CellPress OPEN ACCESS

Trends in Ecology & Evolution

Special issue: Animal behaviour in a changing world

Opinion

Emerging technologies for behavioral research in changing environments

Iain D. Couzin[1,2,*,@] and Conor Heins[1,2]

The first response exhibited by animals to changing environments is typically behavioral. Behavior is thus central to predicting, and mitigating, the impacts that natural and anthropogenic environmental changes will have on populations and, consequently, ecosystems. Yet the inherently multiscale nature of behavior, as well as the complexities associated with inferring how animals perceive their world, and make decisions, has constrained the scope of behavioral research. Major technological advances in electronics and in machine learning, however, provide increasingly powerful means to see, analyze, and interpret behavior in its natural complexity. We argue that these disruptive technologies will foster new approaches that will allow us to move beyond quantitative descriptions and reveal the underlying generative processes that give rise to behavior.

Highlights

Behavioral change is essential to facilitate effective response to environmental change, yet the multiscale nature of behavior, and difficulties in establishing how animals perceive and respond to their environment, have limited our ability to achieve a quantitative and predictive understanding of animal behavior in the wild.

Major, ongoing technological advances, especially in machine learning, provide increasingly powerful means to acquire, identify, and interpret animal behavior and to do so in ways that embrace its natural complexity.

We need to move beyond statistical descriptions of behaviors to approaches that allow us to also infer the generative processes by which behaviors arise.

A rigorous and predictive science of behavior will likely prove essential for the development of effective conservation strategies and policies in the face of increasingly-rapid anthropogenic change.

Technology offers new ways to study behavior

Animal behavior impacts biological processes across vast scales, from the genetic, endocrine, and neural interactions within organisms, to the interactions among organisms, and the functional complexity of groups, populations, and ecosystems. Thus, understanding the causes and consequences of behavior and how it impacts eco-evolutionary processes, requires quantitative means to assess its inherently multiscale and dynamical nature.

It is the complex and unpredictable nature of behavior that so engages the mind when observing animals and yet also makes it so challenging to study. This has resulted in a bias towards the study of a subset of 'tractable' behaviors (i.e., those that are amenable to human observation and/or interpretation). Emerging and evolving technologies, however, are progressively allowing researchers to develop and utilize approaches that do not ignore the complexities and dynamical relationships of behavior in the natural world, but rather actively embrace them. Such approaches, which we discuss here, will facilitate the study of a much broader range of species in natural, or naturalistic, environments. In doing so the behavioral sciences will increasingly be able to contribute to efforts to conserve biodiversity in changing environments.

Methods to locate, identify, and estimate the body posture of animals

Traditionally, obtaining behavioral data, such as movement and/or identifiable (i.e., stereotyped) behaviors, including grooming, agonistic interactions, and feeding, has been achieved via human observation, either directly, or from video. In recent years, however, advances in machine learning, and particularly in deep learning (Box 1), provide tools that greatly complement, or even replace, the observational stage of video analysis as well as identification of behaviors, alleviating important bottlenecks in behavioral research.

In order to track animals in a scene, they must first be detected. For simple scenes, as in many laboratory settings, classical computer vision processing steps, such as **background subtraction** (see Glossary) and **blob detection**, can be employed to **segment** animals from their background. If

[1]Department of Collective Behaviour, Max Planck Institute of Animal Behavior, Konstanz, Germany
[2]Centre for the Advanced Study of Collective Behaviour & Department of Biology, University of Konstanz, Germany

*Correspondence: icouzin@ab.mpg.de (I.D. Couzin).
[@]Twitter: @icouzin

Box 1. Machine learning and deep learning for behavioral studies

Alongside the rapid growth in computing resources and the volumes of data provided by new acquisition methods, advances in machine learning have revolutionized the way behavioral biologists both acquire and interpret behavioral data [52–54]. One of the most successful approaches is deep learning, artificial neural networks with multiple 'deep' layers, which can automatically learn arbitrarily complex input–output relationships directly from data [2,3]. In the context of computer vision applications like object detection, segmentation, or tracking, deep learning can be combined with large labeled datasets to automatically detect and infer the identity of, for example, an individual animal from high-dimensional data sources like video [10–14,55] or audio [55,56]. In particular, these networks automatically learn the disparate statistical features of the data (changes in lighting, brightness, pose) and thus are much more robust than classical computer vision pipelines, typically consisting of cascades of processing steps like filters and feature-detectors. One major drawback of deep learning, and particularly supervised deep learning, is that it often requires large, labeled datasets to sufficiently train.

On the side of interpretation, unsupervised and self-supervised learning techniques can be used to automatically identify latent statistical structure in behavior; this can be combined with clustering and domain expertise to identify and study the multiscale structure of behavior. This approach differs from classical approaches in ethology and life sciences in general, which often rely on hand-crafted features and thresholds to segregate behavior into different categorical classes, which themselves come with inbuilt assumptions about which aspects of behavior are meaningful. Instead, unsupervised learning approaches, especially when combined with deep neural networks, make minimal assumptions about the causal structure of behavior and thus can maximally extract information from the raw data. Behavioral biologists can leverage these methods in combination with domain expertise to discover factors or features of variation that might remain undetectable by conventional, human-biased methods [15–19,25–33]. Although unsupervised learning does not require labels to train, the most powerful unsupervised learning techniques (which use deep neural networks) often also require large amounts of training data.

animals are in complex scenes, such as those with unpredictable or changing light conditions, deep neural networks, like **convolutional neural networks (CNNs)** [1–3] or **vision transformers (ViTs)** [4,5], are more appropriate, but these need to be trained with human-provided labels (i.e., a dataset of 'correct answers', e.g., labelled images, need to be provided to allow the algorithm to learn). Since generating large, labelled datasets is time-consuming, methods like model-assisted labelling and **active learning** have been developed to increase the efficiency of training and can thus greatly reduce this cost [6,7]. Fully automated methods, known as **video-object segmentation** and **tracking (VOS/VOST)**, are also becoming increasingly powerful [8,9]. These algorithms automatically locate nonrigid targets in complex scenes, a ubiquitous challenge for self-driving cars, for example, and offer great promise for behavioral applications.

