-value.

4.7.10 Repeatability

We calculate the repeatability R of the network age consisting of repeated measurements over several days of an individual I as $R(I) = Var_p/(Var_i + Var_p)$ with Var_i being the variance of the network age of an individual measured over the available days and Var_p being the variance of the mean network ages of a control group. The control group consists of all bees inside the same age span as I on all days on which the network age values for I were collected. A repeatability close to 1 means that the individual variance is low compared to the population variance. A repeatability close to 0 means that the individual variance outweighs the population variance.

4.7.11 Network age transition clustering

In order to cluster the transitions of different individuals in a cohort, we first collect the network ages for every individual in a feature vector where each entry corresponds to the individual's network age for one day. We then do a linear inter- and extrapolation for missing values (e.g. due to absence or the individual dying). For the cohort of bees, we calculate the euclidean distance between each individual's feature vector. Then we perform a hierarchical clustering using Ward's

method (Ward, 1963) using the Python library scikit-learn (Salvatier et al., 2016) and extract the first three clusters. See Figure 4.19 for an example of the clustering. See Figure 4.20 for the network age development of different cohorts and Figure 4.21 for all bees.

We fixed the number of clusters to three for visualization purposes as that is the minimum number that showed a lagged transition from low to high network age for all cohorts. In hierarchical clustering, cutting off the dendrogram of the agglomerative clustering at a deeper level and thus increasing the number of clusters will further subdivide the existing clusters. Figure 4.22 gives an example with N=5 clusters.

4.7.12 Quantifying when bees first split into distinct network age modes

We used KMeans to cluster the network age distribution of every day into two distinct clusters corresponding to the two modes. Then we check for every bee that we observed at least once as a young bee below the age of 6 (N=1079) at which age she first gets assigned to the higher cluster (mean=12.33, 95% CI = [6, 25.7], median=11, N=572). We ignore bees that are never assigned to the upper cluster (e.g. because we do not observe them for a long enough timespan). See Figure 4.23 for the distribution of biological ages.

4.7.13 Definition of circadian rhythmicity

The motion velocity of a bee was determined by dividing the euclidean distance between two consecutive detections by the time passed (a multiple of $\frac{1}{3}$ seconds). Any duplicate IDs were discarded. The velocity was median filtered with a kernel size of 3 to remove outliers. See Figure 4.24 for an example of the velocity of a specific bee over multiple days.

Lomb-Scargle periodograms were computed for all individuals remaining in the dataset at any time in the period from 2016-07-20 to 2016-09-18. For each day and individual, the Lomb-Scargle periodogram was calculated on the motion velocities over an interval of three days, i.e. including the preceding and following day. The circadian activity was confirmed as strong peaks at a period of 1 day. Lomb-Scargle periodograms were computed using the Astropy package (Collaboration et al., 2018). See Figure 4.24 for an example of a bee's velocity and the resulting Lomb-Scargle periodogram.

In the following analyses, we reduced computational load by fitting a single sine wave of fixed frequency = 1/d (least-squares fit). For each fit, we extract the power as $P(f) = 1 - (SSE_{sine}/SSE_{constant})$ with SSE_{sine} being the residuals (sum of squared errors) of the sine fit and $SSE_{constant}$ being the residuals (sum of squared errors) of a constant model assuming the mean of the data.

The power, hence, reflects how much of the velocity variation can be explained by the sinusoidal oscillation, or circadian rhythm.

4.7.14 Targeted embedding using CCA

We show that we can extract targeted embeddings from the spectral factors of the interaction networks that are better in predicting other properties of the individuals. To compute those targeted embeddings, we exactly follow the methodology outlined in Methods: Network age - CCA, but for each property (days until death, time of peak activity, circadian rhythm, day- and nighttime velocities) we extract a one-dimensional embedding from the spectral factors that maximizes the correlation with this property using canonical-correlation analysis.

4.7.15 Prediction of other behavior-related measures

We follow the same methodology as described in Methods: Task prediction models and bootstrapping to evaluate how well the various age measures explain various properties of the individuals with a slight adjustment: We choose the identity function as the link function of the generalized linear model and as the activation function of the neural network's final layer. We model the residual distribution as a normal distribution with constant variance.

We define the 'days until death' as the number of days left in an individual's life on a given day (for a description of the automatically determined death dates see Methods: Bayesian lifetime model). The time of peak activity and the rhythmicity of daily movement were calculated as described in Methods: Definition of circadian rhythmicity.

For the daytime and nighttime velocities we use the same data as for the circadian rhythmicity (see Methods: Definition of circadian rhythmicity). For the daytime velocities we use the mean of all collected velocities between 09:00-18:00 UTC of the three-day rolling window; for nighttime velocities, 21:00-06:00 UTC.

See Table 4.1a for an overview of the scores of different models and targets.

4.7.16 Future predictability

We evaluate how well we can predict future task allocation using network age and biological age. To ensure that no information leak can occur, we only used supervised information from the past and test it on future data. To do this, we first calculate the spectral factors for the entire dataset as described in Methods: Network age - CCA. The factors are computed for each day separately, hence no information can leak from the past to the future. We then determine the mapping from spectral factors to network age using CCA, but only on a fixed range of days prior to the validation window. We fix the number of days in the train set to 12 days so that we always have approximately the same amount of training data independent of the number of days we predict into the future. Similar to the linear mapping given by CCA, we also determine the parameters of the regression model only on the train dataset from the same fixed time window. We train separate models for all viable ranges of dates and for prediction from one to 11 days into the future. The linear mapping given by the CCA and the predictive models are then applied to the spectral factors on the held-out validation set from a time after the training data set (see Figure 4.25 for an overview about the data handling). For this analysis, we want to evaluate how well we can predict task allocation into the future. We estimate the effect size in R^2_{McF} by calculating the 95% CI for the different

time windows (see Figure 4.25, median improvement in $R^2_{McF} = 0.080$, 95% CI [0.055, 0.090], N=12).

Because the two compared models are not nested, the likelihood ratio test does not apply here. We perform a paired binomial test using the null hypothesis that the improvement in mean squared error in task allocation prediction is zero or less. We find that we can predict the task allocation of an individual seven days into the future with a lower mean squared error (paired binomial test, $p \ll 0.001$, N=55 390) using network age.

This is mostly caused by older bees. For young bees, there is a low amount of variance in task allocation and network age is about as good in describing task allocation for young bees as biological age. Furthermore, we find that we can not reliably predict future task allocation for young bees, suggesting that either the social networks in this study are not predictive for this task or that the future task allocation of young bees is driven by other factors. See Figure 4.26 for an overview of the results.

The predictive power of network age cannot be fully explained by the bimodal distribution of network age and organization of stable 'work groups' within the colony. We perform an additional analysis to reject this repeatability null hypothesis by comparing how well the task allocation prediction using networks works when simply shifting the prediction for the current day into the future. We find that a model fitted to predict future task allocation outperforms this null model considerably (See Figure 4.27).

4.8 Ethics Statement

German law does not require approval of an ethics committee for studies involving insects.

4.9 Data Availability

The data (interaction networks and metadata) that support the findings of this study are available in zenodo with the identifier 10.5281/zenodo.4438013 (Wild et al., 2021a).

4.10 Code Availability

The code is available as open source software under the MIT license in zenodo with the identifier 10.5281/zenodo.4441807 (Wild and Dormagen, 2021).

4.11 Acknowledgements

We are indebted to numerous students, in particular Fernando Wario Vázquez, Franziska Lojewski, Andreas Jörg, Leon Sixt, Hauke Mönck, Maria Sparenberg, Sascha Witte, Alexa Schlegel, Mathis Hocke and Andreas Berg for providing hardware and software parts of the BeesBook system. We thank Peter Knoll and Randolf Menzel for providing the honey bee colony and valuable advice. Giovanni Galizia and Jake Graving gave critical feedback which helped improve the manuscript.

Our work was supported by the HPC-Service of ZEDAT (Freie Universität Berlin) and the North-German Supercomputing Alliance who generously provided computing resources and storage capacity.

DMD received funding from the Andrea von Braun Foundation, and the Elsa-Neumann-Scholarship. MLS is a Simons Foundation Postdoctoral Fellow of the Life Sciences Research Foundation, and received funding from the Heidelberg Academy of Science and the Zukunftskolleg Mentorship Program. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824069. This work was in part funded by the Klaus Tschira Foundation Grant 00.300.2016. IDC and MLS acknowledge support from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC 2117 - 422037984 and IDC acknowledges support from NSF Grant IOS-1355061 and Office of Naval Research Grants N00014-09-1-1074 and N00014-14-1-0635. KST acknowledges support of the Wissenschaftskolleg zu Berlin.

4.12 Competing interests

Authors declare no competing interests.

4.13 Supplementary Information

4.13.1 What is network age?

Contributions of the interaction types on network age

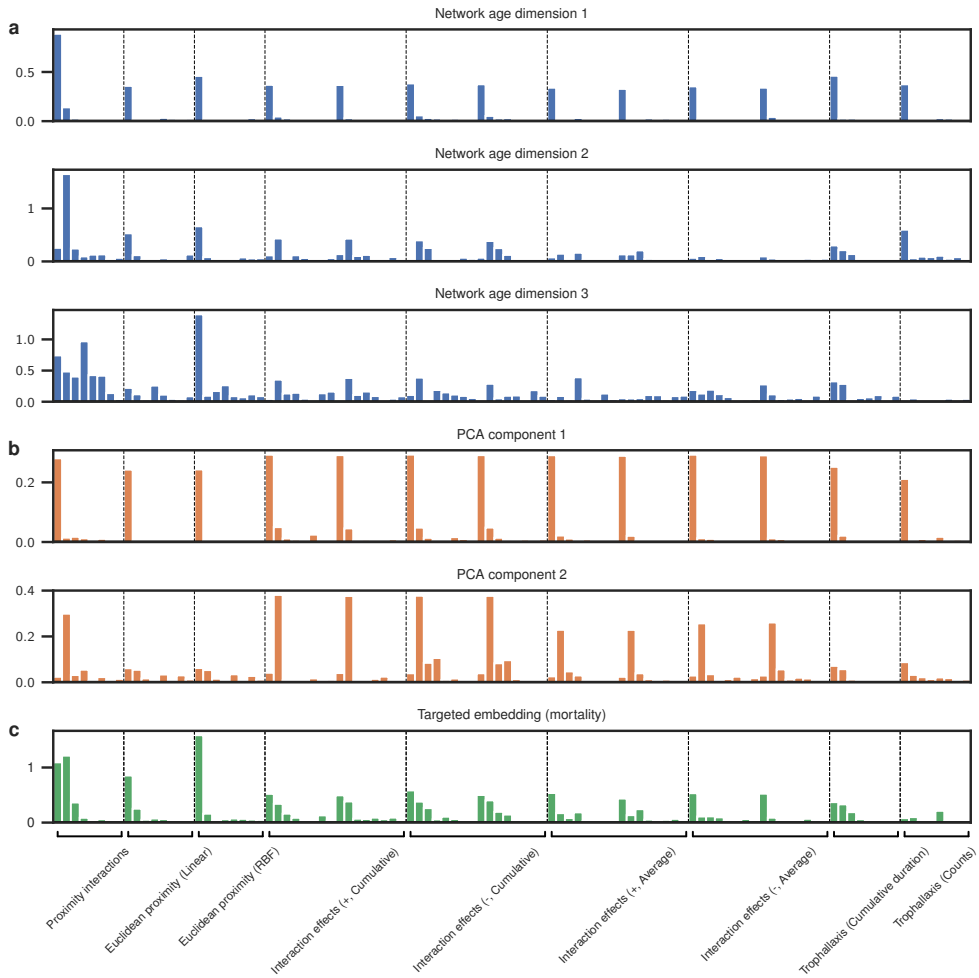

Figure 4.5: **Relative contributions of spectral factors to network age.** For each spectral factor derived from the interaction matrices for the different interaction types, we calculate the relative contributions for the multidimensional variant of network age (**a**, blue), network age PCA (**b**, orange), and the targeted embedding for mortality (**c**, green). A high value signifies that the corresponding factor has a large impact on network age.

To understand the relative contributions of the different interaction types on network age, we can visualize the coefficients of the linear mapping from the spectral factors learned by the CCA

(or PCA, for the network age PCA variant). This also allows us to compare the three dimensions of network age 3D against each other and the PCA variant.

There is a strong collinearity between the spectral decomposition factors derived from the different interaction matrices (because the interaction themselves are strongly correlated, e.g. the number of proximity interactions is correlated with the number of trophallaxis interactions). Therefore, directly interpreting the CCA weights would yield inaccurate results and we first have to decorrelate the spectral embeddings. We apply a PCA to the spectral embeddings, similar to the network age PCA variant, but retain all principal components to which we then apply the CCA. We then have two linear mappings, one from the CCA and one from the PCA. Sequential linear mappings can be combined into one, allowing us to attribute the coefficients from the CCA to the spectral factors, i.e. to measure to what extent each factor contributes to network age:

For spectral factors $F \in \mathbb{R}^{N\times S}$ and the linear mappings $M^{PCA} \in \mathbb{R}^{S\times S}$ and $M^{CCA} \in \mathbb{R}^{S\times 3}$, we calculate the contributions $C \in \mathbb{R}^{S\times 3}$: $C = S^{-1}|M^{PCA}M^{CCA}|$. S is the total number of spectral factors (see Methods: Network age - CCA). N is the number of bee days, i.e. we have one set of spectral factors for each individual per day. We use the absolute value of the contributions because the sign of the spectral factors is not informative.

We see that all interaction types contribute to network age: The first dimension of network age is mainly based on the first spectral factor of each interaction type, with the proximity interactions being slightly more important. Therefore, the first factor of the spectral decomposition alone is already a good representation of task allocation. We see a similar pattern in the first principal component of the PCA, which also explains the similarly high predictiveness of the PCA variant regarding task allocation. In contrast, the second dimension of network age is very different from the PCA's second dimension. This shows that the CCA can select the relevant information from the spectral embeddings.

For the targeted embedding of a bee's mortality, we find that the CCA selects a more diverse set of factors than the first dimension of network age and the first component of the PCA. This shows that different aspects of the social network are relevant for different research questions and explains why the PCA version of network age is much worse when predicting mortality (see Table 4.1).

Location heatmaps

To visually assess how distinctive biological age and network age are with respect to the spatial distribution of individuals, we computed location heatmaps for single-age cohorts tracked over the experimental window (25 days). For each individual within the cohort, we collected their daily biological age, network age, and their positions on the comb (as in Methods: Nest area mapping and task descriptor). Positional data for each day (a 'bee day') was then assigned to a heatmap as follows: For the age heatmaps, we manually defined age intervals based on the age-thresholds given by Seeley (Seeley, 1982) , with an additional split for middle-aged bees for which we observed a high variance. We then assigned bee days according to these thresholds based on the bees' ages on the different days. The network age thresholds for the plots have been set in a way to keep the number of bee days the same for the two plots in each column and can, therefore, differ for different cohorts. For example, the cohort introduced on 2016-08-01 consisted of 123 bees. In the heatmap for age 0-11, there would be N=1394 bee days over the focal period in which the bees were below 12 days

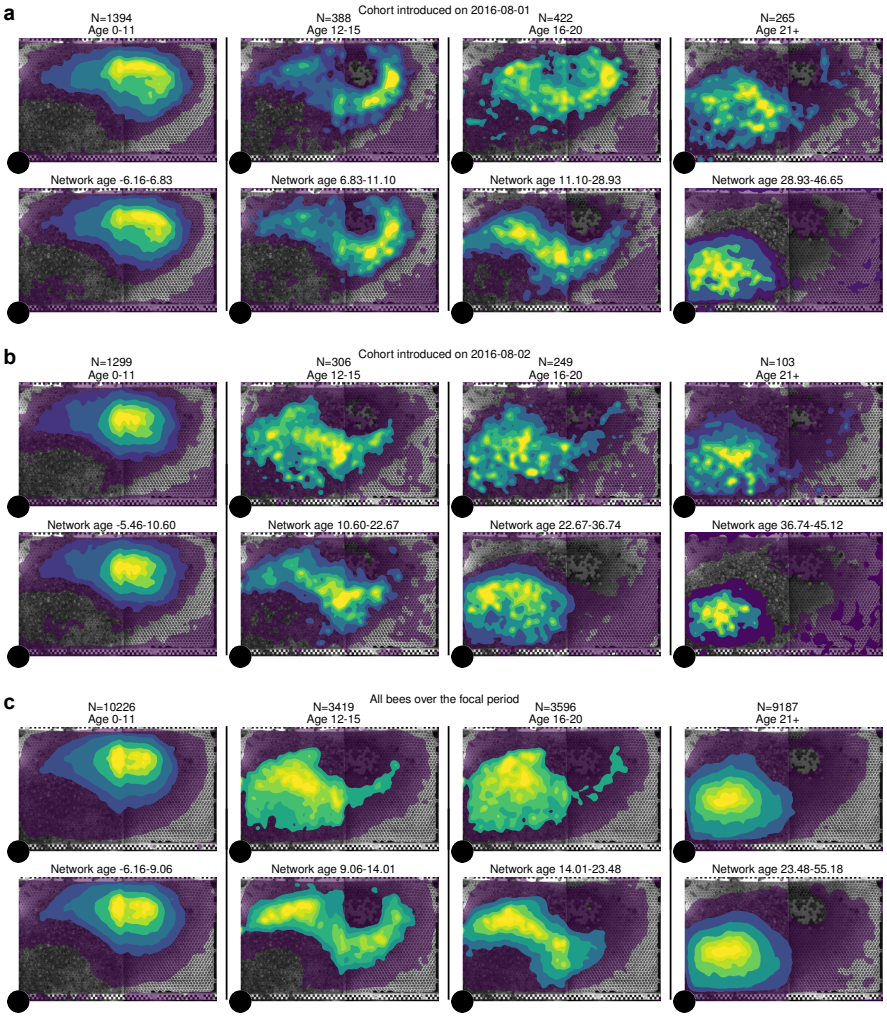

Figure 4.6: **Spatial heatmaps of bees grouped by biological age and network age. a** Cohort of 2016-08-01 (N=123 bees). **b** Cohort from 2016-08-02 (N=119 bees). **c** All bees over the focal period (N=1912 bees). The shaded areas depict density percentiles (brightest to darkest: 99%, 97.5%, 95%, 80%, 70%, 20%).

old (some bees having died). Then, we calculated the first network age thresholds in such a way, that they would also include 1394 bee days.

The shaded areas depict density percentiles (brightest to darkest: 99%, 97.5%, 95%, 80%, 70%, 20%).

See Figures 4.6a and 4.6b for heatmaps showing the distributions of different example cohorts that we introduced at the beginning of the focal period and Figure 4.6c for all bees.

Network age distributions over time

While the network age distribution stays within similar bounds over time, we find that the shape of the distribution (e.g. the proportion of individuals in the lower vs higher percentiles) changes significantly over the experimental window, likely reflecting changes in task allocation (e.g. more individual engaged in brood care initially and more foragers after most brood cells are capped). See Figure 4.7 for distributions over time.

We find no strong evidence of correlations of these changes in distribution with external factors such as the weather, but this may be due to the limited size of the recording window and is an interesting direction for potential future research.

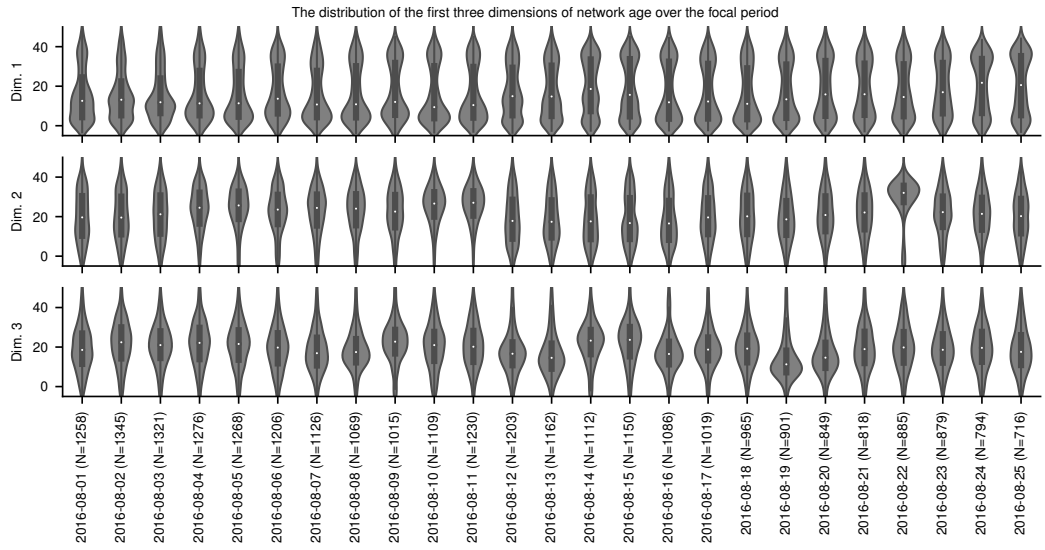

Figure 4.7: **Violin plots of network age distributions over the focal period.** The first three dimensions of network age are depicted. Each violin comprises one dimension of the network age values of all individuals on a given day (the number of individuals on each day is given in the x-axis labels; boxes: center dot, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range).

4.13.2 Network age correctly identifies task allocation

Spatial separation of bees with similar network age

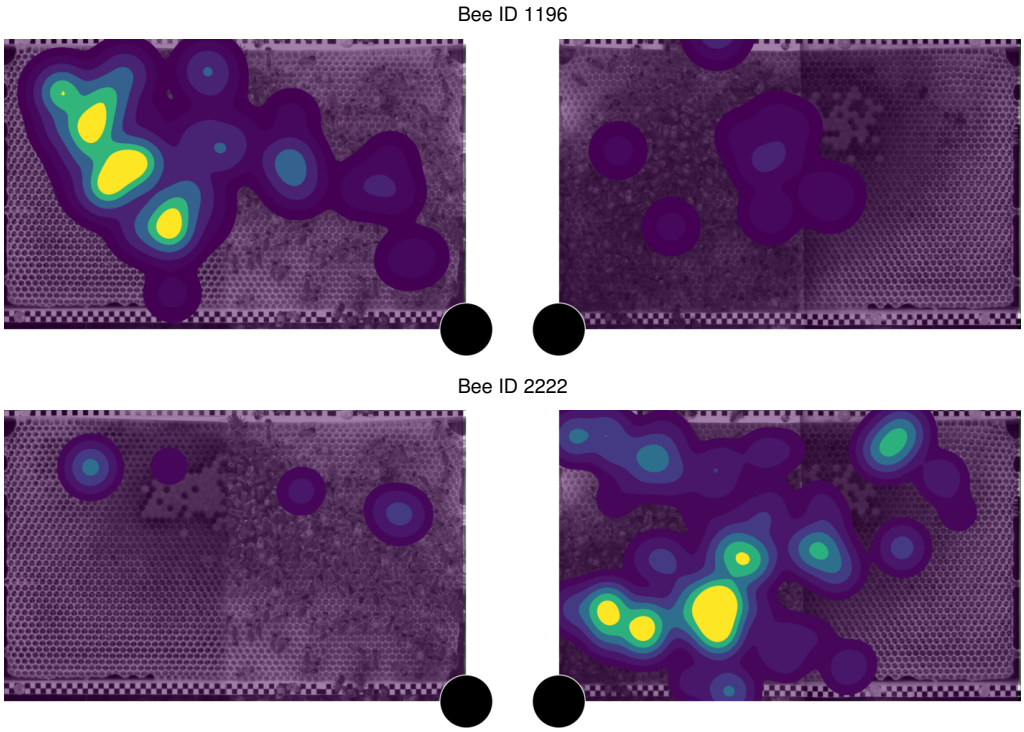

Figure 4.8: **Two bees on 2016-08-05 with similar network ages despite occupying different sides of the comb.** The bee with the ID 1196 is 22 days old and has a network age of 15.09. Her heatmap consists of 1189 samples, each corresponding to one minute on the focal day. The bee with the ID 2222 is 8 days old and has a network age of 14.34. Her heatmap consists of 1284 samples, each corresponding to one minute on the focal day. The black circle marks the location of the entrance. The shaded areas depict density percentiles (brightest to darkest: 99%, 97.5%, 95%, 80%, 70%, 20%).

While the network age of bees is correlated with their locations, we do find examples of bees that have approximately the same network age but occupy different locations on the comb (see Figure 4.8 for an example). We also see that, while two bees having a similar network age is correlated with an increase in proximity interactions, there are many bees with close network ages that do not encounter each other inside the hive (see Figure 4.9).

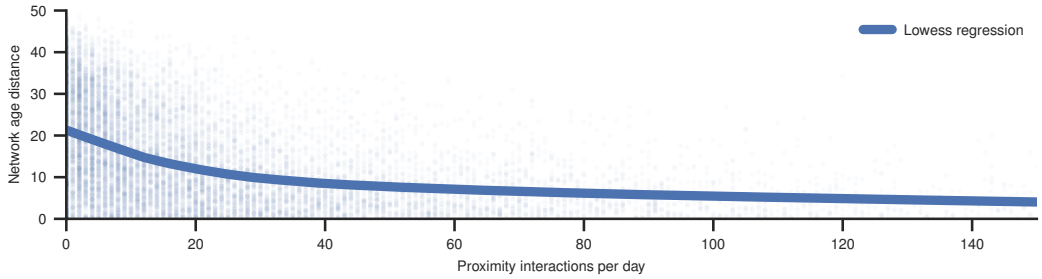

Figure 4.9: **The relationship between distance in network ages and number of proximity interactions for pairs of bees.** Close network ages generally mean more interactions. However, there are many individuals with a close network age and very few interactions. Each point indicates one pair of individuals on one day. The line is given by a lowess regression.

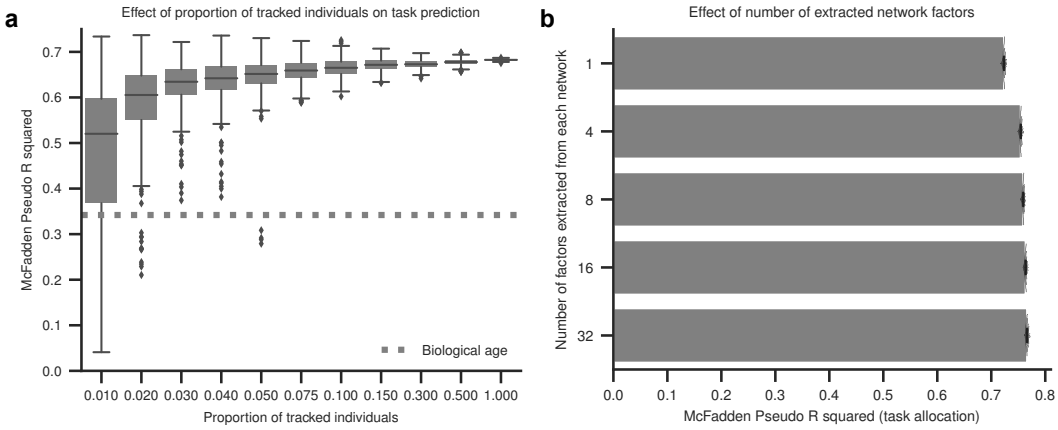

Figure 4.10: **Effects of subsampling the individuals and the number of spectral factors from each interaction network on the task allocation prediction. a** McFadden's R^2 for different fractions of subsampled individuals. While having all bees individually marked and tracked yields the highest median score, tracking only 5% of the individuals yields comparable results when drawing a representative subset. Each box comprises N=128 bootstrap runs (center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range; points, outliers). **b** The influence of the number of dimensions per interaction network extracted using Spectral embedding before applying the CCA on the task allocation prediction scores (using network age 3D). While more dimensions enable the CCA to extract more relevant information, the overall improvement is small. The bars show the mean of N=128 bootstrap samples per dimensionality and the error bars a bootstrap sampled 95% CI of the mean.

Effect of subsampling bees, the dimensionality of spectral embeddings, and only using proximity information on task prediction accuracy

Network age can be computed with a much smaller proportion of marked bees, hence allowing one to study much larger colonies, or reducing the effort for marking individuals. To quantify the effect of sparse sampling, we perform the following analysis. We subsample the individuals entering the analysis and compute interaction networks only for those selected individuals, i.e. only interactions between individuals of this subset are recorded. We then compute spectral factors and network age as described in Methods: Network age - CCA. We find that network age is remarkably robust in this setting. With 5% of the bees tagged, the proportion in variance explained using network age is comparable to the case of using all individuals. Even with only 1% of the individuals used in the analysis, network age still is a better predictor of task allocation than biological age. This suggests that a much larger colony could be analyzed with this method without additional effort, as long as a representative subset of the bees is tracked. See Figure 4.10a for the influence of different subsampling fractions on the results.

Likewise, we also find that the number of spectral factors per interaction mode matrix does not have a strong effect on the proportion of variance explained by the network age extracted from those factors. Using more factors is computationally more expensive and produces statistically better embeddings. However, network age computed using 32 factors explains task allocation only marginally better compared to network age computed using 4 factors (see Figure 4.10b).

We also evaluate variants of network age that can be derived using only the spatial information of detected individuals. To that end, we exactly follow the procedure outlined in Methods: Network age - CCA, but discard all interaction types except for either spatial proximities ('Euclidean proximity' variant, see 'Euclidean proximity networks' in Methods: Social networks) or interaction events derived from spatial proximity ('Proximity events' variant, see 'Proximity interaction network' in Methods: Social networks). We find that network age derived from all interaction types outperforms these proximity based variants in terms of task allocation, but the variant based on proximity events is only marginally worse (see Methods: Task prediction models and bootstrapping) and could be used in future studies if only proximity data is available.

Change of network age mode over a week

We analyze how many bees that are assigned to either of the network age modes on day X are assigned to the same cluster on day X+7 (clusters calculated as in Methods: Quantifying when bees first split into distinct network age modes). Most of the bees in the upper cluster are also in the upper cluster one week later (mean=93% 95% CI [77%, 100%], N=3698 bee days).

Bees in the lower cluster tend to stay there less (mean=56% 95% CI [0%, 96%], N=10 927 bee days) depending on their age (see Figure 4.11).

Variability of task prediction accuracy over time

The task allocation prediction scores are not constant over time for biological age, network age, and the unsupervised (PCA) variant of network age. Here, we show McFadden's Pseudo R^2 scores for individual days. While there is variance in the predictability using network age, we note that on

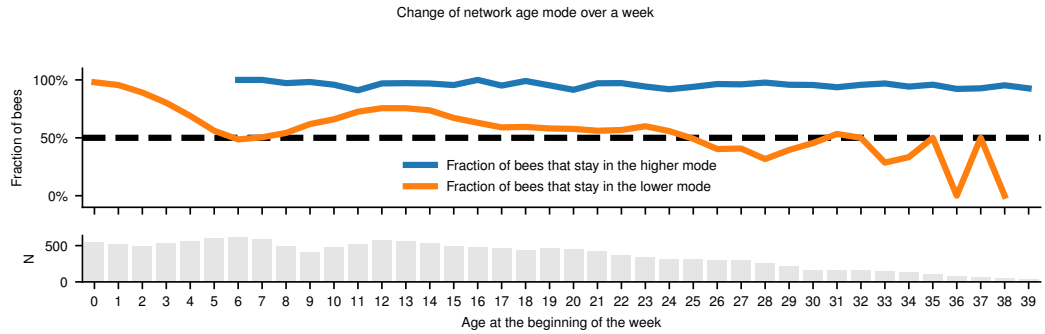

Figure 4.11: **The consistency of network age cluster over the course of a week.** (Top) We considered every individual and day (first day) for which we had data one week later (seventh day). We sorted them by their biological age on the first day (x-axis). Bees that were assigned to the higher network age cluster on the first day make up the blue line, bees from the lower cluster are depicted by the orange line. The y-axis value of each line is the fraction of bees that are still assigned to the same cluster seven days later. A high proportion of bees that are assigned to the higher cluster stay there over the course of one week (blue line). There are more bees that transition from the lower cluster to the higher one, regardless of age (orange line). (Bottom) The number of data points (orange and blue combined) for each age.

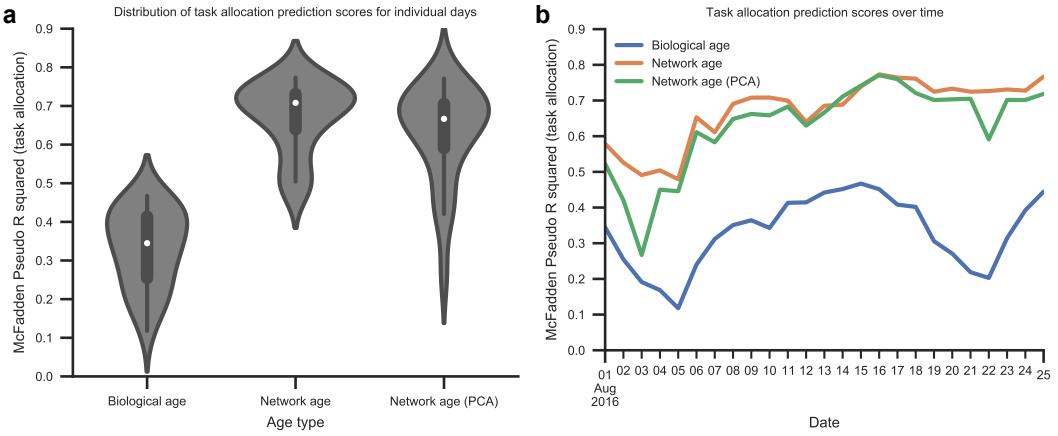

Figure 4.12: **Stability of task allocation prediction over time. a** The distribution of McFadden's R^2 for task allocation prediction for the different days using either network age, the unsupervised network age (PCA), or biological age as predictors (N=25 individual days; boxes: center dot, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range). **b** The same data points plotted over time.

all dates, the models based on network age yield a higher score than the models using biological age.

On all days, the unsupervised PCA variant is more predictive than biological age with comparably small differences to the CCA variant. See Figure 4.12 for the distribution and timeline.

4.13.3 Network age predicts an individual's behavior and future role in the colony

Probability of dying over the course of a week

We have shown that network age is more predictive of the remaining days to live than biological age (see Methods: Prediction of other behavior-related measures). Here we directly compare biologically young but functionally old bees with biologically old but functionally young bees. To define biologically young and old, we orientate ourselves at the caste thresholds proposed by Seeley (Seeley, 1982). To define functionally young and old we use the same thresholds for network age as shown in the heatmaps of Figure 4.3a, because they occupy distinct locations on the comb and therefore hold different functional roles within the colony. However, these exact thresholds are not important as both the effect strength and the significance are high.

We sort the individuals into the two groups based on their network age and biological age. To not count an individual twice, we only consider the last date that a bee was seen in each group. We calculate the probability of dying P as the fraction of bees of each group we do not consider alive after one week (see Methods: Bayesian lifetime model for more information about the death date of a bee). We calculate a confidence interval by drawing 1000 bootstrap samples for each group.

Biologically young but functionally old bees (biological age < 11 days, network age > 30) have a probability of dying over the course of a week of P=80.6% (bootstrap 95% CI=[74.1%, 87.1%], N=139). Biologically old but functionally young bees (biological age > 20, network age < 15) have a significantly lower probability of dying of P=42.1% (bootstrap 95% CI=[37.4%, 47.2%], N=390; χ^2 test of independence $p \ll 0.001$ N=529).

Time spent in task-associated locations as a predictor

To evaluate how well the time spent in the measured task-associated locations itself is able to predict an individual's behavior and future role in the colony, we derive two representations of this spatial information and use them as predictors in the same way as network age (as described in Methods: Prediction of other behavior-related measures, Methods: Future predictability) for predicting mortality and movement patterns.

We use the task descriptor described in Methods: Nest area mapping and task descriptor either directly as the independent variables in the regression models ('Location (4D)' in Table 4.1a) or a 1D representation of this descriptor derived using PCA ('Location (1D)' in Table 4.1a). The 4D descriptor is not directly comparable to network age because of the higher dimensionality, but serves as an upper bound on the information contained in the task descriptor. In terms of dimensionality, the 1D representation can directly be compared to network age.

We find that this spatial information is a better descriptor of an individual's behavior than age, but significantly less informative when compared to network age (Table 4.1d). See Table 4.1 for an overview of the scores and effect sizes.

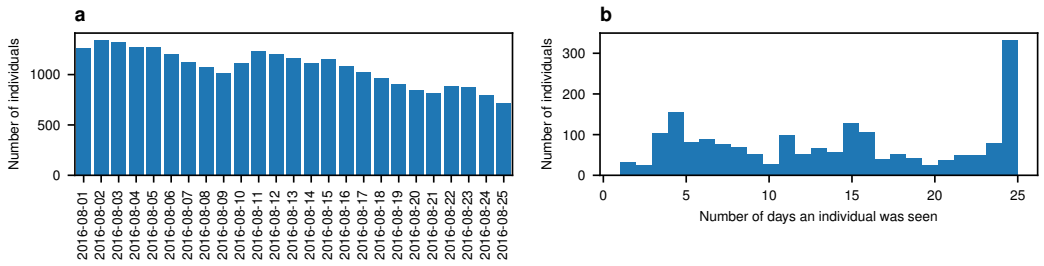

Figure 4.13: **Number of bees observed during our focal period. a** The number of bees per day. Days on which we introduced new bees are indicated by a rise in the population. **b** The number of days any single bee was observed during our focal period. The peak at 25 indicates bees that lived throughout the whole period.

