

# Automated pose estimation in primates

Benjamin Y. Hayden<sup>1</sup>  | Hyun Soo Park<sup>2</sup> | Jan Zimmermann<sup>1</sup> 

<sup>1</sup>Department of Neuroscience, Department of Biomedical Engineering, Center for Magnetic Resonance Research, University of Minnesota, Minneapolis, Minnesota, USA

<sup>2</sup>Department of Computer Science and Engineering, University of Minnesota, Minneapolis, Minnesota, USA

## Correspondence

Benjamin Y. Hayden, Department of Neuroscience, Center for Magnetic Resonance Research, University of Minnesota, Minneapolis, MN, 55455, USA.  
Email: [benhayden@gmail.com](mailto:benhayden@gmail.com)

## Funding information

NIH, Grant/Award Numbers: MH125377, P30DA048742; NSF, Grant/Award Number: 1846031

## Abstract

Understanding the behavior of primates is important for primatology, for psychology, and for biology more broadly. It is also important for biomedicine, where primates are an important model organism, and whose behavior is often an important variable of interest. Our ability to rigorously quantify behavior has, however, long been limited. On one hand, we can rigorously quantify low-information measures like preference, looking time, and reaction time; on the other, we can use more gestalt measures like behavioral categories tracked via ethogram, but at high cost and with high variability. Recent technological advances have led to a major revolution in behavioral measurement that offers affordable and scalable rigor. Specifically, digital video cameras and automated pose tracking software can provide measures of full-body position (i.e., pose) of primates over time (i.e., behavior) with high spatial and temporal resolution. Pose-tracking technology in turn can be used to infer behavioral states, such as eating, sleeping, and mating. We call this technological approach *behavioral imaging*. In this review, we situate the behavioral imaging revolution in the history of the study of behavior, argue for investment in and development of analytical and research techniques that can profit from the advent of the era of *big behavior*, and propose that primate centers and zoos will take on a more central role in relevant fields of research than they have in the past.

## KEYWORDS

big data, deep learning, behavioral tracking, rhesus macaque, primates

## 1 | INTRODUCTION

The Minnesota Zoo in Apple Valley, MN, has the largest public collection of Japanese macaques (*Macaca fuscata*) in the United States (Figure 1). Every morning, the zoo's 27 macaques emerge from their dormitory and enter a large, beautifully architected open enclosure. There are many ways to describe what they do next, but one way to say it is that they proceed to generate an *enormous* amount of data. That is, each of the monkeys moves each of its limbs in a specific way, moves its body position along a particular and often complex path, and interacts with multiple items in the pen. They also interact with each other in complex ways, play, explore, foraging, relax, eat, and so on.

The ability to track and analyze the actions of primates like the ones at the Minnesota Zoo has potential relevance to researchers in

biology and biomedicine, as well as for neuroscience, psychology, comparative biology, and research into animal welfare (e.g., Bliss-Moreau & Rudebeck, 2020; Buffalo et al., 2019; Pereira et al., 2020; Rudebeck et al., 2019; Santos & Rosati, 2015). And yet, nearly all the rich data these monkeys generate largely slips past without being registered.

Instead, characterization of complex primate behavior is limited to what humans can annotate, laboriously, in hand-crafted ethograms. Such measures can only capture information when observers are present, that is, they do not work for behavior that occurs after staff leave for the day, in the evening, or on vacations, or even when the annotators' attention is momentarily diverted (Brando & Buchanan-Smith, 2018). It is also possible to characterize simple behaviors more rigorously using computerized systems. For example,



**FIGURE 1** Camera tracking system at the Minnesota Zoo. Relatively affordable cameras, such as GoPro cameras, when combined with computer vision tracking software, can provide estimates of pose, although they depend on having long-term access and a safe vantage point

we can quantify monkeys' interactions with specialized response systems, such as levers and buttons. Such systems produce data that are several orders of magnitude lower in information rate than the animals' full behavioral repertoire. For example, a measure of gaze direction or preference, typical in many psychological studies, gives one bit of information (left vs. right) per trial. Those data can, of course, be used to test important hypotheses. But they come with a high opportunity cost: they ignore much more data than they capture.

In contrast, behavioral tracking can produce high quantities of data (e.g., 13 body landmarks in three dimensions sampled at 30 Hz, Bala et al., 2020) without human intervention or even oversight. Indeed, without tracking, our ability to collect behaviors with high technical precision and reproducibility is so limited that it might be described as looking at the world through a drinking straw.

This is all changing, and at a rapid pace. A recent series of technological advances have made it possible to collect a good deal of the data these monkeys produce. These technologies are generalizable, meaning that, when appropriately trained, they can provide tracking for any species of primate, including arboreal ones. And not just these monkeys – other species and other locations can also be tracked (Bala et al., 2020; Dankert et al., 2009; Joska et al., 2021; Marshall et al., 2021; Mathis et al., 2020; Walter & Couzin, 2021). We call this technological approach *behavioral imaging*.

We use this term deliberately by comparison with other forms of imaging, which can give detailed information available in the visual domain through the use of advanced technology.

The subject of study in this approach has been called *big behavior* (von Ziegler et al., 2021). By our lights, little behavior refers to measures of behavior that involve highly reduced, low-information tracking of what the animal is doing. Little behavior would include reaction times, preferences expressed through choice, gaze directions, pupil size, and reaches – measures that give a small amount of information per unit time. Little behavior would also include ethograms made by trained annotators – not because they measure small

amounts of information but because that information is filtered through the narrow aperture of human annotation abilities.

Big behavior, by contrast, includes full information about the position of all major joints in the entire body and involves continuous movement over an extended period of time. The extent of that time can be very large. Scientists can now use multiple high-resolution cameras to continuously capture every fine-grained behavior over several weeks, months, and years. They can track pose of both individuals and groups. As with any big data situation, the big behavior revolution raises unprecedented challenges, especially in managing, analyzing, and understanding the data in a fully automatic fashion. However, the benefits it offers are so great that many teams are working to solve them.

The present review will describe how we got to this point, talk about the state of the art, make a few predictions about the near future, and sketch out some of the potential benefits. We will discuss some of the specific scientific problems that big behavior for primates is likely to affect. Likewise, we will argue that primate centers and zoos will take on much larger roles in research than they have in the past.

## 2 | THE CHALLENGES OF BIG BEHAVIOR IN PRIMATES

Behavioral tracking in primates has recently become possible through parallel developments in computer vision, machine learning, and robotics. The story starts at the beginning of the last decade when technical breakthroughs in deep learning-enabled software to recognize objects in an image by use of convolutional neural networks (CNNs, Krizhevsky et al., 2012). Subsequent works improved on the design of CNNs to enable automatic tracking of humans both live and from videos (Cao et al., 2019; Fang et al., 2017; Newell et al., 2016; Wei et al., 2016). That work in turn inspired and facilitated work that allowed for the tracking of animals such as flies, mice, and horses from videos (Marks et al., 2021; Mathis et al., 2018, 2020; Pereira et al., 2019). The tools that allowed these animals to be tracked leveraged the copious capacity of CNNs to learn the visual and geometric relationship between landmark locations (body joints). They also relied on relatively cheap and robust digital cameras and standard computer vision techniques.