Once animals have been detected in a scene, machine learning can be employed to recognize species and even individuals. If lighting conditions can be controlled, multi-animal tracking solutions like idtracker.ai [10] and TRex [11] can automatically learn (and thus maintain) the visual identities of up to 100 unmarked conspecifics and with diverse species (e.g., insects, fish, and mammals). Presently, individual recognition tasks in the wild require extensive training [12–14], but the progressive development of algorithms capable of compensating for uneven and changing light conditions makes 'field-ready' automated solutions likely in the near future.

While analysis of the movement of animals is sufficient to infer certain behaviors, such as escape and pursuit, a great deal of additional behavioral information can often be obtained by considering the time-varying body posture (pose) of animals. Many behaviors, such as grooming, feeding, and ritualized displays are recognizable due to stereotyped, and often repeated, sequences of postural change. Again, machine learning approaches (first developed to estimate human pose) allow pose estimation of unmarked animals: both 2D [15–17] and multicamera 3D [18] methods have been developed, all of which track **keypoints** on the body (following training via annotation) from which postures are estimated (Figure 1B). Increasingly, however, algorithms are capable of inferring 3D body posture from a single camera perspective when ground truth data are available [19], and even when not, such as by matching the 2D projection of an animal's

Glossary

Acoustic video: a method of imaging that transmits sound pulses and converts the returning echoes into digital images, similar to medical ultrasound. This has advantages over optical imaging in aquatic environments since it can work in the dark, or in turbid water, and can penetrate complex structures, such as allowing tracking of individuals in schools of fish where central individuals would be occluded if using optical video.
Active learning: a class of methods in machine learning where the learning algorithm actively selects data points to learn from. In some architectures, the algorithm may even automatically label queried data with predicted labels or outputs.
Active sensing: the ability to gather sensory information by actively changing the position or properties of sensors. Illustrative examples of active sensing are echolocation and saccadic eye movements.
Background subtraction: a processing step common to many computer vision pipelines, and particularly suited to detecting moving objects, background subtraction typically involves obtaining an image of the background of a scene (i.e. without animals present) and then subsequently calculating the difference (for each pixel) between that reference image and subsequent images in a video sequence.
Bio-loggers: a form of remote measurement that involves the use of animal-borne sensors (also termed 'tags') to track quantities like location, speed, and others. Bio-loggers are used in many fields of study, from animal biology to wildlife conservation and ecology.
Black-box: machine learning models that lack an interpretable internal structure are often described as 'black box' because their predictions may be accurate, but the method by which the algorithm made a decision or prediction is not understandable by a human interpreter. Many deep neural networks are considered 'black-boxes' because they often have many interdependent parameters whose values are difficult to intuitively relate to the network's inputs or outputs. The field of explainable machine learning tries to design algorithms that do not suffer from a lack of interpretability.
Blob detection: a set of computer vision techniques that identify regions in