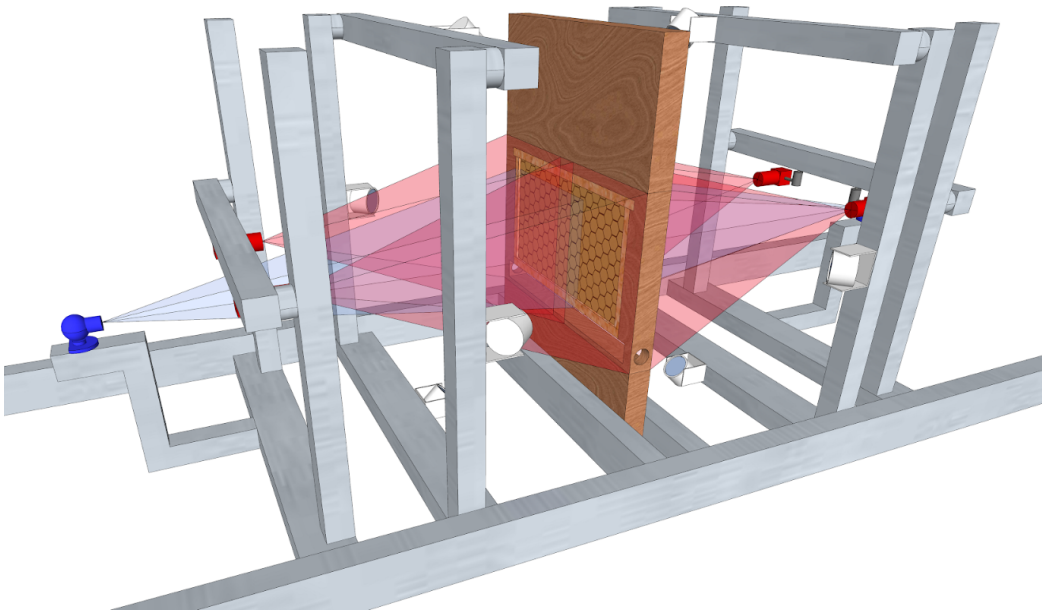

Figure 4.14: **Recording setup** A one-frame observation hive stands in the center of the rig, two high-res cameras (red), and one low-res camera (blue) observe each side of the hive. The low-res cameras only observe the dance floor. Four infrared flashlights were used per side (synchronized to the high-res cameras), constant red lighting illuminated the dance floor. Adapted with permission from (Wario et al., 2015).

Figure 4.15: **Individually marked bees at a feeding site.**

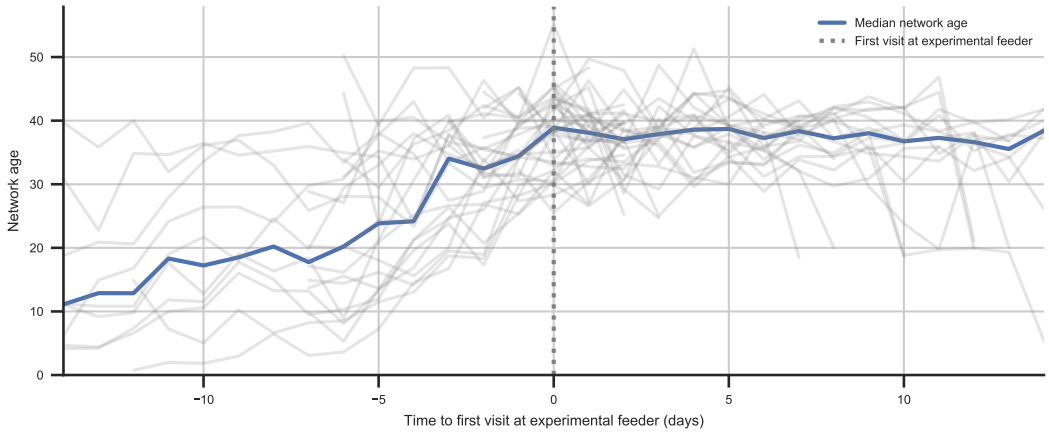

Figure 4.16: **Network age development of new foragers.** For every visiting bee, their first visit to a feeding station was recorded (age 12-40 days, N=40). The plot shows the network age for these bees over the previous and next days centered on their first visit. The grey lines give the network age of individual bees and the blue line indicates the median.

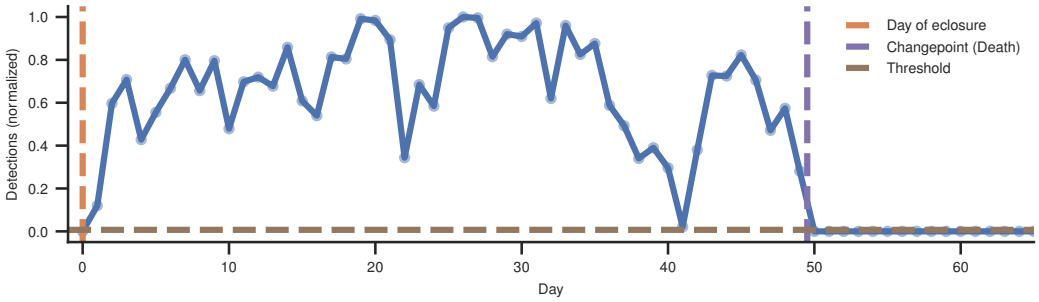

Figure 4.17: **The normalized detection counts of one bee over the recording period following her introduction to the observation hive.** The time of her death (purple dotted line on the right) is the mean from N=1000 Monte Carlo samples from a Bayesian changepoint model based on optimizing a detection threshold that discriminates between days that the bee was alive and days where she was probably dead.

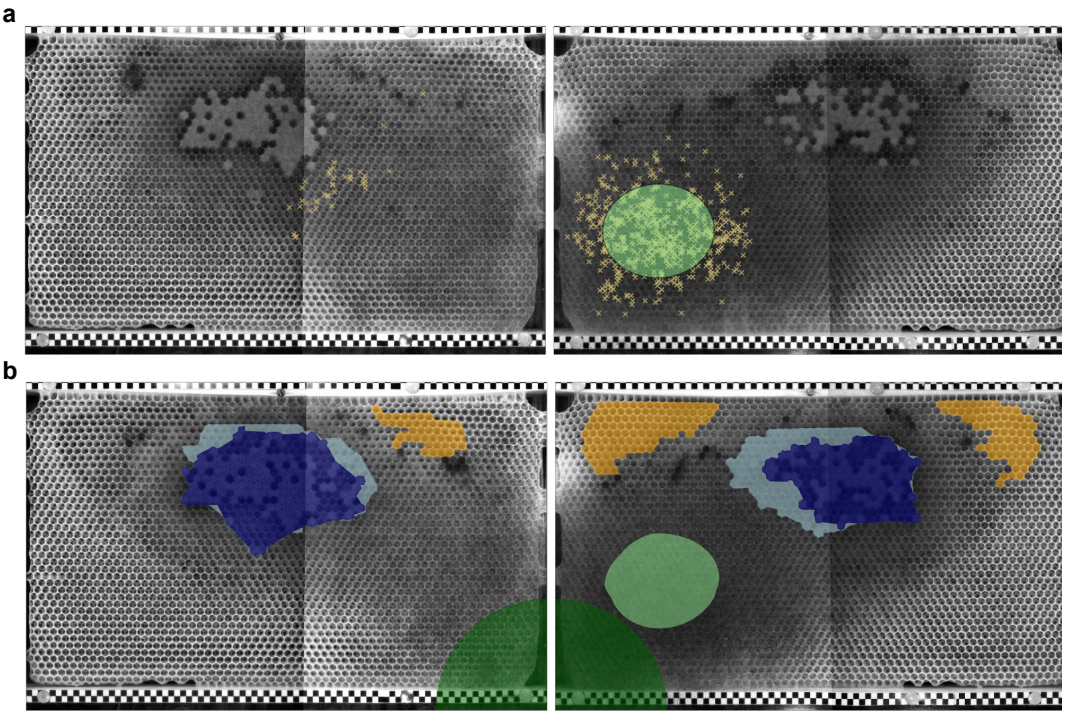

Figure 4.18: **Ground truth data for the location-based task descriptors.** **a** Background-subtracted images of the two sides of the observation hive. The images from the two cameras per side have been stitched together. The markers show the location of the high-confidence waggle run detections for 2016-08-18. Nearly all are on one side of the comb. The green region indicates the dance floor location. **b** Annotations for 2016-08-18, used to generate task descriptors for every bee (dark blue: capped brood, light blue: open brood, yellow: honey, light green: dance floor, dark green: exit region).

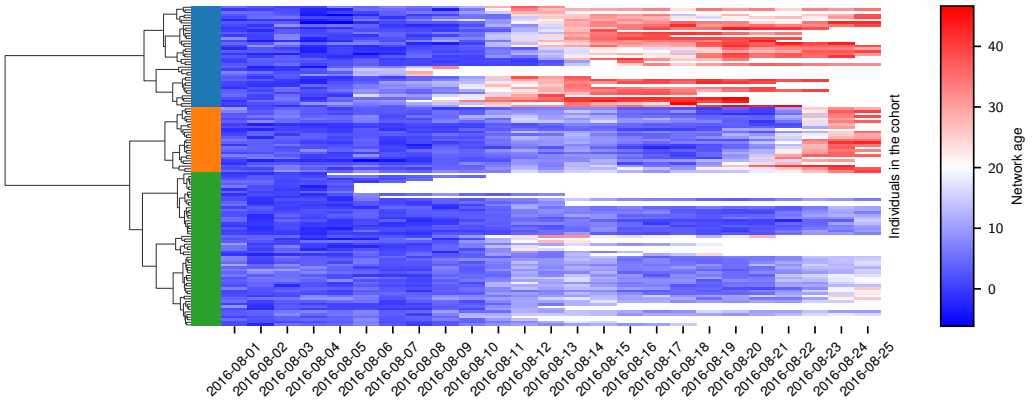

Figure 4.19: **Agglomerative clustering of the network age over time of one cohort of bees (emerged on 2016-08-01)** Missing values (e.g. due to the death of the bee) are white. The three colors (blue, orange, green) on the left side correspond to the clusters (1, 2, 3) in Figure 4.3b.

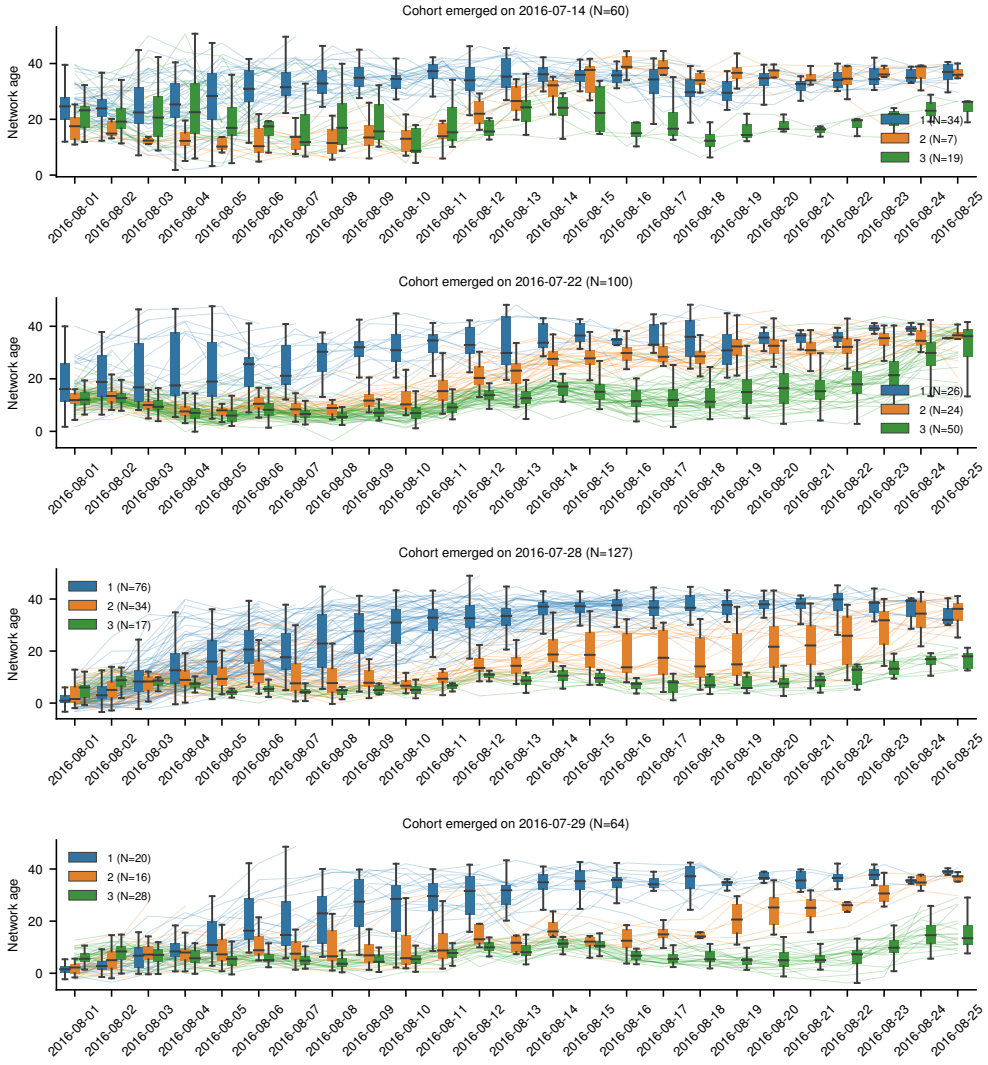

Figure 4.20: **The development of the network age of different cohorts over the focal period (see individual titles for date of emergence and cohort sizes).** The individuals are grouped according to the agglomerative clustering (see legends for the number of individuals in each cohort's clusters). The lines depict the network age of the individual bees and include all data points that are summarized in the boxplots for a given day (center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range).

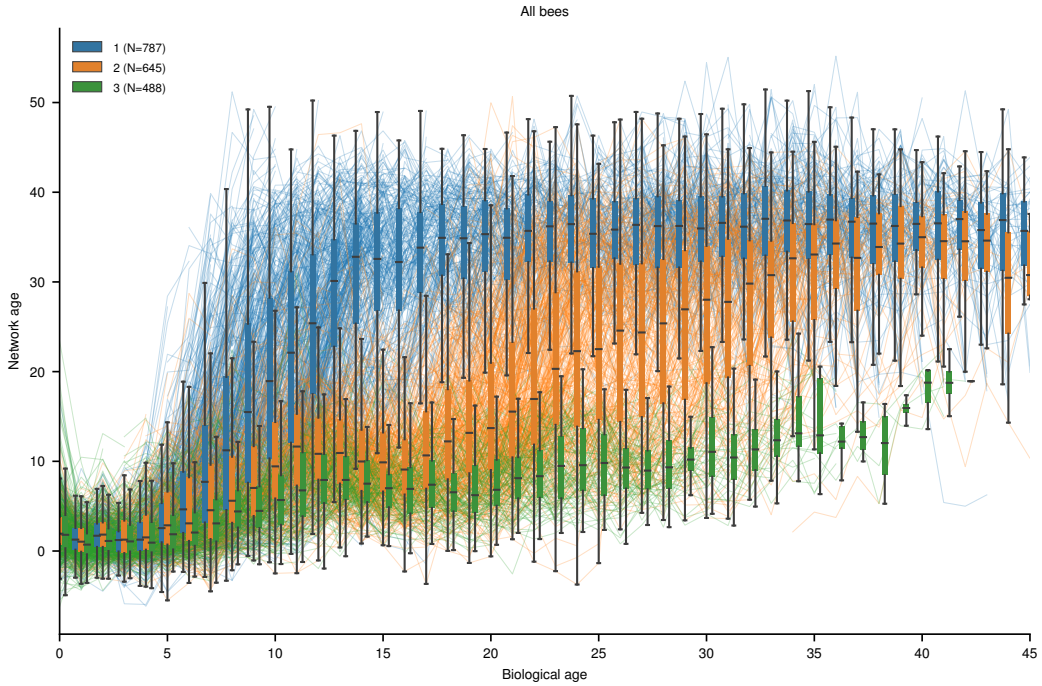

Figure 4.21: **The network age development of all individuals over their age.** To calculate this, we followed a procedure analogous to clustering the cohorts but used network age per age (instead of per date) as the bees' feature vectors for all individuals (N=1920). Note that there was only a subset of 25 days of data available for each individual as some individuals were already in the hive as the focal period began and some were introduced later. The lines depict the network age of individual bees. The number of individuals in each cluster is given in the legend. The boxplots summarize bees belonging to each cluster for a given day (center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range).

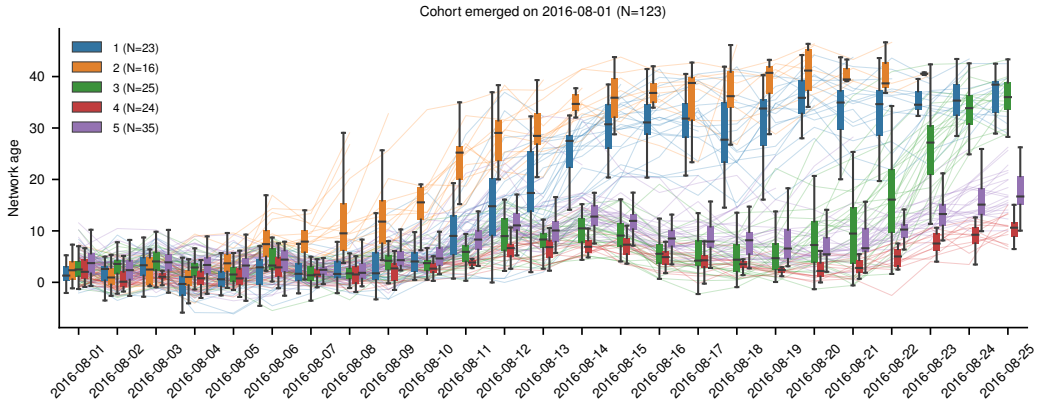

Figure 4.22: **Network age development with higher number of clusters (N=5).** Cutting the dendrogram at a deeper level yields more clusters that further subdivide the previous ones. The number of individuals and the date of emergence is given in the title, the number of individuals in each cluster is given in the legend. The boxplots summarize bees belonging to each cluster for a given day (center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range).

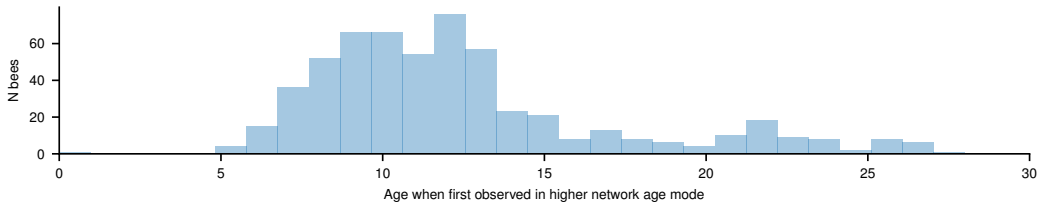

Figure 4.23: **The distribution of the biological age at which a bee was first assigned to the higher network age cluster.** Starting at around age 5-7, the network age distribution becomes bimodal.

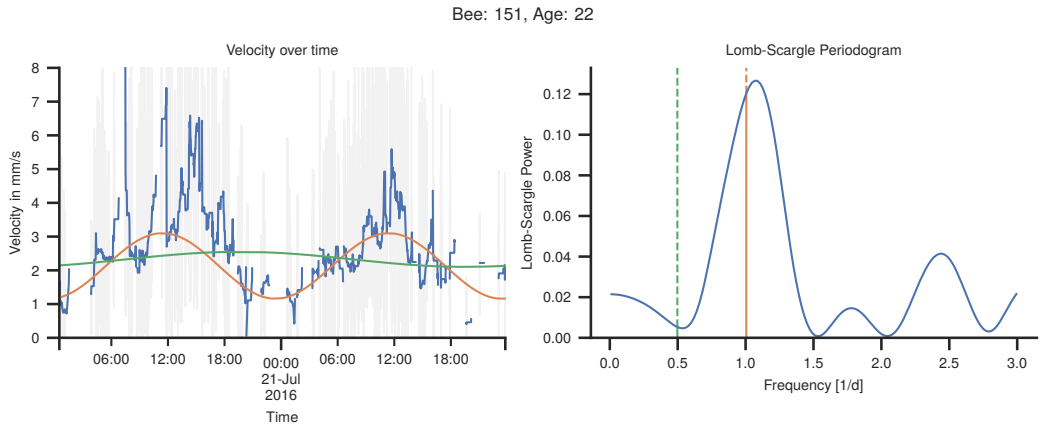

Figure 4.24: **Calculation of circadian rhythmicity from movement data.** (Left) The velocity as raw data (grey line) and smoothed with a median filter (blue line) for one bee over two days. The orange sine wave has a period of one day and was fitted to the velocity data via least-squares fitting. The green line is a similar fit with a period of two days. (Right) The corresponding Lomb-Scargle periodogram with the powers at the one-day frequency (orange) and two-day frequencies (green) highlighted.

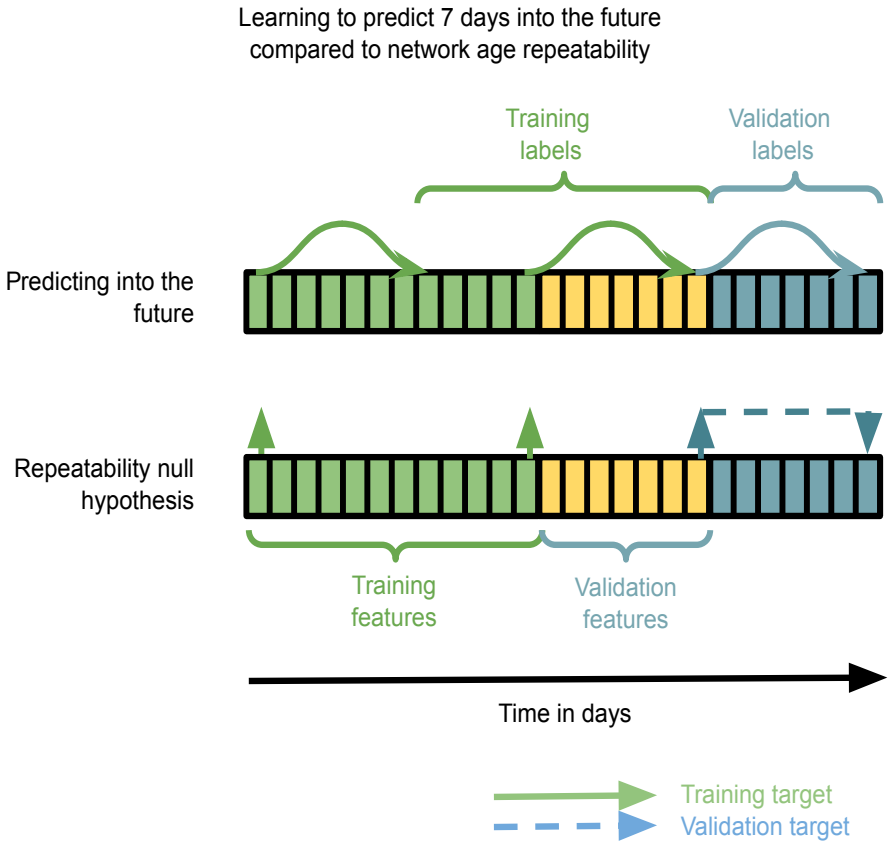

Figure 4.25: **Data handling for cross-validating future predictability.** (Top) Depiction of the process of training a model based on either biological age or network age to predict the future task allocation. We make sure that we do not leak information when evaluating future predictiveness. (Bottom) Data handling for the null hypothesis to check that the predictiveness is not completely explained by the repeatability of network age and task allocation for individual bees.

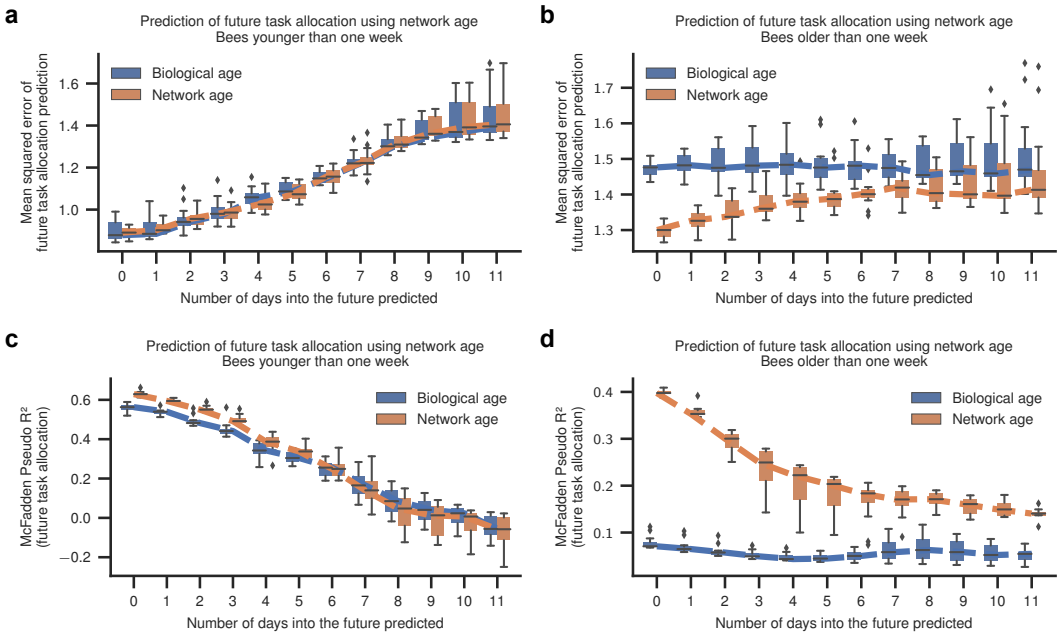

Figure 4.26: **Prediction of future task allocation for young and old bees.** We distinguish between young bees (left column, < 7 days of age) and old bees (right column, > 7 days of age). **a** For young bees, we find neither biological age nor network age to be predictive. **b** For the old bees, network age consistently performs better than age. **c, d** McFadden's R^2_{McF} of the future predictability for young and old bees. Each box comprises N=12 scores from models with N=12 days of training data (center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range; points, outliers).

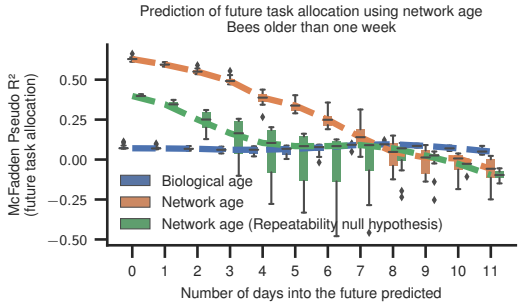

Figure 4.27: **Ruling out repeatability as a cause for predictiveness of network age.** To check that the predictiveness of network age for old bees is not simply due to the bimodal distribution of the network age and its high repeatability, we compare a predictive model based on network age (orange) with the null-hypothesis that the network age of day X is sufficient to describe the coming days without explicitly modeling the changes. Each box comprises N=12 scores from models with N=12 days of training data (center line, median; box limits, upper and lower quartiles; whiskers, 1.5x interquartile range; points, outliers).

(a) R^2_{McF} scores for dependent and independent variables and models

	Days until death	Time of peak activity	Strength of circadian rhythm	Velocity (daytime)	Velocity (night)
Age	0.06 [0.06, 0.07]	0.07 [0.07, 0.08]	0.30 [0.28, 0.30]	0.12 [0.12, 0.13]	0.21 [0.21, 0.23]
Age (nonlinear)	0.07 [-1.60, 0.08]	0.08 [0.07, 0.09]	0.31 [0.29, 0.32]	0.14 [0.13, 0.15]	0.35 [0.34, 0.36]
Network age (PCA)	0.09 [0.08, 0.09]	0.06 [0.06, 0.07]	0.45 [0.43, 0.47]	0.15 [0.14, 0.16]	0.29 [0.27, 0.30]
Location (1D)	0.10 [0.10, 0.11]	0.08 [0.07, 0.08]	0.49 [0.48, 0.49]	0.17 [0.16, 0.17]	0.31 [0.30, 0.33]
Network age (PCA, nonlinear)	0.14 [0.13, 0.14]	0.08 [0.07, 0.09]	0.51 [0.50, 0.52]	0.19 [0.19, 0.20]	0.33 [0.32, 0.34]
Location (1D, nonlinear)	0.15 [0.14, 0.16]	0.09 [0.08, 0.10]	0.51 [0.50, 0.52]	0.21 [0.20, 0.22]	0.34 [0.33, 0.35]
Network age	0.16 [0.16, 0.17]	0.10 [0.09, 0.11]	0.61 [0.61, 0.62]	0.24 [0.23, 0.25]	0.32 [0.31, 0.33]
Age + Network age	0.17 [0.16, 0.17]	0.11 [0.10, 0.11]	0.61 [0.61, 0.62]	0.24 [0.23, 0.25]	0.33 [0.32, 0.34]
Location (4D)	0.21 [0.20, 0.22]	0.10 [0.09, 0.10]	0.59 [0.58, 0.59]	0.28 [0.27, 0.29]	0.32 [0.30, 0.33]
Network age (nonlinear)	0.20 [0.19, 0.20]	0.11 [0.10, 0.11]	0.63 [0.62, 0.64]	0.27 [0.26, 0.28]	0.37 [0.36, 0.38]
Network age (2D)	0.23 [0.22, 0.24]	0.11 [0.10, 0.11]	0.62 [0.61, 0.63]	0.26 [0.25, 0.27]	0.37 [0.36, 0.38]
Location (4D, nonlinear)	0.23 [0.22, 0.24]	0.10 [0.10, 0.11]	0.62 [0.61, 0.62]	0.31 [0.30, 0.31]	0.36 [0.35, 0.37]
Network age (3D)	0.23 [0.22, 0.24]	0.11 [0.10, 0.11]	0.63 [0.62, 0.63]	0.30 [0.29, 0.31]	0.37 [0.36, 0.38]
Targeted embedding (1D)	0.22 [0.21, 0.22]	0.10 [0.10, 0.11]	0.62 [0.61, 0.63]	0.30 [0.30, 0.31]	0.41 [0.40, 0.42]
Targeted embedding (1D, nonlinear)	0.23 [0.22, 0.24]	0.11 [0.10, 0.11]	0.63 [0.63, 0.64]	0.31 [0.30, 0.32]	0.41 [0.40, 0.42]
Network age (2D, nonlinear)	0.27 [0.26, 0.27]	0.12 [0.11, 0.12]	0.63 [0.63, 0.64]	0.29 [0.28, 0.30]	0.40 [0.39, 0.41]
Age + Network age (nonlinear)	0.23 [0.22, 0.23]	0.13 [0.12, 0.13]	0.65 [0.64, 0.65]	0.30 [0.29, 0.31]	0.44 [0.43, 0.45]
Network age (3D, nonlinear)	0.27 [0.26, 0.28]	0.12 [0.11, 0.13]	0.64 [0.64, 0.65]	0.33 [0.32, 0.34]	0.40 [0.39, 0.42]

(b) Improvement in R^2_{McF} (network age vs biological age)

Dependent variable	Effect size [95% CI]
Days until death	0.102 [0.093, 0.110]
Time of peak activity	0.031 [0.021, 0.040]
Strength of circadian rhythm	0.317 [0.308, 0.329]
Velocity (daytime)	0.116 [0.108, 0.127]
Velocity (night)	0.107 [0.093, 0.121]

(c) Improvement in R^2_{McF} (network age vs targeted embeddings)

Dependent variable	Effect size [95% CI]
Days until death	0.050 [0.042, 0.061]
Time of peak activity	0.000 [-0.009, 0.011]
Strength of circadian rhythm	0.009 [0.000, 0.019]
Velocity (daytime)	0.065 [0.052, 0.077]
Velocity (night)	0.083 [0.072, 0.095]

(d) Improvement in R^2_{McF} (network age vs location (1D)))

Dependent variable	Effect size [95% CI]
Days until death	0.061 [0.052, 0.069]
Time of peak activity	0.025 [0.016, 0.034]
Strength of circadian rhythm	0.127 [0.119, 0.138]
Velocity (daytime)	0.073 [0.061, 0.083]
Velocity (night)	0.008 [-0.007, 0.022]

Table 4.1: R^2_{McF} values for different models when predicting different behavior-related metrics. Biological age consistently performs worse than predictions based on network age. Our method can also be applied without spatial information by optimizing the CCA for these metrics, yielding embeddings that are strongly predictive of them while still only using information available from the social networks (see 'Targeted embedding (1D)'). (Bottom) 95% bootstrap confidence intervals of the effect sizes (b) R^2 values for network age and biological age, c) R^2_{McF} values for targeted embeddings and network age, N=128).

	Brood area	Dance floor	Honey storage	Near exit	Combined
Age	0.46 [0.45, 0.47]	0.28 [0.27, 0.29]	0.00 [0.00, 0.00]	0.32 [0.31, 0.33]	0.34 [0.34, 0.35]
Age (nonlinear)	0.51 [0.50, 0.52]	0.35 [0.34, 0.36]	0.09 [0.09, 0.10]	0.41 [0.40, 0.42]	0.38 [0.38, 0.39]
Network age (1% of bees tagged)	0.58 [0.07, 0.79]	0.50 [0.08, 0.68]	0.03 [0.00, 0.21]	0.49 [0.12, 0.73]	0.52 [0.13, 0.70]
Network age (Only euclidean proximities)	0.51 [0.50, 0.52]	0.37 [0.36, 0.38]	0.17 [0.16, 0.18]	0.53 [0.52, 0.54]	0.52 [0.51, 0.52]
Network age (Only euclidean proximities, nonlin...	0.60 [0.60, 0.61]	0.55 [0.54, 0.56]	0.20 [0.18, 0.21]	0.56 [0.55, 0.57]	0.55 [0.55, 0.56]
Network age (1% of bees tagged, nonlinear)	0.62 [0.08, 0.81]	0.58 [0.16, 0.73]	0.16 [0.03, 0.34]	0.55 [0.17, 0.77]	0.55 [0.19, 0.72]
Network age (5% of bees tagged)	0.76 [0.66, 0.83]	0.61 [0.52, 0.68]	0.03 [0.00, 0.08]	0.60 [0.52, 0.69]	0.65 [0.58, 0.70]
Network age (PCA)	0.80 [0.79, 0.80]	0.53 [0.52, 0.54]	0.00 [0.00, 0.00]	0.58 [0.57, 0.59]	0.65 [0.64, 0.65]
Network age (PCA, nonlinear)	0.80 [0.80, 0.80]	0.61 [0.60, 0.62]	0.27 [0.26, 0.28]	0.66 [0.65, 0.66]	0.66 [0.66, 0.67]
Network age (Only proximity events)	0.79 [0.79, 0.80]	0.64 [0.64, 0.65]	0.03 [0.03, 0.03]	0.60 [0.59, 0.60]	0.67 [0.67, 0.68]
Network age (5% of bees tagged, nonlinear)	0.77 [0.63, 0.83]	0.66 [0.58, 0.73]	0.24 [0.15, 0.32]	0.67 [0.57, 0.73]	0.67 [0.59, 0.72]
Network age	0.80 [0.80, 0.81]	0.63 [0.63, 0.64]	0.03 [0.03, 0.04]	0.61 [0.60, 0.62]	0.68 [0.68, 0.69]
Age + Network age	0.81 [0.80, 0.81]	0.63 [0.63, 0.64]	0.10 [0.09, 0.11]	0.61 [0.60, 0.62]	0.69 [0.69, 0.69]
Network age (Only proximity events, nonlinear)	0.81 [0.80, 0.81]	0.68 [0.67, 0.69]	0.32 [0.31, 0.32]	0.68 [0.67, 0.69]	0.70 [0.69, 0.70]
Network age (nonlinear)	0.82 [0.81, 0.82]	0.68 [0.67, 0.69]	0.33 [0.32, 0.33]	0.69 [0.69, 0.70]	0.71 [0.70, 0.71]
Age + Network age (nonlinear)	0.82 [0.82, 0.83]	0.69 [0.68, 0.69]	0.36 [0.35, 0.37]	0.70 [0.69, 0.70]	0.72 [0.71, 0.72]
Network age (2D)	0.84 [0.84, 0.85]	0.65 [0.64, 0.65]	0.46 [0.45, 0.46]	0.64 [0.63, 0.64]	0.73 [0.73, 0.74]
Network age (2D, nonlinear)	0.85 [0.85, 0.85]	0.69 [0.69, 0.70]	0.49 [0.48, 0.50]	0.70 [0.70, 0.71]	0.74 [0.74, 0.75]
Network age (3D)	0.84 [0.84, 0.85]	0.71 [0.70, 0.71]	0.46 [0.45, 0.47]	0.70 [0.70, 0.71]	0.76 [0.75, 0.76]
Network age (3D, nonlinear)	0.86 [0.85, 0.86]	0.75 [0.74, 0.75]	0.50 [0.49, 0.51]	0.76 [0.76, 0.77]	0.77 [0.77, 0.77]

Table 4.2: R^2_{McF} median scores and bootstrapped 95% CI for different combinations of independent variables in the rows and dependent variables in the columns. We evaluated the following independent variables: different dimensionalities of network age ('Network age', 'Network age (2D)', 'Network age (3D)', see Methods: Network age - CCA), network age derived without supervision ('PCA', see Methods: Network age - PCA), with a subsample of the bees ('1% of bees tagged' and '5% of bees tagged', see Section 4.13.2), and based only on spatial proximity ('Euclidean proximities' and 'Proximity events'). The likelihood ratio test was evaluated for the combined models (e.g. 'Age + Network age', see Methods: Statistical comparison of models).

Feeder	GPS coordinate	Distance from hive
F1	52.455726, 13.294224	217 m
F2	52.454866, 13.293092	338 m
F3	52.454547, 13.291737	430 m
F4	52.453682, 13.292841	451 m

Table 4.3: GPS coordinates and distances from the hive for the feeders used in the forager group experiment.