Relative to other animals, nonhuman primates have been much more difficult to track (Bala et al., 2020; Labuguen et al., 2020; Negrete et al., 2021; Testard et al., 2021). There are three reasons for this. First, their body joints are highly flexible and thus generate a large number of distinctive body postures. Each joint has multiple degrees of freedom and these complex joint positions are essential for tracking and identifying activities such as bipedal/quadrupedal locomotion, hanging, and dexterous object manipulation. This fact distinguishes primates from non-primate animals such as rodents and insects – the poses of these animals can be characterized by a small basis set. Second, primates' bodies are covered by thick fur, which is a characteristic failure case of computer vision-based tracking

algorithms. For example, the pelvis location is highly ambiguous while performing a sedentary activity because there are no visually or geometrically salient features to identify the pelvis joint. Third, primates' body movements are fundamentally three dimensional (3D). The range of motion of bodies spans full three-dimensional space, and more importantly, the motion often involves 3D interactions with objects, peers, and environments. This dimensionality raises two major problems – (1) considerable occlusion and (2) broad variation in appearances.

We have proposed that, given these difficulties, nonhuman primate tracking cannot be achieved using existing tracking paradigms (Bala et al., 2020). In particular, off-the-shelf solutions like DeepLabCut need orders of magnitude more data than they do use other species, and so bespoke solutions currently work best. We need innovations in both hardware and software that are tailored to reflect the primate characteristics. For the readers' benefit, we next describe several of the needed innovations and several of the major factors that must be considered and the obstacles that must be overcome.

## 2.1 | 2D versus 3D representation

Primates move in three dimensions. Although other animals also move in three dimensions, primates tend to make use of all three quite a bit more than other popular model organisms. A rat's major axes of movement are on a plane along the ground and a worm's behavior is well studied in a petri dish. By contrast, a monkey moves vertically, but, more importantly, its limbs move with much more freedom relative to its central body axis.

In most computer vision solutions, tracking occurs on the XY plane, without regard to depth (i.e., the Z plane); doing so is much easier to implement and costs little with plane-moving species. Continuing with the rodent example, a typical rodent tracking system uses an overhead camera and tracks XY locations of landmarks. This 2D representation, however, shows limited expressibility to describe the behaviors of primates due to their 3D body movement over 3D scenes. The 2D representation is a camera projection of 3D behaviors where its location drastically varies as the viewpoint changes. A 3D representation (XYZ) is a viable solution that requires a specialized system such as a depth camera or a multiview camera system. For instance, a system of multiview cameras is used to develop OpenMonkeyStudio that enables tracking 3D body movements of macaques (Bala et al., 2020, Figure 2). The resulting 3D representation is invariant to the viewpoint change, which allows modeling coherent behavioral clusters compared with the 2D representation as shown in Figure 3.

## 2.2 | Target-specific model versus generalizable model

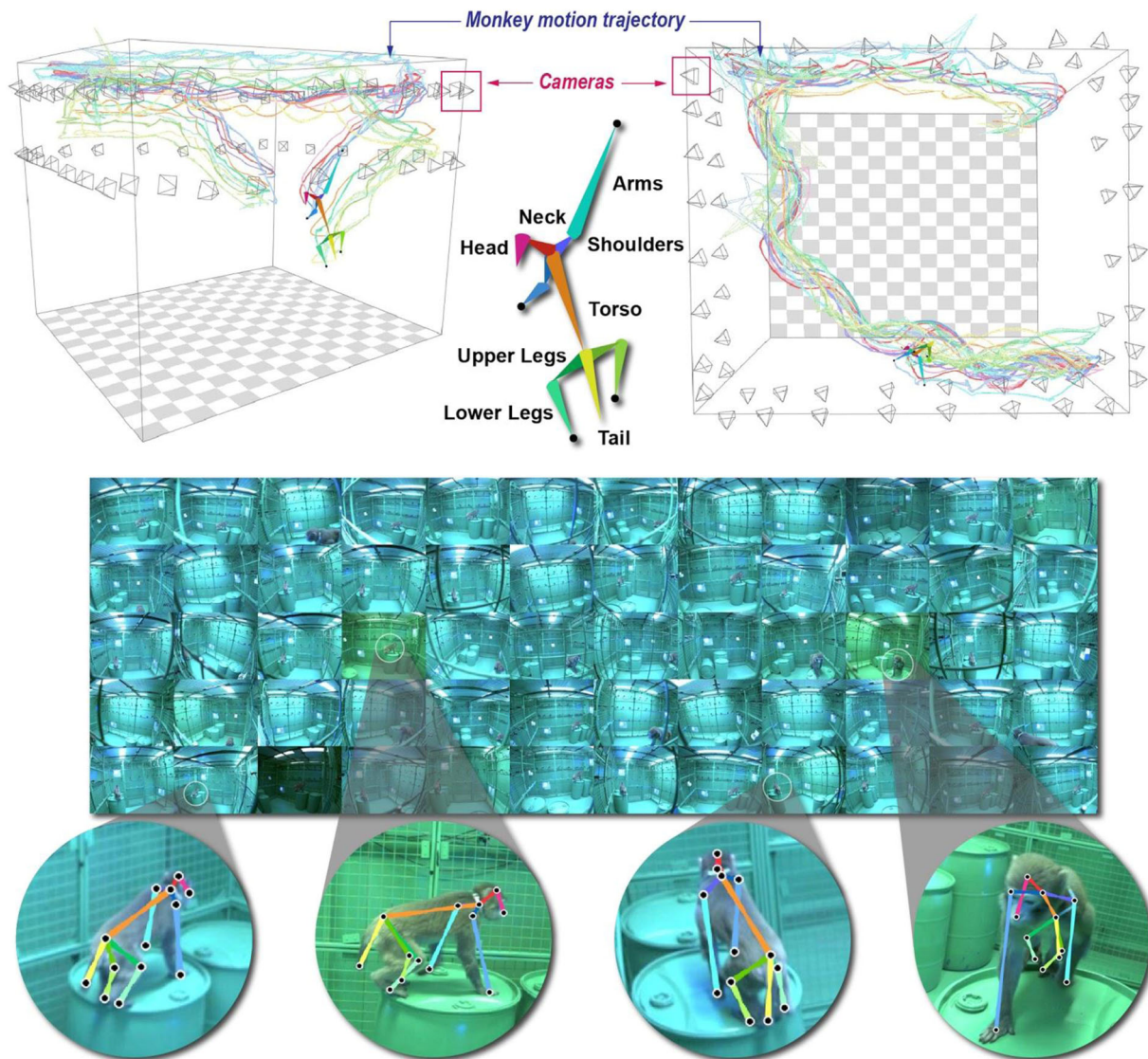
The ability of a CNN model to track primates is limited by our ability to train the model. Our ability to train the model is limited by the size

and quality of the training set. Thus, the success of tracking algorithms is predicated on the existence of large, accurate, and well-curated datasets. For instance, large scale annotated data sets such as COCO and MPII have been used to train CNNs to reliably detect human poses (Andriluka et al., 2014; Lin et al., 2014). However, such large data sets do not exist for most animal species.

This data challenge has been addressed by training a target-specific model facilitated by annotation tools. For example, user interfaces provided by DeepLabCut and LEAP have enabled effective pose annotations by use of a very small number of annotations (as few as 10–1000) in a video (Mathis et al., 2018; Pereira et al., 2019). These methods, which make use of transfer learning or its variants, work for relatively simple organisms like flies, worms, and mice (e.g., Mathis et al., 2018; Nath et al., 2019).

Despite its remarkable performance, this approach is not suited to track the behaviors of nonhuman primates because visual appearance significantly varies as a function of viewpoint, pose, and species (Bala et al., 2020). As a result, a CNN trained using a data set from one viewpoint does not generalize to that from another viewpoint. Instead of learning such target-specific models, we have proposed a new paradigm that aims to learn a generalizable model by collecting a large, curated data set. This curated data set must include diverse images across species, background, illuminations, poses, viewpoints, and interactions, where the size of data is comparable to the human data size (order of millions). The MacaquePose data set that is made of 17,000 annotated internet images (Labugen et al., 2020). It is an inspiration for the OpenMonkeyChallenge project (Yao et al., 2021). It includes a large variety of appearances, poses, and viewpoints of macaques in the wild, which can be combined with state-of-the-art pose detection models such as DeepLabCut and LEAP. The performance was unprecedented: no existing macaque detection model was able to achieve that level of accuracy and generalization, due to the scale of the data. Nonetheless, it still lags behind the accuracy of human pose estimation. OpenMonkeyChallenge is designed to address this limitation by creating a 6 fold-bigger-data set across multiple species to enable learning a shared representation between them (Yao et al., 2021).