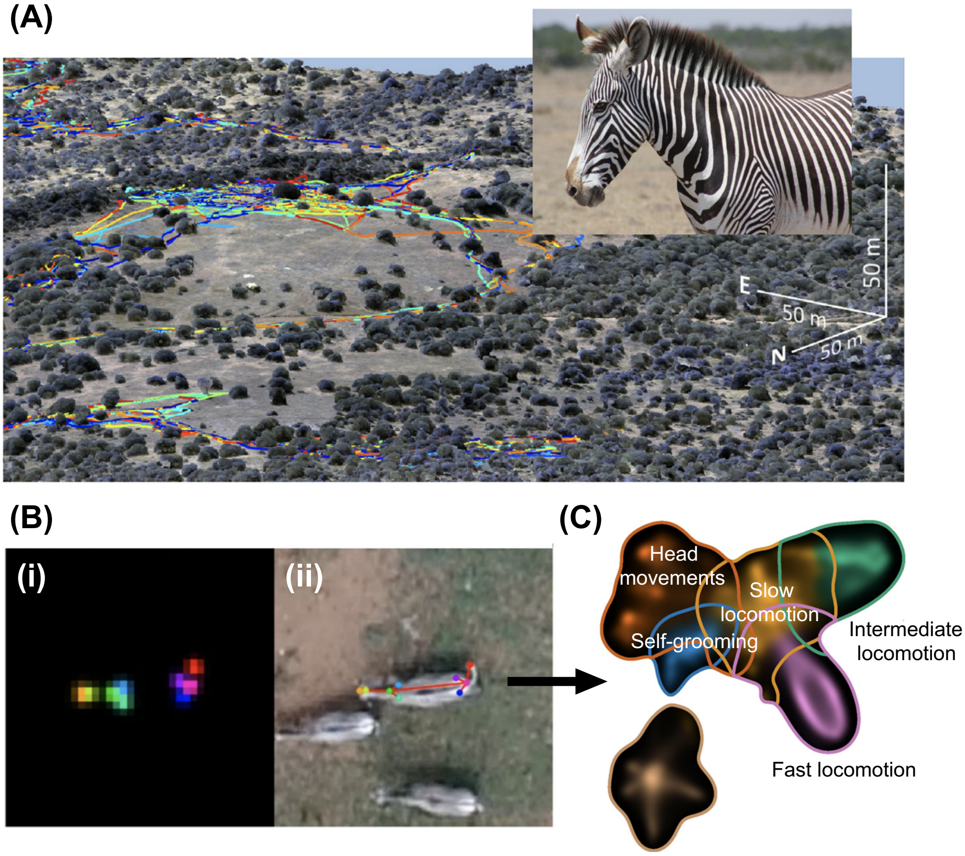

Figure 1. Analyzing behavior from drone-based video. (A) Structure-from-motion algorithms allow detailed 3D environmental mapping from drone-based video and geo-referenced paths of animals [37] such as Grevy's zebra (*Equus grevyi*, inset), obtained using machine learning (Box 1). (B,i) In addition to locating animals in a scene, machine learning can estimate the locations of markerless keypoints. (B,ii) This allows estimates of body posture over time [15]. (C) Dimensionality reduction and clustering methods distill multiscale structure from postural data to identify specific behaviors, as well as transitions between them [25–33,52–54,60]. These panels represent an illustrative example only and are modified from [15,23,37,60].

outline and texture to a 3D articulated model of the species [20], or by employing **neural radiance fields** to infer nonrigid 3D structures, including animals' bodies, directly from video sequences [21].

Before turning to how such data can be analyzed to infer behavior, it is worth noting that there exist additional means of acquiring video information aside from visible light. Similar machine learning approaches can be applied to **thermal imaging**, **LiDAR** depth imaging (where recent developments in solid state devices offer great promise for behavioral research), and in the aquatic environment via acoustic imaging (including **acoustic video** [22]). In addition, machine learning now makes possible a very different approach for remote localization and behavioral inference: **seismic sensing**. This has recently been shown to allow automated species identification, localization, and even quantification of certain behaviors, offering a means to conduct long-term studies in appropriately instrumented areas (Box 2). These advances could greatly inform, and be complemented by, the more detailed, yet inherently focal, analyses achievable from video sources in the field, such as from drones [23].

an image whose pixels are counted as 'similar' (in terms of intensity, color, texture, etc.) and typically in close proximity to other such pixels. Blob detection is commonly used to identify initial candidate objects ('blobs') that may then be sorted into objects after additional filtering and analysis.

Causal inference: an interdisciplinary field and arguably a subdiscipline of statistics that hopes to disentangle the causal relations between at least two random variables (i.e., to enable statistically valid claims about how changes in one variable depend on changes in the other). Many 'causally agnostic' statistical inference techniques (e.g., ordinary least squares regression) do not provide statistical guarantees of causal inference, but only inference of statistical association or correlation.

Clustering: clustering algorithms partition a set of datapoints into discrete clusters or groups, with the underlying assumption that points within a cluster are more similar to each other than to those in other clusters. The choice of similarity metric used for clustering depends on the dataset and the task at hand. Along with dimensionality reduction, cluster analysis is common in exploratory data analysis pipelines, where the goal is often to extract simple, interpretable insights from complex high-dimensional datasets.

Collective memory: in the context of collective animal behavior, collective memory refers to the storage of information by a social group on a longer timescale than is encoded, or detectable, by an individual member of the group.

Convolutional neural networks (CNNs): a class of artificial neural networks that are particularly well-suited to processing datasets with space- or time-invariant features (e.g., images, audio). CNNs are distinguished by a special 'shared connectivity' in their network architecture, where the input data are processed through cascades of filters (or convolutions) that learn to detect higher-order features across different parts of the input; applications include tracking and/or classifying objects in video and speech recognition.

Dimensionality reduction: many modern data-heavy disciplines (especially in the natural sciences) regularly handle large amounts of data that are comprised of many distinct but potentially correlated channels; in this setting, dimensionality reduction is often

Box 2. Seismic sensing can allow species identification, localization, and behavioral classification in terrestrial environments

With major benefits including its noninvasive nature, ability to record for long periods of time, and with seismometer costs falling, ground-based vibrational analyses may be uniquely well-suited to wildlife monitoring in changing environments. For localization via triangulation, sufficiently strong signals must be present at a sufficient number of receivers and thus the scale and accuracy of localization will typically depend on the size of the animals, the density of the seismic array, and the suitability of the substratum for transmitting seismic information. A recent study in Africa [56] demonstrated proof of principle using machine learning to demonstrate accurate discrimination of multiple species from seismic cues alone, including elephants (*Loxodonta africana*), giraffe (*Giraffa reticulata*), zebra (*Equus quagga*), hippo (*Hippopotamus amphibius*), hyena (*Crocuta crocuta*), leopard (*Panthera pardus pardus*), and dik-dik (*Madoqua kirkii*) (Figure I). In addition, they demonstrated accurate localization of elephants as well as their gait (walking or running) and whether they were actively communicating via 'rumbles'. As with other instrumented methods, improved energy efficiency and reduced costs will provide the opportunity for increasingly dense, autonomous arrays to be installed with increased ease (below the surface and with remote connectivity), allowing for continued improvements in performance, and thus the diversity of terrestrial species that could be monitored via seismic cues.

Seismic approaches may also prove valuable to the study of animal behavior in the subterranean environment; which, despite its ecological importance, has received comparatively very little attention, due to the practical limitations of conventional approaches. More broadly, our understanding of animal behavior is heavily biased towards those species which have been easiest to observe and great benefits could come from encouraging researchers to develop and apply emerging technologies to the study behavior in those environments that have, to-date, tended to be avoided.