Date	Location	Forager bees observed at feeder
2016-08-01	Hive-F1	135, 199, 217, 220, 228, 233, 253, 319, 392, 539, 727, 818, 845, 860, 1110
2016-08-02	F1	135, 217, 233, 253, 648, 845, 860
2016-08-03	F1	135, 228, 233, 253, 392, 648, 845, 860
2016-08-04	F1	220, 228, 233, 392, 648, 710, 885
2016-08-05	F1	135, 220, 228, 233, 392, 644, 648, 710, 769, 885, 931
2016-08-08	F1-F2	199, 319, 337, 392, 555, 644, 710, 769, 885, 1593
2016-08-09	F1-F2	199, 319, 337, 392, 555, 644, 648, 710, 769, 785, 885, 912, 1122, 1362, 1518, 1593, 2031
2016-08-10	F1-F2	199, 319, 337, 392, 648, 769, 885, 1122, 1362, 1593, 2031
2016-08-11	F2	199, 319, 337, 392, 644, 769, 785, 885, 1122, 1362, 1518, 1593, 2031
2016-08-12	F2	199, 319, 337, 392, 769, 785, 885, 1122, 1362, 1518, 1593, 2031
2016-08-14	F2	199, 319, 337, 392, 555, 644, 710, 769, 785, 885, 1232, 1362, 1518, 1593, 2031
2016-08-16	F2	199, 319, 337, 710, 769, 885, 912, 1122, 1362, 1518, 1593, 2031
2016-08-17	F2	199, 319, 337, 644, 710, 769, 885, 912, 1122, 1518, 1593, 2031
2016-08-18	F2	199, 319, 337, 644, 769, 885, 912, 1122, 1518, 2031
2016-08-19	F2	319, 710, 885, 1593, 2031
2016-08-22	F2-F3	319, 644, 710, 885, 1593, 2031
2016-08-23	F3	1180, 1197, 1232, 1471, 1593, 1662, 1714, 1799, 2031, 2106, 2984
2016-08-24	F3	1180, 1197, 1232, 1471, 1593, 1714, 1793, 1799, 2031, 2106, 2984
2016-08-25	F4	1180, 1197, 1714, 1793, 1799, 2031, 2106

Table 4.4: Tag IDs and locations observed forager bees during the forager group experiment.

5 Behavioral variation across the days and lives of honey bees

Michael L. Smith [1,2,3,4,†,*], Jacob D. Davidson [1,2,3,†,*], Benjamin Wild [5], David M. Dormagen [5], Tim Landgraf [5], Iain D. Couzin [1,2,3,*]

[1] Department of Collective Behaviour, Max Planck Institute of Animal Behavior, 78464 Konstanz, Germany.
[2] Department of Biology, University of Konstanz, 78464 Konstanz, Germany.
[3] Centre for the Advanced Study of Collective Behaviour, University of Konstanz, 78464 Konstanz, Germany.
[4] Department of Biological Sciences, Auburn University, 36849 Auburn AL, USA.
[5] Department of Mathematics and Computer Science, Freie Universität Berlin, 14195 Berlin, Germany.

[*] Corresponding authors: mls0154@auburn.edu, jdavidson@ab.mpg.de, icouzin@ab.mpg.de
† Equal contribution

5.1 Preface

The previous works discussed in Chapter 3 and Chapter 4 were based solely on data collected from the BeesBook recording system in Berlin. While this approach allowed for the creation and analysis of honey bee tracking datasets that were unprecedented in their scope, one major limitation was that it did not allow for the generalization of findings beyond this one recording setup. To address this issue, a fruitful collaboration was established, resulting in the development of another recording system in the *Max Planck Institute of Animal Behavior* in Konstanz. This new automated tracking system was used in the paper titled *Behavioral Variation Across the Days and Lives of Honey Bees,* in which we obtained long-term data on more than 4,100 bees across 16 age-matched cohorts tracked within an observation hive for more than 50 days throughout a summer.

In addition to obtaining new data, we also introduced a novel analysis framework to describe behavioral variation at different timescales, with a focus on the individuals' space use within the colony. Our analysis revealed that, at the scale of a single day, bees exhibited varying patterns

in their use of space and movement, with older and younger bees showing different behavioral patterns. When we investigated the bees' entire lifetimes, we found that some bees consistently exhibited different movement patterns and transitioned to dance floor/outside activities at different ages. Overall, this work provides a more comprehensive understanding of honey bee behavior across different temporal scales, and highlights the importance of multi-site data collection for generalizable findings in the field of animal behavior. It is worth noting that the hardware design and software of the automated tracking system are open access and open source, thus enabling other research groups to replicate and expand upon this work in different locations.

This chapter was previously published as:
Smith et al. (2022a) — Michael L. Smith, Jacob D. Davidson, Benjamin Wild, David M. Dormagen, Tim Landgraf, and Iain D. Couzin. Behavioral variation across the days and lives of honey bees. *iScience*, 25(9):104842, 2022a. URL https://doi.org/10.1016/j.isci.2022.104842

5.2 Author Contributions

Methodology: MLS, JDD, BW, DD, TL, IDC; Investigation: MLS, JDD, TL, IDC; Fieldwork: MLS; Software: BW, DMD, TL; Visualization: JDD; Supervision: TL, IDC; Writing-original draft: MLS, JDD; Writing-review and editing: MLS, JDD, BW, DMD, TL, IDC.

5.3 Summary

In honey bee colonies, workers generally change tasks with age (from brood care, to nest work, to foraging). While these trends are well-established, our understanding of how individuals distribute tasks during a day, and how individuals differ in their lifetime behavioral trajectories, is limited. Here, we use automated tracking to obtain long-term data on 4,100+ bees tracked continuously at 3Hz, across an entire summer, and use behavioral metrics to compare behavior at different timescales. Considering single days, we describe how bees differ in space use, detection, and movement. Analyzing the behavior exhibited across their entire lives, we find consistent inter-individual differences in the movement characteristics of individuals. Bees also differ in how quickly they transition through behavioral space to ultimately become foragers, with fast-transitioning bees living the shortest lives. Our analysis framework provides a quantitative approach to describe individual behavioral variation within a colony from single days to entire lifetimes.

5.4 Introduction

Social insect colonies are comprised of individual organisms that form a cooperative entity to propagate their genes (Seeley, 1989b; Wilson and Sober, 1989; Smith and Szathmary, 1995). To survive, grow, and reproduce, a colony must navigate the same biotic and abiotic challenges as unicellular and multicellular organisms, but coordination must now occur at the level of individual workers (Hölldobler et al., 2009). Social insect colonies lack centralized control, but across the ants, bees, termites, and wasps, tasks are instead self-organized among workers, whether genetically, physiologically, spatially, or temporally (Gordon, 1996; Mersch et al., 2013; Seeley, 1982; Beshers and Fewell, 2001; Oster and Wilson, 1978; Porter and Tschinkel, 1985; Jeanne, 1986; Jeanne et al., 1988; Frumhoff and Baker, 1988; Robinson et al., 1989; Fewell and Page, 1993; O'Donnell and Jeanne, 1995; Naug and Gadagkar, 1998; Oldroyd and Fewell, 2007; Robinson et al., 2009a; Jandt and Dornhaus, 2009; Baudier et al., 2020). Understanding how individuals combine to form a collective provides insights into the evolutionary drivers of organization across biological scales (Smith and Szathmary, 1995; Davidson et al., 2021).

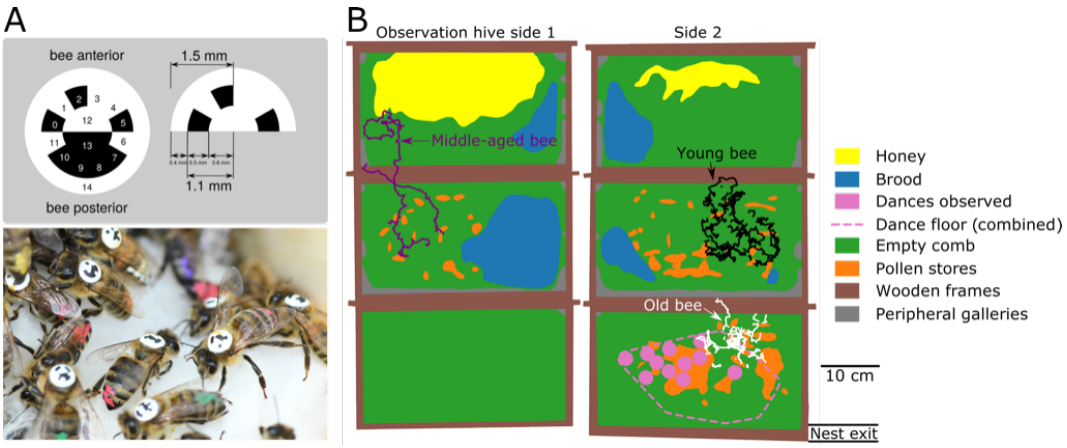

Figure 5.1: **Long-term honey bee tracking** (A) Bees were individually marked with barcodes, and tracked using the BeesBook tracking system (Boenisch et al., 2018a). (B) An example map of the observation hive, with colors to denote different nest substrates. Dots overlaid on the map show trajectories of three representative bees with short trajectories selected from 11 August 2018: (black) young bee, age 6 days; (purple) middle-aged bee, age 16 days; (white) old bee, age 26 days. Nest exit/entrance at the lower right corner.

A key challenge for highly-integrated collective systems, such as eusocial insects, is how to allocate tasks among the individual units. While a fixed allocation strategy may be efficient in stable environments, a flexible approach allows colonies to respond to changing conditions (Gordon, 2014, 2016). Responsive (and decentralized) changes in task allocation can arise, for example, from individuals with different response thresholds for task-specific stimuli (Bonabeau et al., 1997), individuals selecting tasks based on current need or availability (Tofts, 1993; Jeanne L, 1996), state-dependent probabilities to switch or remain in a current task (Gordon, 1999; Goldsby et al., 2012), age, developmental, or physiological task engagement (Seeley, 1982; Cook et al., 2019; Robinson et al., 1989; O'Donnell and Jeanne, 1993), or a combination of these mechanisms (Johnson, 2010). These mechanisms can also depend on the type of task: non-specialized tasks may be distributed widely among colony members, whereas tasks requiring certain physiological abilities may be restricted to specific individuals (Johnson, 2003; Robinson et al., 2009b). Across social insect species, how and when tasks are allocated among individuals represents a balance between robustness and flexibility in colony function (Charbonneau and Dornhaus, 2015).

In colonies of the Western honey bee *Apis mellifera* individuals perform different tasks according to multiple factors, including developmental state, genetics, and behavioral feedback mediated by social interactions (Huang et al., 1994; Beshers and Fewell, 2001; Robinson, 2002; Grozinger et al., 2007; Johnson, 2008a; Cook and Breed, 2013; Cook et al., 2019; Wild et al., 2021b). This gives rise to a general tendency for young bees to care for brood in the center of the nest, middle-age bees to perform various tasks throughout the nest, and old bees to forage outside and advertise food sites with waggle dances on the dance floor (Seeley, 1982). Within these general trends, individuals may switch between tasks, or perform multiple different tasks in a day; therefore, individual behavior is better described with "task-repertoires" — groups of tasks that are similar behaviorally and/or

spatially (Seeley, 1982; Johnson, 2010). Although task repertoires vary with age, an age-based categorization does not account for variation among individuals throughout their lives, or how previous social and/or environmental experiences may influence task allocation (Beshers and Fewell, 2001; Jeanson and Weidenmüller, 2014; Wild et al., 2021b).

While previous studies have relied on human observation to assign behavior to individuals using ethograms (e.g. Lindauer (1952); Seeley (1982); Seeley and Kolmes (1991); Johnson (2003); Siegel et al. (2013); Smith et al. (2017); Perez and Johnson (2019)), recent advances in automated tracking make it possible to extract behavioral metrics beyond the scope and scale of human observation (e.g. continuous location and instantaneous speed) (Wario et al., 2015; Mersch et al., 2013; Crall et al., 2015; Gernat et al., 2018; Wild et al., 2018; Bozek et al., 2021; Crall et al., 2018; Jones et al., 2020; Richardson et al., 2021). This allows one to move from general trends to detailed, long-term, quantification of behavior. The use of quantitative metrics to characterize behavior enables a data-driven approach to investigate the causes and consequences of individual variability and inter-individual differences across timescales.

In this study, we present data and analyze the behavior of 4,100+ honey bees across 16 age-matched cohorts tracked within an observation hive for 50+ days throughout a summer (July-October 2018). We define an analysis framework using behavioral metrics calculated from the motion data that quantify bees' space use, detection, and movement. We use this framework to examine behavioral variation among age-matched bees, as well as variation in the behavioral trajectories of individuals over lifetimes. This analysis framework enables a quantitative comparison of the behavior of thousands of individuals at different timescales.

5.5 Results

5.5.1 Long-term tracking of individually-marked bees

We tagged and tracked over 4,100 individuals, 3 times per second (3 Hz), day and night for 50+ days during summer 2018 using the using the BeesBook tracking system (Boenisch et al., 2018a) (Figure 5.1A). Newborns were introduced to the 3-frame observation hive every 4-6 days, in cohorts of 200-600 bees. Each time a new cohort was introduced, we recorded the comb contents in the observation hive (as in Smith et al. 2016) to map the honey stores, brood nest, and dance floor. The dance floor is an area typically near the next exit, where foragers advertise food sites with waggle dances (Seeley, 1995). These content maps allow us to determine the context of the spatio-temporal patterns of activity exhibited by bees throughout their lives, in the context of their changing social and structural nest environment (Figures 5.1B and S1).

To quantify the activity of individual bees on a given day, we compute multiple behavioral metrics, which describe space use (time on honey, brood areas, and dance floor, and median exit distance), detection (time observed, time outside, number of outside trips, and number of dance floor visits), and movement (median speed, speed circadian coefficient, dispersion, and fraction of the nest visited). See Figure 5.3 for a visual depiction of these metrics, and see Methods and Table 5.1 for a complete description of how each is computed.

At any given time, bees on honey storage and brood areas tend to be younger than bees on the dance floor (Figure 5.2A). This trend is consistent with the well-established sequence of young

workers performing within-nest tasks, and old workers foraging outside (Seeley, 1982; Robinson, 1992). As individuals age, they spend more time on the dance floor, but we observe considerable differences among bees within the same age-matched cohort on any given day (Figures 5.2B, S2).

5.5.2 Individual behavior during a single day

In this section we examine single-day behavioral variation. We use the term "behavioral day" to refer to the behavior of a single bee on a single day. Quantitatively, a behavioral day refers to the behavioral metrics shown in Figure 5.3 calculated on a single day for a single bee. Note that individual bees have multiple "behavioral days" that make up their life, and may exhibit different behavior on different days – in this section we describe differences in individual behavioral days irrespective of individual identity, and then in the next section we use the known identity of each bee to compare how individuals change their behavior over time. For consistency in comparing behavior, we focus the analysis on the 50-day period during which new cohorts were added every 4-6 days. In this time period, the dataset includes a total of 53,032 behavioral days, which are from 4,193 tracked bees.

We use principal component analysis (PCA), clustering, and visualization methods to describe the space of behavioral variation (Valletta et al., 2017). Note that this does not assign specific activities (e.g. fanning) to individuals over time, like an ethogram, but instead uses behavioral metrics computed from the barcode-tracking data to identify patterns and similarities in behavior among behavioral days. The behavioral metrics include time on honey, brood, and dance floor areas, exit distance, time observed, time outside, number of outside trips, number of dance floor visits, median speed, speed circadian coefficient, dispersion, and fraction of the nest visited; these 12 metrics represent space use, detection, and movement and are graphically depicted in Figure 5.3 and defined in full detail in Table 5.1.

PCA extracts the dominant axes of behavioral variation, i.e. the relative weightings of the behavioral metrics that explain the largest percentage of variance in the data matrix. To perform PCA, we first arrange the data in a matrix structure, normalize each metric so that all can be compared in the same standardized units, and then calculate the PCA decomposition. In the day-data matrix M_{ij}, each row $i = 1...53,032$ is for a single behavioral day (i.e. a single tracked bee on a single day), and the columns $j = 1...12$ are for each behavioral metric. The data matrix is normalized following standard procedures by subtracting the mean and dividing by the standard deviation of each column. With this, the total variance is simply the matrix norm and is equal to the number of metrics, i.e. $||M_{ij}|| = 12$, and the percentage of variance explained can be computed using the remaining variance after subtracting a particular pattern from the data matrix (Valletta et al., 2017; Jolliffe, 2002). Because of the normalization, positive/negative weightings in the PCA components represent higher/lower values of a metric with respect to the average across all behavioral days.

We find that the first 3 principal components represent important axes of variation among behavioral days. The first PCA component explains the largest amount of variance (28.8%), and is strongly weighted by space use: in particular, time on the dance floor and low exit distance (Figure 5.4A). The second PCA component accounts for 19.8% of the variance and is strongly weighted by fraction of the nest visited and dispersion, which are two complementary metrics that reprents

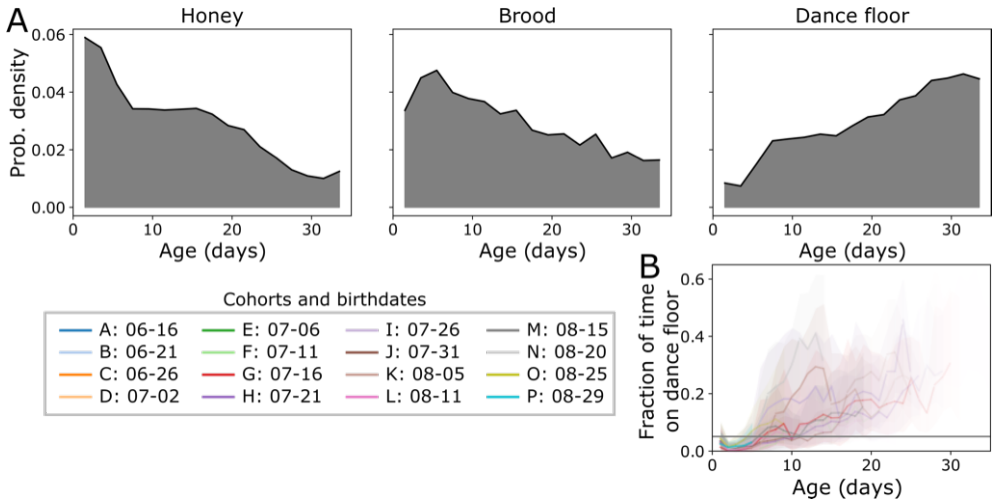

Figure 5.2: **Bee nest usage histograms and changes with age**. See also Figure S2. (A) Substrate usage histograms with respect to age. (B) Cohort distributions of dance floor usage with age. Colors represent the different cohorts, ordered chronologically by birthday, with corresponding alphabetical names. Lines show the mean and the shaded area shows the standard deviation across bees in each cohort. The transparency is proportional to the fraction of bees in a cohort that lived to a certain age.

how wide-ranging a bee is, regardless of where it tends to be located in the nest (see Figure 5.3 and Table 5.1 for descriptions of how these metrics are defined). The third PCA component (12.3% of the total variance) is most strongly weighted by speed and time spent on brood.

The first 3 components explain 60.9% of the total variance of all behavioral metrics included in the normalized data matrix (M_{ij}), with the first component accounting for 28.8%. In comparison, the total amount of variance explained by age alone is only 9.2%, and that by age & cohort is only 17% (see Table 5.2). This demonstrates that although there are indeed consistent trends among bees as they age, as well as differences across cohorts, age and cohort alone are insufficient to account for the full range of behavioral variation in the data. We therefore use a more general method to describe the dominant daily behavioral differences among bees.

To visualize the multi-dimensional range of variation in the data matrix (M_{ij}), we use t-distributed stochastic neighbor embedding (t-SNE; Maaten and Hinton 2008). Each "data point" in this embedding is a single behavioral day. Because it is initialized with the first two PCA components, the global structure of the t-SNE embedding corresponds to the first two PCA components, but the local structure can represent higher-order PCA components (Figure 5.4B). This embedding is therefore a useful visual aid to display the range of single-day behavioral variation in a compact yet meaningful way.

We apply Ward hierarchical clustering (Ward, 1963; Valletta et al., 2017) to group behavioral days with similar behavioral metrics. Examining the dendogram structure and the within-cluster variance as a function of number of clusters, we find that distinct behavioral clusters do not exist in our data (Figure S3) However, with this clearly in mind, we can still employ clustering tools

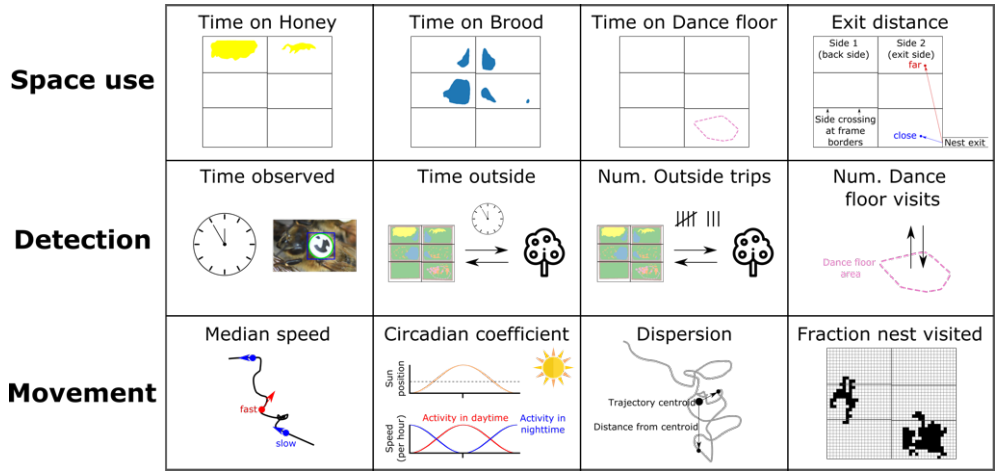

Figure 5.3: **Metrics used to quantify behavior of tracked bees**. We use 12 behavioral metrics to quantify the behavioral of individual bees. These are grouped in metrics describing space use (time spent on honey, brood, and dance floor areas, and median exit distance), detection (time observed, time spent outside the nest, number of outside trips, and number of dance floor visits), and movement (median speed, speed circadian coefficient, dispersion, and fraction of the nest visited. See Table 5.1 and Methods for further details on how each metric is computed.

to aid in the comparison, visualization, and thus the understanding, of the trends we see in the behavioral metrics of bees on different days (Figure 5.4C). To this end, we focus on the 5-cluster result as a practically-useful grouping to describe dominant trends in the data, as represented by differences in the first 3 PCA components.

Each data point in Figure 5.4B represents the behavioral metrics calculated for a single bee on a single day; a "behavioral day" (see Figure 5.3 for behavioral metrics). Cluster 1 (blue) represents behavioral days that have a high dance floor use and time outside, with higher activity during the day (higher than average speed circadian coefficient), and the highest median age. Similar to cluster 1, cluster 2 (orange) represents behavioral days in which bees spent time close to the exit; however, in contrast to cluster 1, cluster 2 behavioral days have particularly high number of outside trips and dance floor visits, with visits to multiple areas of the nest, not just the dance floor and exit frame - see histograms in Figure 5.4C. Cluster 3 (green) represents behavioral days in which bees have the highest distance to the exit and time spent on honey areas, in comparison with other clusters. Cluster 4 (red) represents behavioral days associated with middle-aged bees, with metrics representing slow, localized behavior, mostly on empty-comb "border" regions of the nest (hence the lower than average time on both honey and brood areas). Cluster 5 (purple) represents behavioral days with higher than average speed and dispersion, as well as a higher than average time on brood areas.

These groupings illustrate the dominant differences among behavioral days. While cluster 1 and cluster 2 behavioral days represent similar space use, they have differences in detection and movement; both represent behavioral days where bees exited the nest, but cluster 1 represents more

time outside, while cluster 2 represents more trips and visits to places in the nest other than just the dance floor. Clusters 3, 4, and 5 are similar in detection (these behavioral days are mostly inside the nest), but differ in space use and movement. Cluster 3 behavioral days are farthest from the exit and have the highest average time on honey areas, while cluster 4 behavioral days have high time spent on areas that border brood and honey, and cluster 5 has high time spent on brood areas. Cluster 3 and 4 behavioral days have lower movement speeds, while cluster 5 has higher speeds. Although there is much overlap in the age distributions, cluster 3 behavioral days have the lowest median age (4 days), clusters 2, 4, and 5 have intermediate values (median ages of 9, 8, and 10 days, respectively), and cluster 1 the highest median age (15 days) (Figure 5.4D).

5.5.3 Behavioral variation over lifetimes

In this section we use individual identities to compare lifetime behavioral trajectories among individual bees. Bees are known to change their behavior over time due to internal processes such as physiological development, interactions with other bees, and environmental factors (Johnson, 2003, 2010; Robinson, 1992; Wild et al., 2021b; Amdam and Omholt, 2003). To quantitatively compare how different bees change behavior as they age, we use a procedure similar to that used for behavioral days. However, now instead of a single day, each data point represents the entire life of an individual bee — we refer to this as a "bee-life". Quantitatively, a bee-life is defined by the behavioral metrics for each day of the bee's life (i.e. the series of behavioral days that make up an individual's life).

We again use PCA and clustering to describe behavioral variation among bee-lives. To do this, we arrange the data into a three-dimensional matrix form to represent individual behavioral metrics for each day of a bee's life. The bee-life data matrix is of the form $B_{\alpha tj}$, where α is for individual bees, t is an index over the days of the bee's life, and j is for the different behavioral metrics (these are the same as used in the per-day analysis; see Figure 5.3). The life-PCA decomposition considers each bee α as a single input entry; the components can thus represent both consistent lifetime differences in behavioral metrics, as well as changes in behavior over time (Figures 5.5A, S4A). Note that due to the high-dimensional input (i.e all behavioral metrics over multiple days), each PCA component in the lifetime analysis represents a comparatively smaller fraction of the variance, as compared to the per-day analysis shown previously in Figure 5.4. The first two life-PCA components nonetheless represent strong trends in the data (Figure S4B), and we focus our interpretations on these first two life-PCA components.

The first life-PCA component predominantly represents an overall difference in movement activity across the whole lifetime of a bee (Figures 5.5A, S4A). Positive projections of lifetime data onto the first life-PCA component represent bees with higher movement activity (higher dispersion and fraction of nest visited), and negative projections represent lower movement activity than average. The second life-PCA component represents differences in space use - both averaged across a lifetime, as well as changes with age (Figures 5.5A, S4A). Positive projections of lifetime data onto the second life-PCA component represent an early transition to increased dance floor use and time outside, while negative projections represent a lesser or later increase in these metrics.

Applying hierarchical clusterical to group bee-lives with similar behavior, the dendogram structure (Figure S4C) and the within-cluster variance (Figure S4D) suggest a continuous range of

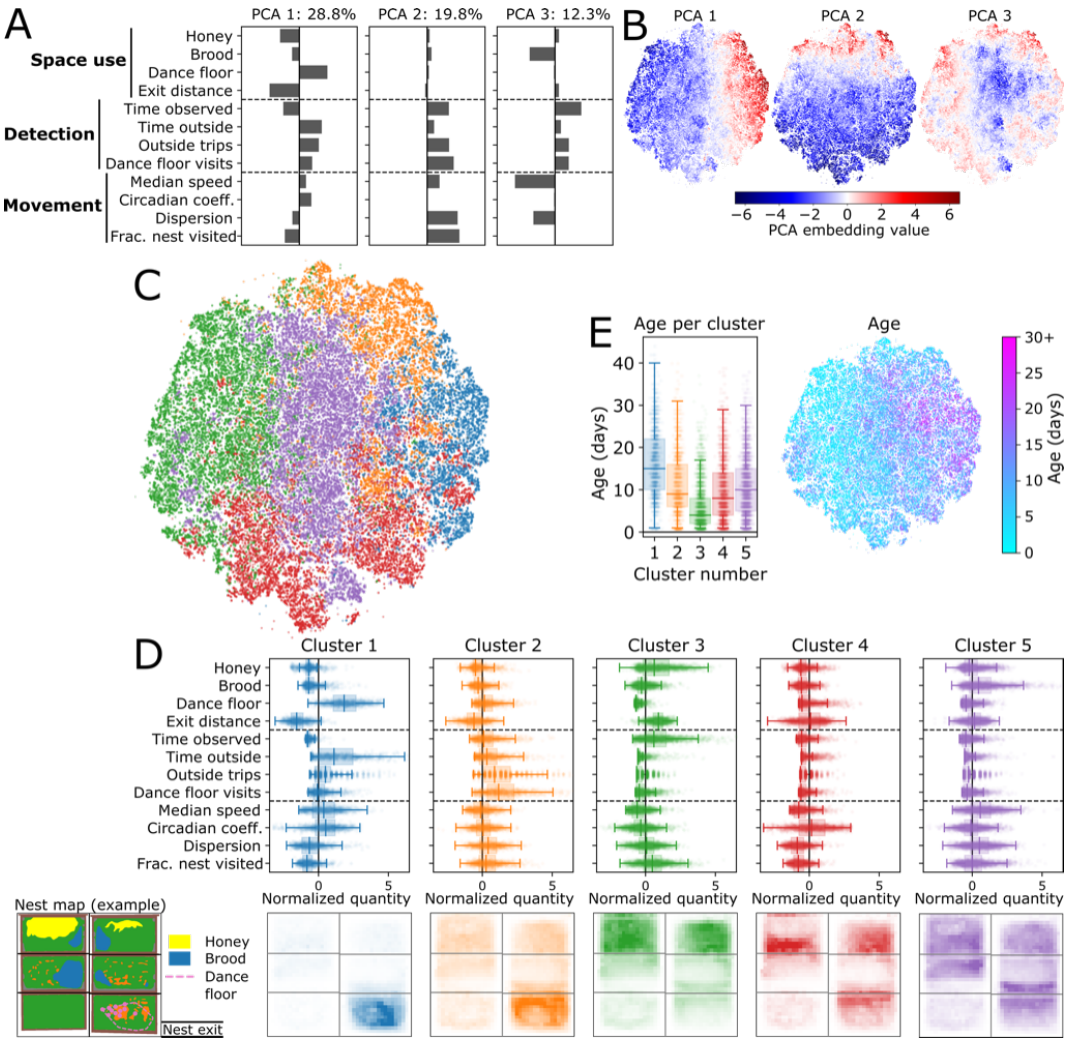

Figure 5.4: **Differences in observed single day behavior.** See also Figure S3. (A) The first three components from the PCA decomposition of individual bee behavioral metrics on a given day. (B) t-SNE embedding of behavioral days, colored by the projection values along each PCA component dimension. The t-SNE is initialized with the first two PCA component projections, and therefore the global structure of the t-SNE embeddings aligns with these projections. (C) Distributions of behavioral days using 5 clusters, showing behavioral metrics and average nest location histograms. Colors highlight the different groups on the t-SNE embedding (center). Nest histograms reflect the layout of the observation hive, sample shown at the bottom right for reference, and in Figures 5.1B, S1. (D) Age distributions of each behavioral day cluster.

variation in lifetime behavior. As with the per-day results, we therefore use clustering as a descriptive tool. We use a 5-cluster grouping to describe, compare, and visualize differences in the lifetime behavior of bees along the dominant two life-PCA components (Figure S4E). To visualize these groups, we show each bee as a single data point projected on the first two life-PCA components (Figure 5.5B), the lifespan of bees in each cluster (Figure 5.5C), the distribution of values of behavioral metrics averaged over the lifetime of bees in each cluster (Figure 5.5D), and the average behavioral changes with time as projected onto the behavioral day embedding (Figure 5.5E). Life cluster 1 includes bees with average movement activity levels, but with late transitions to dance floor/outside activity. Conversely, life cluster 3 bees are similar in their average movement activity (i.e. a similar projection onto life PCA 1 compared to cluster 1 bees), but show an earlier transition to dance floor/outside activity. The average projection on the behavioral day embedding further demonstrates this trend: while bees in clusters 1 and 3 on average move through similar regions of behavioral space (both move "left-to-right", going roughly through the middle of the behavioral day embedding space), life cluster 3 bees change their behavior at younger ages than life cluster 1 bees (Figure 5.5E).

A similar comparison differentiates life cluster 4 and 5 bees; although bees in both of these clusters have higher than average movement activity over their lifetimes (i.e. positive projections onto life PCA 1), life cluster 4 bees transitioned at older ages to dance floor/outside activity in comparison with life cluster 4 bees. Life cluster 2 bees have the lowest movement activity (negative projection onto life PCA 1), and average transition ages. Bees in life clusters 3 and 5 - i.e. those that transitioned to dance floor/outside activities at younger ages - had shorter lifespans than bees in the other life clusters (Figure 5.5C).

While there is no correlation between the two PCA projections (this is true by definition of the PCA calculation), the points representing the individual bees are not uniformly distributed on the PCA axes (Figure 5.5B). High movement activity bees (life clusters 4&5; positive projection on life-PCA 1) have a higher variance in observated rates of transitioning, compared to slow/localized bees (life cluster 1; negative projection on life-PCA 1). This means that while overall movement characteristics are not correlated with when a bee transitions to outdoor-related activities, high movement activity bees tend to be at the extremes of the distribution (i.e. the earliest or the latest to transition bees are fast). Slow/localized bees have average ages of behavioral transitions and a lower variance among individuals.

The results of the bee-life analysis show that individuals exhibit lifelong behavioral consistencies: those with positive projections onto life PCA 1 (e.g. those in life cluster 4) tend to move faster and visit more areas of the nest over their entire lives, and individuals with negative projections onto life PCA 1 (e.g. those in life cluster 2) tend to move more slowly and visit less areas of the nest over their entire lives (5.5D). We also find that individuals differ in the timing of their developmental transitions (Figure 5.5E): individuals with positive projections onto life PCA 2 (e.g. those in life clusters 3 and 5) transition at younger ages to outside activities in comparison with individuals with negative projections onto life PCA 2 (e.g. those in life clusters 1 and 4). Moreover, we see that the timing of behavioral transitions is related to longevity: individuals who transitioned to outside activities at earlier ages lived shorter lives (Figure 5.5C).

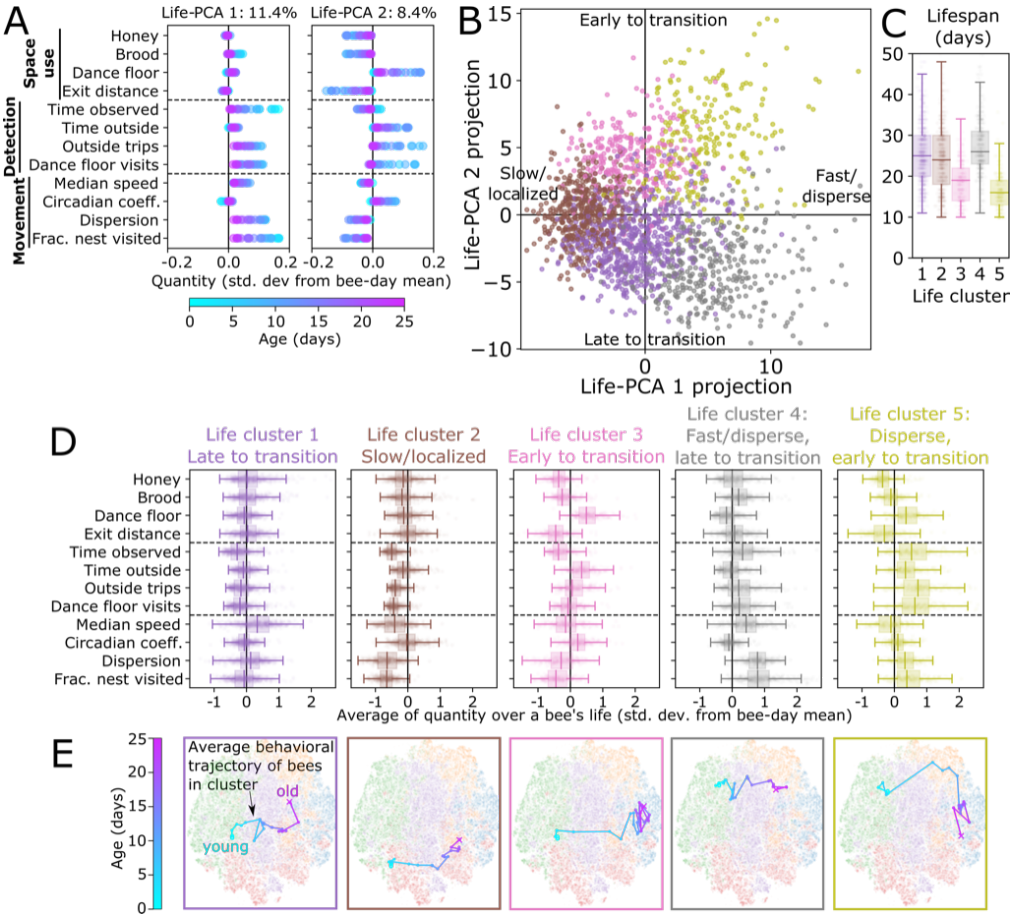

Figure 5.5: **Behavioral differences across a bee's entire life.** See also Figure S4. (A) PCA decomposition of bee-lives shows the dominant modes for how the behavioral metrics change over time. Plots show each PCA mode plotted in terms of behavioral metrics, using normalized quantities with the same units as in Figure 5.4 (i.e. zero represents the mean of the behavioral metric across all behavioral days). Points are colored by the age corresponding to each metric (see Figure S4A for an expanded plot). (B) A plot of individual bee-lives projected onto the first two PCA modes. Each point represents the life of a single bee. The colors correspond to a 5-cluster division, identified via Ward hierarchical clustering, and the labels describe the different life-PCA axes. (C) Distributions of number of days lived for bees in each life cluster. Note that bees are only included if they lived at least 10 days. (D) The distribution of the lifetime average of each behavioral metric among individual bees, grouped by bee-life cluster. The x-axis is the average of each quantity during a bee's life, in units of standard deviations from the mean of all behavioral days. (E) The average lifetime behavioral "trajectory" of behavioral days in each life cluster, projected onto the behavioral day embedding space. This is determined by averaging the metrics of bees of a certain age in each cluster and projecting these averages onto the behavioral day t-SNE embedding shown in Figures 5.4C, (Poličar et al. (2019); see Methods). Points and connecting lines are colored by age.

5.6 Discussion

Using individual tracking data from 4,100+ honey bees, we calculated behavioral metrics from the motion data and defined an analysis framework to describe behavioral variation at different timescales. At the timescale of a single day, bees differed their space use, detections, and movement, as quantified by the behavioral metrics shown in Figure 5.3. Although some behavioral patterns are more associated with older bees (e.g. behavioral day cluster 1), and others with younger bees (e.g. behavioral day cluster 3), we see considerable overlap in the age distributions associated with different behavioral days (Figure 5.4). Looking at the entire lives of individuals, bees predominantly differed in their movement patterns (speed/dispersion; Life-PCA 1), and the age at which they transitioned to dance floor/outside activities (Life-PCA 2) (Figure 5.5). We found that across entire lifetimes, some individuals exhibit consistently different movement characteristics – in particular, consistently higher (or lower) dispersion across nest areas over their entire lives (Figure 5.5).