Once the large data set is collected, a generalizable CNN can be trained to detect primate poses from out-of-sample videos regardless of camera configuration, background, and species. Such networks can also be trained to use models of the animals to draw inferences about positions of landmarks that are occluded or that are not in the original training set (Bala et al., 2020). For example, we have collected more than 100,000 images with 17 landmark annotations portraying primates in natural habitats. The images can be obtained from the internet, national primate research centers, and from zoos, and include 26 species (20 monkeys and 6 apes) as shown in Figure 4. This is, by far, the largest collection for nonhuman primates. With this data set, we recently opened a new benchmark challenge called OpenMonkeyChallenge (<http://openmonkeychallenge.com/>) that facilitates an annual competition to develop a generalizable pose detection model. We believe that this data set, and others like it, will be crucial for developing widely used models for tracking nonhuman primates.



**FIGURE 2** Automated pose-tracking software, such as OpenMonkeyStudio (Bala et al., 2020), can provide high-quality tracking of poses in primates. The OMS system is based on multiview capture, which can bypass problems associated with occlusion

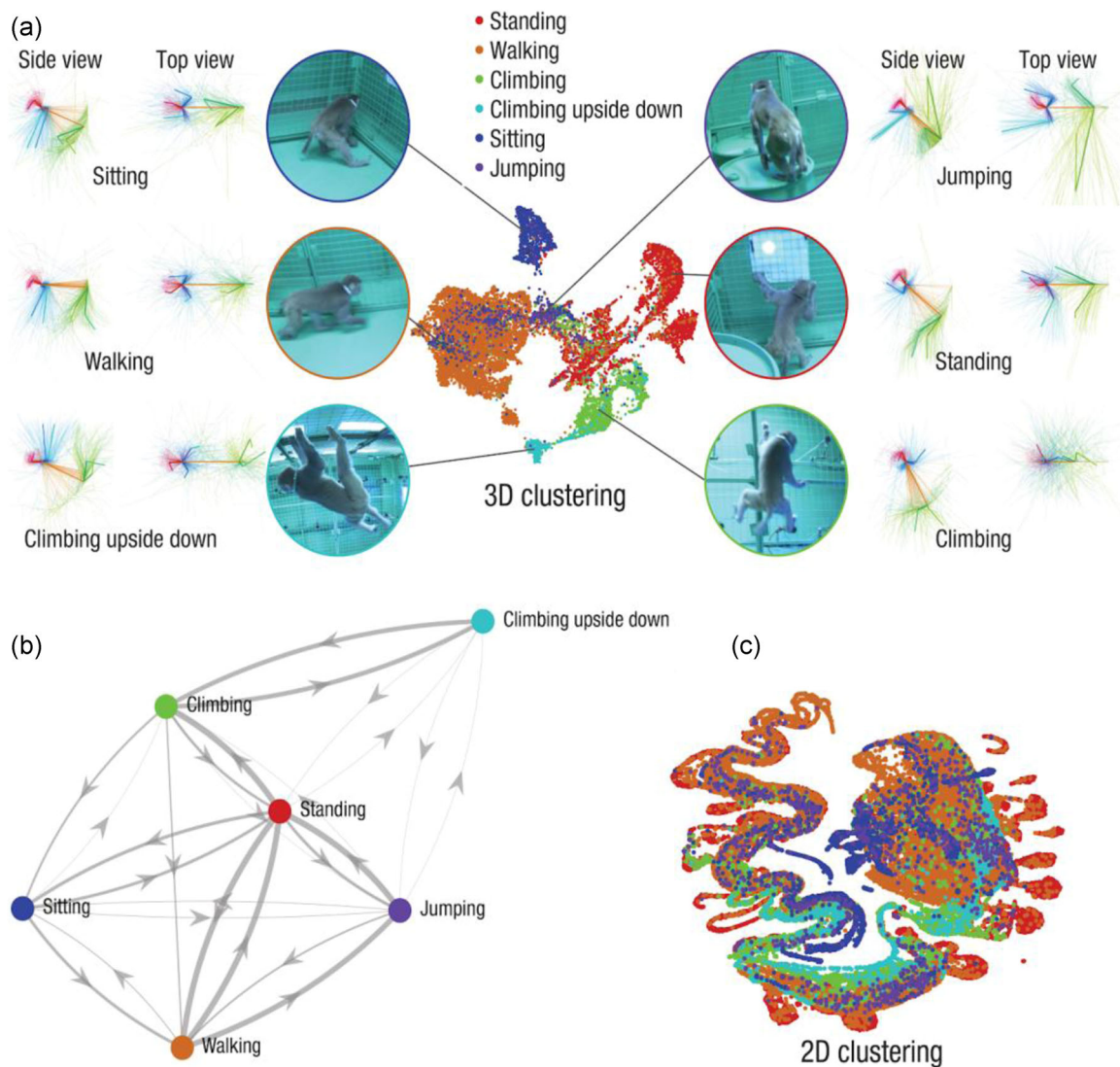
### 2.3 | Multi-view camera system

Most existing animal tracking approaches are designed to draw pose inferences from an image stream from a single camera. Due to the 3D nature of primate behaviors, multiview camera systems provide several benefits. These include:

- *Disocclusion*: in tracking, occlusion is the main source of pose detection error. In particular, when a primate interacts with objects and other primates, those other things will cover some of the relevant landmarks. For example, occlusion is often an unavoidable problem with arboreal primates, such as gibbons, because the trees in which they spend time naturally cover their landmarks. Indeed, one side of the animal's body will almost always occlude the other, just as the limbs will often occlude the more central parts of the body. Some of this occlusion can be mitigated by a

multiview camera system; even when the cameras do not offer views of the occluded parts, the parallax views they provide can greatly facilitate inferences about the true positions of the relevant landmarks. The benefits of disocclusion are particularly valuable when considering social interactions, which are often naturally occluded by the agents involved in the interaction.

- *Robust estimation*: Multiple simultaneous observations reinforce robust estimation. The multiview images of a common primate are visually and geometrically distinctive, and therefore, aggregating the detections from these multiview images allows more accurate and robust estimation in the presence of measurement noise.
- *Data augmentation*: Multiview images provide trivial data augmentation. Annotations from two viewpoints can be reconstructed in 3D, and in turn, projected onto other images, which forms additional annotations without manual effort. This leads to significant improvement on annotation efficiency (Bala et al., 2020; Marshall et al., 2021).