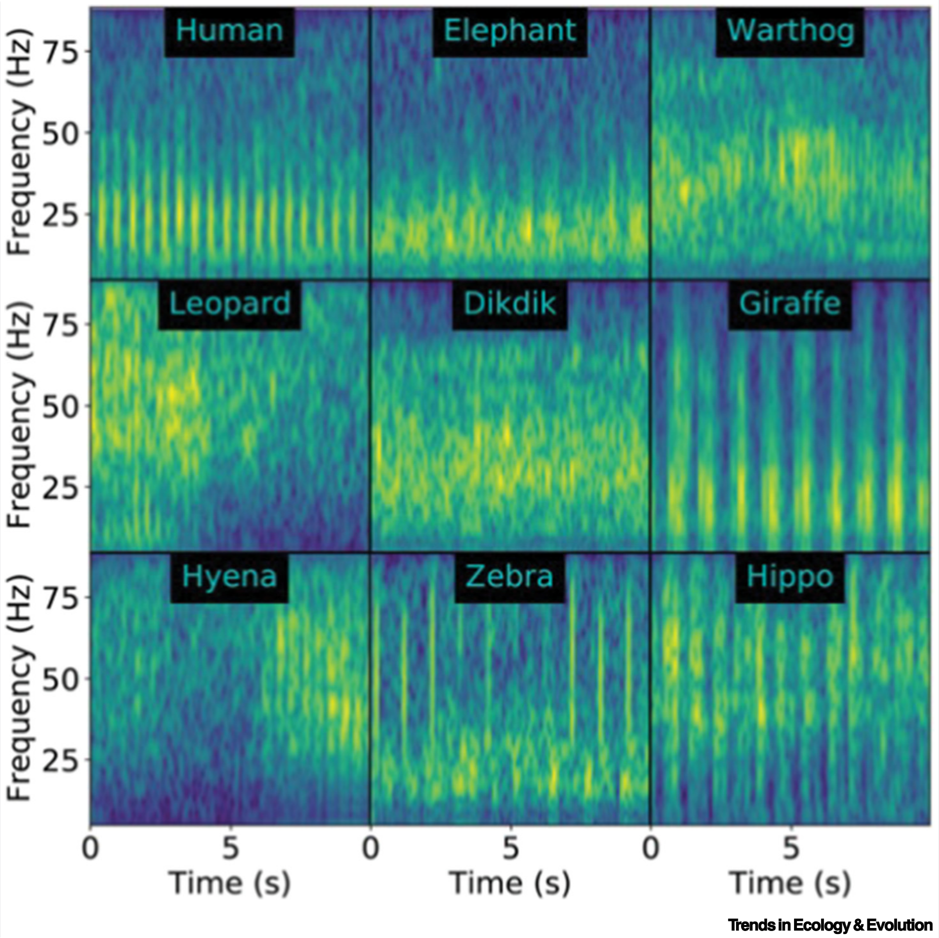

Figure I. Machine learning achieves accurate species recognition from seismic spectrograms, from [56].

used to process such large, multidimensional datasets by reducing them to a smaller set of variables (dimensions) that still maintain sufficient information for the task at hand.
Explainable machine learning: also known as explainable artificial intelligence (or XAI), explainable machine learning is a field focused on creating machine learning models whose decisions and predictions are understandable by humans. It aims to show the user what features in the input data (or transformations thereof) are relevant to its decision-making. This allows more transparency with regard to the assumptions and decisions of otherwise black-box algorithms.
Keypoints: in the context of pose-estimation and kinematics, keypoints are coordinates that correspond to fixed positions on the body and are tracked over time. Keypoints are often targeted to a small set of particular body parts like joints, whose relative changes over time can be used to identify characteristic movement patterns such as corresponding to locomotory, feeding, or mating behaviors.
Kinematics: in the context of movement analysis, kinematics describes the geometry and dynamics of postural variables, typically focusing on the movements of a single individual. When measuring animal behavior, these postural variables might be the output of a pose-estimation algorithm, consisting of measurements like the relative positions or angles of postural keypoints (e.g., the positions of joints over time).
LiDAR: an acronym for 'light detection and ranging' or 'laser imaging, detection, and ranging', LiDAR is a method for using time of flight of reflected laser light to measure ranges (distances) between the sensor and a surface.
Movement ecology: an interdisciplinary research paradigm that applies statistical inference techniques to movement data, to study the drivers and impacts of organismal movement, ranging from the movements of microorganisms to large-scale migratory patterns.
Neural radiance fields: a novel neural network architecture that automatically learns to visually render volumetric (i.e., 3D) scenes by training only on a sparse set of 2D images of the scene viewed at particular distances and angles.
Resource selection analysis: a set of techniques used in movement ecology that rely on statistical models to estimate the effects of variables-of-interest on

Furthermore, for the study of many airborne animals, non-echolocating aquatic animals, and animals that range widely, researchers must catch and instrument animals with **bio-loggers** in order to track them and infer their behavior (Box 3). Even where remote methods are possible, on-animal instrumentation can provide highly valuable synergistic data, such as regarding physiology [24].

From postural kinematics to behavior

Analysis of the **kinematics** of movement and body posture allows considerable insight into the behavior of many animals, especially since automated methods can be employed to seek multiscale patterns indicative of specific behavioral actions. This process relies on **dimensionality reduction** and **clustering**, typically employing methods from machine learning and probabilistic inference [25–28]. These analyses directly learn compressed representations of the structure of the data without human supervision, hence they are often termed 'unsupervised' or 'self-supervised' algorithms. However, as with any statistical method, they are still biased by the particular data used to train them (e.g., sampling biases) and representational assumptions. Barring these biases, these methods are relatively agnostic to the data provided and, if applied correctly, are highly effective at finding whatever structures may exist, irrespective of the scale(s) over which they occur. Thus, they are well-suited to the analysis of time-varying behavioral data, such as of body posture (but also other data types, such as acoustic data) where they can reveal repeated, stereotyped kinematic 'motifs' (analogous to 'syllables' in acoustic communication) (Figure 1C), as well as how sequences of motifs are strung together to create more complex behaviors.