Behavioral differences among individuals may enable eusocial insect colonies to be flexible in response to changing conditions, yet robust to the maintenance of other colony functions (Jandt and Gordon, 2016; Garrison et al., 2018). Individual tracking of bumblebees has revealed consistent differences in movement activity (Crall et al., 2018; Jandt and Dornhaus, 2009) – in particular, in the overall spatial area occupied by an individual (i.e. dispersion). Other work has shown, for example, that bumblebees differ in thermoregulation response thresholds (Jandt and Dornhaus, 2014), ants show consistent differences in exploratory behavior (Maák et al., 2020), and honey bees differ in dance activity in response to the same food source (George and Brockmann, 2019). It is important to note that the colony response is an emergent outcome of the many individuals, where each individual also adjusts their behavior in response to the behavior of others (e.g. Ulrich et al. (2021)). In general, the distribution of individual behavioral traits within a eusocial insect colony is expected to affect colony function, because the colony is the reproductive unit that selection acts upon (Jeanson and Weidenmüller, 2014; Jandt and Gordon, 2016). However, the effect of inter-individual variation may depend on the specific function. For example, while the effect of inter-individual differences in response thresholds on overall bumblebee colony thermoregulation behavior is unclear (Jandt and Dornhaus, 2014), variation in body size among bumble bee workers in a colony has been linked to enhanced comb production (Holland et al., 2021), and other work with ants has demonstrated that the distribution of individual traits affects colony foraging behavior (Kolay et al., 2020). To understand the effects of inter-individual variation on colony performance, it is therefore important to consider both the specific colony function as well as the ecological context (Davidson et al., 2021; Gordon, 2016).

It is well-known that there is a genetic basis for behavior in honey bees (Robinson et al., 1989; Fewell and Page, 1993; Calderone and Page, 1988; Page and Robinson, 1991; Calderone and Page, 1991; Junca et al., 2019; George et al., 2020), which likely also applies to lifetime behavior. The cohorts used in this study came from naturally-mated colonies; each source colony has a different queen, and while some cohorts came from the same source colony (see Figure S5), workers in a given source colony also represent multiple different patrilines (queens mate with 12 ± 6 drones; (Tarpy et al., 2004)). To examine precisely the extent to which our results have a genetic basis, future work could compare behavior from single-drone inseminated queens, or use genomic sequencing to determine each worker's patriline (Junca et al., 2019). Patriline diversity is important

for colony-level function (Jones et al., 2004; Seeley and Tarpy, 2007; Mattila and Seeley, 2007; Mattila et al., 2012)); whether a diversity in "bee-lives" (i.e. differences in movement characteristics and behavioral transitioning ages; Figure 5.5) contributes to colony function is unknown.

In our analysis, we find that bees differ in both movement characteristics and the age at which they transition to spending time on the dance floor and outside of the nest (Figures 5.5, S4). Previous work has noted how age is not the only factor that determines task allocation and behavioral transitions; social interactions, colony state, and environmental conditions also play a role (Johnson, 2010; Beshers and Fewell, 2001; Jeanson and Weidenmüller, 2014; Wild et al., 2021b). While we see differences in space use with age, in our analysis of movement characteristics, we find that average speed tends to increase with age but dispersion does not (Figure S2. For example, while age explains 9.2% of the variance of all metrics together, age explains only 0.6% of the variance in dispersion. The amount explained by age is 14.4% for speed, and as much as 23.1% for median exit distance (see Table 5.2). We also note that precocious foraging, which is similar to the "early-to-transition" individuals that we observe, can be induced via hormone treatments (Robinson et al., 1989), infection (Woyciechowski and Moroń, 2009), colony demography (Huang and Robinson, 1996), or even pesticide exposure Hesselbach et al. (2020), but here we see that such individuals exist even in unmanipulated colonies, similar to (Wild et al., 2021b). In wasps, differences in the age at which individuals transition to different tasks has also been observed (Jeanne et al., 1988). Across cohorts, individuals from cohort N did tend to show more early-to-transition behavior than bees in other cohorts (Figure S5) but further experiments would be needed to show whether such differences are driven by genetic or environmental factors.

Our study uses a large observation hive (3-frames; 7252 cm^2 of surface area), which is larger and can house more bees than other studies using automated tracking of honey bees (e.g. Bozek et al. (2021); Jones et al. (2020); Wild et al. (2021b,b). It is possible that nest size influences task allocation or transition rates; for example, workers in smaller colonies may transition between tasks more frequently (Jeanne, 1986; Dornhaus et al., 2012)). The observation hive was designed to mimic natural conditions and provide sufficient space for spatially separated comb-use areas (e.g. a dance floor that does not overlap with brood). Still, it is smaller than a natural nest (mature natural nests can have 13369 ± 1174 cm2 of comb surface area; Smith et al. (2016)). We note that a systematic comparison of how nest structure influences behavior should consider not only size, but also nest geometry (e.g. Pinter-Wollman (2015)).

Previous work has used ethograms to define categorical age-based labels such as nurses, middle-aged bees, and foragers (Lindauer, 1952; Seeley, 1982; Seeley and Kolmes, 1991; Johnson, 2010, 2008a,b). While such labels have the advantage of being easy to interpret, manually assigning behavioral tasks has multiple disadvantages, including: limited reproducibility (ethogram interpretations depend on the observer), behavioral descriptions must fit into pre-defined categories, and scaling issues (tracking multiple bees simultaneously, or over long time-periods, can be infeasible). Although automated tracking methods address these issues, simple trajectory data may not always be of direct biological or functional relevance (Krause et al., 2013). In the current study, for example, we incorporate maps of the nest structure to extract additional biological information for a given spatial positioning (e.g. the individual is located atop brood, versus on the dance floor). With honey bees, tasks are often location-specific, such that, for example, bees found on the brood area are typically doing brood care (Seeley, 1982). However, using location to infer task is an

assumption, and some tasks, such as fanning, may not be location-specific. This is an inherent tradeoff with high-throughput methods like automated tracking. An important area for future work is to compare and relate the results of automated tracking methods, to approaches that use ethograms to manually assign behavior and task repertoires (e.g. cell cleaning, fanning, waggle dances) (Lindauer, 1952; Seeley, 1982; Smith et al., 2017; Perez and Johnson, 2019; Mattila et al., 2012).

Recent work has combined barcode tracking with supervised machine learning methods to automatically identify specific behavioral events (Jones et al., 2020; Gernat et al., 2023). These approaches apply convolutional neural networks (CNNs) to video data to identify a specific behavior of interest (e.g. egg-laying), which can be associated with the known identities of tracked bees through the barcode positions. Gernat et al. (2023) trained their CNN to detect trophollaxis events, and Jones et al. (2020) to detect egg-laying events and when bees exited for outside trips. These are supervised methods which require training and specified behavior to identify, and thus have focused on a few types of behavioral events which could be reliably identified. Alternatively, recent work has combined general methods of pose estimation with barcode tracking and applied this to bumblebees (Smith et al., 2022b); such pose estimation data could be used with unsupervised methods in order to identify complex behavioral patterns without training or a-priori specification (Berman et al., 2014; Graving and Couzin, 2020). In contrast to these approaches, which use smaller colonies and shorter tracking periods of 2-7 days (Jones et al., 2020; Gernat et al., 2023; Smith et al., 2022b), in this study we extract only trajectory data from barcode tracking, which enables the analysis of thousands of bees during their entire lifetimes in a timespan of several months. Future work can merge these approaches or choose the methods most appropriate to specific bioloical questions, by combining aspects of supervised identification of behavioral events, unsupervised behavioral classification from pose estimation, and behavioral metrics calculated from trajectory data.

Automated tracking makes it possible to obtain long-term datasets for thousands of individuals, making it possible to investigate individual variation at an unprecedented scale. Our long-term tracking results present a detailed picture of how individuals in a colony differ in their behavior from day-to-day and over entire lifetimes, and establishes an analysis framework that can quantify these differences and how they may contribute to colony function.

5.6.1 Limitations of the study

In this study, we analyzed the data of thousands of honeybees tracked using barcodes in an observation hive over an entire summer. While we examine variation among behavioral days and across the lifetimes of individual bees, we note that the metrics used to quantify behavior are restricted to quantities that can be calculated from the trajectory data (see Figure 5.3 for behavioral metrics). As such, these metrics do not directly represent biologically-relevant behavioral patterns, such as foraging, cell cleaning, or fanning, that are typically identified with manual observation. Future work could examine how the behavioral metrics calculated from trajectory data are correlated with such manual assignments of behavior.

Although we examined the behavior of thousands of bees from multiple age-matched cohorts, our data are from a single observation hive over a single summer. Given that the colony had free

access to forage outside, and that behavior can change with environmental factors, we can expect results to differ quantitatively from year-to-year. Nonetheless, we expect that observed qualitative trends would be similar for such a repeated experiment. Future work would be needed to test the repeatability and robustness of the observed trends, given the colony-level sample size.

5.7 Acknowledgements

Thank you to Giovanni Galizia for providing access to his rooftop apiary, to Jan Peters for expert colony maintenance, to Dagmar Olalere and Katja Anderson for essential logistical support, to Jake Graving for critical feedback, and to Louisa Neubauer, Sinje Tigges, Christine Bauer, and Jayme Weglarski for patiently helping to tag bees.

5.8 Declaration of interests

Authors declare that they have no competing interests.

5.9 Funding

This work was supported by:
Simons Foundation Postdoctoral Fellowship of the Life Sciences Research Foundation (MLS)
Heidelberger Akademie der Wissenschaften under the WIN (wissenschaftlichen Nachwuchs) program (MLS, JDD)
Deutsche Forschungsgemeinschaft (German Research Foundation) under Germany's Excellence Strategy EXC 2117-422037984 (IDC)
Office of Naval Research N00014-19-1-2556 (IDC)
HPC-Service of ZEDAT (Freie Universität Berlin) (BW, DMD, TL)
North-German Supercomputing Alliance (BW, DMD, TL)
This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 860949 (IDC)
European Union's Horizon 2020 research and innovation program (grant no. 824069) (DMD, TL)
Klaus Tschira Foundation (grant no. 00.300.2016) (BW, TL)
Andrea von Braun Foundation and the Elsa-Neumann-Scholarship (DMD)
Zukunftskolleg Mentorship Program (MLS, TL)

5.10 STAR Methods

5.10.1 Resource availability

Lead contact

Information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Michael L. Smith (mls0154@auburn.edu).

Materials availability

The study did not generate new unique reagents

Data and code availability

- The data and code needed to reproduce the results in this publication are available at: github.com/jacobdavidson/bees_lifetimetracking_2018data.
- The full dataset associated with tracking bees during summer 2018 is deposited at Zenodo and is publicly available. The DOI is doi.org/10.5281/zenodo.6045860 and is also listed in the Key Resources table. This full dataset includes the x-y trajectories, behavioral metrics calculated at different time intervals (1-hour, 5 minute, and 1 minute intervals), and comb maps.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

5.10.2 Experimental model and subject details

Newborn worker bees were sourced from colonies headed by naturally mated queens from the University of Konstanz apiary. Individual age-matched cohorts were selected from eight different source colonies: cohorts A, H, M from colony c1; cohort D from colony c2; cohorts B, I, N from colony c3; cohort L from colony c4; cohorts C, E, K from colony c5; cohorts F, G from colony c6; cohort O from colony c7; cohorts J, P from colony c8.

5.10.3 Method details

Observation hive and nest maps

This research was conducted at the University of Konstanz, Germany (47.6894N, 9.1869E). On 10 June 2018, the observation hive was installed with a single queen, 2,000 unmarked workers, and three frames of mixed brood and honey ("Deutsche-Normal" frames: 395 x 225 mm, observation hive: 490 x 742 mm; note that this observation hive is the largest, to date, to be used for automated tracking in honey bees (Wario et al., 2015; Bozek et al., 2021; Boenisch et al., 2018a; Wild et al., 2021b)). From 16 July to 3 Septempter 2018, every 4-6 days, we individually marked and introduced 200-600 newborn honey bees to the observation hive (total bees tagged: 5,343).

Although tracking data was obtained continuously until 9 October 2018, we perform our analysis on a focus observation period of 16 July - 3 September, during which new cohorts were regularly introduced. Newborns were hatched overnight in an incubator kept at 34°C and 50 %RH, and marked the following morning with individual BeesBook tags (Wario et al., 2015; Boenisch et al., 2018a). Tags are printed on paper and attached to the thorax of bees, and remain attached for their whole lives. From 16 July to 3 Sept 2018 (50 days) we recorded the observation hive at 3 frames per second using four Basler acA4112-20um cameras fitted with Kowa LM25XC lenses and the recording software Motif (Loopbio GmbH). The colony was illuminated with infrared light (850nm 3W LED's), which is invisible to honey bees (Peitsch et al., 1992). The entire recording rig (observation hive, cameras, lighting) was kept in the dark, to mimic the natural conditions of the honey bee nest. Workers had free access to forage outside, through a entrance tunnel (2-cm diameter). To keep track of the colony's weight, the observation hive was kept on a scale which logged its weight every hour (10g sensitivity, Wolf Waagen GmbH). To create a map of the nest, every 4-6 days we traced the contents of the observation hive onto plastic sheets by outlining the following: honey storage, pollen storage, brood, empty comb, wooden frames, peripheral galleries, and dances observed on the dance floor (as in (Smith et al., 2016); Figures 5.1B, S1). These plastic sheets were then scanned with an architectural scanner (Ruch-Medien, Konstanz), and digitized. By overlaying the bee trajectories upon the maps, we determined what type of nest environment an individual experienced (Figure 5.1B).

5.10.4 Quantification and statistical analysis

Data processing and behavioral metrics

Using the BeesBook system (https://github.com/BioroboticsLab/pipeline), the raw image data were processed to detect and decode the individually marked bees (Wild et al., 2018; Boenisch et al., 2018a; Wild et al., 2021b). For each individual, its tag id, id detection confidence, position, and orientation were tracked over time, and stored in a PostgreSQL database. The death date of each marked individual was estimated using a Bayesian changepoint model (as in (Wild et al., 2021b)). This method accounts for a low rate of erroneous detections in bees that have already died, and time periods when individuals are observed less frequently or not at all (e.g. while foraging). An individual's death date was used as a cutoff for including data in subsequent calculations.

We chose metrics that represent space use within the nest (time on honey, brood, or dance floor, and exit distance), detection (time observed, time outside, number of outside trips, and number of dance floor visits), and movement/spatial localization (speed, circadian coefficient, dispersion, fraction of nest visited). Although some of these metrics are correlated (Figure S3A), they nonetheless represent different aspects of behavior, and we use the approach of combining multiple different metrics in order to obtain results that are robust to inclusion of specific metrics, as well as any particular parameter choices associated with each metric.

We processed the trajectory data to obtain the quantities used in the subsequent analyses by first averaging over 1-hour time bins and saving the quantities of interest for each individual bee. The 1-hour bins were used to speed up processing the large amount of data. All data points used in the analysis were above a detection confidence threshold of 0.8, and we calculated behavioral metrics for each bee that had a minimum of 10 detections in that hour. For time observed, number

of outside trips, and number of dance floor visits, the per-day value is a sum across hours. The circadian coefficient is determined using the per-hour median speed over the coarse of a day. For the other 8 metrics, the per-day quantity is calculated as a weighted average across the hours in the day, where weightings are done according to the amount of time observed in that hour.

Table 5.1 shows a summary with definitions of all metrics used. Further details regarding calculations of substrate usage, circadian coefficient, and trips are described here.

Substrate usage is calculated using the comb substrate maps shown in Figure S1, grouping together capped and young brood into a single category. Note that dances were observed only within a limited time range (pink circles in Figure S1), but all occurred in a similar area. Defining the dance floor based on only direct observations would be overly restrictive, so we defined the dance floor area using a convex hull that contains all dances over the entire observation period (dashed pink line in Figure S1). Because the comb contents changed over time, and were not measured each day, we calculated substrate usage by a weighted average from values calculated using the substrate maps on the measurement days before and after the day in consideration. To illustrate this procedure, consider the day July 18, which has the closest measurement days of July 16 and 21. Denote the comb map from July 16 as A, and the comb map on July 21 as B. We first use the trajectory coordinates of the bee on July 18 to calculate two different approximate usage fractions: S_i^A, which is the fraction of time spent on substrate i as determined using map A, and S_i^B, which is the fraction of time spent on substrate i as determined using map B. The estimated substrate usage fraction for July 18 is calculated is calculated as a weighted average of these values:

$$f_i = w_A S_i^A + w_B S_i^B,$$

where for this example the weights are $w_A = 0.6$ and $w_B = 0.4$, because the first comb measurement day is closer than the second to July 18. The nest comb contents over time were also determined by this same linear interpolation method between nest content measurement days.

The circadian coefficient is calculated as the correlation of median speed over the day with a daily rhythm that follows the sun. We approximate the daily rhythm with a sine curve of $\sin((h - m)/(2\pi))$, where h is the hour of the day, and m is chosen so that the maximum of the curve coincides with the highest sun position of the day, which was approximately 13:30 CEST during the observation period. The circadian coefficient is then calculated as

$$C = \frac{\sum_{h=1}^{24} s_h \sin\left(\frac{h-m}{2\pi}\right)}{\sum_{h=1}^{24} s_h},$$

With this normalization the coefficient satisfies $-1 \leq C \leq 1$, where the extreme values only occur if the bee is not observed for the whole day. Positive values represent higher speed or only being observed during the day, while negative values represent higher speed or only being observed at night.

A bee's barcode is not always detected when it is in the observation hive, for example if the bee is upside-down or in a dense crowd of other bees. Because of this, we used both detection and exit distance to estimate when a bee was outside. The time outside and number of outside trips are estimated by first calculating the time observed and median exit distance in 1-minute bins over the coarse of a day. A bee is then estimated to have exited the nest in a time bin t_{exit} if the

Metric	Definition and description
Honey, Brood, Dance floor	Fraction of observed time spent on these substrates, as defined using the comb maps of Figure S1. For days when the comb was not measured, we used a weighted average with the closest measurement days.
Exit distance	Median shortest path distance to the exit (which is located in the lower right corner), accounting for possible routes to switch sides, but not adding any extra distance for a switch of sides
Time observed	Total time observed in a day, calculated as the total number of detections with confidence interval over the 0.8 detection threshold, divided by the frame rate of 3 frames per second.
Time outside	An estimate of the total amount of the time a bee spends outside during a day.
Number of outside trips	An estimate of the number of times a bee exited the nest in a day.
Number of Dance floor visits	The number of times that a bee entered the dance floor from another substrate
Median speed	Median speed during time observed, omitting instances where the bee switched sides of the comb, as well as when the time between detections was > 1 second
Circadian coefficient	A representation of how activity levels change with the time of day; positive values represent higher observed speeds during the day, while negative values represent higher speeds at night.
Dispersion	Root mean square distance from the centroid of the x-y coordinates, calculated by considering motion in a 2D plane in the hive (i.e. neglecting whether the bee was detected on the front or the back of the observation hive).
Fraction nest visited	After dividing the nest area into discrete spatial bins of 2cm$\times$2cm (the same grid size used in the spatial histograms shown in Figure 5.4), this is the fraction of bins with at least one detection. Note that the body size of a bee is approx. 1 cm.

Table 5.1: Behavioral metrics used in the analysis. See text for precise descriptions of how substrate usage (honey, brood, dance floor), circadian coefficient, and time outside/outside trips are calculated.

time observed in t_{exit} is less than a threshold of t_{obs}=2 sec., and if the median exit distance in time bin $t_{exit} - 1$ is less than a threshold d_{exit}=18.75 cm (1500 pixels). The bee is considered to have re-entered in bin t_{enter} if the time observed in t_{enter} is greater than or equal to t_{obs}. The values t_{obs} and d_{exit} are analysis parameters, and the results can depend strongly on the choice of d_{exit}; we choose the value of 18.75cm to represent a feasible median exit distance for a bee traveling to the exit during a 1-minute period. With these results, we determine multiple instances of exit and re-entry times during the course of a day, and use this to calculate the number of outside trips (the number of times a bee is estimated to have exited the nest), as well as the time outside.

Note that dispersion and fraction of the nest visited are two complementary metrics which both represent how wide-ranging each bee is, regardless of where it tends to be located in the nest. While dispersion and fraction of the nest visited give similar results for continuous exploratory movement, they can yield different trends for other cases; for example, bursty movement can yield high fraction of nest visited yet low dispersion, while directed, straight-line back-and-forth movement can yield low fraction of the nest visited yet high dispersion.

PCA and clustering on single day metrics

Using the behavioral metrics (Table 5.1), we create a data matrix M_{ij}, where each row i represents one behavioral day, and columns $j = 1...12$ are the different quantities. A behavioral day is only included if that bee was alive on the given day and had more than 1,000 detections over the whole day. This represents a total time observed of 5.5 minutes during a day; using this removes 7201 behavioral days with few detections (results are qualitatively similar whether these are included or not). In addition, we do not include bees on the first day they were introduced, because on this day there were not observed for a full 24 hours. With this criteria $i = 1...53,032$ behavioral days are included in the analysis. Although the total number of tagged bees was 5,343, the bees in cohorts A-F were tagged before filming began, and some died before 16 July. Due to this, and after filtering, we include data from a total of 4,193 unique bees in the analysis (the number is 4,229 before filtering for few detections).

Note that the nest contents – in particular the size of the honey and brood areas – change over time (Figure S1). We account for these changes in order to focus on variation among the activity of bees in the nest at a given time, instead of changes in substrate usage that result from a different nest composition. For honey and brood areas, we account for this by subtracting the nest content fraction from the individual bee substrate usage fraction for each day. The dance floor is unaffected, since it is defined as the same area over the course of the observation period.

Following standard procedures, we normalized the data matrix $\mathbf{M}$ so that the column mean is zero and the column standard deviation is 1. We then performed principal component analysis (PCA) on the resulting matrix to obtain the components shown in Figure 5.4A. The result of PCA is a matrix U_{ij}, where i represents behavioral days and $j = 1...12$ for the PCA components (corresponding to the total number of behavioral metrics).

Next, we perform Ward hierarchical clustering, implemented in Python in the package `scipy.cluster.hierarchy`, to obtain the results shown in Figures 5.4 and S3.

Ward clustering minimizes the overall within-cluster variance. We write this as variance fraction: for n clusters, this is calculated as

$$W(n) = \frac{1}{\|\mathbf{M}\|^2} \sum_{q=1}^{n} \sum_{i \in q} \sum_{k} \Big(M_{ik} - \langle M_{jk} \rangle_{j \in q} \Big)^2, \tag{5.1}$$

where $\langle \cdot \rangle_{j \in q}$ represents an average over the indices j that are elements of cluster q, and $\|\mathbf{M}\|^2 = \sum_{i,j} M_{ij}^2$ is the squared magnitude of the data matrix. This is shown in Figure S3.

We use t-SNE embedding (Maaten and Hinton, 2008) implemented in openTSNE (Poličar et al., 2019), with parameters of perplexity=30 and n_iter=1000, and initial conditions set by the first two PCA dimensions to obtain the behavioral day embeddings shown in Figure 5.4. This package enables the mapping of new data to existing embeddings, which we used to show the average bee-life trajectories on top of the behavioral day embedding (Figure 5.5E).

Variance fraction explained

To compute the fraction of the total variance explained by age, cohort, or a combination of factors, we use the same procedure as in Eq. 5.1, but instead generalize to use some grouping $\{G\}$ instead of a certain number of clusters. The grouping $\{G\}$ can be defined to include bees of a certain age, bees of a certain cohort, or both of these (bees of a certain cohort having a certain age). The variance fraction for all metrics is then calculated as

$$W(\{G\}) = \frac{1}{\|\mathbf{M}\|^2} \sum_{q \in \{G\}} \sum_{i \in q} \sum_{k} \Big(M_{ik} - \langle M_{jk} \rangle_{j \in q} \Big)^2. \tag{5.2}$$

For a certain metric k, this is simply

$$W(\{G\}, k) = \frac{1}{\sum_i M_{ik}^2} \sum_{q \in \{G\}} \sum_{i \in q} \Big(M_{ik} - \langle M_{jk} \rangle_{j \in q} \Big)^2. \tag{5.3}$$

Bee-life

We use the behavioral metrics (Figure 5.3) computed over multiple days in order to compare the lifetime behavioral trajectories of individual bees. The results in Figure 5.4 treat each day for each bee separately, and each row of $\mathbf{M}$ represents one behavioral day. Building on this notation, we know that a bee's life is made up of multiple behavioral days. To ask about bee-lives with similar patterns and changes of activity as a bee ages, we filter and transform the data, and perform PCA on the behavioral metrics of each bee over time.

Individual bees have different lifespans; because of this, we did not include all bees in the lifetime analysis, but only those that were observed for at least 10 days. To compare lives we also need to a set a maximum value of the number of days to compare; we use a maximum of 25 days as value that is representative of the lifetime behavioral changes of bees. Since PCA cannot be performed if values

Metric	Variance explained by grouping (percentage) Age&Cohort	Age	Cohort
Honey	11.0	3.9	5.8
Brood	7.7	2.9	1.7
Dance floor	26.6	19.3	2.7
Exit distance	28.7	23.1	2.6
Time observed	27.1	17.1	14.2
Time outside	15.3	10.6	2.1
Num. outside trips	12.6	5.9	2.1
Num. dance floor visits	12.7	2.7	4.6
Median speed	27.0	14.4	6.8
Speed circadian coeff.	10.4	6.1	2.4
Dispersion	9.9	0.6	6.8
Fraction nest visited	14.5	3.7	10.1
All data and metrics	17.0	9.2	5.2

Table 5.2: **Amount of single-day behavioral variation explained by cohort and age**. This is calculated by subtracting the conditional average of each metric from the normalized behavioral day data matrix, where the conditional averages are calculated using groupings of age&cohort, age, or cohort. Values are shown for each metric using Eq. 5.3, as well as all metrics combined (Eq. 5.2.

are missing (which occurs, for example, after a bee has died), we use per-age average values of each metric to fill in missing values of the behavioral metrics for the purposes of PCA and clustering.

The tensor $B_{\alpha tj}$ is used to represent bee-lives, where α is for individual bees, t is an index over the days in the bee's life, which goes from 0 to l_α, where l_α is the total number of days the bee lived, and $j = 1...12$ is an index over the component values of $\mathbf{M}$. To analyze how different one bee's life is from another's, we must consider that all bees did not live for the same number of days. Because of this, we use a parameter $A_{max} = 25$ for the maximum age used in the bee-life analysis. Because some bees did not live a total of A_{max} days, and even if a bee was alive there could be some days where it was not detected by the tracking system, we only include bees for the lifetime analysis that had $D_{min} = 10$ days or more in the behavioral day data matrix. With these criteria, and also only keeping bees from cohort G onward, i.e. bees with birthdates within the observation period, we include 2027 bees in the bee-life analysis. We note that A_{max} and D_{min} are analysis parameters and quantitatively affect results, although we found that different values of these parameters lead to qualitatively similar interpretations in the differences among bee-lives. We used averages to fill in values of the bee-life matrix for the purposes of PCA and clustering, because PCA cannot be calculated on a matrix that has missing values. Let $h_{tj} = \langle B_{\alpha tj} \rangle_\alpha$, where the notation $\langle \cdot \rangle_\alpha$ represents an average over the index α, denote the average behavioral metrics for each day of the lives of bees that were observed. For a bee that was dead or not observed on day t of its life, we fill these values by setting $B_{\alpha tj}|_{\text{bee } \alpha \text{ dead or not observed on day } t} = h_{tj}$. We use h_{tj} instead of zeros to fill values for the bee-life distance metric, because although the column average of $\mathbf{M}$ is zero, the average conditional on the age of the bee is nonzero, and therefore filling with zeros would bias the results. After this filtering and processing, we use the bee-life matrix $B_{\alpha tj}$ as input to PCA and clustering, to obtain the results shown in Figure 5.5.

We obtain that PCA 1 explains 11.4% of the total variance, PCA 2 explains 8.4% of the total variance, and further components explain a smaller fraction of the total variance (Figures 5.5, S4). Note that because the input is high-dimensional, with $A_{max} * 12 = 300$ columns, the fraction of the variance explained by any single mode is relatively small, with an average at 0.31%, and therefore the first two modes represent strong patterns in the data because they are very high above this average variance fraction.

We use $B_{\alpha tj}$ as input to Ward hierarchical clustering obtain the clusters shown in Figure 5.5, and use a 5 cluster grouping to highlight differences along the first two dominant life PCA modes (Figure S4).

5.11 Supplemental figures

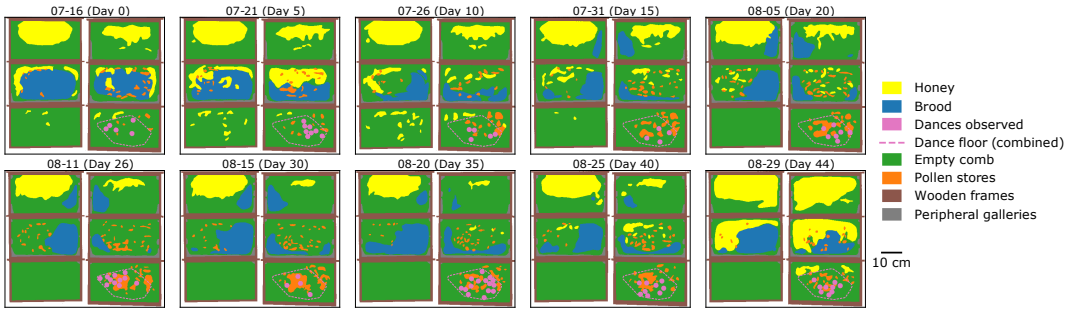

Figure S1: **Comb contents over the observation period.** Figure shows the nest contents and tracings. The pink dashed line in each shows the "combined dance floor", which is defined as a convex hull that contains the locations of all observed dances. The relative locations of the nest contents remained stable throughout the experiment, with honey stored at the top of the nest, brood reared in the center, and a dance floor at the bottom of the nest, near the nest entrance.

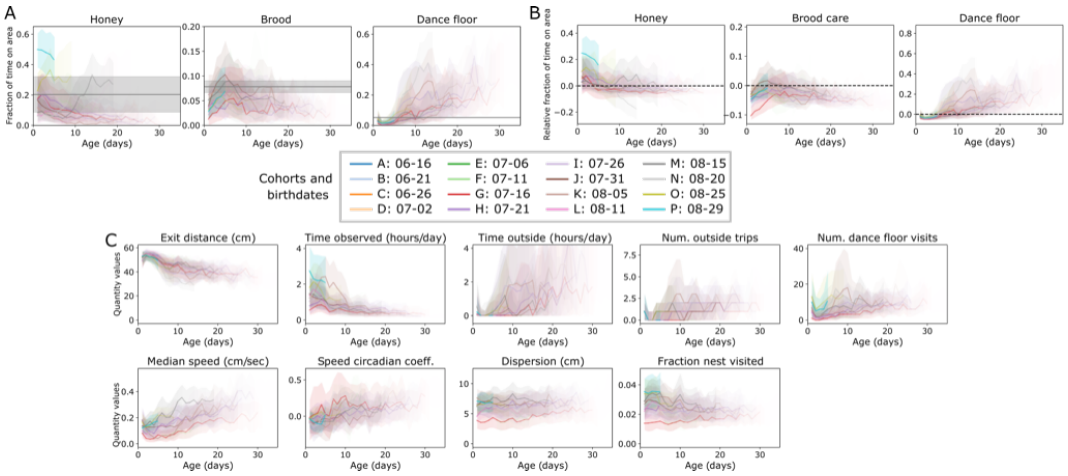

Figure S2: **Substrate and other quantities with age.** See also Figure 5.2. Cohorts are indicated by the different colors for each plot. (A) Fraction of time spent on honey and brood. The gray line and shaded area shows the mean and standard deviation of the amount of honey or brood in the nest over time. (B) Fraction of time on honey and brood relative to nest contents, determined by subtracting the average contents of the nest for a given day. (C) Other behavioral metrics with age.

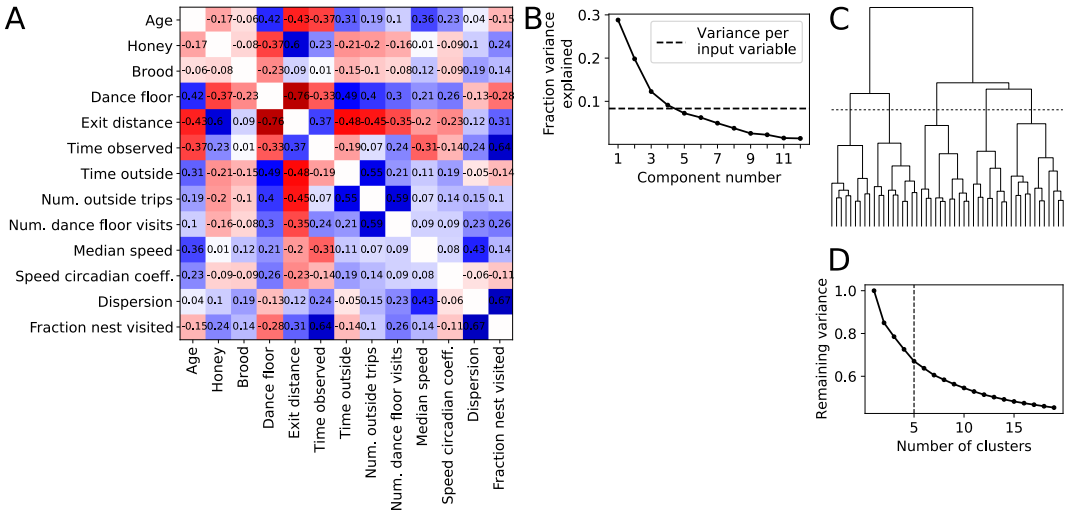

Figure S3: **Correlation between single-day quantity values, PCA variance explained, and behavioral day clustering results.** (A) Pearson (ranked) correlation coefficient shown, for all behavioral metrics as well as age. Blue indicates a positive correlation, and red indicates a negative correlation; the values of the correlation coefficient are shown for each pair of quantities. (B) Variance explained per behavioral day PCA component. Dashed line shows the variance per input column of M_{ij}, i.e. the contribution of each behavioral metric to the total variance. (B) Clustering dendogram and (D) average remaining variance as a function of the number of clusters (Eq. 5.1). In both, the dashed line shows the distance cut-off for a 5-cluster division.

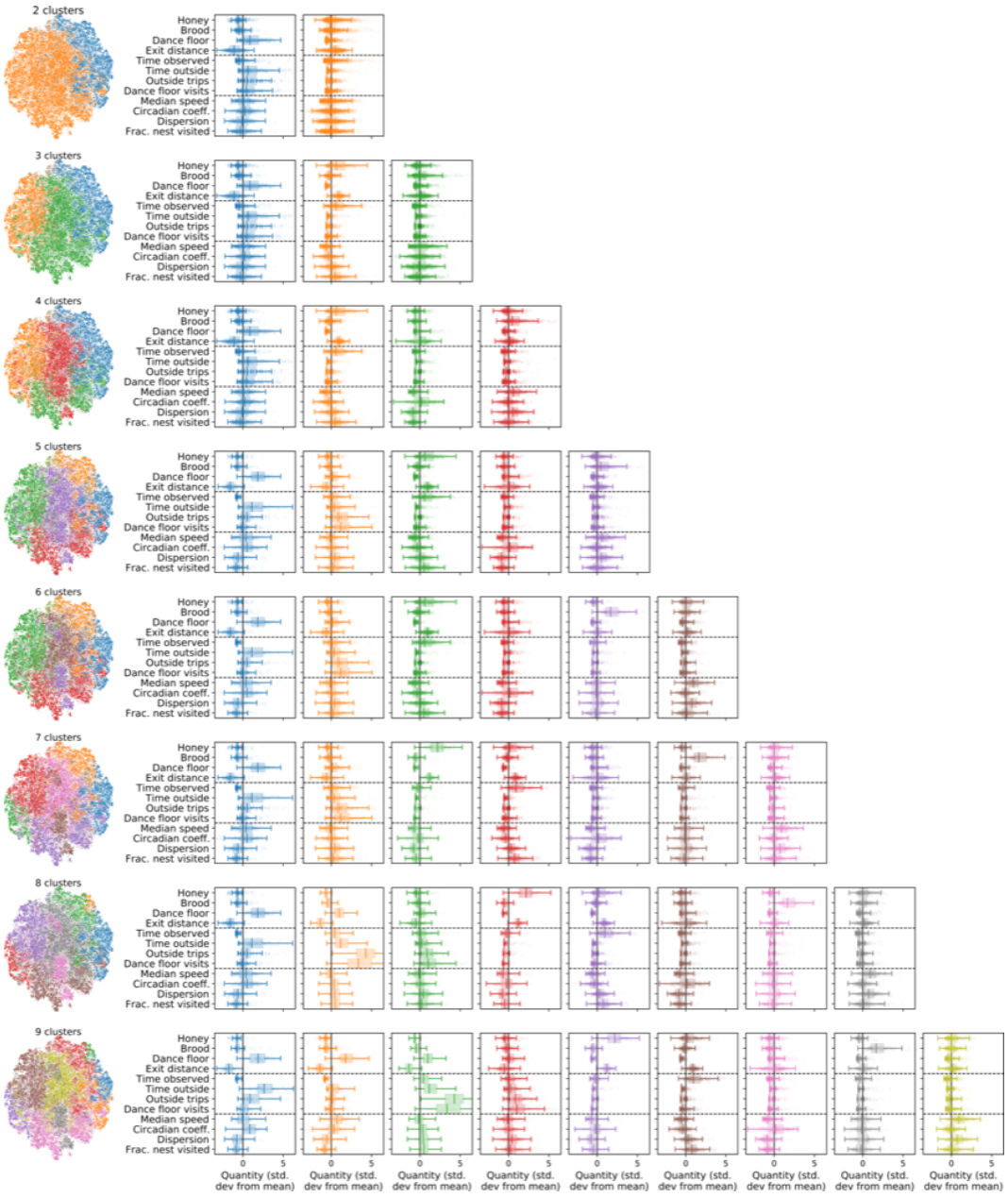

Figure S4: **Behavioral day embeddings and metrics for different numbers of clusters.** Each row corresponds to results for a certain number of clusters as obtained using Ward hierarchical clustering, and colors highlight the different clusters. (Left) t-SNE embeddings, where each point represents the behavioral metrics calclated for a single bee on a single day (see Figure 5.3 for metrics). (Right) distributions of behavioral metrics for behavioral days in each cluster. See Figure S3C for the clustering dendogram. Figure 5.4 in the main text shows results for 5 clusters.