**FIGURE 3** With OpenMonkeyStudio, we compare two-dimensional (2D) and 3D representations in macaques (Bala et al., 2020). We compared our systems' ability to recognize semantic actions (standing, walking, climbing, climbing supine, sitting, and jumping). (a) The poses are clustered by using UMAP. Each cluster is correlated with specific actions. (b) With the clusters, we recognize actions using the  $k$  nearest neighbor search and visualize the transitions between actions. (c) In contrast, the 2D representation provides the clusters that are driven by the pose and viewpoint

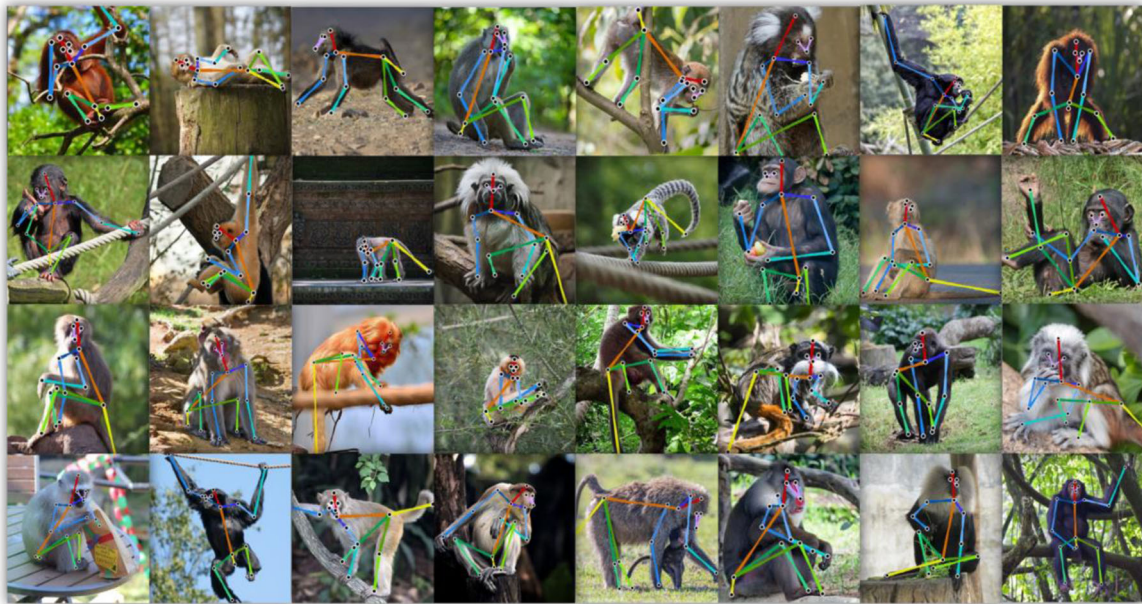
For instance, in OpenMonkeyStudio, two to three images are manually annotated and nearly 50–60 multiview images are automatically annotated using the multiview geometry. This results in a generalizable model trained by a large, annotated data set (~200K images).

- **3D representation:** Multiview geometry offers a 3D representation. This is different from the 3D lifting approach that reconstructs 3D pose from a single view image of which reconstruction is defined up to scale and orientation (Günel et al., 2019). With multiview images, full metric scale reconstruction can be obtained.

## 2.4 | Identifying individuals

Primates are social animals. Indeed, their social nature makes them an appealing model organism for humans, including both healthy social

development and diseases associated with impaired social function, such as autism and Alzheimer's Disease. The fact that they are social means that any technology deployed to understand their behavior must work with groups. Indeed, both zoos and primate centers often have groups as a standard organizational unit. Standard pose tracking software can typically be extended to include tracking multiple individuals (see above and Bala et al., 2020). However, these systems need to have advanced features to ensure *individuation*. That is, systems may obtain good performance on pose estimation through interpolation and deduction, but these tricks can lead to confusion over identity. Identity confusion can cause critical problems with experiments when the goal is to look at changes in individuals longitudinally. Consider, for example, a study to monitor the success of a physical therapy regime for an individual monkey who has a limp. If the tracking system confused him with a conspecific, it will appear



**FIGURE 4** Images from our new data set, OpenMonkeyChallenge, illustrating the range of species, individuals, and backgrounds

that the limp has healed spontaneously, when in reality, it is just tracking a different individual. Even if there is a small probability of misidentification, this probability may be effectively compounded over time – possibly to near certainty – in long-lasting studies.

Individuation is a deceptively tricky problem and is one that is poorly suited for pose tracking systems. The reason is that, without secondary checks, the method is susceptible to catastrophic failures – even rare loss of coverage can lead to confusion and major problems with data. Thus, individuation often can benefit from complementary approaches, such as identification of individuals by face, body shape, movement type, or other distinguishing features. For example, one possible solution is to use technology such as radio frequency identification (RFID) tags to define individuals. That in turn would require placing RFID tags onto individuals' bodies, which may be impractical. Finally, a third approach involves humans checking the results of the tracking system to ensure that individuation is maintained. This approach is costly but may still be orders of magnitude more achievable than full pose tracking by humans. There is some work on the use of trackers in primates (e.g., Gazes et al., 2019; Gelardi et al., 2020).

## 2.5 | General technical issues

In addition to these problems there are several basic technical problems that need to be overcome and whose difficulty should not be understated. We list a few of them here. Anyone interested in tracking primates must have enough cameras to cover the area of interest. Typically, tracking is considerably more useful if the user can achieve 3-D tracking, which in turn means multiple (often four or more) cameras. In either case, if the recording is longitudinal, these installations will need to be weatherproofed and will need to be

protected from the prying hands of the primates themselves. This level of coverage may be daunting at large zoo exhibits and even more so in field cages at primate centers.

These cameras will need to be synchronized to the precision of a single frame and must maintain synchronization across the period of recording. All cameras must be calibrated, both intrinsically and extrinsically, and this calibration must take place often (typically at least once a day). The cameras will typically generate large volumes of data (often over 1 TB per hour), and this data stream must be managed and stored. That information must be transported from the recording site to the location where the servers are stored. Often it is impractical to build them near the site where the primates live, so long high-quality fiber-optic wires, possibly requiring laying new wiring, may be unavoidable. The need for storage can be reduced if some computations can occur online or at the site of recording, but this, in turn, raises other technical challenges. Ultimately, these things can be costly – a simple web cam will likely not offer the full benefits of tracking. Moreover, the installation must be coordinated with the stakeholders of the site, and if it is a zoo or primate center, they may have other needs that supersede those of the scientists who want to do the tracking.

## 3 | AUTOMATED BEHAVIORAL STATE IDENTIFICATION (“ETHOGRAMMING”)

Automated behavioral sampling and identification (briefly, “automated ethogramming”) procedures offer the promise of using a computer to perform the process of identification of behavior, rapidly, cheaply, and on a massive scale (Marks et al., 2021; Voloh et al., 2021). Because it is much cheaper and faster, automated

behavioral identification promises many things that are uneconomic with human annotators, such as massive throughput testing, or monitoring behavior twenty-four hours per day, or testing unlikely hypotheses that would be too costly to test with standard approaches.

Since it is a new technology, it is difficult to predict how behavioral imaging will be best used. Here, we will make some educated guesses. However, like many others, we believe that because of its ability to tell us about behavior, the promise is high (Calhoun & El Hady, 2021; Calhoun & Hayden, 2015; Datta et al., 2019; Krakauer et al., 2017). Specifically, we envision that longitudinal and quantifiable analysis from years-long recordings of primates will enable us to discover the reproducible behavioral patterns and characteristics, which has been impossible from small-scale data analysis. Pose (see above) can tell us where the landmarks and major joints are positioned in space, but the key dependent variable for many studies is what behavior the animal is performing (Anderson & Perona, 2014). In other words, the ideal output of behavioral imaging is the category of behavior that describes the animal at every moment in time. Such a time series would be the starting point for a great deal of research.