In doing so it has been shown that even relatively simple animals, such as the fruit fly (*Drosophila melanogaster*), exhibit over 100 stereotyped motor behaviors, such as types of grooming, locomotion, and limb-, wing-, and proboscis-extension states [29]. In other species, such as juvenile zebrafish (*Danio rerio*) and rodents, some behaviors tend to exist as relatively discrete states, whereas others are found to exist on a multiscale continuum [30–32]. Analysis of the transitions between behavioral states/motifs in insects [29], fish [31], and rodents [30,32], reveals a hierarchically organized, modular, structure (thus animals may have to transition through some behavioral states to reach others). This may provide important insights into how the brain represents and organizes behavior both within and across species [33].

Researchers are only just beginning to apply this methodology to the study of behavior in the wild, such as the 3D kinematics of locomotion in cheetahs [34]. However, this is sure to change, with a

Box 3. Instrumenting animals with bio-loggers is an important means to obtain behavioral information from many wild animals

A complimentary approach, which is sometimes the only feasible means to obtain behavioral data, is to instrument animals with 'bio-logging' devices. These typically include GPS (although reverse-GPS, such as the ATLAS system [57] offers energy and weight savings, albeit at the cost of a limited range) and increasingly, not least due to increased efficiency of electronics and improved battery and solar cell technology, additional sensors. These include gyroscopes, inertial measurement units, and magnetometers which provide subsecond movement information [24,29,38,45,46,55,57,58], as well as data relevant to the decisions made by animals such as on-board acoustic [40,58], and sometimes video, recordings. Researchers are actively working on how to optimally integrate such complimentary data streams, as well as how to best apply machine learning (Box 1) to associate the kinematic sequences obtained from bio-loggers to specific, validated behaviors; thus allowing post-hoc, or even on-board, behavioral categorization [57–59]. While much less rich than posture analyses, this approach will increasingly inform us about important features of behavior in the wild.

In the aquatic environment, where GPS can only be employed near the surface, inertial measurements, combined with depth sensors, provide extremely valuable information about subsurface movements. Furthermore, even very small animals, such as fish, can be instrumented with acoustic transmitters ('pingers'), which due to the considerable speed and range of acoustic propagation underwater, allows relatively long-range acoustic sensing and thus triangulation via receivers in known locations [58]. Similar methodology also allows triangulation of naturally vocalizing (and thus not necessarily instrumented) animals, such as cetaceans, where the content of the vocalization, increasingly decoded using machine learning methods, can be highly informative about behavior [59].

spatial movement patterns, like habitat selection or the consumption of nutritional and energetic resources in space. These covariates often include environmental features but can also incorporate the movements of conspecifics; more sophisticated resource selection techniques take into account the feedback between the drivers of animal movement and the movement itself.

Segment: image segmentation is a category of computer vision methods that segregate pixels in an input image, often spatially proximal ones, into different groups or 'segments' based on some similar property relevant to the given computer vision task. For example, many blob detection methods are examples of segmentation because they assign groups of neighboring pixels into discrete segments that correspond to objects or blobs.

Seismic sensing: a recording method that measures ground-vibrations and has been primarily used in the geophysics and earth sciences, for example, to detect and quantify geophysical phenomena like earthquakes. See Box 2 for an example of multispecies detection and classification based on seismic data.

Sensorimotor transformation: this describes the transformation of information from sensory modalities to motor outputs. Animals with nervous systems implement these transformations by means of various sequential and recurrent computations implemented in distributed neural circuits.

Structure-from-motion (SfM): refers to the problem of reconstructing a 3D representation of an object or surface from a collection of 2D images over time (which must be taken at slightly different angles in a sequence). In the context of inferring the 3D surface of natural environments, SfM algorithms can be applied to sequences of, for example, drone-based video to map the topography and textural features of the environment.

Thermals: columns of warm air rising from the earth's surface; can be used as a source of lift generation for flying animals such as migratory birds.

Thermal imaging: a method of imaging that detects natural infrared radiation emitted from surfaces and objects. Due to the increase in emission of thermal radiation with temperature, thermal imaging can be used to visualize relative variations in temperature.

number of benchmark datasets for animal pose estimation in the wild having recently been made available [35].

Measuring changing environments

Technological advances in remote sensing, from imaging by drones and manned aircraft, to satellite-based acquisition, provide a wealth of relevant information about the natural environments in which animals are embedded, including estimates of 3D topography, vegetation structure and nutritional quality, soil moisture, precipitation, winds, temperature gradients, currents, etc. [36]. Such data are invaluable for relating local structure, and environmental fluctuations, to animals' decision-making.

Since many decisions depend on fine-scale environmental features, drone-based imaging (e.g., video or LiDAR) is likely to be particularly important for many behavioral studies. Video alone is sufficient to establish [via **structure-from-motion (SfM)** algorithms] detailed mapping of the 3D topography of the environment, from which other features, such as vegetation, trails, water bodies, rocks, etc. can be identified (Figure 1A). In combination with tracking and body posture estimation, as described earlier, researchers can computationally reconstruct the animals' visual perspectives and thus obtain valuable information about their time-varying perspective of contemporaneous social and physical environmental features [37].

In some cases it is possible to utilize the animals themselves as a distributed array of living environmental sensors [38]. The motion of birds contains information about the 3D structure of **thermals** [39] and on-board acoustic recordings can be used to infer the dynamic social environment of bats [40]. On-board video (e.g., optical/thermal/LIDaR) can similarly provide important information.