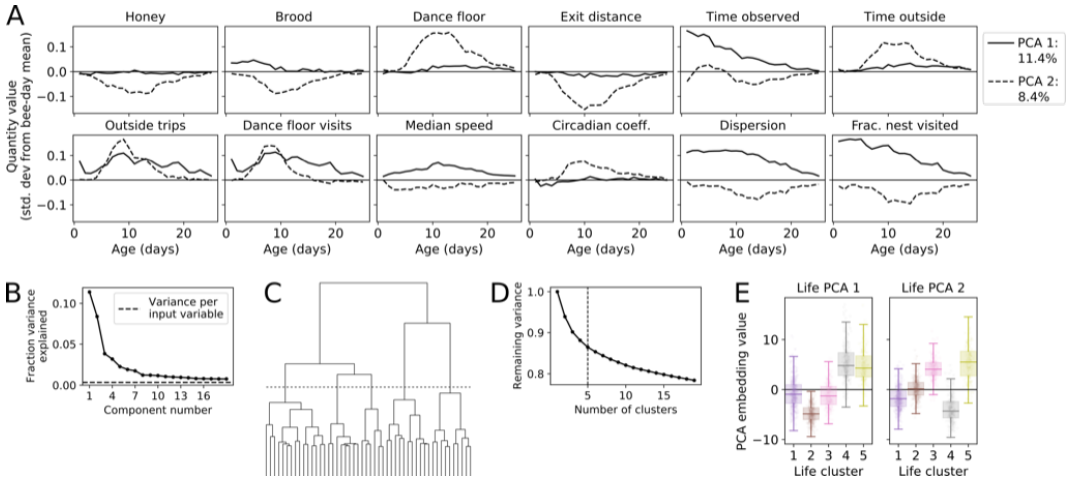

Figure S5: **Lifetime PCA decomposition and clustering.** (A) The first two life-PCA modes plotted in terms of behavioral metrics with age, using normalized quantities with the same units as Figure 5.4 (i.e. zero represents the mean of a certain behavioral metric across all behavioral days). (B) Variance explained per bee-life PCA component. Dashed line shows the variance per input column of $B_{\alpha tj}$, i.e. the contribution of each behavioral metric j for a bee with age t days to the total variance. (C) Clustering dendogram and (D) average remaining variance as a function of the number of clusters. In both, the dashed line shows the distance cut-off for a 5-cluster division. (E) Life-PCA embedding values, obtained by projecting lifetime behavioral metrics onto the lifetime PCA decomposition shown in A and Figure 5.5A.

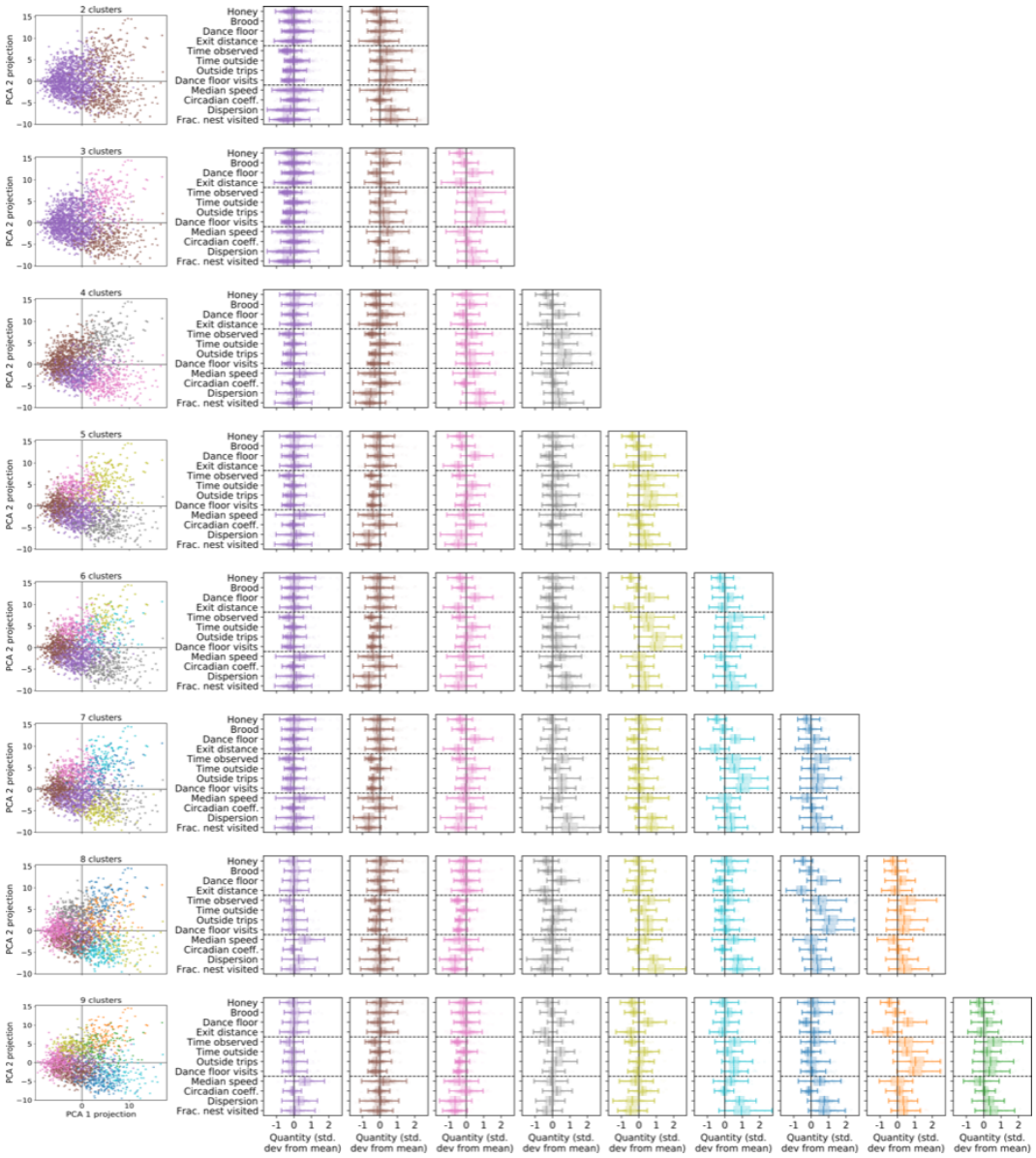

Figure S6: **Lifetime embedding and average metrics for different numbers of clusters.** Each row corresponds to results for a certain number of clusters as obtained using Ward hierarchical clustering, and colors highlight the different clusters. (Left) Bee-life data projected onto the first two PCA modes, where each point represents the life of a single bee. (Right) Distributions of the lifetime average of each behavioral metric among individual bees, grouped by cluster. As in Figure 5.5D, where results for 5 clusters are shown, the x-axis is the average of each quantity during a bee's life, in units of standard deviations from the mean of all behavioral days. Colors highlight the different groupings on the embeddings shown at the left. See Figure S4C for the clustering dendogram. Figure 5.5 in the main text shows results for 5 clusters.

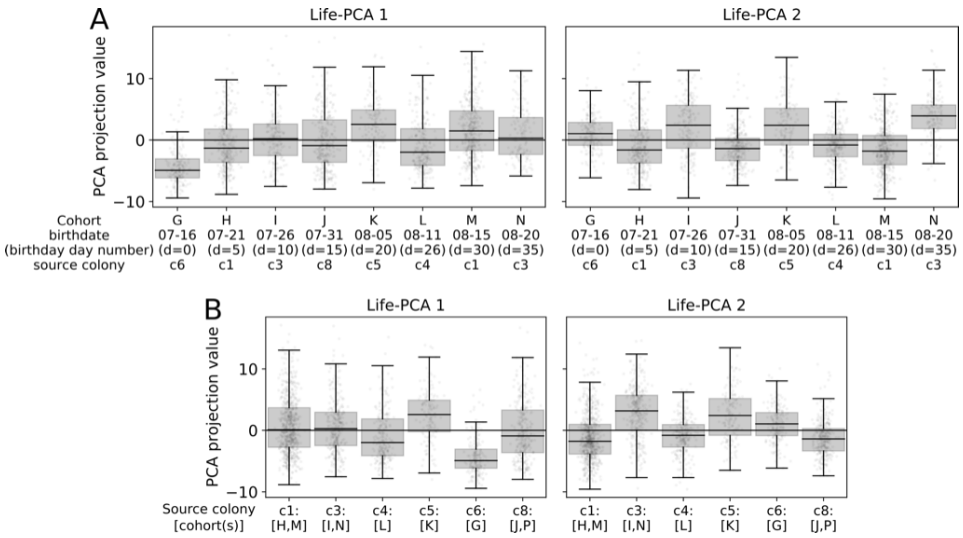

Figure S7: **Cohort and source colony distributions of lifetime behavior.** The distribution of life-PCA embeddings for bees that were included in the lifetime analysis. Note that only cohorts from G onward are included in the lifetime analysis, because these bees have birthdates within the observation period. The life-PCA modes 1 and 2 are shown in Figure 5.5. (A) Per-cohort distributions of lifetime PCA 1 and 2 projections. Cohorts are sorted chronologically, with birthdate and associated source colony shown in the label. (B) Per-source colony distributions of lifetime PCA 1 and 2. Associated cohorts are listed in the label.

6 Learning to embed lifetime social behavior from interaction dynamics

Benjamin Wild [1,*], David M. Dormagen [1], Michael L. Smith [2,3], Tim Landgraf [1,*]

[1] Department of Computer Science, Freie Universität, Berlin, Germany.
[2] Department of Biological Sciences, Auburn University, Auburn, AL 36849, USA.
[3] Department of Collective Behaviour, Max Planck Institute of Animal Behavior, Konstanz, Germany.

[*] Corresponding authors: b.w@fu-berlin.de, tim.landgraf@fu-berlin.de

6.1 Preface

Chapter 3 introduced a tracking system that allowed for the creation of honey bee datasets with high accuracy and unprecedented scale. In Chapter 4, this tracking system was used to derive and analyze data acquired in the Berlin recording setup, while in Chapter 5, the same system was used in the setup in Konstanz. These two works established new quantitative approaches to understand individual behavior in the colony based on their social network and behavioral metrics.

However, these two approaches were not suitable for analyzing data acquired from several colonies or over extended periods in a combined analytical framework. To address this issue, the paper entitled *Learning to Embed Lifetime Social Behavior from Interaction Dynamics* introduces a new method to identify the functional roles of individuals in social groups over time, using several datasets of lifetime trajectories of individually-marked honey bees acquired using the recording setups in Berlin and Konstanz.

The method is based on the common factors that partially determine the roles individuals take and uses this dependency as an implicit bias to learn a temporally consistent representation of functional roles from social interaction networks. This work introduces a principled temporal matrix factorization model that jointly learns the average developmental path and structured variations of individuals in the social network over their entire lives.

The resulting embeddings are biologically relevant and consistent over time, allowing for the comparison of individuals' functional roles regardless of when or in which colony they lived.

This chapter was previously published as:
(Wild et al., 2022) — Benjamin Wild, David M. Dormagen, Michael L. Smith, and Tim Landgraf. Learning to embed lifetime social behavior from interaction dynamics. In *Volume 2 of the Proceedings of the joint 12th International Conference on Methods and Techniques in Behavioral Research and 6th Seminar on Behavioral Methods held online May 18-22 2022*. 12th International Conference on Methods and Techniques in Behavioral Research and 6th Seminar on Behavioral Methods, 2022. URL https://doi.org/10.6084/m9.figshare.20066849.v2

6.2 Author contributions

Conceptualization: BW, TL; Methodology: BW; Software: BW; Resources, supervision: TL; Project administration: TL; Data curation: BW, DMD; Writing: BW, DMD, MLS, TL; Visualization: BW.

6.3 Abstract

Interactions of individuals in complex social systems give rise to emergent behaviors at the group level. Identifying the functional role that individuals take in the group at a specific time facilitates understanding the dynamics of these emergent processes. An individual's behavior at a given time can be partially inferred by common factors, such as age, but internal and external factors also substantially influence behavior, making it difficult to disentangle common development from individuality. Here we show that such dependencies on common factors can be used as an implicit bias to learn a temporally consistent representation of a functional role from social interaction networks. Using a unique dataset containing lifetime trajectories of multiple generations of individually-marked honey bees in two colonies, we propose a new temporal matrix factorization model that jointly learns the average developmental path and structured variations of individuals in the social network over their entire lives. Our method yields inherently interpretable embeddings that are biologically relevant and consistent over time, allowing one to compare individuals' functional roles regardless of when or in which colony they lived. Our method provides a novel quantitative framework for understanding behavioral heterogeneity in complex social systems.

6.4 Introduction

Animals living in groups often coordinate their behavior, resulting in emergent properties at the group level. The dynamics of the inter-individual interactions produce, for example, the coherent motion patterns of flocking birds and shoaling fish, or the results of democratic elections in human societies. In many social systems, individuals differ consistently in how, when, and with whom they interact. The way an individual participates in social interactions and therefore contributes to the emergence of group-level properties can be understood as its functional role within the collective (Farine and Whitehead, 2015; Gordon, 2010; Krause et al., 2015; Pinter-Wollman et al., 2014).

Technological advances have made it possible to track all individuals and their interactions, ranging from social insects to primate groups (Mersch et al., 2013; Gernat et al., 2018; Boenisch et al., 2018a; Mathis et al., 2018; Graving et al., 2019; Pereira et al., 2019). These methods produce datasets that have unprecedented scale and complexity, but identifying and understanding the functional roles of the individuals within their groups has emerged as a new and challenging problem in itself. Social network analysis of interaction networks has proven to be a promising approach because interaction networks are comparatively straightforward to obtain from tracking data, and the networks represent each individual in the global context of the group (Krause et al., 2015; Pinter-Wollman et al., 2014; Brask et al., 2021; Wild et al., 2021b).

In most social systems, the way individuals interact changes over time, due to new experiences, environmental changes, or physiological conditions. Furthermore, groups themselves also tend to change, both in size and composition (Gordon and Mehdiabadi, 1999; Davidson and Gordon, 2017; Naug, 2008; Aplin et al., 2015; Sendova-Franks et al., 2010; Ament et al., 2010). Despite these changes over time, an objective measure of the functional role should identify individuals that serve a similar function (e.g. a guard versus a forager). Unfortunately, we are now facing a recursive definition of function: We are trying to derive the function of an individual from the network, but the network itself is also a function of the individuals' behavior (and other factors). Still, consider a group-living species in which only a subset of individuals engage in nursing duties. If we analyze the networks of different groups of the same species in different environmental conditions and group sizes, we still expect an objective measure of function to be shared among individuals engaged in nursing, regardless of these confounding factors. How can we extract such an objective measure from a constantly changing network of interactions without a fixed frame of reference?

In many social systems, individuals share common factors that partially determine the roles they take. For example, an individual's age can have a strong influence on behavior. In humans, factors such as socioeconomic status are comparatively easy to measure yet determine behavior and, therefore, interactions to a large extent. If individuals take on roles partially determined by a common factor, can we use this dependency to learn an objective measure of function? Here, we show that such common factors are a powerful inductive bias to learn semantically consistent functional descriptors of individuals over time, even in highly dynamic social systems.

In recent years, methods that automatically learn semantic embeddings from high-dimensional data have become popular. These methods map entities into a learned vector space. For example, in natural language models, a word can be represented as a vector, such that specific regions in the manifold of learned embeddings correspond to words with similar meaning. Similarly, recommender systems can learn meaningful embeddings of users and items, for example, movies, such that similar entities cluster in the manifold of learned embeddings (Frome et al., 2013; Asgari and Mofrad, 2015; Camacho-Collados and Pilehvar, 2018; Nelson et al., 2019).

Such embeddings are usually learned from the data without additional supervision. In recommender systems, a movie's genre is usually not given in a dataset of user ratings, yet the genre can be identified given the learned embeddings (Koren et al., 2009). This capability of learning embeddings from raw data and using them in downstream tasks is desirable in datasets of social interactions, where raw data is often abundant but labels are hard to acquire. Furthermore, embeddings are interpretable. For example, vector arithmetic of word embeddings can be used to understand how semantic concepts the natural language model has learned from the data relate to

each other (Mikolov et al., 2013). For entities that change over time, trajectories of embeddings can be analyzed, i.e., how one entity changes within the learned manifold of embeddings. Such analyses can, for example, reveal how environmental conditions such as resource availability affect behavioral changes within the group (Richardson et al., 2021; Smith et al., 2022a).

Most real-world networks have a hierarchical organization with overlapping communities, and thus soft community detection algorithms are often used to group and describe entities (Richardson et al., 2021; Palla et al., 2005; Ahn et al., 2010). Non-negative matrix factorization (NMF) is a principled and scalable method to learn embeddings from data that can be represented in matrix form, such as interaction networks. NMF has an inherent soft clustering property and is therefore well suited to derive embeddings from social interaction networks (Ding and He, 2005). If the embeddings allow us to predict relevant behavioral properties, they serve our understanding as *semantic* representations.

In symmetric non-negative matrix factorization (SymNMF), the dot products of any two individuals' embeddings (*factor vectors*) reconstruct their interaction affinity (Wang et al., 2011; Shi et al., 2015), see Figure S1 a and b). However, this algorithm has no straightforward extension in temporal settings where the interaction matrices change over time. The interaction matrices at different time points can be factorized individually, but there is no guarantee that the embeddings stay semantically consistent over time. The dot product is permutation invariant, therefore factorization can result in different embeddings depending on the optimization method being used, or noise in the data. Consider the hypothetical case of two groups of animals of the same species with two tasks, guards and nurses. Factorizing the interaction matrices of both groups will likely reveal two clusters, but there is no guarantee that the same cluster will be assigned to the same task for both groups. The same problem can occur in the case of only one group with new animals emerging and some dying over time without any changes in the distribution of tasks on the group level. In this case, the embeddings are not semantically consistent over time. The prediction of relevant behavioral properties will deteriorate, and individuals cannot be meaningfully compared against each other.

Several approaches to extend NMF to temporal settings have been proposed in a variety of problem settings. Previous work proposed factorization methods for time series analysis (Yu et al., 2016; Mackevicius et al., 2019), while others focus on the analysis of communities that are determined by their temporal activity patterns (Gauvin et al., 2014). Jiao and coworkers consider the case of communities from graphs over time and enforce temporal consistency with an additional loss term (Jiao et al., 2017). Several previous works represent network embeddings as a function of time (Yu et al., 2017a) and (Yu et al., 2017b), but the meaning of these embeddings can still shift over time. Temporal matrix factorization is similar to the tensor decomposition problem, which has many proposed solutions, see review by (Kolda and Bader, 2009). In particular, time-shifted tensor decomposition methods have been used in multi-neuronal spike train analysis, when recordings of multiple trials from a population of neurons are available (Mørup et al., 2008; Williams, 2020).

We approach this problem in the honey bee, a popular model system for studying individual and collective behavior (Elekonich and Roberts, 2005). Honey bees allocate tasks across thousands of individuals without central control, using an age-based system: young bees care for brood, middle-aged bees perform within-nest labor, and old bees forage outside (Seeley, 1982; Johnson, 2010). While age is a good predictor for the task of an average bee, individuals often deviate drastically from

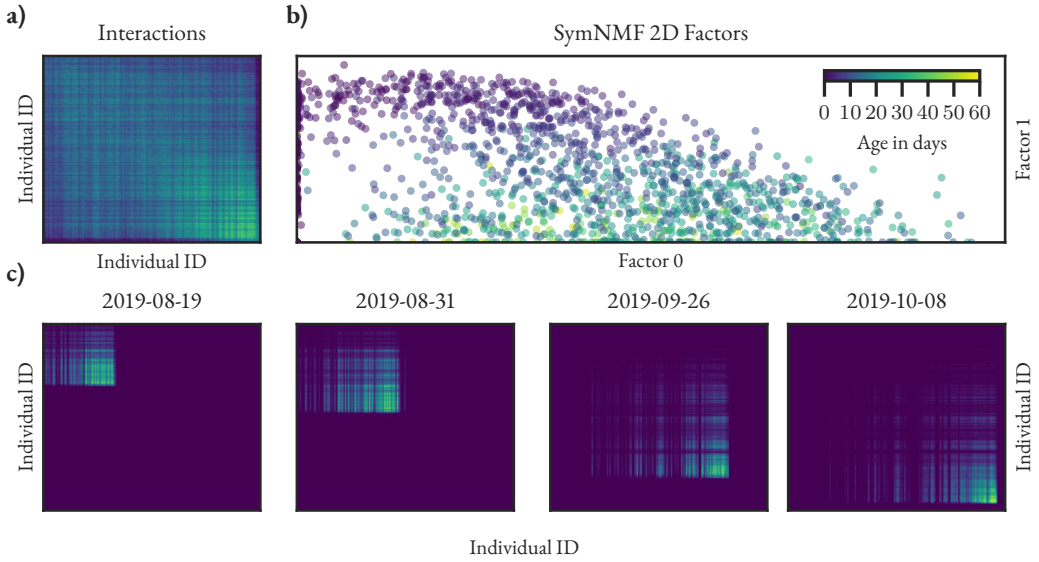

Figure S1: For a daily snapshot of a temporal social network, symmetric NMF is able to extract meaningful factor representations of the individuals. Colors represent the interaction frequencies of all individuals (**a**). The age-based division of labor in a honey bee colony is clearly reflected in the two factors - same-aged individuals are likely to interact with each other (**b**). For long observation windows spanning several weeks, the social network changes drastically as individuals are born, die, and switch tasks (**c**). Here, we investigate how a representation of temporal networks can be extracted, such that the factors representing individuals can be meaningfully compared over time, and even across datasets.

this common developmental trajectory due to internal and external factors. Honey bee colonies are also organized spatially: brood is reared in the center, honey and pollen are stored at the periphery, and foragers offload nectar near the exit. Therefore, an individual's role is partially reflected in its location, which provides the unique opportunity to evaluate whether learned embeddings based on the interaction data alone are meaningful.

A recent work proposes a method based on spectral decomposition to extract a semantic embedding (*Network age*) from honey bee interaction matrices and shows that these embeddings can be used to predict task allocation, survival, activity patterns, and future behavior (Wild et al., 2021b). The method proposed here is conceptually similar but solves several remaining challenges. Here, we introduce Temporal NMF (TNMF), which yields consistent semantic embeddings even for individuals from disjoint datasets, for example, data from different colonies, or for long-duration recordings that contain multiple lifetime generations.

TNMF jointly learns a) a functional form of the average trajectory of embeddings along the common factor, b) a set of possible functional deviations from the average trajectory, and c) for each individual, a soft-clustering assignment (*individuality embedding*) to these deviations. We show that these representations can be learned in an unsupervised fashion, using only interaction

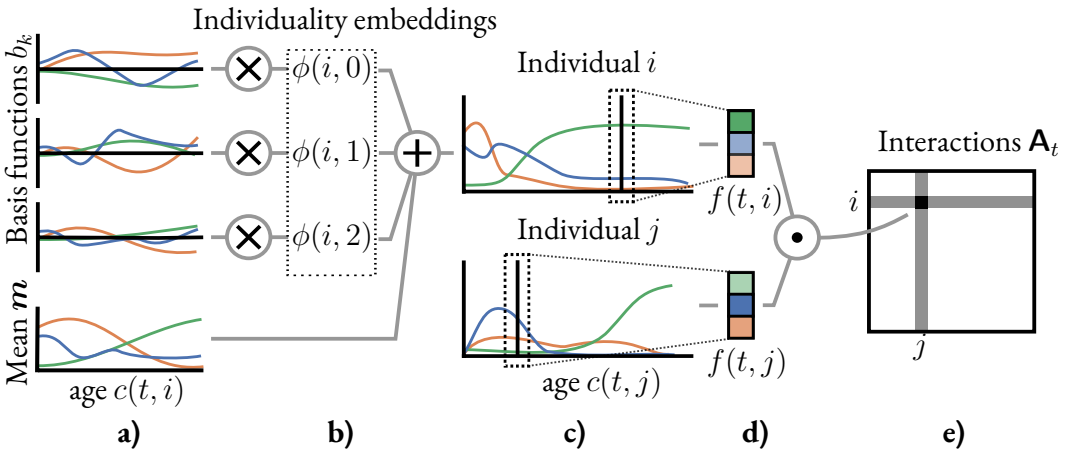

Figure S2: Overview of the method: We learn a parametric function describing the *mean life trajectory* $\boldsymbol{m}(c(t,i))$ and a set of basis functions of individual variation $\boldsymbol{b}(c(t,i))$, where $c(t,i)$ is the age of individual i at time t (**a**). For each individual, an embedding is learned consisting of one scalar per basis function that scales the contribution of the respective basis function - this vector of weights makes up the *individuality embedding* of an individual (**b**). The mean trajectory $\boldsymbol{m}(c(t,i))$ plus a weighted sum of the basis functions $\boldsymbol{b}(c(t,i))$ constitute the *lifetime trajectory* of each individual (**c**). At each time point, factors can be extracted from the individual lifetime trajectories (**d**) to reconstruct the interaction affinity between individuals (**e**). Note that the lifetime trajectories are functions of the individuals' ages, while interactions can occur at any time t.

matrices of the individuals over time. We analyze how well the model is able to disentangle common development from individuality using a synthetic dataset. Furthermore, we introduce a unique dataset containing lifetime trajectories of multiple generations of individually-marked honey bees in two colonies. We evaluate how well the embeddings learned by TNMF capture the semantic differences of individual honey bee development by evaluating their predictiveness for different tasks and behaviorally relevant metrics compared to several baseline models proposed in previous works.

6.5 Materials and Methods

6.5.1 Temporal NMF algorithm

SymNMF factorizes a matrix $\boldsymbol{A} \in \mathbb{R}_+^{N\times N}$ such that it can be approximated by the product $\boldsymbol{F}\boldsymbol{F}^T$, where $\boldsymbol{F} \in \mathbb{R}_+^{N\times M}$ and $M \ll N$:

$$\hat{\boldsymbol{F}} = \underset{\boldsymbol{F}\geq 0}{\operatorname{argmin}} \left\| \boldsymbol{A} - \boldsymbol{F}\boldsymbol{F}^T \right\|^2 \quad \boldsymbol{A}_{i,j} \approx \boldsymbol{f}(i)\cdot\boldsymbol{f}(j)^T \quad \boldsymbol{f}(i) = \boldsymbol{F}_{i,:} \quad \boldsymbol{f}(i) \in \mathbb{R}_+^M \tag{6.1}$$

When applied to social networks, $\boldsymbol{f}(i)$ can represent the role of an entity within the social network $\boldsymbol{A}$ (Wang et al., 2011; Shi et al., 2015) - however, in temporal settings, factorizing the matrices for different times separately will result in semantically inconsistent factors.

Here we present a novel temporal NMF algorithm (*TNMF*) which extends SymNMF to temporal settings in which $\mathbf{A} \in \mathbb{R}_+^{T\times N\times N}$ changes over time t. We assume that the entities $i \in \{0, 1, \ldots, N\}$ follow to some extent a common trajectory depending on an observable property (for example the age of an individual). We represent an entity at a specific point in time t using a factor vector $\boldsymbol{f}^+(t, i)$ such that

$$\hat{\mathbf{A}}_{t,i,j} = \boldsymbol{f}^+(t,i) \cdot \boldsymbol{f}^+(t,j)^T \qquad \hat{\mathbf{A}} \in \mathbb{R}_+^{T\times N\times N} \quad \boldsymbol{f}^+(t,i) \in \mathbb{R}_+^M \tag{6.2}$$

In contrast to SymNMF, we do not directly factorize $\mathbf{A}_t$ to find the optimal factors that reconstruct the matrices. Instead, we decompose the problem into learning an average trajectory of factors $\boldsymbol{m}(c(t,i))$ and structured variations from this trajectory $\boldsymbol{o}(t,i)$ that depend on the observable property $c(t,i)$:

$$\boldsymbol{f}(t,i) = \boldsymbol{m}(c(t,i)) + \boldsymbol{o}(t,i) \qquad \boldsymbol{f}^+(t,i) = \max(0, \boldsymbol{f}(t,i)) \tag{6.3}$$

$$c : \mathbb{N}^{T\times N} \to \mathbb{N} \quad \boldsymbol{m} : \mathbb{N} \to \mathbb{R}_+^M \quad \boldsymbol{o} : \mathbb{N}^{T\times N} \to \mathbb{R}^M$$

This decomposition is an inductive bias that allows the model to learn semantically consistent factors for entities, even if they do not share any data points (e.g., there is no overlap in their interaction partners), as long as the relationship between functional role and $c(t,i)$ is stable. Note that in the simplest case $c(t,i) = t$, *TNMF* can be seen as a tensor decomposition model, i.e. the trajectory of all entities is aligned with the temporal dimension t of $\mathbf{A}$. In our case, $c(t,i)$ maps to the age of individual i at time t.

While many parameterizations for the function $\boldsymbol{o}(t,i)$ are possible, we only consider one particular case in this work: We learn a set of *individuality basis functions* $\boldsymbol{b}(c(t,i))$ (shared among all entities) that define a coordinate system of possible individual variations and the *individuality embeddings* ϕ, which capture to what extent each basis function applies to an entity:

$$\boldsymbol{o}(t,i) = \sum_{k=0}^{K} \phi_{i,k} \cdot b_k(c(t,i)) \qquad \phi : \mathbb{R}^{N\times K} \quad b_k : \mathbb{N}^T \to \mathbb{R} \tag{6.4}$$

where K is the number of learned basis functions. This parameterization allows us to disentangle the forms of individual variability (*individuality basis functions*) and the distribution of this variability (*individuality embeddings*) in the data.

We implement the functions $\boldsymbol{m}(c(t,i))$ and $\boldsymbol{b}(c(t,i))$ with small fully connected neural networks with non-linearities and several hidden layers. The parameters θ of these functions and the entities' embeddings ϕ are learned jointly using minibatch stochastic gradient descent:

$$\hat{\theta}, \hat{\phi} = \underset{\theta,\phi}{\operatorname{argmin}} \left\| \boldsymbol{A} - \hat{\boldsymbol{A}} \right\|^2 \tag{6.5}$$

Note that non-negativity is not strictly necessary, but we only consider the non-negative case in this work for consistency with prior work (Wang et al., 2011; Shi et al., 2015). Furthermore, instead of one common property with discrete time steps, the factors could depend on multiple continuous properties, i.e. $c : \mathbb{R}^{T\times N} \to \mathbb{R}^{P}$, e.g. the day and time in a intraday analysis of social networks.

We find that the model's interpretability can be improved using additional regularization terms without significantly affecting its performance. We encourage sparsity in both the number of used factors and individuality basis functions by adding L_1 penalties of the mean absolute magnitude of the factors $\boldsymbol{f}(t, i)$ and basis functions $\boldsymbol{b}(c(t, i))$ to the objective. We encourage individuals' lifetimes to be represented with a sparse embedding using an L_1 penalty of the learned *individuality embeddings* ϕ.

We also introduce an optional adversarial loss term to encourage the model to learn embeddings that are semantically consistent over time, i.e. to only represent two entities that were present in the dataset at different times with different embeddings if this is strictly necessary to factorize the matrices $\mathbf{A}$. We jointly train a discriminative network $d(\phi_i)$ that tries to classify the time of the first occurrence of all entities based on their *individuality embeddings* ϕ. The negative cross-entropy loss of this model is added as a regularization term to equation 6.5 in a training regime similar to generative adversarial networks (Goodfellow et al., 2020). Note that a high cross-entropy loss of the discriminative network $d(\phi_i)$ implies that the distribution of *individuality embeddings* ϕ is consistent over time.

We implemented the model using PyTorch (Paszke et al., 2019) and trained it in minibatches of 256 individuals for 200 000 iterations with the Adam optimizer (Kingma and Ba, 2015). We calculate the reconstruction loss $\left\|\mathbf{A}_t - \hat{\mathbf{A}}_t\right\|^2$ only for valid entries, i.e., we mask out all matrix elements where one of the individuals is not alive at the given time t. The code of our reference implementation is publicly available: [Link redacted during peer review to preserve anonymity]

6.5.2 Data

Synthetic data

We created synthetic datasets using a generative model of interactions based on a common latent trajectory of factors and groups with structured variations from this trajectory. We compute the number of interactions between two individuals as the dot product of their latent factors and additive Gaussian noise. Using these datasets we can evaluate whether the model successfully converges and is able to correctly identify which individual belongs to which latent group, even in the presence on high amounts of observational noise. While we believe that such a latent structure exists in most complex social systems, it is not directly observable, and thus, for data from a real system, we can only evaluate the model on proxy measures (see section 6.5.3) that are observable.

We model a common lifetime trajectory of factors using a smoothed Gaussian random walk in $\mathbb{R}^{+}$ with $\sigma_{\text{walk}} = 1$ for the steps of the random walk and $\sigma_{\text{smoothing}} = 10$ for the Gaussian smoothing kernel. See Figure S3 a) for one example of a generated lifetime trajectory with three factors. We then randomly create latent groups by creating smoothed Gaussian random walks that define how these groups differ from the common lifetime trajectory. See Figure S3 b) for the

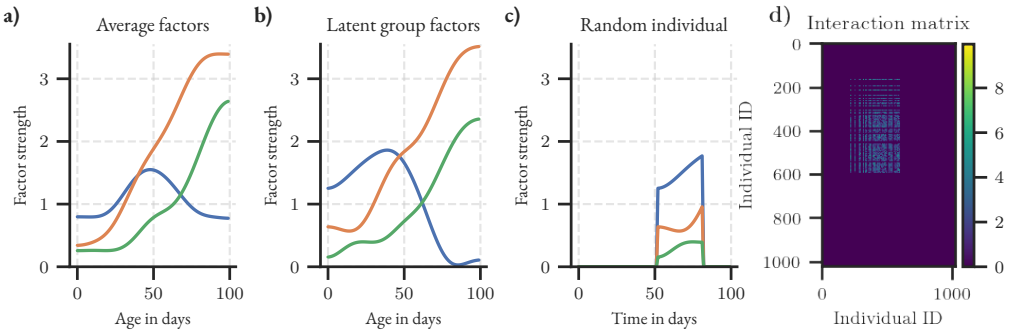

Figure S3: Example of one synthetic dataset. **a)** Common lifetime trajectory of all entities. **b)** The lifetime trajectory of one latent group. **c)** The factors of one individual in the dataset of the latent group visualized in **b**. **d)** Generated interaction matrix for one day.

lifetime trajectory of one latent group. For each group, we also define different expected mean lifetimes. We set the average lifetime of an entity to 30 days with a standard deviation of 10 days. We then randomly assign 1024 individuals to those latent groups and also assign random dates of emergence and disappearance of these individuals in the dataset. We then compute the individual factor trajectories for each individual, as can be seen in Figure S3 c). Finally, for 100 days of simulated data, we generate interaction matrices by computing the dot products of the factors of all individuals (Figure S3 d).

We then measure how well the *individuality embeddings* ϕ of a fitted model match the true latent groups from the generative model using the adjusted mutual information score (Nguyen et al., 2009). Furthermore, we measure the mean squared error between the ground truth factors and the best permutation of the factors $\boldsymbol{f}^{+}$. We evaluate the model on 128 different random synthetic datasets with increasing Gaussian noise levels in the interaction tensor.

Honey bee data

Honey bees are an ideal model system with a complex and highly dynamic social structure. The entire colony is observable most of the time. In recent years, technological advances have made it possible to automatically track individuals in entire colonies of honey bees over long periods of time (Wario et al., 2015; Gernat et al., 2018; Boenisch et al., 2018a). We analyze a dataset obtained by tracking thousands of individually marked honey bees at high temporal and spatial resolution, covering entire lifespans and multiple generations.

Two colonies of honey bees were continuously recorded over a total of 155 days. Each individual was manually tagged at emergence, so the date of birth is known for each bee. Timestamps, positions, and unique identifiers of all (N=9286) individuals from these colonies were obtained using the BeesBook tracking system (Wario et al., 2015; Boenisch et al., 2018a; Wild et al., 2021b). See Table 6.1 for dates and number of individuals. Temporal affinity matrices were derived from this data as follows: For each day, counts of proximity contact events were extracted. Two individuals were defined to be in proximity if their markers' positions had an euclidean distance of less than 2 cm for at least 0.9 seconds. The daily affinity between two individuals i and j based on their

counts of proximity events $p_{t,i,j}$ at day t was then computed as: $\mathbf{A}_{t,i,j} = \log(1 + p_{t,i,j})$, $\mathbf{A} \in \mathbb{R}^{N_t \times N_i \times N_i}$, where N_t is the number of days and N_i the number of individuals in the dataset.

The datasets also contains labels that can be used in proxy tasks (see section 6.5.3) to quantify if the learned embeddings and factors are semantically meaningful and temporally consistent.

The datasets are open access and available under the *Creative Commons Attribution 4.0 International* license: [Link redacted during peer review to preserve anonymity].

Dataset	Dates	Days	Individuals	Interaction pairs
BN16	2016-07-23 to 2016-09-17	56	2443	43 174 748
BN19	2019-07-25 to 2019-11-01	99	6843	167 366 381

Table 6.1: The honey bee datasets contain the number of proximity-inferred interactions extracted from tracking data of all individuals in two long-term recordings spanning a total of 155 days and 9286 individuals.

In both datasets, we define $c(t, i)$ as the age in days of an individual i at time t.

6.5.3 Evaluation

Reconstruction: We measure how well the original interaction matrices $\mathbf{A}$ can be reconstructed from the factors. We do not require the model to reconstruct the interaction matrices as well as possible because we only use the reconstruction as a proxy objective to learn a meaningful representation. Still, a high reconstruction loss could indicate problems with the model, such as excessive regularization.

Consistency: We measure to what extent the *individuality embeddings* ϕ change over time. For each model, we train a multinomial logistic regression model to predict the source cohort (date of birth) and calculate the area under the ROC curve ($\text{AUC}_{\text{cohort}}$) using a stratified 100-fold cross-validation with scikit-learn (Pedregosa et al., 2011). The baseline models do not learn an individuality embedding; therefore we compute how well the model can predict the cohort using the mean factor representation of the individuals over their lives. We define consistency as $1 - \text{AUC}_{\text{cohort}}$ of this linear model. Note that a very low temporal consistency would indicate that the development of individual bees changes strongly between cohorts and colonies, which we know not to be true.