For illustrative purposes, we consider a couple of examples. Suppose we have developed a potential animal model of clinical depression and wish to validate it (e.g., see work by Shively & Willard, 2012). It could be validated by comparing the amount of time the animal is immobile, sleeping, or stays in its dormitory – behaviors that would have clear face validity to symptoms of human depression. As noted above, these behaviors can, in principle, be scored by a human observer, but only in small quantities because of the cost. Another example would be anxiety – behavioral identification could be used to determine whether the animal shows behaviors with face validity to human anxiety disorders such as excessive grooming, fastidiousness, unusual behaviors around doorways – behaviors with face validity to human anxiety. Moreover, the ability to monitor for both types of behavior efficiently means that we could then engage in unsupervised classification of possible depression analogs by the presence or absence of specific symptoms and/or by their cross-symptomology with anxiety homologs. This clustering could then differentiate different subtypes of depression correlates.

Another example would be in the form of precursors to drug abuse. Monkeys are an important model organism for drug abuse for several reasons, including their long lifespan, their closeness to human body size (compared to rodents), and the similar architecture of their prefrontal cortex. Within the world of addiction, one important question is how strongly one can predict susceptibility to drug abuse based on past behaviors. This kind of search is essentially a needle in a haystack type of search – we must have access to large amounts of highly quantified (and ideally unbiased) measures of behavior that we can analyze and correlate with subsequent behavior following drug exposure. Behavioral imaging with state identification provides an ideal input to such studies.

More generally, behavioral states could be used as a dependent variable in studies of how some intervention affects the way the animal distributes its behavioral state over time. Or they could be

used as an independent variable, such as in studies of how an animal's behavior affects that of a conspecific. Identifying behavior based on pose is a nontrivial problem and remains an area of active research in many species. Most existing work has been done for animals like flies, worms, and mice. In primates, automated behavioral identification is relatively unexplored. However, the general principles developed for other animals ought to generalize relatively straightforwardly. In general, such approaches take the raw pose data, subject it to some preprocessing (such as centering the animal and rotating it to a canonical orientation), and then performing some unsupervised procedure, such as t-SNE or UMAP, to identify behavioral states.

Automated behavioral state identification is an alternative to hand-scored behavioral sampling procedures that produce ethograms. Automating the process addresses the key limitations in this behavioral sampling approach by providing a computational solution that is cost-effective, consistent, fast, accurate, and fine-grained (Anderson & Perona, 2014).

- (1) *Cost-effectiveness*: the human-scored ethogram approach involves highly trained human annotators who watch live animals or videos and mark down the behaviors they observe based on a predetermined rubric. These humans are skilled in identifying behaviors; indeed, the need to carefully train these observers is a major cost of the ethogram approach and one that makes it very expensive to implement. Being able to do that takes education and training and does not scale with large datasets. Big behavior automates behavioral annotations, which reduces the cost of training, management, and labor.
- (2) *Consistency and speed*: different human raters tend to have different subjective criteria, meaning the studies are often not reproducible with the same data if the raters are no longer available (Anderson & Perona, 2014; Levitis et al., 2009). Even well-trained individual human observers can often be somewhat inconsistent. This is especially likely to be a problem with novel and unusual primate behaviors, or ones that are poorly characterized.
- (3) *Accuracy*: human raters are intrinsically fallible – they get bored, especially after many hours of video footage, they use criteria that change subtly over time, they have blind spots, systematic patterns of behavior that they are less likely to detect. Consider for example how often even professional sports referees make mistakes or even disagree, and that in sports, categories like fair and foul are rigorously defined.
- (4) *Bias*: human raters tend to exhibit systematic biases (Kardish et al., 2015). Indeed, careful experimental work has shown that human observers interpret behaviors through their own expectations (Tuytens et al., 2014). More broadly, a great deal of behavioral identification is subjective, meaning it may be difficult to evaluate bias so that it can be reduced or avoided; even ostensibly objective measures ought to be blinded, and seldom are (Holman et al., 2015).
- (5) *Longitudinal perspective*: Humans are especially poor at detecting rare behaviors, and at detecting the kinds of behaviors that involve gradual change over long periods of time (Biggs

et al., 2014; Kryszczuk & Boyce, 2002; Wolfe et al., 2005). Automated ethogramming offers the opportunity to take the long view, to accumulate data over long periods, and draw inferences about rare and slowly changing behaviors. We anticipate this benefit may be especially important for developmental trajectories, such as skill-learning, aging, and adolescence.

- (6) *Annotation resolution*: Humans have a sensitivity limit. We can only detect a certain number of behaviors and can only detect a certain granularity of behavior. For example, it may take some effort to be able to identify the different gaits of a running horse, especially subtly different ones. Moreover, it's possible that there are some patterns in the data that are real and measurable, but humans simply are not adapted to notice them without help. For example, some behaviors may reflect certain patterns of simultaneous activity across multiple limbs and may only be detectable and classifiable after certain dimensionality reduction processes that humans lack the cognitive capacity to perform.

## 4 | WHAT BEHAVIORAL IMAGING OF PRIMATES OFFERS BIOLOGY AND BIOMEDICINE

Behavioral imaging of primates can be of great benefit to all manner of scientists – from those interested in the biology of primates to those who are interested in using primates as model organisms for disease. It is important to note that these categories are not mutually exclusive, nor does biomedical research necessarily make use of invasive procedures on unhealthy animals. Indeed, there is a great deal of important biomedical research that can and should be done on healthy animals; this study would seek to understand the biology of these animals without any use of disease models or simulacra of disease processes. Thus, the study of healthy animals poses an opportunity for biomedicine as well as for biology.

### 4.1 | Improvements in diagnosis

Diagnosis typically requires specialized expertise in humans; in animals, who cannot talk and often disguise their symptoms, it's more difficult. Many conditions can only be diagnosed following conspicuous presentation of behavioral symptoms. Many diagnoses require advanced veterinary intuition. In many cases, animals deliberately camouflage their symptoms. For these reasons, diagnosis is likely to be a major focus for behavioral imaging. This is not to say that computers will monitor all behavior and replace veterinary expertise. Instead, we imagine that behavioral imaging will provide a complementary measure that will boost trained medical opinion.

First, it will provide much more information, over a longer period of time. Second, it will offer the ability to identify subtle and hidden behavioral patterns, that may be ignored or missed. For instance, a monkey may slightly drop her shoulder while walking when her arm is wounded. Such differences represent symptoms, which have not

been coded and listed as a behavioral marker by the experts. Third, it will offer acute measures, with greater quantity, accuracy, and sensitivity than human observers can provide. Finally, it will provide a ready and quantifiable control group, including individuals in the same facility, to say whether the individual in question differs from their peers.

### 4.2 | Finer-grained diagnoses

Many diseases are clusters of distinct conditions with somewhat different symptom profiles and different best treatments. To give a well-known example, there are many effective psychopharmaceutical treatments for clinical depression, and different patients respond differently to different ones. Trying each one takes several months, so if diagnosis could be sped up, a great deal of suffering could be eliminated. It is hoped that finer-grained diagnosis, based on behavioral presentation, would provide an indicator of which treatment should be tried first. The ability to characterize behavior with high bandwidth promises greater information that can be clustered and specific disease subtypes identified. Even modest improvements in the classification of diseases into different subtypes, which in turn can motivate the specialized study of treatment.

To give one well-known example, deep brain stimulation (DBS) treatment for depression in humans is sometimes focused on the subgenual anterior cingulate sulcus (Lozano et al., 2008; Mayberg et al., 2005). This is a brain region whose responses are known to be associated with depression, with negative outcome monitoring, and with depression severity (Alexander et al., 2019; Azab & Hayden, 2018; Drevets et al., 2008; Maisson et al., 2021; Roberts & Clarke, 2019). However, despite promising early treatment results, follow-up studies have been equivocal. It has been proposed that the core issue is that we are treating two different diseases with different anatomical bases (Drysdale et al., 2017; McGrath et al., 2013). Improved diagnosis in humans would help us test these hypotheses faster, and more valid animal models (see above) would help even more.