Many animals are not only influenced by environmental features, but in turn their behavior shapes their environment. For example, grazing animals consume, as well as potentially trample, vegetation. This influences the feeding and movement decisions of others, resulting in complex feedbacks that can include the emergent establishment of trail systems that connect foraging areas and watering holes across substantial spatial scales. Remote sensing could allow such poorly understood, yet clearly ubiquitous forms of multispecies, and likely trans-generational, **collective memory** and decision-making to be studied in diverse ecosystems.

From data to decision-making

Once we have obtained environmental data, which for field studies will often be highly incomplete, we face the challenge of establishing how the animals perceive, and respond to, their environment. This is particularly problematic since behavioral decision-making is also a multiscale process and typically depends on animals integrating information from multiple modalities; vision, olfaction, tactile, etc. Not only will many of these streams be inaccessible to measurement by researchers, but even when they are present, we often lack a sufficiently precise understanding of the sensory physiology of focal species to be able to infer how the animals perceive their environment. In addition, each sensory modality depends on different physical properties, each of which is susceptible to changes in the environment: wind or currents may strongly impact acoustic and olfactory information, but not visual information, whereas fog or turbidity impact the scattering and attenuation of light to a much greater degree than it impacts acoustic propagation, and so on. Furthermore, decision-making depends on individuals' internal physiological states, such as stress, hunger, and previous experience (e.g., learning). Thus, there is a growing need for the integration of new technologies that can measure the physiology of the animals under study [24].

A powerful suite of statistical tools have been developed, especially in the field of **movement ecology** [41], to describe how movement covaries with environmental features (e.g., **resource**

Video-object segmentation and tracking (VOS/VOST): video-object segmentation and video-object tracking, respectively, refer to the problems of identifying (single or multiple) objects from video and tracking their identities over time. These can also mutually inform one another. Solutions for VOS/VOST have application in a wide range of disciplines, from safe autonomous driving to tracking diverse species in the wild.
Vision transformers (ViTs): a particular type of transformer designed for image data. Transformers are a powerful class of artificial neural networks that have achieved state-of-the-art performance in language modeling and sequence prediction tasks. Transformers are powered by a mechanism called 'attention' that renders them particularly effective. Recently, this architecture has been applied to image data in the form of ViTs, which have emerged as a competitive alternative to convolutional neural networks (CNNs) in image- and video-related processing tasks.

selection analysis [42]). The vast majority of this work focuses on describing the statistical properties of movement, but relatively little work, by contrast, considers how such statistics arise; that is, the internal (biological) and external (environmental) processes by which they are generated. For example, it is known that many animals employ both local 'area-restricted' search as well as switching to a much broader 'global search' when seeking food in unpredictable environments [43]. A traditional movement ecology approach would be to model this as a stochastic switching process, drawing from different turning angle distributions to represent each search type. While this correctly captures the observed statistics of animal search, these purely descriptive approaches do not necessarily inform us why these patterns arise. A complimentary, explanatorily motivated approach could consider how such statistical properties arise from underlying generative biological processes or normative principles. For the specific example of area-restricted and global search, it has been found that this behavior naturally emerges if animals move to constantly minimize informational uncertainty in unpredictable environments (a form of **active sensing**) and further that such maximally informative foraging can be implemented by a relatively simple underlying computational substrate [44].

In reality, however, it is seldom possible to know the 'currency' that motivates animal decisions. Movement may, for example, be heavily constrained by the energetic costs of moving, resulting in birds moving to exploit thermals [45,46] or terrestrial animals following trails. Furthermore, multiple factors are likely at play simultaneously: trails provide decreased resistance to movement, but they also encode valuable information about the environment, thus their utility will likely depend on both energetic and informational factors. Similarly, many bird species both exploit, and provide, social information when moving within, and between, thermals [39,45], which directly impacts energetics [39,45,46], but also their capacity to learn navigational routes [47].

From prediction to causal inference

If we can apply machine learning to create useful compressed representations of animals' movements, and we can compute a similarly compressed representation of important sensory inputs, it is also then possible to employ information-based measures to infer how sensory inputs causally impact behavior. The brain has, after all, evolved to perform dimensionality reduction, to make sense of complex sensory streams and translate them to relatively low-level behavioral decisions. Doing so in practice, however, proves to be very hard.

If our focus is to predict what an animal will do, such as where it will move when experiencing novel, or changing, environmental conditions, we can train a **black-box** machine learning algorithm to optimally predict animals' movement. Similar to the problems of animal tracking and body posture reconstruction, predictive accuracy matters; but an understanding of how the neural network solves this problem is usually irrelevant to the goal of building a powerful predictive model.

However, if we want to understand how an animal perceives and makes decisions within its environment, this level of behavioral description is insufficient. To move from prediction to causal explanation, we must consider the level of biological organization relevant to understanding. For instance, we could leverage the ability of deep neural networks to learn arbitrary input–output mappings and dissect the trained network's learned representations as a sandbox system for representing the 'space of computations' sufficient to perform a given task. In combination with known biological constraints for a given system, this neural network 'dissection' approach could be used to infer the possible linear/nonlinear functions involved in transforming sensory information to motor outputs. In contrast to other approaches, under this scheme the deep neural network would not be interpreted as a one-to-one representation of the underlying 'neural wetware' used by a given system; instead, the low-dimensional space of functions learned by the network could be used as a hypothesis-generating tool to identify possible **sensorimotor transformations** used by the organism at the algorithmic or computational levels [48]. **Explainable machine learning** and **causal inference** [49–51] could, therefore, provide vital toolkits to achieve an improved computational understanding of animal behavior, without committing to the biological details of how real neural circuitry implements a given computation.