Mortality and Rhythmicity: We evaluate how well a linear regression model can predict the mortality (number of days until death) and circadian rhythmicity of the movement (Wild et al., 2021b) (R^2 score of a sine with a period of 24 h fitted to the velocity over a three-day window). These metrics are strongly correlated with an individual's behavior (e.g. foragers exhibit strong circadian rhythms because they can only forage during the daytime; foragers also have a high mortality). We follow the procedure given in (Wild et al., 2021b) and report the 100-fold cross-validated R^2 scores for these regression tasks.

Time spent on different nest substrates: For a subset of the data, from 2016-08-01 to 2016-08-25, nest substrate usage information is also available. This data contains the proportion of time each individual spends in the brood area, honey storage, and on the dance floor. This data was

previously published and analyzed (Wild et al., 2021a,b). The task of an honey bee worker is strongly associated with her spatial distribution in the hive. We therefore expect a good representation of the individuals' functional role to correlate with this distribution.

For this data, we expect the factors $\boldsymbol{f}^+$ and *individuality embeddings* ϕ to be semantically meaningful and temporally consistent if they reflect an individual's behavioral metrics (mortality and rhythmicity) and if they do not change strongly over time (measured in the consistency metric).

6.5.4 Baseline models

Biological Age: Task allocation in honey bee is partially determined by temporal polyethism. Certain tasks are usually carried out by individuals of about the same age, e.g. young bees are usually occupied with nursing tasks. We therefore use the age of an individual as a baseline descriptor.

Symmetric NMF: We compute the factors that optimally reconstruct the original interaction matrices using the standard symmetric NMF algorithm (Shi et al., 2015; Kuang et al., 2015), for each day separately, using the same number of factors as in the TNMF model.

Optimal permutation SymNMF: We consider a simple extension of the standard SymNMF algorithm that aligns the factors to be more consistent over time. For each pair of subsequent days, we consider all combinatorial reorderings of the factors computed for the second day. For each reordering, we compute the mean L_2 distance of all individuals that were alive on both days. We then select the reordering that minimizes those pairwise L_2 distances and greedily continue with the next pair of days until all factors are aligned. Furthermore, we align the factors across colonies (where individuals cannot overlap) as follows: we run this algorithm for both datasets separately and align the resulting factors by first computing the mean embedding for all individuals grouped by their ages. As before, we now select from all combinatorial possibilities the reordering that minimizes the L_2 distance between the embeddings obtained from both datasets.

Tensor decomposition: We also compare against a constrained non-negative tensor decomposition model with symmetric factors $\boldsymbol{F} \in \mathbb{R}_+^{N\times M}$ and temporal dynamics constrained to the diagonals, i.e. $\mathbf{D} \in \mathbb{R}_+^{T\times M\times M}$ and $\mathbf{D}_t = \mathrm{diag}(\boldsymbol{d}_t)$, $\boldsymbol{d}_t \in \mathbb{R}_+^M$.

$$\hat{\mathbf{A}}_t = \boldsymbol{F}\mathbf{D}_t\boldsymbol{F}^T \tag{6.6}$$

$$\hat{\boldsymbol{F}}, \hat{\mathbf{D}} = \underset{\boldsymbol{F},\mathbf{D}}{\operatorname{argmin}}\, T^{-1} \sum_{t=0}^{T} \left\| \boldsymbol{A}_t - \hat{\boldsymbol{A}}_t \right\|^2 \tag{6.7}$$

Temporal NMF models: We evaluate variants of the temporal symmetric matrix factorization algorithms proposed by (Jiao et al., 2017) and (Yu et al., 2017a).

For the tensor decomposition and temporal NMF baselines, we follow the procedure given above for the *Optimal permutation SymNMF* to find the optimal reordering to align the factors obtained by applying models to the two datasets separately.

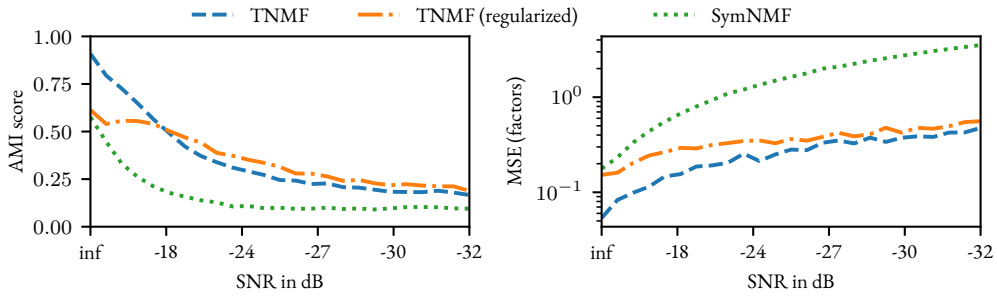

Figure S4: AMI score and mean squared error between true factors and the best permutation of learned factors for increasing noise levels. The median values over 128 trial runs are shown.

6.6 Results

6.6.1 Synthetic data

We find that for low levels of noise, our model can identify the truth group assignments with high accuracy, and are still significantly better than random assignments even at very high levels of noise (see figure S4). Note that for this experiment, we evaluated a model with the same hyperparameters as used in all plots in the results section (see Table 6.2) and a variant without explicit regularization except the L_1 penalty of the learned *individuality embeddings* ϕ ($\lambda_{\text{embeddings}}$, because this regularization is required to meaningfully extract clusters), which was set to 0.1.

6.6.2 Honey bees

Mean lifetime model: The model learns a sparse representation of the developmental trajectory of a honey bee in the space of social interactions. Only two factors are effectively used (they exceed the threshold value of 0.01). These factors show a clear trend over the life of a bee, indicating that the model captures the temporal aspects of the honey bee division of labor (See Figure S5).

Interpretability of factors: To understand the relationship between the factors and division of labor, we calculate how the factors map to the fraction of time an individual spent on the brood area, honey storage, or dance floor (where foragers aggregate). Time spent on these different substrates is a strong indicator of an individual's task. The factor f_1, which peaks at young age (Figure S5), correlates with the proportion of time spent in the brood area, while a high f_0 indicates increased time spent on the dance floor. Therefore, the model learned to map biologically relevant processes.

Individuality basis functions and individuality embeddings: Due to the regularization of the embeddings, the model learns a sparse set of *individuality basis functions*. As encouraged by the model, most individuals can predominantly be described by a single basis function. That means that while each honey bee can collect a unique set of experiences, most can be described with a few common *individuality embeddings* which are consistent across cohorts and colonies. In the

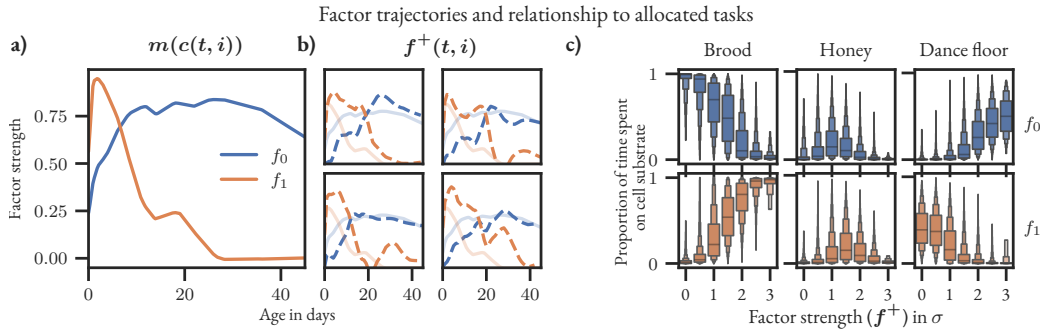

Figure S5: **a)** Mean lifetime trajectories according to $\boldsymbol{m}(c(t,i))$. The model learns a sparse representation of the functional position of the individuals in the social network. f_0 (blue) mostly corresponds to middle-aged and older bees, and f_1 (orange) predominantly describes young bees. Only factors with a mean magnitude of at least 0.01 are shown. **b)** Even though the model uses only these two factors, it is still expressive enough to capture individual variability, as can be seen in randomly sampled individuals' lifetime trajectories. **c)** The individual factors $\boldsymbol{f}^+$ and the proportion of time the individuals spent on different nest substrates. The strong correlation indicates that the learned factors are a good representation of the individuals' roles in the colonies. Note that the factors have been divided by their standard deviation here for ease of comparability.

context of honey bee division of labor, the basis functions are interpretable because the factors correspond to different task groups. For example, $b_{12}(c(t,i))$ (accounting for $\approx 10.7\%$ of the individuals) describes workers that occupy nursing tasks much longer than most bees.

Evaluation: We verify that the learned representations of the individuals are meaningful (i.e., they relate to other properties of the individuals, not just their interaction matrices) and semantically consistent over time and across datasets using the metrics described in the section *Evaluation*. We compare variants of our model with different adversarial loss scaling factors and factor L_1 regularizations, the baseline models, and the individuals' ages. We expect a good model to be temporally consistent and semantically meaningful. All variants of our model outperform the baselines in terms of the semantic metrics *Mortality* and *Rhythmicity*, except for the (Yu et al., 2017a) model, which performs comparably well in the *Mortality* metric. The adversarial loss term further increases the *Consistency* metric without negatively affecting the other metrics. A very strong adversarial regularization (see row with $\lambda_{\text{adv}} = 1$ in Table 6.2) prevents the model from learning a good representation of the data. See Table 6.2 for an overview of the results.

Scalability: The functions $\boldsymbol{m}(c(t,i))$ and $\boldsymbol{b}(c(t,i))$ are learned neural networks with non-linearities. The objective is non-convex and we learn the model parameters using stochastic gradient descent. Optimization is therefore slower than the standard NMF algorithms that can be fitted using algorithms such as Alternating Least Squares (Kim et al., 2014). We found that the model converges faster if the reconstruction loss of the age based model $\boldsymbol{m}(c(t,i))$ is additionally minimized with the main objective in equation 6.5. Due to the minibatch training regime, our method should scale well in larger datasets. Small neural networks were sufficient to learn the

Model					
Method	Variant	$\left\|\boldsymbol{A}-\hat{\boldsymbol{A}}\right\|^2 \downarrow$	Consistency ↑	Mortality ↑	Rhythmicity ↑
Age	-	-	-	0.02	0.20
SymNMF	Vanilla	0.9	0.18	0.01	0.02
SymNMF	Optimal permutation	0.9	0.12	0.09	0.35
Tensor decomposition	-	1.36	0.03	0.06	0.09
DNMF (Jiao et al., 2017)	$\gamma = 0.1$	0.9	0.19	0.02	0.05
DNMF (Jiao et al., 2017)	$\gamma = 1$	1.15	0.15	0.01	0.04
s-TMF (Yu et al., 2017a)	$\beta = 0.01, d = 5$	1.59	0.03	0.17	0.06
TNMF	No regularization	1.21	0.17	0.30	0.48
TNMF	$\lambda_{\text{adv}} = 0, \lambda_{\text{f}} = 0.01$	1.26	0.18	0.10	0.40
TNMF	$\lambda_{\text{adv}} = 0.1, \lambda_{\text{f}} = 0.01$	1.28	0.35	0.20	0.42
TNMF	$\lambda_{\text{adv}} = 1, \lambda_{\text{f}} = 0.01$	1.88	0.5	0.03	0.25
TNMF	$\lambda_{\text{adv}} = 0, \lambda_{\text{f}} = 0.1$	1.31	0.19	0.09	0.38
TNMF	$\lambda_{\text{adv}} = 0.1, \lambda_{\text{f}} = 0.1$	1.33	0.37	0.10	0.42

Table 6.2: The evaluation metrics for TNMF and the baseline models described in section 6.5.3. Note that the SymNMF model reconstruction loss can be seen as a lower bound for the matrix factorization models considered here, and imposing a temporal structure or regularization causes all models to explain less variance in the data. However, for all models except TNMF this does not result in a significant increase of the other metrics. The underlined model is used in all plots in the results section.

functions $\boldsymbol{m}(c(t,i))$ and $\boldsymbol{b}(c(t,i))$ in our experiments. Most of the runtime during training is spent on the matrix multiplication $f^+(t,i) \cdot f^+(t,j)^T$ and the corresponding backwards pass.

Tradeoff between temporal consistency and semantic meaningfulness: We performed a grid search over the hyperparameters λ_{f}, λ_{adv}, λ_{basis}, and $\lambda_{\text{embeddings}}$ (see Table 6.3) to evaluate whether models can only be either semantically meaningful or temporally consistent. For this analysis, we define *Semantic meaningfulness* as the sum of the *Rhythmicity* and *Mortality* metrics introduced in section 6.5.3. We find that models that are very temporally consistent fail to learn semantically meaningful information. Interestingly, the models with the best tradeoff between the two metrics are almost as semantically meaningful as those models with low temporal consistency and the highest semantic meaningfulness. This analysis suggests that regularization encourages the model to only represent different individuals differently if this is strictly necessary to factorize the data.

6.7 Discussion

Temporal NMF factorizes temporal matrices with overlapping and even disjoint communities by learning an embedding of individuals as a function of a common factor, such as age, and a learned representation of the individuals' individuality. This explicit dependency on a common factor that partially determines the function of an individual constitutes an inductive bias. We show that the model learns semantically consistent representations of individuals, even in challenging cases, such as the datasets analyzed in this work.

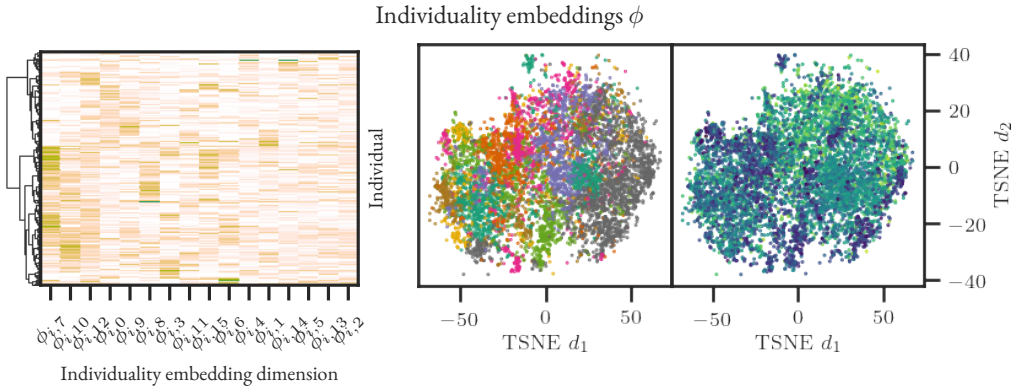

Figure S6: **Left:** Hierarchical clustering of individuality embeddings: Most individuals strongly correspond to a single individuality basis function, making it easy to cluster their lifetime social behavior (i.e. each individual has a high value in a single dimension for their individuality embedding). Because each cluster is strongly associated with a specific individuality basis function, and because each basis function is interpretable (Figure S5), these blueprints of lifetime development can also be intuitively understood and compared. **Right:** TSNE plots of the individuality embeddings colored by cluster (left) and the maximum circadian rhythmicity of an individual during her lifetime (right), indicating that the embeddings are semantically meaningful.

The individual components of the model are straightforward to visualize and interpret. The learned individuality embeddings ϕ can be understood as soft-cluster assignments relating to the whole lifetime of an individual, while the factor vectors $\boldsymbol{f}^{+}(t, i)$ can be interpreted as cluster assignments of the individuals at a specific point in time, i.e. two individuals with similar factor vectors are likely to interact if they exist in the same group at the same time. Furthermore, the model encourages sparsity, making the results easier to interpret because the model only uses as many factors and clusters as necessary.

We identified a crucial trade-off that comes with temporal consistency: For a specific point in time, the ability to predict behaviorally relevant attributes will likely be worse for a model that learns temporally consistent representations compared to a non-consistent model with the same capacity. Conversely, in more challenging cases, e.g. when taking long periods of time or data from disjoint communities into consideration, temporally consistency is indispensable for a good representation. Furthermore, we found that models can be temporally consistent, semantically meaningful, or both; selecting the correct model requires an inductive bias, but regularization of the model also influences the results.

On the honey bee dataset, TNMF obtains biologically meaningful lifetime trajectories with promising prospects for experimental application. TNMF may help advance our understanding of the colony function and the interplay between environmental factors and individual and collective responses. The method presented here offers a way to investigate the impact of stress factors, such as pesticides, parasitic mites, and agricultural monoculture, on the social structure of colonies.

6.8 Ethics Statement

German law does not require approval of an ethics committee for studies involving insects.

6.9 Supporting information

6.9.1 Model details and hyperparameters

Regularization terms

$$R_{\text{embeddings}} = \lambda_{\text{embeddings}} N_i^{-1} \sum_{i=0}^{N_i} \sum_{k=0}^{K} |\phi_{i,k}| \tag{6.8}$$

$$R_{\text{f}} = \lambda_{\text{f}} N_i^{-1} \sum_{i=0}^{N_i} \sum_{t=0}^{N_t} f^{+}(t,i) \tag{6.9}$$

$$R_{\text{basis}} = \lambda_{\text{basis}} N_a^{-1} \sum_{a=0}^{N_a} \sum_{k=0}^{K} |b_k(a)| \quad N_a = 60 \tag{6.10}$$

where N_a can be any number higher than the oldest individual in the dataset at any time.

$$R_{\text{adv}} = \lambda_{\text{adv}} N_i^{-1} \sum_{i=0}^{N_i} \log\left(\frac{\exp(\boldsymbol{d}(\phi_i)_{c[i]})}{\sum_d^{N_d} \exp(\boldsymbol{d}(\phi_i)_d)} \right) \tag{6.11}$$

where $\boldsymbol{d}(\phi_i)$ is the probability distribution returned by the discriminative network, $c[i]$ the day the entitity i emerged in the dataset, and N_d the number of days in the dataset.

See appendix 6.9.2 for an ablation study of the effect of these regularization term on the results.

Network architecture

We use the following neural network architecture for the functions $\boldsymbol{m}(c(t,i))$, $\boldsymbol{b}(c(t,i))$, and $\boldsymbol{d}(\phi_i)$:

$$\text{Linear}(N_{in}, N_h) \rightarrow \text{LReLU} \rightarrow \underbrace{\text{Linear}(N_h, N_h) \rightarrow \text{LReLU}}_{N_l\text{-times}} \rightarrow \text{Linear}(N_h, N_{out})$$

where *Linear* is an affine transformation $f(x) = Ax + b$ and $\alpha = 0.3$ for the Leaky ReLU activation function. For $\boldsymbol{m}(c(t,i))$ and $\boldsymbol{b}(c(t,i))$: $N_{in} = 1$ (the individuals' ages). For $\boldsymbol{m}(c(t,i))$: $N_{out} = M$ and for $\boldsymbol{b}(c(t,i))$: $N_{out} = MK$. For $d(\phi_i)$: $N_{in} = K$ and $N_{out} = N_{\text{labels}}$.

Hyperparameters

The scaling factors for the regularization losses (see Table 6.3) were manually selected by increasing each factor until it prevented the model from converging (i.e. the reconstruction loss of the full

Table 6.3: Hyperparameters used in the evaluated models (if not stated otherwise)

Parameter	Value	Description
N_l	3	Number of hidden layers
N_h	64	Hidden layer size
M	8	Number of factors
K	16	Number of individuality basis function
N_{labels}	100	Number of cohorts
N_{batch}	128	Minibatch size
N_{steps}	100 000	Number of training iterations
λ_{f}	0.1	Factor L_1 regularization
λ_{adv}	0.1	Factor L_1 regularization
λ_{basis}	0.01	Basis function L_1 regularization
$\lambda_{\text{embeddings}}$	0.1	Embedding L_1 regularization

model $\boldsymbol{f}^+(t, i)$ did not improve on the age model $\boldsymbol{m}$). This initial set of hyperparameters was then manually refined such that each regularization loss was still effective (e.g. the factor regularization loss L_f reduced the total number of factors effectively used by the model). Overfitting was not a concern because the model is fitted unsupervised and the goal of the hyperparameter selection was to find a set of parameters that is sparse and interpretable, and not to increase the predictive capabilities of the learned factors.

Model fitting

The model was fitted using a single GPU (GeForce RTX 2080 Ti). A training run consisting of 200 000 minibatches finished in about six hours. Due to overhead in data loading and preprocessing, up to three training runs could be executed in parallel without negatively affecting the runtime.

Algorithm 1: Training loop

for $b = 0$ **to** N_{steps} **do**
  Draw minibatch of N_{batch} random individuals
  Compute $\hat{\mathbf{A}}_{t,i,j}$ $\forall t$ for individuals i, j in minibatch
  Compute model training loss: equation 6.5 + regularization
  Update parameters for $\boldsymbol{m}(c(t, i))$, $\boldsymbol{b}(c(t, i))$, and $\boldsymbol{d}(\phi_i)$
  Compute $\hat{l}_i = \boldsymbol{d}(\phi_i)$ for individuals i in minibatch
  Compute discriminator training loss
  Update parameters for $\boldsymbol{d}(\phi_i)$
end for

6.9.2 Ablation study

As highlighted in section 6.6, the regularization terms R_{f} and R_{adv} improve the semantic meaningfulness and temporal consistency of the model. Here we present an ablation study that shows the effects of the regularization terms on the sparseness and interpretability of the learned factors and embeddings, and how the adversarial term influences the distribution of the learned embeddings. While the regularization terms increase the methodological complexity of the model, we argue that they improve the interpretability of the results.

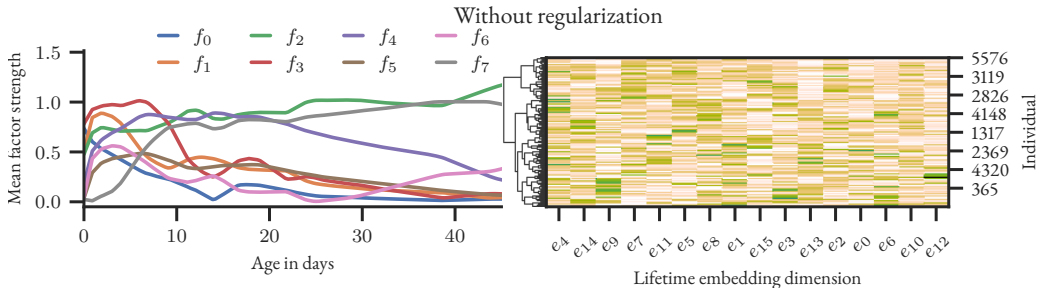

Figure S7: A model trained without the regularization terms $R_{\text{embeddings}}$, R_{f}, R_{basis}, and R_{adv}. The model uses all eight factors, and many of them are strongly correlated (left). Most individuals personality offsets are a function of multiple embeddings (right).

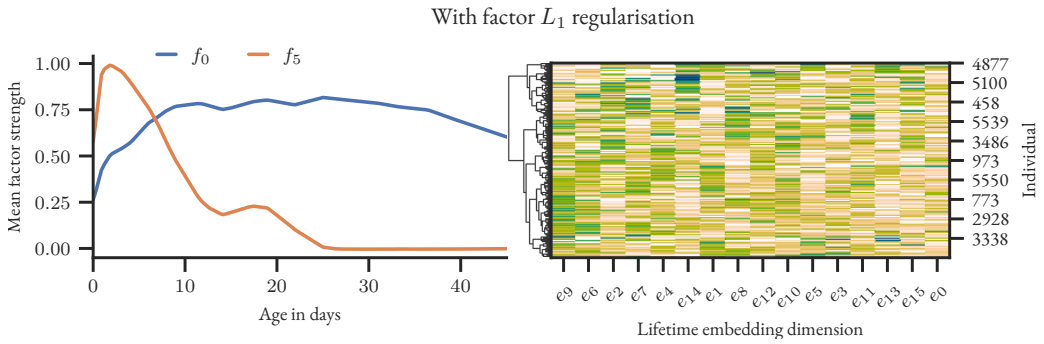

Figure S8: A model trained with only the regularization term R_{f}. The factor trajectories are now sparse and uncorrelated (left), most individuals personality offsets are a still a function of multiple embeddings (right).

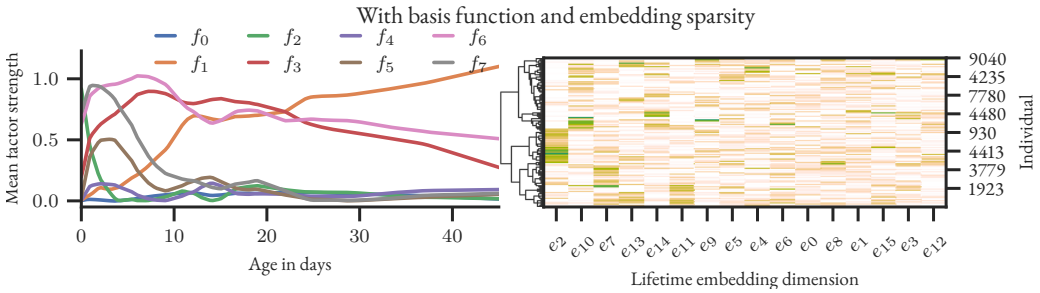

Figure S9: The regularization terms $R_{\text{embeddings}}$ and R_{basis} introduce sparseness in the embeddings (right), and also slightly decorrelate the factor trajectories (left).

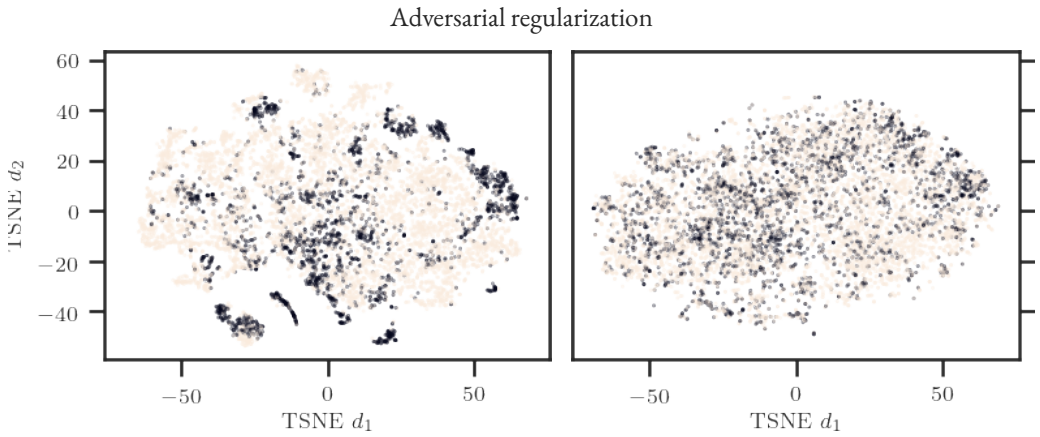

Figure S10: Scatter plots of the individuality embeddings ϕ (reduced to two dimensions using TSNE) for two models trained without (left) and with adversarial (right) regularization. The color encodes the dataset of the individuals (Dark = BN16, Bright = BN19). The model with adversarial regularization learns to embed the individuals from two different colonies that never interacted with each other in a joint individuality embedding space.

6.9.3 Baseline models

We implemented SymNMF, and models proposed in (Jiao et al., 2017) and (Yu et al., 2017a) in PyTorch and compare them to our method. Following the notation given in (Yu et al., 2017a), we list the degree of the fitted polynomial as d, and the regularization parameter as β. For the model proposed by (Jiao et al., 2017), we list the regularization as γ. For SymNMF, the *Aligned* variant refers to the *Aligned symmetric NMF* described in section 6.5.4.

We only evaluate the non-negative and symmetric variants of the models for consistency with our method. For all baselines, we only list the hyperparameters with the best results that we were able to obtain.

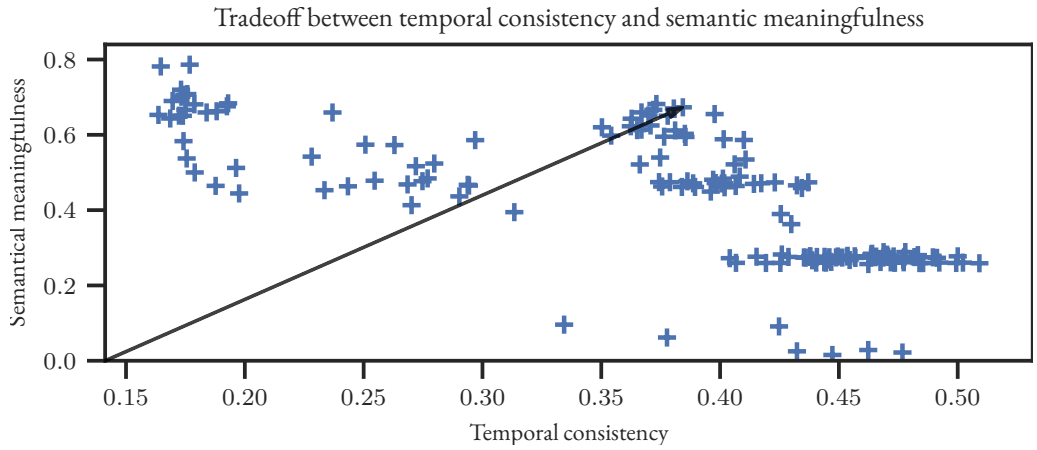

Figure S11: A grid search over the hyperparameters reveals that models can be temporally consistent, semantically meaningful, or both. Semantic meaningfulness is the sum of the *Rhythmicity* and *Mortality* metrics introduced in section 6.5.3. The arrow points to the model with the best tradeoff.

Temporal reordering of factors for baseline models

Algorithm 2: Temporal reordering of factors for baseline models

for $d = 0$ **to** N_{days} **do**
  $F_{\text{result}} < -$
  if $d == 0$ **then**
    $F_{\text{previous\_reordered}} \leftarrow$ Precomputed SymNMF factors of day 0
  else
    $F_{\text{current}} \leftarrow$ Precomputed SymNMF factors of day d
    min_loss $\leftarrow 0$
    $F_{\text{best}} \leftarrow F_{\text{current}}$
    for all permutations p of orderings of factors $[0..M]$ **do**
      $F_{\text{current\_reordered}} \leftarrow F_{\text{current}}$ reordered by permutation p
      if $(F_{\text{current\_reordered}} - F_{\text{previous\_reordered}})^2 <$ min_loss **then**
        min_loss $\leftarrow (F_{\text{current\_reordered}} - F_{\text{previous\_reordered}})^2$
        $F_{\text{best}} \leftarrow F_{\text{current\_reordered}}$
      end if
    end for
    $F_{\text{previous\_reordered}} \leftarrow F_{\text{best}}$
  end if
  $F_{\text{result}}.insert(F_{\text{previous\_reordered}})$
end for
return F_{result}

6.9.4 Learned basis functions and individual trajectories

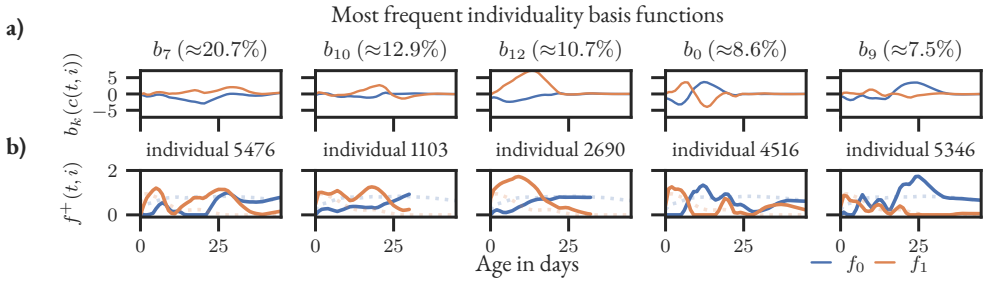

Figure S12: **a)** Magnitude of factor offsets for the five most common individuality basis functions over age $b_k(c(t,i))$. The percentage of individuals that most strongly correspond to the individual basis functions is shown in the column titles. More than 60% of the individuals strongly correspond to one of the five basis functions shown here. **b)** Because the basis functions describe *individuality offsets* from the mean lifetime trajectory, it may be easier to interpret them by visualizing individual examples. For each of the basis functions (top row), we show a lifetime trajectory of an individual that corresponds to that basis function (bottom row). Note that individuals can die or disappear at any time (solid lines). The mean lifetime trajectories are shown as dotted lines in the background.

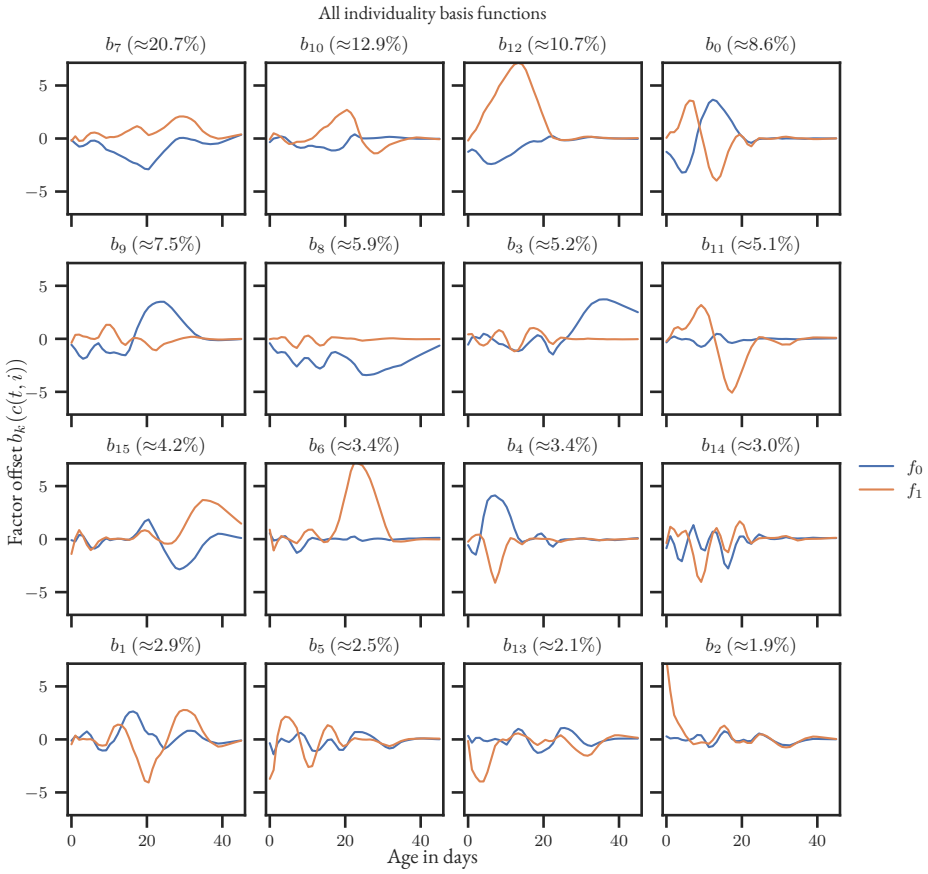

Figure S13: Magnitude of factor offsets for all learned individuality basis functions over age $b_k(c(t,i))$.

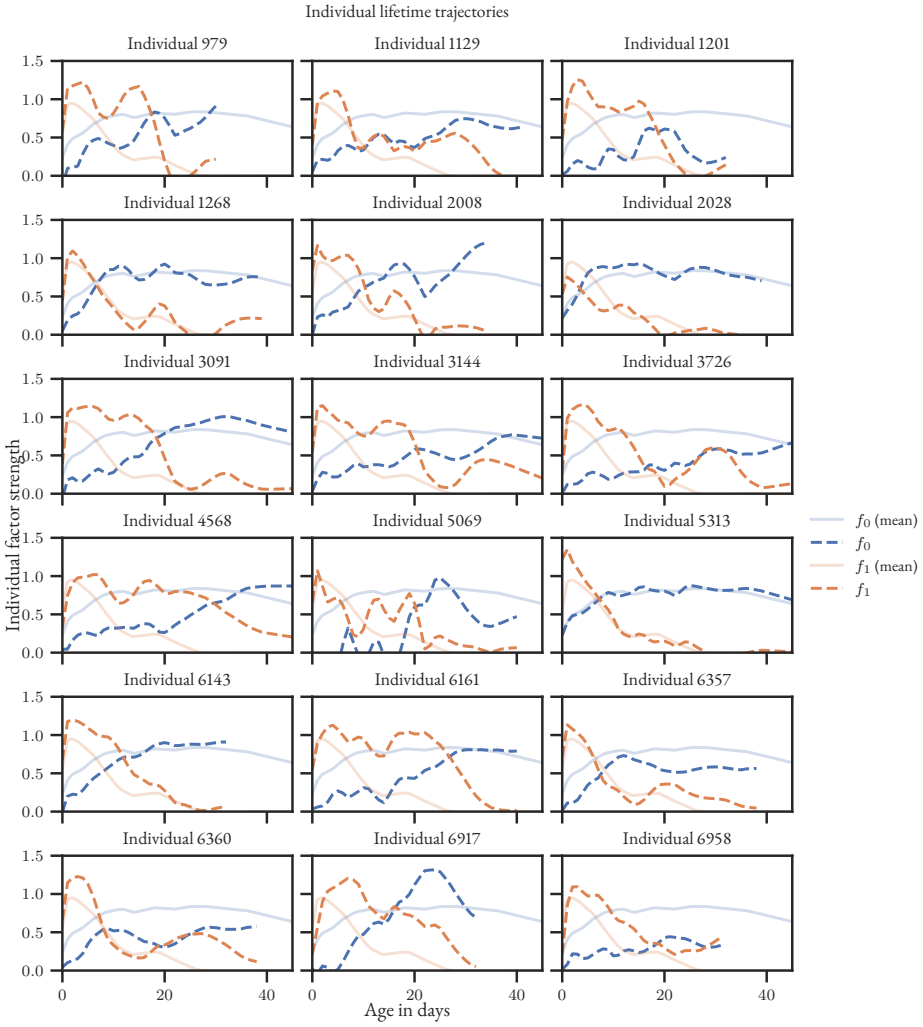

Figure S14: Individual lifetime trajectories: 18 individuals that lived for at least 35 days were randomly sampled and their factors $f(t, i)$ were computed over all t, constituting their lifetime trajectories.

7 Discussion

This thesis entitled *Individuality in the hive* investigates the emergence of collective behavior in social animals, focusing on honey bee colonies as a model system. It poses two primary research questions: how does collective behavior arise in animal societies without central organization, and how do individual behaviors contribute to collective outcomes? To address these questions, this study introduces methods to track individual honey bees and analyze their behaviors and interactions within the colony. The following discussion reviews the key findings emerging from this work.