### 4.3 | Validating animal models of disease

Primate models of disease are crucial to modern biomedical research. In many cases, models are imperfect or are only relevant to a subset of symptoms. In other cases, the validity of primate models is unknown. For example, the limitations of models are especially well delineated in the case of psychiatric illness, which often relies on personal reports. Consider that many of the major criteria for depression and obsessive-compulsive disorder rely on subjective descriptions of feelings (Beck et al., 1988; Goodman et al., 1989). Ascertaining the validity of an animal disease model can be a surprisingly ad hoc procedure, based on a superficial assessment of major symptoms (Geyer & Markou, 1995; Koob & Zimmer, 2012). In many cases, establishing a valid animal model for some disease or



symptom class takes many years - even decades - to result in a consensus across the field. Despite these problems, some diseases, or some symptom classes may be characterized by a specific set of behaviors and behavioral patterns; it is possible that the particular constellation of behaviors can serve as an ethological fingerprint of the disease or symptom. While this possibility is speculative, it cannot readily be tested until we have behavioral tracking. Optimistically, these fingerprints can be compared across species to test specific hypotheses about the behavioral validity of specific models.

#### 4.4 | Improvements in treatment

Many diseases have well-defined treatments that require parameter-setting (Johnson et al., 2013). For example, Parkinson's Disease is well treated by DBS although the parameter space is very large. The optimization of parameters takes place through what is more or less a gradient descent procedure, that is, through trial and error (Schiff, 2010). The slow part is the assessment of state in the patient. Primate models can improve that process, but they still have a slow assessment stage - often even slower in primates because they lack the verbal modality. Big behavior offers a solution to this problem - it is both high bandwidth and computerized, so it can be made very fast, and potentially improve treatment time by orders of magnitude. It is possible - albeit speculative at this point - that behavioral imaging can lead to rapid closed-loop parameters for treatment. We can test specific parameters and read out the effects of those parameters - whether they are stimulation parameters in DBS, anatomical positioning of stimulation, different drug doses, or any other of a large number of possibilities.

#### 4.5 | Replacement for other dependent variables

To be a bit more speculative, behavioral imaging could potentially offer new dependent variables that could reduce the need for invasive measures, leading to benefits in welfare, and increases in the efficiency of research. A good deal of research may require some measure of the efficacy of some intervention, but not depend on any particular measure. For example, consider a team of researchers who want to rapidly screen several dozen candidates for headache medicine, most of which will be inert. In that case, behavioral imaging may provide information sufficient to move on to the second stage. For example, animals without headaches may be more active overall, may engage in more social behaviors, may spend more time in well-lit areas, and so on. And it may do so at a price that is much cheaper, and with less intervention, than other measures (such as measures of internal physiological functioning).

Or, to give an example drawn from the psychological literature, scientists are sometimes interested in knowing the way that an animal will respond to seeing a reflection of itself in the mirror (Gallup et al., 2002). For example, a researcher may want to know whether an animal differentiates seeing itself in the mirror from seeing

another conspecific. This study may in turn be biased by observation or may be slow and may benefit from running dozens of animals instead of a few. Automated tracking could make high-throughput mirror testing feasible. It could also make it possible to compare many species in a larger clade, rather than forcing psychologists to pick one or two exemplar species.

Indeed, in such cases, big behavior offers the possibility of performing such experiments in animals that are inaccessible to other measures, including rare and endangered animals, and highly intelligent animals, such as apes. Even for standard laboratory primates, behavioral imaging may be an order of magnitude cheaper than other measures (see above) and may therefore be a preferred alternative. For example, it may require less specialized equipment, trained technicians, or built laboratory environments. Finally, even if it is not cheaper, it may have fewer welfare costs than invasive measures, and may be preferred for that reason.

#### 4.6 | Natural behaviors

Finally, behavioral imaging allows scholars to monitor naturalistic - or at least relatively naturalistic behaviors. When primates perform tasks that resemble those for which they have evolved, their behavior is more likely to reflect their normal response biases, and thus to be more ethologically valid, and likely more interpretable (Hayden, 2018; Pearson et al., 2014). For example, we have argued that both risk and impulsivity measures in primates are biased by standard laboratory tasks (Blanchard et al., 2014; Eisenreich et al., 2019; Hayden, 2016). While we may currently lack even strong hypotheses about the types of scientific questions that can best be asked by studying natural behavior at scale, the best ideas often come after the data are collected and while they are being analyzed. And analysis of new data or old data is better scales is often the best way to drive the development of novel theories, hypotheses, and tests. For these reasons we believe that automated recording of natural behavior is likely, by itself, to produce important insights.

### 5 | WHAT BEHAVIORAL IMAGING OFFERS ZOOS AND PRIMATE CENTERS

Much of the behavioral imaging work to be done can occur at specialized sites such as zoos and primate centers. These places have their own agendas and their own group of engaged stakeholders. While research is often a major goal, it is not the only one, and even when research is a priority, there are many researchers with competing demands. It is worth exploring, therefore, how the benefits that behavioral imaging offers can also benefit zoos and primate centers.

Indeed, many of these benefits are overlapping. For example, zoos and primate centers have a great interest in keeping their animals healthy, and that means that any improvements in diagnosis and detection of latent disease states can help them in their own mission.

Consider, for example, the benefits offered by an always-on behavioral tracking system that identifies pose for all primates in an enclosure, and then quantifies behavioral state. This system could perform monitoring at night and holidays when staff is absent and send warnings to on-call veterinary staff. It could also serve as a tracking history for an animal that needs diagnosis and ask how long a particular behavior has been evident, how long it has taken to progress whether this animal is showing improvement, and so on. It could free up time for the veterinary staff to focus on other issues, such as enrichment, that would directly improve the lives of the animals on site, thus leading to secondary welfare benefits.

Another example would come from enrichment interventions. Zoo and primate center staff typically have at hand multiple possible enrichment opportunities, but very little direct data from their own subjects about which enrichment opportunities have the greatest desired effect. And the desired effect may be species dependent, and it may be individual dependent, and may even be context dependent. When an animal is anxious, the best enrichment may be different than when the animal is recovering from a broken leg. The ability to measure behavior and automatically evaluate the efficacy of specific enrichment interventions is likely to lead to fast advances in welfare.

It is important to note that these benefits go hand in hand with the research benefits – the tracking that benefits welfare and health can also produce data that is of interest to scientists, thus making scientific exploration more readily available to scholars interested in it.

Finally, it is worth noting that there are potential conservation benefits to behavioral imaging. Many insights obtained from artificial environments can be used to gain insights into how animals behave in their natural environments and develop insights that can be used to protect them in those environments. Second, they can be used to test hypotheses about how best to reintegrate animals into natural environments, and speed and improve that process. And third, by improving welfare in unnatural environments, imaging can help keep animals healthy and productive, so they (or their progeny) are ready to return to the environment at a later time.

## 6 | LIMITATIONS OF BEHAVIORAL IMAGING

The present review is meant to highlight the virtues of behavioral imaging. Of course, it has many limitations as well. First, it is nontrivial to implement pose tracking in any environment, even in a laboratory. It is undoubtedly more difficult in environments tailored to other goals. Successful installation of a working system may require multiple computer scientists and electrical engineers. This level of complexity is probably unavoidable right now, but may be made simpler in the future as tracking systems – both software and hardware – improve.

Second, there is a basic limit on the amount of information that can be gained solely from behavioral imaging. A good deal of what goes on inside an animal is not visible to the outside world. For

example, an animal hiding in the corner may be afraid, maybe depressed, maybe tired, or may just be actively observing the environment from a safe vantage point. It is very possible even in theory that there is no way given the behavior alone to draw a robust inference about the animal's true state. Indeed, animals may expend effort to hide or conceal their status, such as if they are injured. That is, there may be some information – perhaps a great deal – of information that is not detectable by behavior alone.