Concluding remarks

Assessing behavior in changing environments is more important than ever in the face of human-induced rapid environmental changes. We are in the midst of a global biodiversity crisis; if we do not understand how the decisions, and movements, of organisms shape, and in turn are shaped by, ecological processes, we cannot hope to develop an evidence-based strategy for mitigating anthropogenic impacts. Great potential can be realized when scientific disciplines coevolve with technology and by intensifying this synergy we have prescient opportunities to explore the rich multiscale nature of behavior and the central role it plays in ecological processes (see Outstanding questions).

Acknowledgments

We thank Jake Graving, Blair Costelloe, Kamran Safi, and Ben Koger for valuable feedback. This work was supported by the Office of Naval Research (to I.D.C.; grant #N00014-64019-1-2556), the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement (to I.D.C.; #860949), the Struktur- und Innovationsfonds für die Forschung (SI-BW) of the State of Baden-Württemberg, the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy-EXC 2117–422037984 (to I.D.C.), and the Max Planck Society.

Declaration of interests

No interests are declared.

References

1. Le Cun, Y. *et al.* (1989) Handwritten digit recognition: applications of neural network chips and automatic learning. *IEEE Commun. Mag.* 27, 41–46
2. LeCun, Y. *et al.* (2015) Deep learning. *Nature* 521, 436–444
3. Goodfellow, I. *et al.* (2016) *Deep Learning*, MIT Press
4. Dosovitskiy, A. *et al.* (2020) An image is worth 16x16 words: transformers for image recognition at scale. *arXiv* Published online October 22, 2020. https://doi.org/10.48550/arXiv.2010.11929
5. Khan, S. *et al.* (2021) Transformers in vision: a survey. *ACM Comput. Surv.* 54, 1–41
6. Cohn, D.A. *et al.* (1996) Active learning with statistical models. *J. Artif. Intell. Res.* 4, 129–145
7. Jeon, J. *et al.* (2003) Automatic image annotation and retrieval using cross-media relevance models. In *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 119–126, Association for Computing Machinery
8. Luiten, J. *et al.* (2018) Premvos: proposal-generation, refinement and merging for video object segmentation. In *Asian Conference on Computer Vision*, pp. 565–580, Springer
9. Yao, R. *et al.* (2020) Video object segmentation and tracking: a survey. *ACM Trans. Intell. Syst. Technol.* 11, 1–47
10. Romero-Ferrero, F. *et al.* (2019) idtracker. ai: tracking all individuals in small or large collectives of unmarked animals. *Nat. Methods* 16, 179–182
11. Walter, T. and Couzin, I.D. (2021) TRex, a fast multi-animal tracking system with markerless identification, and 2D estimation of posture and visual fields. *eLife* 10, e64000
12. Schofield, D. *et al.* (2019) Chimpanzee face recognition from videos in the wild using deep learning. *Sci. Adv.* 5, eaaw0736
13. Ferreira, A.C. *et al.* (2020) Deep learning based methods for individual recognition in small birds. *Methods Ecol. Evol.* 11, 1072–1085
14. Guo, S. *et al.* (2020) Automatic identification of individual primates with deep learning techniques. *iScience* 23, 101412
15. Graving, J.M. *et al.* (2019) DeepPoseKit, a software toolkit for fast and robust animal pose estimation using deep learning. *eLife* 8, e47994
16. Lauer, J. *et al.* (2022) Multi-animal pose estimation, identification and tracking with DeepLabCut. *Nat. Methods* 19, 496–504
17. Pereira, T.D. *et al.* (2022) SLEAP: a deep learning system for multi-animal pose tracking. *Nat. Methods* 19, 486–495
18. Dunn, T.W. *et al.* (2021) Geometric deep learning enables 3D kinematic profiling across species and environments. *Nat. Methods* 18, 564–573
19. Gosztolai, A. *et al.* (2021) LiftPose3D, a deep learning-based approach for transforming two-dimensional to three-dimensional poses in laboratory animals. *Nat. Methods* 18, 975–981
20. Zuffi, S. *et al.* (2019) Three-D safari: learning to estimate zebra pose, shape, and texture from images in the wild. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 5359–5368, IEEE
21. Yang, G. *et al.* (2022) Banmo: building animatable 3d neural models from many casual videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2863–2873, IEEE
22. Handegard, N.O. *et al.* (2012) The dynamics of coordinated group hunting and collective information transfer among schooling prey. *Curr. Biol.* 22, 1213–1217
23. Koger, B. *et al.* (2022) Multi-animal behavioral tracking and environmental reconstruction using drones and computer vision in the wild. *bioRxiv* Published online July 2, 2022. https://doi.org/10.1101/2022.06.30.498251
24. Williams, H.J. *et al.* (2021) Future trends in measuring physiology in free-living animals. *Philos. Trans. R. Soc. B* 376, 20200230
25. Berman, G.J. (2018) Measuring behavior across scales. *BMC Biol.* 16, 1–11
26. Mathis, M.W. and Mathis, A. (2020) Deep learning tools for the measurement of animal behavior in neuroscience. *Curr. Opin. Neurobiol.* 60, 1–11

Outstanding questions

Can we achieve automated methods for acquiring, identifying, and interpreting animal behavior?

Does a comprehensive reconstruction of the behavioral state space, and probabilistic interstate transition pathways, give deep insight into how behavior is structured within the brain?

Are there principled ways with which to infer causal relationships between sensory input and behavioral decisions?

To what degree will it become possible to infer internal (hidden) states from high-dimensional observational data?

Will routine use of neural recording technology ever be possible for wild animals and, if so, when?