The first paper, entitled *Tracking All Members of a Honey Bee Colony Over Their Lifetime Using Learned Models of Correspondence* (Chapter 3), introduces a novel method for automatically tracking individual honey bees within a colony, which enables the observation of large numbers of individuals and their interactions over extended periods. The proposed system uses machine learning techniques to identify and distinguish between individuals and employs an elaborate tracking algorithm, resulting in a unique and extensive trajectory dataset for investigating honey bee collective behavior. The paper demonstrates the potential of computer vision and machine learning techniques for addressing the challenges of tracking bees within a colony. It also establishes foundational methods for subsequent works that utilize this dataset to measure space usage, movement patterns, and social interactions of every individual in a honey bee colony.

The second paper, entitled *Social networks predict the life and death of honey bees* (Chapter 4), explores the relationship between individual behaviors and the emergence of collective behavior in honey bee colonies. This paper proposes a method to evaluate the social network of honey bees and its impact on their development and behavior. This method introduces a concise descriptor termed *network age*, which effectively predicts task allocation, survival, activity patterns, and future behavior. A detailed analysis of the developmental trajectories of multiple cohorts of individuals in a natural setting using *network age* identifies distinct developmental pathways and critical life changes. Along with this work, the first-ever large-scale honey bee tracking dataset was released, opening up a broad range of future studies (Wild et al., 2021a). The third paper, entitled *Individual variation in behavior and task allocation in honeybee colonies* (Chapter 5), focuses on the behavioral variation of individuals with the collective across their days and lives. The study uses behavioral metrics to compare the individual's behavior at different timescales and describes how bees differ in space use, detection, and movement at the scale of a single day and how some bees consistently exhibit different movement patterns and transition to dance floor and outside activities at different ages throughout their entire lifetimes. This quantitative analysis of individual behavioral variation within a colony offers insights into the relationship between individual behaviors of honey bees and how they may contribute to colony function. The data used in this study was likewise made publicly accessible as an open-access data set (Smith et al., 2022c).

The fourth paper, entitled *Learning to Embed Lifetime Social Behavior from Interaction Dynamics* (Chapter 6), addresses a significant limitation of previous analyses in this thesis. While the previous studies created and analyzed honey bee tracking datasets of unprecedented scale and detail, they were each limited to a single recording setup and used methods unsuitable for analyzing data acquired from multiple colonies or over extended periods. To address this issue, the paper introduces a new method that identifies the functional roles of individuals in social groups over time using several datasets of lifetime trajectories of individually marked honey bees acquired from different recording setups. The method uses a temporal matrix factorization model that jointly learns the average developmental path and structured variations of individuals in the social network over their entire lives. The resulting embeddings allow for the comparison of individuals' functional roles regardless of when or in which colony they lived and provide a novel quantitative framework for understanding behavioral heterogeneity in complex social systems across time and space. This work represents a significant methodical advance in the study of collective behavior in honey bee colonies and provides a principled method to investigate the underlying mechanisms and principles that govern the emergence of such behavior.

In summary, the thesis has investigated the emergence of collective behavior in complex animal societies without central organization and the relationship between individual behaviors and the global behavior of the collective. By using honey bees as a model organism, the thesis has demonstrated the potential of computer vision and machine learning techniques in tracking and analyzing individual honey bees' movements and interactions, revealing insights into the social organization of honey bee hives and their collective behavior. Specifically, the thesis has introduced novel methods for automatically tracking individual honey bees, assessing their social network, analyzing their behavioral variation, and identifying their functional roles in social groups over time. These methods provide a foundation for future studies and contribute to a better understanding of the importance of studying individual behaviors to understand collective phenomena.

While this thesis significantly advances our understanding of emergent behavior in social animals, it is not without limitations. One of the main limitations is the strong focus on methods and individual behavior. While these methods are applied to explain collective behavior, future studies could more directly study collective behavior to provide a comprehensive understanding of its underlying mechanisms. Additionally, the study relies solely on observational data, and no specific interventions or further experiments are conducted to study specific behaviors or phenomena in a controlled setting. Another limitation is the reliance on honey bees as the sole model organism to investigate collective behavior. While honey bees are an excellent model organism, the findings may not be entirely generalizable to other social animals. Moreover, the study's setup, which relies on computer vision and identification markers attached to individuals' thoraxes, means that not all bees can be detected at all times, particularly outside the colony. Additionally, the system cannot capture many sensory modalities relevant to bee social behavior, such as olfactory signals or vibrations. Another significant area for improvement in the study is the technological barrier to employing these methods. Setting up the recording system, tracking algorithm, and accessing the necessary storage and compute resources is not straightforward and requires collaboration between biologists and computer scientists. Likewise, data analysis using the proposed methods is technically challenging due to the scope of the data and the complexity of data analysis. Thus, analyzing the large and complex datasets requires expertise from both biologists and computer

scientists, leading to results based on data acquired from only two recording sites in Berlin and Konstanz. While the data collected from these sites provides valuable insights, the results may not be entirely generalizable to other contexts or settings. In conclusion, while this thesis provides valuable insights into the emergent behavior of social animals, it is essential to recognize its limitations. The study's focus on methods and individual behavior, its reliance on observational data, and the use of honey bees as the sole model organism may limit the generalizability of the findings. Additionally, the technological and data analysis challenges associated with the study may prevent the broad adoption of these methods in other settings. Nonetheless, the thesis provides a solid foundation for future research to build upon, and the study's findings have significant implications for understanding the mechanisms underlying collective behavior.

There are several potential directions for future work building on the limitations and contributions of this thesis. First, while honeybees have been used as a model organism in this study, it would be beneficial to apply the presented methods to investigate the collective behavior of other social animals, such as ants or termites, to assess the generalizability of the methods and findings. Second, while this thesis provides a comprehensive analysis of honey bee behavior in a natural setting, future studies could explore the effects of specific interventions or controlled experiments on the emergent behavior of a colony. For instance, researchers could investigate the effect of removing specific individuals or modifying the social network on the collective behavior of a colony. Such interventions would help identify the key factors that drive the emergence of collective behavior. Likewise, the effect of external stressors such as heat or insecticides on the colony's social network and task allocation could be measured (Insolia et al., 2022). Furthermore, biomimetic robots could be used to directly interfere and interact with, e.g., the honey bee waggle dance communication, thus allowing researchers to directly interfere in the communication of forage sites and measure the effect on the social interaction networks (Landgraf et al., 2012; Koenig et al., 2020). Third, future work could explore the potential of other sensory modalities, such as vibrations or olfactory signals, to complement the visual tracking system employed in this thesis (Uthoff et al., 2023). This could provide a more complete picture of the social interactions within a colony and offer insights into the role of non-visual cues in the emergence of collective behavior. Furthermore, the effects of genetic and transcriptomic variation on individual behavior and the social interactions within the colony could be directly measured using modern sequencing techniques (Jones et al., 2020). Fourth, while the methods proposed in this thesis represent a significant technological advancement, there is still room for improvement in accuracy and efficiency. For example, incorporating additional machine learning techniques or developing new tracking algorithms that can handle more complex environments could further enhance the quality of the data and reduce the computational burden of data analysis, e.g., by directly learning the optimal design of identification markers to improve detection accuracy in the specific recording setup and conditions in an observation hive. Finally, while the results of this study shed light on the mechanisms underlying collective behavior in honeybee colonies, further research is needed to investigate the potential applications of these methods. For instance, circadian rhythms are of fundamental importance in many biological processes (Kondo et al., 1994; Bell-Pedersen et al., 2005), yet much of the existing research studying these rhythms has been limited by small sample sizes, short observation periods, and unrealistic conditions (Bloch et al., 2013). Furthermore, methods developed in this thesis could be applied beyond the research of collective intelligence: For example, the Temporal Nonnegative Matrix

Factorization (TNMF) introduced in Chapter 6 could be applied to many settings in which matrix factorization is commonly used, such as recommender systems, network analysis, audio processing, bioinformatics, and more.

To conclude, this thesis has provided insights into the emergence of collective behavior in social animals, and introduced novel methods for automatically tracking individual honey bees, assessing their social network, analyzing their behavioral variation, and identifying their functional roles. The presented findings have shed light on the social organization of honey bee hives and their collective behavior, demonstrating the importance of studying individual behaviors to understand collective phenomena in complex systems.

Bibliography

Damien R. Farine and Hal Whitehead. Constructing, conducting and interpreting animal social network analysis. *Journal of Animal Ecology*, 84(5):1144–1163, 2015.

Deborah M. Gordon. *Ant Encounters: Interaction Networks and Colony Behavior*. Princeton University Press, 2010.

Jens Krause, Richard James, Daniel W. Franks, and Darren P. Croft. *Animal Social Networks*. Oxford University Press, 2015. ISBN 978-0-19-967904-1.

Noa Pinter-Wollman, E. A. Hobson, J. E. Smith, A. J. Edelman, D. Shizuka, S. de Silva, J. S. Waters, S. D. Prager, T. Sasaki, G. Wittemyer, J. Fewell, and D. B. McDonald. The dynamics of animal social networks: Analytical, conceptual, and theoretical advances. *Behavioral Ecology*, 25(2): 242–255, 2014. doi: 10.1093/beheco/art047.

Bert Hölldobler, Edward O Wilson, et al. *The superorganism: the beauty, elegance, and strangeness of insect societies*. WW Norton & Company, 2009.

Deborah M. Gordon. The organization of work in social insect colonies. *Nature*, 380(6570): 121–124, 1996. doi: 10.1038/380121a0.

Chris Tofts and Nigel R. Franks. Doing the right thing: Ants, honeybees and naked mole-rats. *Trends in Ecology & Evolution*, 7(10):346–349, 1992. doi: 10.1016/0169-5347(92)90128-x.

James FA Traniello and Rebeca B. Rosengaus. Ecology, evolution and division of labour in social insects. *Animal behaviour*, 53(1):209–213, 1997.

Noa Pinter-Wollman, Ashwin Bala, Andrew Merrell, Jovel Queirolo, Martin C. Stumpe, Susan Holmes, and Deborah M. Gordon. Harvester ants use interactions to regulate forager activation and availability. *Animal Behaviour*, 86(1):197–207, 2013. doi: 10.1016/j.anbehav.2013.05.012.

Noa Pinter-Wollman, Roy Wollman, Adam Guetz, Susan Holmes, and Deborah M. Gordon. The effect of individual variation on the structure and function of interaction networks in harvester ants. *Journal of The Royal Society Interface*, 8(64):1562–1573, 2011. doi: 10.1098/rsif.2011.0059.

Thomas D. Seeley. The tremble dance of the honey bee: Message and meanings. *Behavioral Ecology and Sociobiology*, 31(6):375–383, 1992. doi: 10.1007/bf00170604.

Deborah M. Gordon and Natasha J. Mehdiabadi. Encounter rate and task allocation in harvester ants. *Behavioral Ecology and Sociobiology*, 45(5):370–377, 1999. doi: 10.1007/s002650050573.

Jacob D. Davidson and Deborah M. Gordon. Spatial organization and interactions of harvester ants during foraging activity. *Journal of The Royal Society Interface*, 14(135):20170413, 2017. doi: 10.1098/rsif.2017.0413.

Lauren E. Quevillon, Ephraim M. Hanks, Shweta Bansal, and David P. Hughes. Social, spatial, and temporal organization in a complex insect society. *Scientific Reports*, 5:13393, 2015. doi: 10.1038/srep13393.

Joffrey Planckaert, Stamatios C. Nicolis, Jean-Louis Deneubourg, Cédric Sueur, and Olivier Bles. A spatiotemporal analysis of the food dissemination process and the trophallactic network in the ant Lasius niger. *Scientific Reports*, 9(1):1–11, 2019. doi: 10.1038/s41598-019-52019-6.

Karl von Frisch. *The Dance Language and Orientation of Bees.* Harvard University Press, 1967.

Benjamin P Oldroyd. What's Killing American Honey Bees? *PLoS Biology*, 5(6):e168, 2007. doi: 10.1371/journal.pbio.0050168.

Eva Crane. *The World History of Beekeeping and Honey Hunting*. Taylor & Francis, 1999. ISBN 978-0-415-92467-2.

Fernando Wario, Benjamin Wild, Margaret J. Couvillon, Raúl Rojas, and Tim Landgraf. Automatic methods for long-term tracking and the detection and decoding of communication dances in honeybees. *Frontiers in Ecology and Evolution*, 3, 2015. doi: 10.3389/fevo.2015.00103.

Deborah M. Gordon. The Ecology of Collective Behavior. *PLoS Biology*, 12(3), 2014. doi: 10.1371/journal.pbio.1001805.

Federico Rossi, Saptarshi Bandyopadhyay, Michael T. Wolf, and Marco Pavone. Multi-agent algorithms for collective behavior: A structural and application-focused atlas. *CoRR*, abs/2103.11067, 2021.

Haibin Duan, Mengzhen Huo, and Yanming Fan. From animal collective behaviors to swarm robotic cooperation. *National Science Review*, 10(5):nwad040, 2023.

Herbert Blumer. Social problems as collective behavior. *Social problems*, 18(3):298–306, 1971.

Andy Clark. Whatever next? predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(3):181–204, 2013. doi: 10.1017/s0140525x12000477.

Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language models are few-shot learners. In Hugo Larochelle, Marc'Aurelio Ranzato, Raia Hadsell, Maria-Florina Balcan, and Hsuan-Tien Lin, editors, *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020.

Dhruba Naug. Structure of the social network and its influence on transmission dynamics in a honeybee colony. *Behavioral Ecology and Sociobiology*, 62(11):1719–1725, 2008. doi: 10.1007/s00265-008-0600-x.

David Baracchi and Alessandro Cini. A Socio-Spatial Combined Approach Confirms a Highly Compartmentalised Structure in Honeybees. *Ethology*, 120(12):1167–1176, 2014. doi: 10.1111/eth.12290.

Jacobus C. Biesmeijer and Thomas D. Seeley. The use of waggle dance information by honey bees throughout their foraging careers. *Behavioral Ecology and Sociobiology*, 59(1):133–142, 2005. doi: 10.1007/s00265-005-0019-6.

Margaret J. Couvillon, Roger Schürch, and Francis L. W. Ratnieks. Waggle Dance Distances as Integrative Indicators of Seasonal Foraging Challenges. *Plos One*, 9(4):e93495, 2014. doi: 10.1371/journal.pone.0093495.

Jens Krause, Stefan Krause, Robert Arlinghaus, Ioannis Psorakis, Stephen Roberts, and Christian Rutz. Reality mining of animal social systems. *Trends in Ecology & Evolution*, 28(9):541–551, 2013. doi: 10.1016/j.tree.2013.06.002.

Anthony I. Dell, John A. Bender, Kristin Branson, Iain D. Couzin, Gonzalo G. de Polavieja, Lucas P.J.J. Noldus, Alfonso Pérez-Escudero, Pietro Perona, Andrew D. Straw, Martin Wikelski, and Ulrich Brose. Automated image-based tracking and its application in ecology. *Trends in Ecology & Evolution*, 29(7):417–428, 2014. doi: 10.1016/j.tree.2014.05.004.

Mayank Kabra, Alice A. Robie, Marta Rivera-Alba, Steven Branson, and Kristin Branson. JAABA: interactive machine learning for automatic annotation of animal behavior. *Nature Methods*, 10(1):64–67, 2013. doi: 10.1038/nmeth.2281.

Eyrun Eyjolfsdottir, Kristin Branson, Yisong Yue, and Pietro Perona. Learning recurrent representations for hierarchical behavior modeling. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net, 2017.

Christina Blut, Alessandro Crespi, Danielle Mersch, Laurent Keller, Linlin Zhao, Markus Kollmann, Benjamin Schellscheidt, Carsten Fülber, and Martin Beye. Automated computer-based detection of encounter behaviours in groups of honeybees. *Scientific Reports*, 7(1):17663, 2017. doi: 10.1038/s41598-017-17863-4.

Danielle P. Mersch, Alessandro Crespi, and Laurent Keller. Tracking individuals shows spatial fidelity is a key regulator of ant social organization. *Science*, 340(6136):1090–1093, 2013.

James D. Crall, Nick Gravish, Andrew M. Mountcastle, and Stacey A. Combes. BEEtag: A Low-Cost, Image-Based Tracking System for the Study of Animal Behavior and Locomotion. 10(9):e0136487, 2015. doi: 10.1371/journal.pone.0136487.

Tim Gernat, Vikyath D. Rao, Martin Middendorf, Harry Dankowicz, Nigel Goldenfeld, and Gene E. Robinson. Automated monitoring of behavior reveals bursty interaction patterns and rapid spreading dynamics in honeybee social networks. *Proceedings of the National Academy of Sciences*, 115(7):1433–1438, 2018. doi: 10.1073/pnas.1713568115.

Benjamin Wild, Leon Sixt, and Tim Landgraf. Automatic localization and decoding of honeybee markers using deep convolutional neural networks. *arXiv preprint*, pages 1–20, 2018.

Martin Lindauer. Ein beitrag zur frage der arbeitsteilung im bienenstaat. *Zeitschrift für vergleichende Physiologie*, 34(4):299–345, 1952.

Thomas D. Seeley. Adaptive significance of the age polyethism schedule in honeybee colonies. *Behavioral Ecology and Sociobiology*, 11(4):287–293, 1982. doi: 10.1007/bf00299306.

Thomas D. Seeley and Steven A. Kolmes. Age polyethism for hive duties in honey bees – illusion or reality? *Ethology*, 87(3-4):284–297, 1991. doi: 10.1111/j.1439-0310.1991.tb00253.x.

Brian R. Johnson. Organization of work in the honeybee: A compromise between division of labour and behavioural flexibility. *Proceedings of the Royal Society B: Biological Sciences*, 270 (1511):147–152, 2003. doi: 10.1098/rspb.2002.2207.

Adam J. Siegel, M. Kim Fondrk, Gro V. Amdam, and Robert E. Page. In-hive patterns of temporal polyethism in strains of honey bees (Apis mellifera) with distinct genetic backgrounds. *Behavioral Ecology and Sociobiology*, 67(10):1623–1632, 2013. doi: 10.1007/s00265-013-1573-y.

Michael L Smith, Phoebe A Koenig, and Jacob M Peters. The cues of colony size: how honey bees sense that their colony is large enough to begin to invest in reproduction. *Journal of Experimental Biology*, 220(9):1597–1605, 2017. doi: 10.1242/jeb.150342.

Adrian A. Perez and Brian R. Johnson. Task repertoires of hygienic workers reveal a link between specialized necrophoric behaviors in honey bees. *Behavioral Ecology and Sociobiology*, 73(9), 2019. doi: 10.1007/s00265-019-2731-7.

Josefine Bohr Brask, Samuel Ellis, and Darren P Croft. Animal social networks: an introduction for complex systems scientists. *Journal of Complex Networks*, 9(2):cnab001, 2021.

Matthew J. Hasenjager, William Hoppitt, and Ellouise Leadbeater. Network-based diffusion analysis reveals context-specific dominance of dance communication in foraging honeybees. *Nature Communications*, 11(1):1–9, 2020. doi: 10.1038/s41467-020-14410-0.

Katarzyna Bozek, Laetitia Hebert, Yoann Portugal, Alexander S Mikheyev, and Greg J Stephens. Markerless tracking of an entire honey bee colony. *Nature communications*, 12(1):1733, 2021.

Brian R. Johnson. Division of labor in honeybees: form, function, and proximate mechanisms. *Behavioral Ecology and Sociobiology*, 64(3):305–316, 2010. doi: 10.1007/s00265-009-0874-7.

Walter Marcelo Farina. The interplay between dancing and trophallactic behavior in the honey bee Apis mellifera. *Journal of Comparative Physiology A*, 2000. doi: 10.1007/s003590050424.

M. B. Girard, H. R. Mattila, and T. D. Seeley. Recruitment-dance signals draw larger audiences when honey bee colonies have multiple patrilines. *Insectes Sociaux*, 58(1):77–86, 2011. doi: 10.1007/s00040-010-0118-x.

James D. Crall, Nick Gravish, Andrew M. Mountcastle, Sarah D. Kocher, Robert L. Oppenheimer, Naomi E. Pierce, and Stacey A. Combes. Spatial fidelity of workers predicts collective response to disturbance in a social insect. *Nature Communications*, 9(1):1201, 2018. doi: 10.1038/s41467-018-03561-w.

Beryl M Jones, Vikyath D Rao, Tim Gernat, Tobias Jagla, Amy C Cash-Ahmed, Benjamin E R Rubin, Troy J Comi, S Bhogale, S S Husain, C Blatti, Martin Middendorf, Saurabh Sinha, S Chandrasekaran, and G. E. Robinson. Individual differences in honey bee behavior enabled by plasticity in brain gene regulatory networks. *eLife*, 9:e62850, 2020.

Thomas O Richardson, Tomas Kay, Raphael Braunschweig, Opaline A Journeau, Matthias Ruegg, Sean Mcgregor, Paolo De Los Rios, and Laurent Keller. Ant behavioral maturation is mediated by a stochastic transition between two fundamental states. *Current Biology*, 31:1–8, 2021. doi: 10.1016/j.cub.2020.05.038.

Zachary Y. Huang, G. E. Robinson1, and D. W. Borst. Physiological correlates of division of labor among similarly aged honey bees. *Journal of Comparative Physiology A*, 174(6):731–739, 1994. doi: 10.1007/bf00192722.

Samuel N Beshers and Jennifer H Fewell. Models of division of labor in social insects. *Annual review of entomology*, 46(1):413–440, 2001.

Gene E Robinson. Genomics and integrative analyses of division of labor in honeybee colonies. *The American Naturalist*, 160(S6):S160–s172, 2002.

Christina M. Grozinger, Yongliang Fan, Shelley E.R. Hoover, and Mark L. Winston. Genome-wide analysis reveals differences in brain gene expression patterns associated with caste and reproductive status in honey bees (apis mellifera). *Molecular Ecology*, 16(22):4837–4848, 2007. doi: 10.1111/j.1365-294X.2007.03545.x.

Brian R Johnson. Within-nest temporal polyethism in the honey bee. *Behavioral Ecology and Sociobiology*, 62(5):777–784, 2008a.

Chelsea N. Cook and Michael D. Breed. Social context influences the initiation and threshold of thermoregulatory behaviour in honeybees. *Animal Behaviour*, pages 1–7, 2013. doi: 10.1016/j.anbehav.2013.05.021.

Chelsea N. Cook, Thiago Mosqueiro, Colin S. Brent, Cahit Ozturk, Jürgen Gadau, Noa Pinter-Wollman, and Brian H. Smith. Individual differences in learning and biogenic amine levels

influence the behavioural division between foraging honeybee scouts and recruits. *Journal of Animal Ecology*, 88(2):236–246, 2019. doi: 10.1111/1365-2656.12911.

Raphaël Jeanson and Anja Weidenmüller. Interindividual variability in social insects - proximate causes and ultimate consequences. *Biological Reviews*, 89(3):671–687, 2014. doi: 10.1111/brv.12074.

Andrea Frome, Gregory S. Corrado, Jonathon Shlens, Samy Bengio, Jeffrey Dean, Marc'Aurelio Ranzato, and Tomás Mikolov. Devise: A deep visual-semantic embedding model. In Christopher J. C. Burges, Léon Bottou, Zoubin Ghahramani, and Kilian Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States*, pages 2121–2129, 2013.

Ehsaneddin Asgari and Mohammad R. K. Mofrad. Continuous Distributed Representation of Biological Sequences for Deep Proteomics and Genomics. *Plos One*, 10(11):e0141287, 2015. doi: 10.1371/journal.pone.0141287.

Jose Camacho-Collados and Mohammad Taher Pilehvar. From word to sense embeddings: a survey on vector representations of meaning. *Journal of Artificial Intelligence Research*, 63(1):743–788, 2018. doi: 10.1613/jair.1.11259.

Walter Nelson, Marinka Zitnik, Bo Wang, Jure Leskovec, Anna Goldenberg, and Roded Sharan. To Embed or Not: Network Embedding as a Paradigm in Computational Biology. *Frontiers in Genetics*, 10, 2019. doi: 10.3389/fgene.2019.00381.

Yehuda Koren, Robert Bell, and Chris Volinsky. Matrix Factorization Techniques for Recommender Systems. *Computer*, 42(8):30–37, 2009. doi: 10.1109/mc.2009.263.

Tomas Mikolov, Wen-tau Yih, and Geoffrey Zweig. Linguistic regularities in continuous space word representations. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 746–751, Atlanta, Georgia, 2013. Association for Computational Linguistics.

Franziska Boenisch, Benjamin Rosemann, Benjamin Wild, David Dormagen, Fernando Wario, and Tim Landgraf. Tracking all members of a honey bee colony over their lifetime using learned models of correspondence. *Frontiers in Robotics and AI*, 5, 2018a. URL https://doi.org/10.3389/frobt.2018.00035.

Karl von Frisch. *Tanzsprache und Orientierung der Bienen*. Springer, 1965. ISBN 978-3-642-94916-6.

Thomas D. Seeley. *Honeybee Democracy*. Princeton University Press, Princeton, New Jersey, 2010. ISBN 978-0-691-14721-5.

Gene E Robinson. Regulation of division of labor in insect societies. *Annual review of entomology*, 37:637–665, 1992. doi: 10.1146/annurev.ento.37.1.637.

Christoph Grüter, Luis E Acosta, and Walter M Farina. Propagation of olfactory information within the honeybee hive. *Behavioral Ecology and Sociobiology*, 60(5):707–715, 2006.

M. S. Balbuena, J. Molinas, and W. M. Farina. Honeybee recruitment to scented food sources: correlations between in-hive social interactions and foraging decisions. *Behavioral Ecology and Sociobiology*, 66(3):445–452, 2012. doi: 10.1007/s00265-011-1290-3.

Christoph Grüter and Walter M. Farina. Past Experiences Affect Interaction Patterns Among Foragers and Hive-Mates in Honeybees. *Ethology*, 115(8):790–797, 2009. doi: 10.1111/j.1439-0310.2009.01670.x.

Christoph Grüter and Francis L. W. Ratnieks. Honeybee foragers increase the use of waggle dance information when private information becomes unrewarding. *Animal Behaviour*, 81(5):949–954, 2011. doi: 10.1016/j.anbehav.2011.01.014.

Joaquín Goyret and Walter M. Farina. Non-random nectar unloading interactions between foragers and their receivers in the honeybee hive. *Naturwissenschaften*, 92(9):440–443, 2005. doi: 10.1007/s00114-005-0016-7.

Rodrigo De Marco and Walter Farina. Changes in food source profitability affect the trophallactic and dance behavior of forager honeybees (Apis mellifera L.). *Behavioral Ecology and Sociobiology*, 50(5):441–449, 2001. doi: 10.1007/s002650100382.

Monica Raveret Richter and Keith D. Waddington. Past foraging experience influences honey bee dance behaviour. *Animal Behaviour*, 46(1):123–128, 1993. doi: 10.1006/anbe.1993.1167.

Gordon J. Berman, Daniel M. Choi, William Bialek, and Joshua W. Shaevitz. Mapping the stereotyped behaviour of freely moving fruit flies. *Journal of The Royal Society Interface*, 11(99):20140672, 2014. doi: 10.1098/rsif.2014.0672.

AlexanderB. Wiltschko, MatthewJ. Johnson, Giuliano Iurilli, RalphE. Peterson, JesseM. Katon, StanL. Pashkovski, VictoriaE. Abraira, RyanP. Adams, and SandeepRobert Datta. Mapping Sub-Second Structure in Mouse Behavior. *Neuron*, 88(6):1121–1135, 2015. doi: 10.1016/j.neuron.2015.11.031.

Fernando Wario, Benjamin Wild, Raúl Rojas, and Tim Landgraf. Automatic detection and decoding of honey bee waggle dances. *PloS one*, 12(12):e0188626, 2017.

Thomas M Thompson. *From error-correcting codes through sphere packings to simple groups*. Number 21. Cambridge University Press, 1983.

Fernando Wario. A computer vision based system for the automatic analysis of social networks in honey bee colonies. PhD dissertation, Freie Universität Berlin, 2017.

Ingemar J. Cox. A review of statistical data association techniques for motion correspondence. *International Journal of Computer Vision*, 10(1):53–66, 1993. doi: 10.1007/bf01440847.

Yi Wu, Jongwoo Lim, and Ming-Hsuan Yang. Online object tracking: A benchmark. In *2013 IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, June 23-28, 2013*, pages 2411–2418. IEEE Computer Society, 2013. doi: 10.1109/CVPR.2013.312.

Wenhan Luo, Junliang Xing, Anton Milan, Xiaoqin Zhang, Wei Liu, and Tae-Kyun Kim. Multiple object tracking: A literature review. *Artif. Intell.*, 293:103448, 2021. doi: 10.1016/j.artint.2020.103448.

Margrit Betke and Zheng Wu. Data association for multi-object visual tracking. *Synthesis Lectures on Computer Vision*, 6(2):1–120, 2016. doi: http://dx.doi.org/10.2200/S00713ED1V01Y201603COV009.

Alfonso Pérez-Escudero, Julián Vicente-Page, Robert C Hinz, Sara Arganda, and Gonzalo G de Polavieja. idTracker: tracking individuals in a group by automatic identification of unmarked animals. *Nature Methods*, 11(7):743–748, 2014. doi: 10.1038/nmeth.2994.

Hjalmar S. Kühl and Tilo Burghardt. Animal biometrics: quantifying and detecting phenotypic appearance. *Trends in Ecology & Evolution*, 28(7):432–441, 2013. doi: 10.1016/j.tree.2013.02.013.

Naiyan Wang and Dit-Yan Yeung. Learning a deep compact image representation for visual tracking. In Christopher J. C. Burges, Léon Bottou, Zoubin Ghahramani, and Kilian Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States*, pages 809–817, 2013.

Benjamin Rosemann. Ein erweiterbares system für experimente mit multi-target tracking von markierten bienen. Master Thesis, Freie Universität Berlin, 2017.

Franziska Boenisch. Feature Engineering and Probabilistic Tracking on Honey Bee Trajectories. Bachelor Thesis, 2017.

Bo Wu and Ramakant Nevatia. Detection and tracking of multiple, partially occluded humans by bayesian combination of edgelet based part detectors. *Int. J. Comput. Vis.*, 75(2):247–266, 2007. doi: 10.1007/s11263-006-0027-7.

Chang Huang, Bo Wu, and Ramakant Nevatia. Robust object tracking by hierarchical association of detection responses. In David A. Forsyth, Philip H. S. Torr, and Andrew Zisserman, editors, *Computer Vision - ECCV 2008, 10th European Conference on Computer Vision, Marseille, France, October 12-18, 2008, Proceedings, Part II*, volume 5303 of *Lecture Notes in Computer Science*, pages 788–801. Springer, 2008. doi: 10.1007/978-3-540-88688-4_58.

L. Wang, N. H. C. Yung, and L. Xu. Multiple-Human Tracking by Iterative Data Association and Detection Update. *IEEE Transactions on Intelligent Transportation Systems*, 15(5):1886–1899, 2014. doi: 10.1109/tits.2014.2303196.

Thomas Fasciano, Hoan Nguyen, Anna Dornhaus, and Min C. Shin. Tracking multiple ants in a colony. In *2013 IEEE Workshop on Applications of Computer Vision (WACV)*, pages 534–540. Ieee, 2013. ISBN 978-1-4673-5054-9. doi: 10.1109/wacv.2013.6475065.

Harold W. Kuhn. The Hungarian method for the assignment problem. 2:83–97, 1955.

Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Mach. Learn.*, 20(3):273–297, 1995. doi: 10.1007/BF00994018.

Tin Kam Ho. Random decision forests. In *Third International Conference on Document Analysis and Recognition, ICDAR 1995, August 14 - 15, 1995, Montreal, Canada. Volume I*, pages 278–282. IEEE Computer Society, 1995. doi: 10.1109/ICDAR.1995.598994.

Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vander-Plas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Edouard Duchesnay. Scikit-learn: Machine learning in python. *J. Mach. Learn. Res.*, 12:2825–2830, 2011. doi: 10.5555/1953048.2078195.

John C. Platt. Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods. In *Advances in Large Margin Classifiers*, pages 61–74. MIT Press, 1999.

Kent A. Spackman. Signal detection theory: Valuable tools for evaluating inductive learning. In *Proceedings of the Sixth International Workshop on Machine Learning*, pages 160–163. Morgan Kaufmann Publishers Inc., 1989. ISBN 978-1-55860-036-2.

Jakob Mischek. Probabilistisches Tracking von Bienenpfaden. Master Thesis, Freie Universität Berlin, 2016.

Brent Komer, James Bergstra, and Chris Eliasmith. Hyperopt-sklearn: Automatic hyperparameter configuration for scikit-learn. In Stéfan van der Walt and James Bergstra, editors, *Proceedings of the 13th Python in Science Conference, SciPy 2014, Austin, Texas, July 6 - 12, 2014*, pages 32–37. scipy.org, 2014. doi: 10.25080/Majora-14bd3278-006.

Franziska Boenisch, Benjamin Rosemann, Benjamin Wild, Fernando Wario, David Dormagen, and Tim Landgraf. BeesBook Recording Season 2015 Sample, 2018b.

Lance Rice, Anna R. Dornhaus, and Min C. Shin. Efficient training of multiple ant tracking. In *2015 IEEE Winter Conference on Applications of Computer Vision, WACV 2015, Waikoloa, HI, USA, January 5-9, 2015*, pages 117–123. IEEE Computer Society, 2015. doi: 10.1109/WACV.2015.23.

Benjamin Wild, David M. Dormagen, and Tim Landgraf. Social networks predict the life and death of honey bees - data, 2021a.

Benjamin Wild, David M. Dormagen, Adrian Zachariae, Michael L. Smith, Kirsten S. Traynor, Dirk Brockmann, Iain D. Couzin, and Tim Landgraf. Social networks predict the life and death of honey bees. *Nature Communications*, 12(1):1110, 2021b. URL https://doi.org/10.1038/s41467-021-21212-5.

Ioannis Psorakis, Stephen J. Roberts, Iead Rezek, and Ben C. Sheldon. Inferring social network structure in ecological systems from spatio-temporal data streams. *Journal of the Royal Society, Interface*, 9(76):3055–3066, 2012. doi: 10.1098/rsif.2012.0223.

David Lusseau and M. E. J. Newman. Identifying the Role That Animals Play in Their Social Networks. *Proceedings: Biological Sciences*, 271:S477–s481, 2004.

Dorothy L. Cheney, Joan B. Silk, and Robert M. Seyfarth. Network connections, dyadic bonds and fitness in wild female baboons. *Royal Society Open Science*, 3(7):160255, 2016. doi: 10.1098/rsos.160255.

Lucy M. Aplin, Damien R. Farine, Julie Morand-Ferron, Andrew Cockburn, Alex Thornton, and Ben C. Sheldon. Experimentally induced innovations lead to persistent culture via conformity in wild birds. *Nature*, 518(7540):538–541, 2015. doi: 10.1038/nature13998.

Nicolas Claidière, Emily J. E. Messer, William Hoppitt, and Andrew Whiten. Diffusion Dynamics of Socially Learned Foraging Techniques in Squirrel Monkeys. *Current Biology*, 23(13):1251–1255, 2013. doi: 10.1016/j.cub.2013.05.036.

Jessica C. Flack, Michelle Girvan, Frans B. M. de Waal, and David C. Krakauer. Policing stabilizes construction of social niches in primates. *Nature*, 439(7075):426–429, 2006. doi: 10.1038/nature04326.

Ajay Mehra, Andrea L. Dixon, Daniel J. Brass, and Bruce Robertson. The Social Network Ties of Group Leaders: Implications for Group Performance and Leader Reputation. *Organization Science*, 17(1):64–79, 2006. doi: 10.1287/orsc.1050.0158.

Cedric Sueur and Odile Petit. Organization of Group Members at Departure Is Driven by Social Structure in Macaca. *International Journal of Primatology*, 29(4):1085–1098, 2008. doi: 10.1007/s10764-008-9262-9.

Ariana Strandburg-Peshkin, Danai Papageorgiou, Margaret C. Crofoot, and Damien R. Farine. Inferring influence and leadership in moving animal groups. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1746):20170006, 2018. doi: 10.1098/rstb.2017.0006.

Sara Brin Rosenthal, Colin R. Twomey, Andrew T. Hartnett, Hai Shan Wu, and Iain D. Couzin. Revealing the hidden networks of interaction in mobile animal groups allows prediction of complex behavioral contagion. *Proceedings of the National Academy of Sciences*, 112(15):4690–4695, 2015. doi: 10.1073/pnas.1420068112.

Ana B. Sendova-Franks, Rebecca K. Hayward, Benjamin Wulf, Thomas Klimek, Richard James, Robert Planqué, Nicholas F. Britton, and Nigel R. Franks. Emergency networking: Famine relief in ant colonies. *Animal Behaviour*, 79(2):473–485, 2010. doi: 10.1016/j.anbehav.2009.11.035.

Alex Gomez-Marin, Joseph J. Paton, Adam R. Kampff, Rui M. Costa, and Zachary F. Mainen. Big behavioral data: Psychology, ethology and the foundations of neuroscience. *Nature Neuroscience*, 17(11):1455–1463, 2014.

Stanley S. Schneider and Lee A. Lewis. The vibration signal, modulatory communication and the organization of labor in honey bees, *Apis mellifera*. *Apidologie*, 35(2):117–131, 2004. doi: 10.1051/apido:2004006.

Z. Y. Huang and G. E. Robinson. Regulation of honey bee division of labor by colony age demography. *Behavioral Ecology and Sociobiology*, 39(3):147–158, 1996. doi: 10.1007/s002650050276.

Gro V. Amdam and Stig W. Omholt. The hive bee to forager transition in honeybee colonies: The double repressor hypothesis. *Journal of Theoretical Biology*, 223(4):451–464, 2003. doi: 10.1016/s0022-5193(03)00121-8.

Kate E. Ihle, Robert E. Page, Katy Frederick, M. Kim Fondrk, and Gro V. Amdam. Genotype effect on regulation of behaviour by vitellogenin supports reproductive origin of honeybee foraging bias. *Animal Behaviour*, 79(5):1001–1006, 2010. doi: 10.1016/j.anbehav.2010.02.009.