Third, even among behavior that is expressed, some of it requires high-quality tracking that may be impractical in zoo environments. To give a few examples, there may be important information available about the animal's internal state that is visible in facial expressions, fine hand movements, and gaze directions. Indeed, eyes can give information about the locus of attention, about intentions, about priorities, and about beliefs; these things can in turn provide information about the animal's internal state. Even the pupil size of the animal can give a great deal of information about the animal's beliefs, expectations, surprise, reward function, attention, arousal, and urges. This is information that is available in behavior but extremely difficult to extract in practice.

Finally, it is worth emphasizing again that zoos and primate centers have many competing goals, and that benefiting scientist is not the only one. As such, behavioral imaging must satisfy the constraints imposed by the competing goals of the recording site. These can include unobtrusiveness, limits to a number of cameras, limits to power, and storage. It can also involve political calculations, which may involve coordination among stakeholders.

## 7 | NEW ROLES FOR PRIMATE CENTERS AND ZOOS IN THE ERA OF BIG BEHAVIOR

A great deal of primate research takes place in the laboratory. In contrast, zoos have historically been used for research less than they could. And while primate centers typically have a strong research role, this role conflicts with their other roles, including as a supplier of primates, and with their own research agenda, which may conflict with that of any given scientist. We believe that both zoos and primate centers' resources are poised to become more useful in the era of big behavior.

The underutilization of primate centers and zoos stems from several specific factors that, on balance, give the laboratory distinct advantages. These include the fact that laboratories offer freedom to perform invasive experiments, give access to research animals full time, and the ability to use animals solely for research. Laboratories also offer the ability to focus on any of a small number of well-studied research organisms, with the benefits that standardization of research animals brings. While primate centers have some of these benefits, the primates located there are often needed for breeding, and competition for access to them can be strong. We believe that behavioral imaging mitigates some of the limitations of zoos and primate centers, and will make them more attractive locations to perform research.



**FIGURE 5** The orangutan enclosure at the Toronto Zoo allows for panoramic viewing of the research subjects, who have a large and enriched space to move in. Such enclosures are particularly useful for behavioral imaging

### 7.1 | Zoos and primate centers are designed for observing natural behavior

Zoos are designed with public audiences in mind and are architected to maximize unobtrusive viewing from many vantage points (Figure 5). Indeed, they are laid out to have viewing in a way that does not bother animals and is also safe from their interference. This is ideal for big behavior. It allows measuring natural behaviors including social interactions with minimal interference (see above). While primate centers are often not designed for the public, they typically have the observational capacity for another reason – to allow staff to rapidly scan the housing for checking on animals and ascertaining if there are any problems. This is an advantage over many research laboratory environments, which are often constructed generically and without observation in mind. Many laboratories have tight space restrictions, which can result in imaging that is difficult or distorted. This is also a major advantage of zoos and primate centers relative to field research sites, which pose many practical problems for observation. For example, field sites are often remote and far from power and repair facilities, and studies can last multiple weeks.

### 7.2 | Zoos and primate centers have many individuals

While laboratory studies of primates typically focus on a few individuals (in many cases, two per study), zoos and primate centers have access to many more. Even in zoos, in which numbers may be limited for a given species, this number is augmented by the national network of exchange of animals between zoos. Of course, laboratories can compensate for low numbers with a high degree of training and instrumentation per subject. But with tracking, these needs can be reduced. The larger number of individuals is important because it can reduce Type II errors, can improve generalizability, and can facilitate individual difference studies. Moreover, the relative ease of collecting behavioral imaging data means that data from multiple zoos can in principle be combined, thereby further increasing

numbers. In contrast, laboratory research is supported solely by funding for the research itself, so the cost of adding additional animals to a study is borne by the research funds.

### 7.3 | Zoos (and sometimes primate centers) have many species

Individual laboratories specialize in a handful of model organisms and focus on a deep search into the biology of those individuals. This lack of species diversity is a limit for biomedical research laboratories, as it reduces generalizability. In contrast, zoos typically have a large number of species. This breadth of species has several benefits. Perhaps the greatest is for comparative studies. Indeed, the lack of variety of species has been identified as a limitation of major biological and biomedical research in the modern era, and a limit on the generalizability of the resulting science. The variety is limited for practical reasons by the complexity of having to learn the intricacies of multiple animal species. Zoos already bear this cost. We can only imagine the benefits to the research of having access to a large number of species, each of which can be selected for a particular study tailored to the specific needs of that study (cf. Stevens et al., 2005).

### 7.4 | Zoos and primate centers have excellent records about their animals

Zoos and primate centers have access to their animals for long periods of time, which makes it possible to answer important questions about their ancestry, behavioral history, their DNA, endocrine profile, and so on. This is an advantage relative to field studies, where animal subjects are often catch-as-catch-can. Even in cases where the same individuals are followed for many years, there is an enormous cost to developing this knowledge, and that cost makes it harder to explore new species or groups of individuals. The record-keeping in zoos and primate centers is also better than the laboratory, where most animals are procured from suppliers with unknown details, or, in the case of primates from breeding sites where this information is often unavailable.

### 7.5 | Zoo and primate center environments are designed for welfare

Zoos and primate centers have a keen interest in keeping their animals happy and unstressed. Aside from statutory and regulatory reasons, these places have a strong incentive to do this. Zoo audiences like seeing happy animals, and happy animals are more usable by primate center researchers and purchasers. In contrast, laboratory environments are designed to optimize data collection. As a result, zoos provide opportunities for ethologically valid behavior. That type of behavior is invaluable. For example, psychiatric studies often rely

on the assumption that animal models are psychiatrically normal - deviations introduce biases which can, in turn, reduce the efficacy of treatment effects. Even nonpsychiatric studies, however, can be hampered by the introduction of stressors that serve as uncontrolled variables (at best) or as competitors to variables of interest.

## 8 | CONCLUSION

Behavioral imaging in primates is poised to provide many research benefits to biologists and medical researchers. The data such methods provide is greatly enriched compared to data derived from earlier methods, such as preference, reaction time, and gaze direction, which involves low amount of information about animals' internal states. This is not to minimize the value of such methods. Measure of such low information behavior has been critically important for many studies and has led to many important insights. Indeed, that approach has been a mainstay of our own labs. Our labs, among others, have been interested in studying more complex and naturalistic behaviors and have begun to do so, although, still using very simple measures like positions of single landmarks and gaze (Yoo et al., 2020; Yoo, Hayden, et al., 2021; Yoo, Tu, et al., 2021). Clearly, big questions can be asked with conventional approaches.

However, it's still looking at the world through a keyhole. With new technologies and new approaches, we are now able to open the door and step into the world and see behavior fully. It's still hard to predict all the effects that change will cause. That's partly because most of the analyses to study these data have not been invented. They will require new mathematical techniques that are only beginning to be delineated. Moreover, it's not clear how much of the animals' internal states leak out through their behavior. Nonetheless, we believe that these changes will lead to great advances in our understanding of the machinery of living things.

### AUTHOR CONTRIBUTION

**Benjamin Yost Hayden:** conceptualization (equal); supervision (equal).

**Hyun Soo Park:** conceptualization (equal); resources (equal).

### ETHICS STATEMENT

This article is a broad overview/review paper, and no original research was done for it. However, all research described herein complied with the American Society of Primatologists Ethical Principles for the Treatment of Non-Human Primates.