Can remote sensing be integrated with multiscale modeling in such a way as to compensate for important gaps in our understanding of how animals perceive their environments?

To what degree do processes of emergent sensing, driven by interactions within and/or between species, explain individuals' response to changing environments?

Can we achieve the major and sustained interdisciplinary effort that is required to achieve the predictive science of animal behavior in natural environments necessary for effective evidence-based conservation of species and ecosystems?

27. Hausmann, S.B. *et al.* (2021) Measuring and modeling the motor system with machine learning. *Curr. Opin. Neurobiol.* 70, 11–23
28. Bialek, W. (2022) On the dimensionality of behavior. *Proc. Natl. Acad. Sci. U. S. A.* 119, e2021860119
29. Berman, G.J. *et al.* (2016) Predictability and hierarchy in *Drosopila* behavior. *Proc. Natl. Acad. Sci. U. S. A.* 113, 11943–11948
30. Wiltschko, A.B. *et al.* (2015) Mapping sub-second structure in mouse behavior. *Neuron* 88, 1121–1135
31. Johnson, R.E. *et al.* (2020) Probabilistic models of larval zebrafish behavior reveal structure on many scales. *Curr. Biol.* 30, 70–82
32. Wiltschko, A.B. *et al.* (2020) Revealing the structure of pharmacobehavioral space through motion sequencing. *Nat. Neurosci.* 23, 1433–1443
33. Hernández, D.G. *et al.* (2021) A framework for studying behavioral evolution by reconstructing ancestral repertoires. *eLife* 10, e61806
34. Joska, D. *et al.* (2021) AcinoSet: a 3D pose estimation dataset and baseline models for Cheetahs in the wild. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 13901–13908, IEEE
35. Yang, Y. *et al.* (2022) APT-36K: a large-scale benchmark for animal pose estimation and tracking. *arXiv* Published online June 12, 2022. https://doi.org/10.48550/arXiv.2206.05683
36. Zellweger, F. *et al.* (2019) Advances in microclimate ecology arising from remote sensing. *Trends Ecol. Evol.* 34, 327–341
37. Strandburg-Peshkin, A. *et al.* (2017) Habitat and social factors shape individual decisions and emergent group structure during baboon collective movement. *eLife* 6, e19505
38. Ropert-Coudert, Y. and Wilson, R.P. (2005) Trends and perspectives in animal attached remote sensing. *Front. Ecol. Environ.* 3, 437–444
39. Nagy, M. *et al.* (2018) Synchronization, coordination and collective sensing during thermalling flight of freely migrating white storks. *Philos. Trans. R. Soc. B* 373, 20170011
40. Egert-Berg, K. *et al.* (2018) Resource ephemerality drives social foraging in bats. *Curr. Biol.* 28, 3667–3673
41. Nathan, R. *et al.* (2008) A movement ecology paradigm for unifying organismal movement research. *Proc. Natl. Acad. Sci. U. S. A.* 105, 19052–19059
42. Signer, J. *et al.* (2019) Animal movement tools (amt): R package for managing tracking data and conducting habitat selection analyses. *Ecol. Evol.* 9, 880–890
43. Dorfman, A. *et al.* (2022) A guide to area-restricted search: a foundational foraging behaviour. *Biol. Rev.* 97, 2076–2089
44. Calhoun, A.J. *et al.* (2014) Maximally informative foraging by *Caenorhabditis elegans*. *eLife* 3, e04220
45. Flack, A. *et al.* (2018) From local collective behavior to global migratory patterns in white storks. *Science* 360, 911–914
46. Williams, H.J. and Safi, K. (2021) Certainty and integration of options in animal movement. *Trends Ecol. Evol.* 36, 990–999
47. Mueller, T. *et al.* (2013) Social learning of migratory performance. *Science* 341, 999–1002
48. Marr, D. (2010) *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, MIT Press
49. Castelvecchi, D. (2016) Can we open the black box of AI? *Nature* 538, 20–23
50. Holzinger, A. (2018) From machine learning to explainable AI. In *2018 World Symposium on Digital Intelligence for Systems and Machines (DISA)*, pp. 55–66, IEEE
51. Belle, V. and Papantonis, I. (2021) Principles and practice of explainable machine learning. *Front. Big Data* 4, 688969
52. Brown, A.E. and De Bivort, B. (2018) Ethology as a physical science. *Nat. Phys.* 14, 653–657
53. Datta, S.R. *et al.* (2019) Computational neuroethology: a call to action. *Neuron* 104, 11–24
54. Pereira, T.D. *et al.* (2020) Quantifying behavior to understand the brain. *Nat. Neurosci.* 23, 1537–1549
55. Tuia, D. *et al.* (2022) Perspectives in machine learning for wildlife conservation. *Nat. Commun.* 13, 1–15
56. Szenicer, A. *et al.* (2022) Seismic savanna: machine learning for classifying wildlife and behaviours using ground-based vibration field recordings. *Remote Sens. Ecol. Conserv.* 8, 236–250
57. Toledo, S. *et al.* (2020) Cognitive map–based navigation in wild bats revealed by a new high-throughput tracking system. *Science* 369, 188–193
58. Holton, M.D. *et al.* (2021) Animal tag technology keeps coming of age: an engineering perspective. *Philos. Trans. R. Soc. B* 376, 20200229
59. Usman, A.M. *et al.* (2020) Review of automatic detection and classification techniques for cetacean vocalization. *IEEE Access* 8, 105181–105206
60. Graving, J.M. and Couzin, I.D. (2020) VAE-SNE: a deep generative model for simultaneous dimensionality reduction and clustering. *BioRxiv* Published online July 17, 2020. https://doi.org/10.1101/2020.07.17.207993