T. Pankiw and R. E. Page Jr. The effect of genotype, age, sex, and caste on response thresholds to sucrose and foraging behavior of honey bees (Apis mellifera L.). *Journal of Comparative Physiology A*, 185(2):207–213, 1999. doi: 10.1007/s003590050379.

Ricarda Scheiner. Birth weight and sucrose responsiveness predict cognitive skills of honeybee foragers. *Animal Behaviour*, 84(2):305–308, 2012. doi: 10.1016/j.anbehav.2012.05.011.

Ying Wang, Sarah D. Kocher, Timothy A. Linksvayer, Christina M. Grozinger, Robert E. Page, and Gro V. Amdam. Regulation of behaviorally associated gene networks in worker honey bee ovaries. *Journal of Experimental Biology*, 215(1):124–134, 2012. doi: 10.1242/jeb.060889.

Ying Wang, Osman Kaftanoglu, Adam J. Siegel, Robert E. Page, and Gro V. Amdam. Surgically increased ovarian mass in the honey bee confirms link between reproductive physiology and worker behavior. *Journal of Insect Physiology*, 56(12):1816–1824, 2010. doi: 10.1016/j.jinsphys.2010.07.013.

D. Münch, G. V. Amdam, and F. Wolschin. Ageing in a eusocial insect: Molecular and physiological characteristics of life span plasticity in the honey bee. *Functional ecology*, 22(3):407–421, 2008. doi: 10.1111/j.1365-2435.2008.01419.x.

Claudia Dreller, Robert E. Page Jr, and M. Kim Fondrk. Regulation of pollen foraging in honeybee colonies: Effects of young brood, stored pollen, and empty space. *Behavioral Ecology and Sociobiology*, 45(3-4):227–233, 1999. doi: 10.1007/s002650050557.

Thomas D. Seeley. Social foraging in honey bees: How nectar foragers assess their colony's nutritional status. *Behavioral Ecology and Sociobiology*, 24(3):181–199, 1989a. doi: 10.1007/bf00292101.

Kirsten S. Traynor, Yves Le Conte, and Robert E. Page. Age matters: Pheromone profiles of larvae differentially influence foraging behaviour in the honeybee, Apis mellifera. *Animal Behaviour*, 99:1–8, 2015. doi: 10.1016/j.anbehav.2014.10.009.

Seth A. Ament, Ying Wang, and Gene E. Robinson. Nutritional regulation of division of labor in honey bees: Toward a systems biology perspective. *WIREs Systems Biology and Medicine*, 2(5): 566–576, 2010. doi: 10.1002/wsbm.73.

Z. Y. Huang and G. E. Robinson. Seasonal changes in juvenile hormone titers and rates of biosynthesis in honey bees. *Journal of Comparative Physiology B*, 165(1):18–28, 1995. doi: 10.1007/bf00264682.

Amy L. Toth, Sara Kantarovich, Adam F. Meisel, and Gene E. Robinson. Nutritional status influences socially regulated foraging ontogeny in honey bees. *The Journal of Experimental Biology*, 208:4641–4649, 2005. doi: 10.1242/jeb.01956.

Ying Wang, Osman Kaftanoglu, Colin S. Brent, Robert E. Page, and Gro V. Amdam. Starvation stress during larval development facilitates an adaptive response in adult worker honey bees (Apis mellifera L.). *Journal of Experimental Biology*, 219(7):949–959, 2016. doi: 10.1242/jeb.130435.

Thomas D. Seeley. *The Wisdom of the Hive: The Social Physiology of Honey Bee Colonies*. Harvard University Press, 1995. ISBN 978-0-674-95376-5.

Hanna Cholé, Julie Carcaud, Hélène Mazeau, Sylvain Famié, Gérard Arnold, and Jean-Christophe Sandoz. Social Contact Acts as Appetitive Reinforcement and Supports Associative Learning in Honeybees. *Current Biology*, 29(8):1407–1413.e3, 2019. doi: 10.1016/j.cub.2019.03.025.

James C. Nieh. A Negative Feedback Signal That Is Triggered by Peril Curbs Honey Bee Recruitment. *Current Biology*, 20(4):310–315, 2010. doi: 10.1016/j.cub.2009.12.060.

Thomas D. Seeley, P. Kirk Visscher, Thomas Schlegel, Patrick M. Hogan, Nigel R. Franks, and James A. R. Marshall. Stop Signals Provide Cross Inhibition in Collective Decision-Making by Honeybee Swarms. *Science*, 335(6064):108, 2012. doi: 10.1126/science.1210361.

J. Goyret and W. M. Farina. Descriptive study of antennation during trophallactic unloading contacts in honeybees Apis mellifera carnica. *Insectes Sociaux*, 50(3):274–276, 2003. doi: 10.1007/s00040-003-0678-0.

M. L. Smith, M. M. Ostwald, and T. D. Seeley. Honey bee sociometry: Tracking honey bee colonies and their nest contents from colony founding until death. *Insectes Sociaux*, 63(4):553–563, 2016. doi: 10.1007/s00040-016-0499-6.

Mikhail Belkin and Partha Niyogi. Laplacian Eigenmaps for Dimensionality Reduction and Data Representation. *Neural Computation*, 15(6):1373–1396, 2003. doi: 10.1162/089976603321780317.

Ulrike von Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416, 2007. doi: 10.1007/s11222-007-9033-z.

Harold Hotelling. Relations between two sets of variables. *Biometrika*, 28(3-4):321–377, 1936. doi: 10.1093/biomet/28.3-4.321.

Thomas R. Knapp. Canonical correlation analysis: A general parametric significance-testing system. *Psychological Bulletin*, 85(2):410–416, 1978. doi: 10.1037/0033-2909.85.2.410.

Daniel McFadden et al. Conditional logit analysis of qualitative choice behavior. 1973.

Howard Levene. Robust tests for equality of variances. *Contributions to probability and statistics*, pages 278–292, 1960.

Nobuyuki Otsu. A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1):62–66, 1979. doi: 10.1109/tsmc.1979.4310076.

Gene E. Robinson and Francis LW Ratnieks. Induction of premature honey bee (Hymenoptera: Apidae) flight by juvenile hormone analogs administered orally or topically. *Journal of Economic Entomology*, 80(4):784–787, 1987.

Gro V. Amdam, Kari Norberg, M. Kim Fondrk, and Robert E. Page. Reproductive ground plan may mediate colony-level selection effects on individual foraging behavior in honey bees. *Proceedings of the National Academy of Sciences*, 101(31):11350–11355, 2004. doi: 10.1073/pnas.0403073101.

Ricarda Scheiner, Robert E. Page, and Joachim Erber. Sucrose responsiveness and behavioral plasticity in honey bees (Apis mellifera). *Apidologie*, 35(2):133–142, 2004.

Gro V. Amdam and Robert E. Page Jr. The developmental genetics and physiology of honeybee societies. *Animal behaviour*, 79(5):973–980, 2010.

Clint J. Perry, Eirik Søvik, Mary R. Myerscough, and Andrew B. Barron. Rapid behavioral maturation accelerates failure of stressed honey bee colonies. *Proceedings of the National Academy of Sciences*, 112(11):3427–3432, 2015. doi: 10.1073/pnas.1422089112.

Olav Rueppell, Robyn Linford, Preston Gardner, Jennifer Coleman, and Kari Fine. Aging and demographic plasticity in response to experimental age structures in honeybees (Apis mellifera L). *Behavioral Ecology and Sociobiology*, 62(10):1621, 2008. doi: 10.1007/s00265-008-0591-7.

T.O. Richardson, T. Kay, R. Braunschweig, O. A. Journeau, M. Rüegg, S. McGregor, P. De Los Rio, and L. Keller. Ant Behavioral Maturation Is Mediated by a Stochastic Transition between Two Fundamental States. *Current Biology*, in press.

Michael L. Smith. The Honey Bee Parasite Nosema ceranae: Transmissible via Food Exchange? *Plos One*, 7(8):e43319, 2012. doi: 10.1371/journal.pone.0043319.

Antoine Lecocq, Annette Bruun Jensen, Per Kryger, and James C. Nieh. Parasite infection accelerates age polyethism in young honey bees. *Scientific Reports*, 6(1):22042, 2016. doi: 10.1038/srep22042.

Amy C. Geffre, Tim Gernat, Gyan P. Harwood, Beryl M. Jones, Deisy Morselli Gysi, Adam R. Hamilton, Bryony C. Bonning, Amy L. Toth, Gene E. Robinson, and Adam G. Dolezal. Honey bee virus causes context-dependent changes in host social behavior. *Proceedings of the National Academy of Sciences*, 2020. doi: 10.1073/pnas.2002268117.

Hauke J. Mönck. Bioroboticslab/bb_irflash, 2021. URL https://zenodo.org/records/4436467.

John Salvatier, Thomas V. Wiecki, and Christopher Fonnesbeck. Probabilistic programming in Python using PyMC3. *PeerJ Computer Science*, 2:e55, 2016. doi: 10.7717/peerj-cs.55.

Stéfan van der Walt, Johannes L. Schönberger, Juan Nunez-Iglesias, François Boulogne, Joshua D. Warner, Neil Yager, Emmanuelle Gouillart, and Tony Yu. Scikit-image: Image processing in Python. *PeerJ*, 2:e453, 2014. doi: 10.7717/peerj.453.

Hervé Abdi. Singular and Generalized Singular Value Decomposition. In *Encyclopedia of Measurement and Statistics*, pages 908–912. SAGE Publications, Inc., 2007. doi: 10.4135/9781412952644.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Köpf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In Hanna M. Wallach, Hugo Larochelle, Alina Beygelzimer, Florence d'Alché-Buc, Emily B. Fox, and Roman Garnett, editors, *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, pages 8024–8035, 2019.

Joe H. Jr. Ward. Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*, 58(301):236–244, 1963. doi: 10.1080/01621459.1963.10500845.

The Astropy Collaboration, A. M. Price-Whelan, B. M. Sipőcz, H. M. Günther, P. L. Lim, S. M. Crawford, S. Conseil, D. L. Shupe, M. W. Craig, N. Dencheva, A. Ginsburg, J. T. VanderPlas, L. D. Bradley, D. Pérez-Suárez, M. de Val-Borro, T. L. Aldcroft, K. L. Cruz, T. P. Robitaille, E. J. Tollerud, C. Ardelean, T. Babej, M. Bachetti, A. V. Bakanov, S. P. Bamford, G. Barentsen, P. Barmby, A. Baumbach, K. L. Berry, F. Biscani, M. Boquien, K. A. Bostroem, L. G. Bouma, G. B. Brammer, E. M. Bray, H. Breytenbach, H. Buddelmeijer, D. J. Burke, G. Calderone, J. L. Cano Rodríguez, M. Cara, J. V. M. Cardoso, S. Cheedella, Y. Copin, D. Crichton, D. DÁvella, C. Deil, É Depagne, J. P. Dietrich, A. Donath, M. Droettboom,

N. Earl, T. Erben, S. Fabbro, L. A. Ferreira, T. Finethy, R. T. Fox, L. H. Garrison, S. L. J. Gibbons, D. A. Goldstein, R. Gommers, J. P. Greco, P. Greenfield, A. M. Groener, F. Grollier, A. Hagen, P. Hirst, D. Homeier, A. J. Horton, G. Hosseinzadeh, L. Hu, J. S. Hunkeler, Ž Ivezić, A. Jain, T. Jenness, G. Kanarek, S. Kendrew, N. S. Kern, W. E. Kerzendorf, A. Khvalko, J. King, D. Kirkby, A. M. Kulkarni, A. Kumar, A. Lee, D. Lenz, S. P. Littlefair, Z. Ma, D. M. Macleod, M. Mastropietro, C. McCully, S. Montagnac, B. M. Morris, M. Mueller, S. J. Mumford, D. Muna, N. A. Murphy, S. Nelson, G. H. Nguyen, J. P. Ninan, M. Nöthe, S. Ogaz, S. Oh, J. K. Parejko, N. Parley, S. Pascual, R. Patil, A. A. Patil, A. L. Plunkett, J. X. Prochaska, T. Rastogi, V. Reddy Janga, J. Sabater, P. Sakurikar, M. Seifert, L. E. Sherbert, H. Sherwood-Taylor, A. Y. Shih, J. Sick, M. T. Silbiger, S. Singanamalla, L. P. Singer, P. H. Sladen, K. A. Sooley, S. Sornarajah, O. Streicher, P. Teuben, S. W. Thomas, G. R. Tremblay, J. E. H. Turner, V. Terrón, M. H. van Kerkwijk, A. de la Vega, L. L. Watkins, B. A. Weaver, J. B. Whitmore, J. Woillez, and V. Zabalza. The Astropy Project: Building an inclusive, open-science project and status of the v2.0 core package. *The Astronomical Journal*, 156(3):123, 2018. doi: 10.3847/1538-3881/aabc4f.

Benjamin Wild and David M. Dormagen. Bioroboticslab/bb_network_decomposition, 2021. URL https://zenodo.org/records/4441807.

Michael L. Smith, Jacob D. Davidson, Benjamin Wild, David M. Dormagen, Tim Landgraf, and Iain D. Couzin. Behavioral variation across the days and lives of honey bees. *iScience*, 25(9): 104842, 2022a. URL https://doi.org/10.1016/j.isci.2022.104842.

Thomas D Seeley. The honey bee colony as a superorganism. *American Scientist*, 77(6):546–553, 1989b.

David Sloan Wilson and Elliott Sober. Reviving the superorganism. *Journal of Theoretical Biology*, 136(3):337–356, 1989. doi: doi:10.1016/S0022-5193(89)80169-9.

John Maynard Smith and Eors Szathmary. *The major transitions in evolution*. Oxford University Press, 1995.

George F Oster and Edward O Wilson. *Caste and ecology in the social insects*. Princeton University Press, 1978.

Sanford D Porter and Walter R Tschinkel. Fire ant polymorphism: the ergonomics of brood production. *Behavioral Ecology and Sociobiology*, 16(4):323–336, 1985.

Robert L. Jeanne. The organization of work in polybia occidentalis: costs and benefits of specialization in a social wasp. *Behavioral Ecology and Sociobiology*, 19(5):333–341, 1986. doi: 10.1007/bf00295706.

Robert L. Jeanne, Holly A. Downing, and David C. Post. Age Polyethism and Individual Variation in Polybia occidentalis, an Advanced Eusocial Wasp. In *Interindividual Behavioral Variability in Social Insects*, pages 323–357. CRC Press, 1988. ISBN 978-0-429-04047-4.

Peter C Frumhoff and Jayne Baker. A genetic component to division of labour within honey bee colonies. *Nature*, 333:35–38, 1988.

G. E. Robinson, Robert E. Page, Colette Strambi, and Alain Strambi. Hormonal and genetic control of behavioral integration in honey bee colonies. *Science*, 246(4926):109–112, 1989.

J H Fewell and R E Page. Genotypic variation in foraging responses to environmental stimuli by honey bees, apis mellifera. *Experientia*, 49:1106–1112, 1993.

Sean O'Donnell and R. L. Jeanne. The roles of body size and dominance in division of labor among workers of the eusocial wasp polybia occidentalis (olivier) (hymenoptera: Vespidae). *Journal of the Kansas Entomological Society*, 68(1):43–50, 1995.

Dhruba Naug and Raghavendra Gadagkar. The role of age in temporal polyethism in a primitively eusocial wasp. *Behavioral Ecology and Sociobiology*, 42(1):37–47, 1998. doi: 10.1007/s002650050409.

Benjamin P. Oldroyd and Jennifer H. Fewell. Genetic diversity promotes homeostasis in insect colonies. *Trends in Ecology and Evolution*, 22(8):408–413, 2007. doi: 10.1016/j.tree.2007.06.001.

Elva JH Robinson, Ofer Feinerman, and Nigel R Franks. Flexible task allocation and the organization of work in ants. *Proceedings of the Royal Society B: Biological Sciences*, 276(1677): 4373–4380, 2009a.

Jennifer M. Jandt and Anna Dornhaus. Spatial organization and division of labour in the bumblebee bombus impatiens. *Animal Behaviour*, 77(3):641–651, 2009. doi: 10.1016/j.anbehav.2008.11.019.

Kaitlin M. Baudier, Meghan M. Bennett, Madeleine M. Ostwald, Sarah Hart, Theodore P. Pavlic, and Jennifer H. Fewell. Age-based changes in kairomone response mediate task partitioning in stingless bee soldiers (tetragonisca angustula). *Behavioral Ecology and Sociobiology*, 74(10), 2020. doi: 10.1007/s00265-020-02902-4.

Jacob D Davidson, Medhavi Vishwakarma, and Michael L Smith. Hierarchical approach for comparing collective behavior across scales: Cellular systems to honey bee colonies. *Frontiers in Ecology and Evolution*, 9(February):1–11, 2021. doi: 10.3389/fevo.2021.581222.

Deborah M. Gordon. The Evolution of the Algorithms for Collective Behavior. *Cell Systems*, 3 (6):514–520, 2016. doi: 10.1016/j.cels.2016.10.013.

Eric Bonabeau, Andrej Sobkowski, Guy Theraulaz, and Jean-Louis Deneubourg. Adaptive Task Allocation Inspired by a Model of Division of Labor in Social Insects. In *Biocomputing and emergent computation: Proceedings of BCEC97*, pages 36–45. World Scientific Press, 1997. ISBN 978-981-02-3262-7.

Chris Tofts. Algorithms for task allocation in ants. (a study of temporal polyethism: Theory). *Bulletin of Mathematical Biology*, 55(5):891–918, 1993.

Robert Jeanne L. Regulation of nest construction behaviour in polybia occidentalis. *Animal Behaviour*, 52(3):473–488, 1996. doi: 10.1006/anbe.1996.0191.

Deborah M. Gordon. *Ants at Work: How an Insect Society is Organized*. Simon and Schuster, 1999. ISBN 978-0-684-85733-6.

Heather J. Goldsby, Anna Dornhaus, Benjamin Kerr, and Charles Ofria. Task-switching costs promote the evolution of division of labor and shifts in individuality. *Proceedings of the National Academy of Sciences of the United States of America*, 109(34):13686–13691, 2012. doi: 10.1073/pnas.1202233109.

Sean O'Donnell and R. L. Jeanne. Methoprene accelerates age polyethism in workers of a social wasp (polybia occidentalis). *Physiological Entomology*, 18(2):189–194, 1993. doi: 10.1111/j.1365-3032.1993.tb00467.x.

Elva J.H. Robinson, Ofer Feinerman, and Nigel R. Franks. Flexible task allocation and the organization of work in ants. *Proceedings of the Royal Society B: Biological Sciences*, 276(1677):4373–4380, 2009b. doi: 10.1098/rspb.2009.1244.

Daniel Charbonneau and Anna Dornhaus. When doing nothing is something. how task allocation strategies compromise between flexibility, efficiency, and inactive agents. *Journal of Bioeconomics*, 17(3):217–242, 2015. doi: 10.1007/s10818-015-9205-4.

John Joseph Valletta, Colin Torney, Michael Kings, Alex Thornton, and Joah Madden. Applications of machine learning in animal behaviour studies. *Animal Behaviour*, 124:203–220, 2017. doi: 10.1016/j.anbehav.2016.12.005.

I. T. Jolliffe. *Principal Component Analysis*. Springer, New York, 2nd ed. 2002 edition edition, 2002. ISBN 978-0-387-95442-4.

Laurens van der Maaten and Geoffrey Hinton. Visualizing Data using t-SNE. *Journal of Machine Learning Research*, 9(86):2579–2605, 2008.

Pavlin G. Poličar, Martin Stražar, and Blaž Zupan. openTSNE: a modular Python library for t-SNE dimensionality reduction and embedding. Technical report, bioRxiv, 2019.

J. M. Jandt and D. M. Gordon. The behavioral ecology of variation in social insects. *Current Opinion in Insect Science*, 15:40–44, 2016. doi: 10.1016/j.cois.2016.02.012.

Linda Karen Garrison, Christoph Johannes Kleineidam, and Anja Weidenmüller. Behavioral flexibility promotes collective consistency in a social insect. *Scientific Reports*, 8(1):15836, 2018. doi: 10.1038/s41598-018-33917-7.

Jennifer M. Jandt and Anna Dornhaus. Bumblebee response thresholds and body size: does worker diversity increase colony performance? *Animal Behaviour*, 87:97–106, 2014. doi: 10.1016/j.anbehav.2013.10.017.

István Maák, Garyk Roelandt, and Patrizia d'Ettorre. A small number of workers with specific personality traits perform tool use in ants. *eLife*, 9:e61298, 2020. doi: 10.7554/eLife.61298.

Ebi Antony George and Axel Brockmann. Social modulation of individual differences in dance communication in honey bees. *Behavioral Ecology and Sociobiology*, 73(4):41, 2019. doi: 10.1007/s00265-019-2649-0.

Yuko Ulrich, Mari Kawakatsu, Christopher K. Tokita, Jonathan Saragosti, Vikram Chandra, Corina E. Tarnita, and Daniel J. C. Kronauer. Response thresholds alone cannot explain empirical patterns of division of labor in social insects. *PLOS Biology*, 19(6):e3001269, 2021. doi: 10.1371/journal.pbio.3001269.

Jacob G. Holland, Shinnosuke Nakayama, Maurizio Porfiri, Oded Nov, and Guy Bloch. Body Size and Behavioural Plasticity Interact to Influence the Performance of Free-Foraging Bumble Bee Colonies. *Insects*, 12(3):236, 2021. doi: 10.3390/insects12030236.

Swetashree Kolay, Raphaël Boulay, and Patrizia d'Ettorre. Regulation of Ant Foraging: A Review of the Role of Information Use and Personality. *Frontiers in Psychology*, 11, 2020.

Nicholas W. Calderone and Robert E. Page. Genotypic variability in age polyethism and task specialization in the honey bee, apis mellifera (hymenoptera: Apidae). *Behavioral Ecology and Sociobiology*, 22(1):17–25, 1988. doi: 10.1007/bf00395694.

Robert E. Page and Gene E. Robinson. The genetics of division of labour in honey bee colonies. *Advances in Insect Physiology*, 23:117–169, 1991. doi: 10.1016/s0065-2806(08)60093-4.

N. W. Calderone and R. E. Page. Evolutionary genetics of division of labor in colonies of the honey bee (apis mellifera). *American Naturalist*, 138(1):69–92, 1991. doi: 10.1086/285205.

Pierre Junca, Lionel Garnery, and Jean Christophe Sandoz. Genotypic trade-off between appetitive and aversive capacities in honeybees. *Scientific Reports*, 9(1):1–14, 2019. doi: 10.1038/s41598-019-46482-4.

Ebi A. George, Ann-Kathrin Bröger, Markus Thamm, Axel Brockmann, and Ricarda Scheiner. Inter-individual variation in honey bee dance intensity correlates with expression of the foraging gene. *Genes, Brain and Behavior*, 19(1):e12592, 2020. doi: 10.1111/gbb.12592.

D. R. Tarpy, R. Nielsen, and D. I. Nielsen. A scientific note on the revised estimates of effective paternity frequency in apis. *Insectes Sociaux*, 51(2):203–204, 2004. doi: 10.1007/s00040-004-0734-4.

J. C. Jones, Mary R Myerscough, Sonia Graham, and Benjamin P. Oldroyd. Honey bee nest thermoregulation: Diversity promotes stability. *Science*, 305(5682):402–404, 2004. doi: 10.1126/science.1096340.

Thomas D Seeley and David R Tarpy. Queen promiscuity lowers disease within honeybee colonies. *Proceedings. Biological sciences / The Royal Society*, 274(1606):67–72, 2007. doi: 10.1098/rspb.2006.3702.

HR Mattila and TD Seeley. Genetic diversity in honey bee colonies enhances productivity and fitness. *Science*, 317(July):362–364, 2007.

Heather R. Mattila, H. Kern Reeve, and Michael L. Smith. Promiscuous honey bee queens increase colony productivity by suppressing worker selfishness. *Current Biology*, 22(21):2027–2031, 2012. doi: 10.1016/j.cub.2012.08.021.

Michal. Woyciechowski and Dawid Moroń. Life expectancy and onset of foraging in the honeybee (apis mellifera). *Insectes Sociaux*, 56(2):193–201, 2009. doi: 10.1007/s00040-009-0012-6.

Hannah Hesselbach, Johannes Seeger, Felix Schilcher, Markus Ankenbrand, and Ricarda Scheiner. Chronic exposure to the pesticide flupyradifurone can lead to premature onset of foraging in honeybees Apis mellifera. *Journal of Applied Ecology*, 57(3):609–618, 2020. doi: 10.1111/1365-2664.13555.

Anna Dornhaus, Scott Powell, and Sarah Bengston. Group Size and Its Effects on Collective Organization. *Annual Review of Entomology*, 57(1):123–141, 2012. doi: 10.1146/annurev-ento-120710-100604.

Noa Pinter-Wollman. Nest architecture shapes the collective behaviour of harvester ants. *Biology Letters*, 11(10):20150695, 2015. doi: 10.1098/rsbl.2015.0695.

Brian R. Johnson. Global information sampling in the honey bee. *Naturwissenschaften*, 95(6):523–530, 2008b.

Tim Gernat, Tobias Jagla, Beryl M Jones, Martin Middendorf, and Gene E Robinson. Automated monitoring of honey bees with barcodes and artificial intelligence reveals two distinct social networks from a single affiliative behavior. *Scientific reports*, 13(1):1541, 2023.

Matthew A.-Y. Smith, August Easton-Calabria, Tony Zhang, Szymon Zmyslony, Jessie Thuma, Kayleigh Cronin, Cassandra L. Pasadyn, Benjamin L. de Bivort, and James D. Crall. Long-term tracking and quantification of individual behavior in bumble bee colonies. *Artificial Life and Robotics*, 2022b. doi: 10.1007/s10015-022-00762-x.

Jacob M. Graving and Iain D. Couzin. VAE-SNE: a deep generative model for simultaneous dimensionality reduction and clustering. preprint, Animal Behavior and Cognition, 2020.

Dagmar Peitsch, Andrea Fietz, Horst Hertel, John de Souza, Dora Fix Ventura, and Randolf Menzel. The spectral input systems of hymenopteran insects and their receptor-based colour vision. *Journal of Comparative Physiology A*, 170:23–40, 1992. doi: 10.1007/bf00190398.

Benjamin Wild, David M. Dormagen, Michael L. Smith, and Tim Landgraf. Learning to embed lifetime social behavior from interaction dynamics. In *Volume 2 of the Proceedings of the joint 12th International Conference on Methods and Techniques in Behavioral Research and 6th Seminar on Behavioral Methods held online May 18-22 2022*. 12th International Conference on Methods and Techniques in Behavioral Research and 6th Seminar on Behavioral Methods, 2022. URL https://doi.org/10.6084/m9.figshare.20066849.v2.

Alexander Mathis, Pranav Mamidanna, Kevin M. Cury, Taiga Abe, Venkatesh N. Murthy, Mackenzie Weygandt Mathis, and Matthias Bethge. DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nature Neuroscience*, 21(9):1281–1289, 2018. doi: 10.1038/s41593-018-0209-y.

Jacob M Graving, Daniel Chae, Hemal Naik, Liang Li, Benjamin Koger, Blair R Costelloe, and Iain D Couzin. DeepPoseKit, a software toolkit for fast and robust animal pose estimation using deep learning. *eLife*, 8:e47994, 2019. doi: 10.7554/eLife.47994.

Talmo D. Pereira, Diego E. Aldarondo, Lindsay Willmore, Mikhail Kislin, Samuel S.-H. Wang, Mala Murthy, and Joshua W. Shaevitz. Fast animal pose estimation using deep neural networks. *Nature Methods*, 16(1):117–125, 2019. doi: 10.1038/s41592-018-0234-5.

Gergely Palla, Imre Derényi, Illés Farkas, and Tamás Vicsek. Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 435(7043):814–818, 2005. doi: 10.1038/nature03607.

Yong-Yeol Ahn, James P. Bagrow, and Sune Lehmann. Link communities reveal multiscale complexity in networks. *Nature*, 466(7307):761–764, 2010. doi: 10.1038/nature09182.

Chris H. Q. Ding and Xiaofeng He. On the equivalence of nonnegative matrix factorization and spectral clustering. In Hillol Kargupta, Jaideep Srivastava, Chandrika Kamath, and Arnold Goodman, editors, *Proceedings of the 2005 SIAM International Conference on Data Mining, SDM 2005, Newport Beach, CA, USA, April 21-23, 2005*, pages 606–610. SIAM, 2005. doi: 10.1137/1.9781611972757.70.

Fei Wang, Tao Li, Xin Wang, Shenghuo Zhu, and Chris Ding. Community discovery using nonnegative matrix factorization. *Data Mining and Knowledge Discovery*, 22(3):493–521, 2011. doi: 10.1007/s10618-010-0181-y.

Xiaohua Shi, Hongtao Lu, Yangcheng He, and Shan He. Community detection in social network with pairwisely constrained symmetric non-negative matrix factorization. In Jian Pei, Fabrizio Silvestri, and Jie Tang, editors, *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2015, Paris, France, August 25 - 28, 2015*, pages 541–546. ACM, 2015. doi: 10.1145/2808797.2809383.

Hsiang-Fu Yu, Nikhil Rao, and Inderjit S. Dhillon. Temporal regularized matrix factorization for high-dimensional time series prediction. In Daniel D. Lee, Masashi Sugiyama, Ulrike von Luxburg, Isabelle Guyon, and Roman Garnett, editors, *Advances in Neural Information*

Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5-10, 2016, Barcelona, Spain, pages 847–855, 2016.

Emily L Mackevicius, Andrew H Bahle, Alex H Williams, Shijie Gu, Natalia I Denisenko, Mark S Goldman, and Michale S Fee. Unsupervised discovery of temporal sequences in high-dimensional datasets, with applications to neuroscience. *eLife*, 8:e38471, 2019. doi: 10.7554/eLife.38471.

Laetitia Gauvin, André Panisson, and Ciro Cattuto. Detecting the Community Structure and Activity Patterns of Temporal Networks: A Non-Negative Tensor Factorization Approach. *PLoS ONE*, 9(1), 2014. doi: 10.1371/journal.pone.0086028.

Pengfei Jiao, Haodong Lyu, Xiaoming Li, Wei Yu, and Wenjun Wang. Temporal community detection based on symmetric nonnegative matrix factorization. *International Journal of Modern Physics B*, 2017. doi: 10.1142/s0217979217501028.

Wenchao Yu, Charu C. Aggarwal, and Wei Wang. Temporally factorized network modeling for evolutionary network analysis. In Maarten de Rijke, Milad Shokouhi, Andrew Tomkins, and Min Zhang, editors, *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining, WSDM 2017, Cambridge, United Kingdom, February 6-10, 2017*, pages 455–464. ACM, 2017a. doi: 10.1145/3018661.3018669.

Wenchao Yu, Wei Cheng, Charu C. Aggarwal, Haifeng Chen, and Wei Wang. Link prediction with spatial and temporal consistency in dynamic networks. In Carles Sierra, editor, *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017*, pages 3343–3349. ijcai.org, 2017b. doi: 10.24963/ijcai.2017/467.

T. Kolda and B. Bader. Tensor Decompositions and Applications. *SIAM Rev.*, 2009. doi: 10.1137/07070111x.

Morten Mørup, Lars Kai Hansen, Sidse Marie Arnfred, Lek-Heng Lim, and Kristoffer Hougaard Madsen. Shift-invariant multilinear decomposition of neuroimaging data. *NeuroImage*, 42(4): 1439–1450, 2008. doi: 10.1016/j.neuroimage.2008.05.062.

Alex H. Williams. Combining tensor decomposition and time warping models for multi-neuronal spike train analysis. *bioRxiv*, page 2020.03.02.974014, 2020. doi: 10.1101/2020.03.02.974014.

Michelle M. Elekonich and Stephen P. Roberts. Honey bees as a model for understanding mechanisms of life history transitions. *Comparative Biochemistry and Physiology. Part A, Molecular & Integrative Physiology*, 141(4):362–371, 2005. doi: 10.1016/j.cbpb.2005.04.014.

Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron C. Courville, and Yoshua Bengio. Generative adversarial networks. *Commun. ACM*, 63 (11):139–144, 2020. doi: 10.1145/3422622.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In Yoshua Bengio and Yann LeCun, editors, *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015.

Xuan Vinh Nguyen, Julien Epps, and James Bailey. Information theoretic measures for clusterings comparison: is a correction for chance necessary? In Andrea Pohoreckyj Danyluk, Léon Bottou, and Michael L. Littman, editors, *Proceedings of the 26th Annual International Conference on Machine Learning, ICML 2009, Montreal, Quebec, Canada, June 14-18, 2009*, volume 382 of *ACM International Conference Proceeding Series*, pages 1073–1080. ACM, 2009. doi: 10.1145/1553374.1553511.

Da Kuang, Sangwoon Yun, and Haesun Park. SymNMF: nonnegative low-rank approximation of a similarity matrix for graph clustering. *Journal of Global Optimization*, 62(3):545–574, 2015. doi: 10.1007/s10898-014-0247-2.

Jingu Kim, Yunlong He, and Haesun Park. Algorithms for nonnegative matrix and tensor factorizations: a unified view based on block coordinate descent framework. *Journal of Global Optimization*, 58(2):285–319, 2014. doi: 10.1007/s10898-013-0035-4.

Michael L Smith, Jacob D Davidson, Benjamin Wild, David M Dormagen, Tim Landgraf, and Iain D Couzin. Honeybee lifetime tracking data 2018, 2022c.

Luca Insolia, Roberto Molinari, Stephanie R. Rogers, Geoffrey R. Williams, Francesca Chiaromonte, and Martina Calovi. Honey bee colony loss linked to parasites, pesticides and extreme weather across the United States. *Scientific Reports*, 12(1):20787, 2022. doi: 10.1038/s41598-022-24946-4.

Tim Landgraf, Michael Oertel, Andreas Kirbach, Randolf Menzel, and Raúl Rojas. Imitation of the honeybee dance communication system by means of a biomimetic robot. In Tony J. Prescott, Nathan F. Lepora, Anna Mura, and Paul F. M. J. Verschure, editors, *Biomimetic and Biohybrid Systems*, pages 132–143, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg. ISBN 978-3-642-31525-1.

Phoebe A. Koenig, Michael L. Smith, Logan H. Horowitz, Daniel M. Palmer, and Kirstin H. Petersen. Artificial shaking signals in honey bee colonies elicit natural responses. *Scientific Reports*, 10(1):3746, 2020. doi: 10.1038/s41598-020-60421-8.

Cassandra Uthoff, Masun Nabhan Homsi, and Martin von Bergen. Acoustic and vibration monitoring of honeybee colonies for beekeeping-relevant aspects of presence of queen bee and swarming. *Computers and Electronics in Agriculture*, 205:107589, 2023. doi: 10.1016/j.compag.2022.107589.

Takao Kondo, Nicholas F. Tsinoremas, Susan S. Golden, Carl Hirschie Johnson, Shinsuke Kutsuna, and Masahiro Ishiura. Circadian Clock Mutants of Cyanobacteria. *Science*, 266(5188):1233–1236, 1994. doi: 10.1126/science.7973706.

Deborah Bell-Pedersen, Vincent M. Cassone, David J. Earnest, Susan S. Golden, Paul E. Hardin, Terry L. Thomas, and Mark J. Zoran. Circadian rhythms from multiple oscillators: lessons from diverse organisms. *Nature Reviews. Genetics*, 6(7):544–556, 2005. doi: 10.1038/nrg1633.

Guy Bloch, Brian M. Barnes, Menno P. Gerkema, and Barbara Helm. Animal activity around the clock with no overt circadian rhythms: patterns, mechanisms and adaptive value. *Proceedings of the Royal Society B: Biological Sciences*, 280(1765):20130019, 2013. doi: 10.1098/rspb.2013.0019.

Zusammenfassung der Ergebnisse

In dieser Dissertation wird das kollektive Verhalten von sozialen Tieren untersucht, wobei Honigbienenkolonien als Modellsystem dienen. Die Arbeit behandelt zwei grundsätzliche Fragestellungen: die Ursprünge des kollektiven Verhaltens in sozialen Tiergesellschaften und die Rolle individueller Verhaltensweisen in dem kollektiven Verhalten. Vier publizierte Arbeiten tragen zur Arbeit bei:

Die erste Arbeit stellt eine auf Machine Learning basierte Methode zur Verfolgung einzelner Honigbienen in einer Kolonie über ihre Lebenszeit vor. Dies ermöglicht die Analyse einer großen Anzahl von Individuen und ihrer Interaktionen und bildet die Grundlage für die nachfolgenden Arbeiten.

Die zweite Arbeit analysiert in das Verhältnis zwischen dem individuellem und kollektiven Verhalten in der Kolonie. Es führt eine Methode namens *network age* ein, mit dem die Aufgabenzuteilung, Überleben und Aktivitätsmuster von Individuuen auf Grundlage ihres sozialen Netwzwerkes vorhergesagt werden kann. Die Ergebnisse zeigen, dass soziale Netzwerke innerhalb der Kolonie die individuelle Entwicklung und das Verhalten erheblich beeinflussen.

Die dritte Arbeit konzentriert sich auf die Verhaltensvariationen einzelner Honigbienen. Mittels quantitativer Analysen wird untersucht, wie individuelle Verhaltensweisen zur Gesamtfunktion der Kolonie beitragen. Die unterschiedlichen Verhaltensmuster von Bienen in innerhalb eins Tages und über ihr gesamtes Leben hinweg werden analysiert.

Die vierte Arbeit stellt eine neue Methode vor, die die Analyse von Daten über mehrere Kolonien über sehr lange Zeiträume ermöglicht. Die Methode ist ein temporales Matrixfaktorisierungsmodell, mit dem die funktionalen Rollen von Individuen in sozialen Gruppen zu identifiziert werden können. Dieser Ansatz ermöglicht die quantitative Analyse von Verhaltensweisen und Rollen von Individuen in komplexen sozialen Systemen über Zeit und Raum hinweg.

Zusammenfassend entwickelt die Arbeit neuartige Methoden zur Verfolgung, Analyse und zum Verständnis individueller und kollektiver Verhaltensweisen in Honigbienenkolonien. Diese Beiträge tragen nicht nur Erforschung des Verhaltens sozialer Insekten bei, sondern bieten auch eine Grundlage für zukünftige Forschungen zum Verständnis kollektiven Verhaltens in komplexen Systemen.