### ORCID

**Benjamin Y. Hayden**  <https://orcid.org/0000-0002-7678-4281>

**Jan Zimmermann**  <http://orcid.org/0000-0002-7678-4281>

### REFERENCES

- Alexander, L., Clarke, H. F., & Roberts, A. C. (2019). A focus on the functions of area 25. *Brain Sciences*, 9(6), 129.
- Anderson, D. J., & Perona, P. (2014). Toward a science of computational ethology. *Neuron*, 84(1), 18–31.
- Andriluka, M., Pishchulin, L., Gehler, P., & Schiele, B. (2014). 2d human pose estimation: New benchmark and state of the art analysis. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 3686–3693).
- Azab, H., & Hayden, B. Y. (2018). Correlates of economic decisions in the dorsal and subgenual anterior cingulate cortices. *European Journal of Neuroscience*, 47(8), 979–993.
- Bala, P. C., Eisenreich, B. R., Yoo, S. B. M., Hayden, B. Y., Park, H. S., & Zimmermann, J. (2020). Automated markerless pose estimation in freely moving macaques with OpenMonkeyStudio. *Nature Communications*, 11(1), 1–12.
- Beck, A. T., Steer, R. A., & Carbin, M. G. (1988). Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. *Clinical Psychology Review*, 8(1), 77–100.
- Biggs, A. T., Adamo, S. H., & Mitroff, S. R. (2014). Rare, but obviously there: Effects of target frequency and salience on visual search accuracy. *Acta Psychologica*, 152, 158–165.
- Blanchard, T. C., Wolfe, L. S., Vlaev, I., Winston, J. S., & Hayden, B. Y. (2014). Biases in preferences for sequences of outcomes in monkeys. *Cognition*, 130(3), 289–299.
- Bliss-Moreau, E., & Rudebeck, P. H. (2020). Animal Models of Human Mood. *Neuroscience & Biobehavioral Reviews*, 120, 574–582.
- Brando, S., & Buchanan-Smith, H. M. (2018). The 24/7 approach to promoting optimal welfare for captive wild animals. *Behavioural Processes*, 156, 83–95.
- Buffalo, E. A., Movshon, J. A., & Wurtz, R. H. (2019). From basic brain research to treating human brain disorders. *Proceedings of the National Academy of Sciences*, 116(52), 26167–26172.
- Calhoun, A., & El Hady, A. (2021). What is behavior? No seriously, what is it? *bioRxiv*.
- Calhoun, A. J., & Hayden, B. Y. (2015). The foraging brain. *Current Opinion in Behavioral Sciences*, 5, 24–31.
- Cao, Z., Hidalgo, G., Simon, T., Wei, S. E., & Sheikh, Y. (2019). OpenPose: Realtime multi-person 2D pose estimation using Part Affinity Fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(1), 172–186.
- Dankert, H., Wang, L., Hoopfer, E. D., Anderson, D. J., & Perona, P. (2009). Automated monitoring and analysis of social behavior in *Drosophila*. *Nature Methods*, 6(4), 297–303.
- Datta, S. R., Anderson, D. J., Branson, K., Perona, P., & Leifer, A. (2019). Computational neuroethology: A call to action. *Neuron*, 104(1), 11–24.
- Drevets, W. C., Savitz, J., & Trimble, M. (2008). The subgenual anterior cingulate cortex in mood disorders. *CNS Spectrums*, 13(8), 663–681.
- Drysdale, A. T., Grosenick, L., Downar, J., Dunlop, K., Mansouri, F., Meng, Y., & Liston, C. (2017). Resting-state connectivity biomarkers define neurophysiological subtypes of depression. *Nature Medicine*, 23(1), 28–38.
- Eisenreich, B. R., Hayden, B. Y., & Zimmermann, J. (2019). Macaques are risk-averse in a freely moving foraging task. *Scientific Reports*, 9(1), 1–12.
- Fang, H. S., Xie, S., Tai, Y. W., & Lu, C. (2017). Rmpe: Regional multi-person pose estimation. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 2334–2343).
- Gallup, G. G., Jr., Anderson, J. R., & Shillito, D. J. (2002). The mirror test. In M. Bekoff, C. Allen, & G. M. Burghardt (Eds.), *The cognitive animal: Empirical and theoretical perspectives on animal cognition* (pp. 325–333). MIT Press.
- Gazes, R. P., Lutz, M. C., Meyer, M. J., Hassett, T. C., & Hampton, R. R. (2019). Influences of demographic, seasonal, and social factors on automated touchscreen computer use by rhesus monkeys (*Macaca mulatta*) in a large naturalistic group. *PLoS One*, 14(4), e0215060.
- Gelardi, V., Godard, J., Paleressompoulle, D., Claidière, N., & Barrat, A. (2020). Measuring social networks in primates: Wearable sensors

- versus direct observations. *Proceedings of the Royal Society A*, 476(2236), 20190737.
- Geyer, M. A., & Markou, A. (1995). Animal models of psychiatric disorders. *Psychopharmacology*, 787–798.
- Goodman, W. K., Price, L. H., Rasmussen, S. A., Mazure, C., Fleischmann, R. L., Hill, C. L., & Charney, D. S. (1989). The Yale-Brown obsessive compulsive scale: I. Development, use, and reliability. *Archives of General Psychiatry*, 46(11), 1006–1011.
- Günel, S., Rhodin, H., Morales, D., Campagnolo, J., Ramdya, P., & Fua, P. (2019). DeepFly3D, a deep learning-based approach for 3D limb and appendage tracking in tethered, adult *Drosophila*. *eLife*, 8:e48571.
- Hayden, B. Y. (2016). Time discounting and time preference in animals: A critical review. *Psychonomic Bulletin & Review*, 23(1), 39–53.
- Hayden, B. Y. (2018). Economic choice: The foraging perspective. *Current Opinion in Behavioral Sciences*, 24, 1–6.
- Holman, L., Head, M. L., Lanfear, R., & Jennions, M. D. (2015). Evidence of experimental bias in the life sciences: Why we need blind data recording. *PLoS Biology*, 13(7), e1002190.
- Johnson, M. D., Lim, H. H., Netoff, T. I., Connolly, A. T., Johnson, N., Roy, A., & He, B. (2013). Neuromodulation for brain disorders: Challenges and opportunities. *IEEE Transactions on Biomedical Engineering*, 60(3), 610–624.
- Joska, D., Clark, L., Muramatsu, N., Jericevich, R., Nicolls, F., Mathis, A., & Patel, A. (2021). AcinoSet: A 3D pose estimation dataset and baseline models for cheetahs in the Wild. *arXiv preprint arXiv:2103.13282*.
- Kardish, M. R., Mueller, U. G., Amador-Vargas, S., Dietrich, E. I., Ma, R., Barrett, B., & Fang, C. C. (2015). Blind trust in unblinded observation in ecology, evolution, and behavior. *Frontiers in Ecology and Evolution*, 3, 51.
- Koob, G. F., & Zimmer, A. (2012). Animal models of psychiatric disorders. *Handbook of clinical neurology*, 106, 137–166.
- Krakauer, J. W., Ghazanfar, A. A., Gomez-Marin, A., Maclver, M. A., & Poeppel, D. (2017). Neuroscience needs behavior: Correcting a reductionist bias. *Neuron*, 93(3), 480–490.
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 25, 1097–1105.
- Kryszczuk, K. M., & Boyce, P. R. (2002). Detection of slow light level reduction. *Journal of the Illuminating Engineering Society*, 31(2), 3–10.
- Labuguen, R., Matsumoto, J., Negrete, S. B., Nishimaru, H., Nishijo, H., Takada, M., & Shibata, T. (2020). MacaquePose: A novel “in the wild” macaque monkey pose dataset for markerless motion capture. *Frontiers in Behavioral Neuroscience*, 14.
- Levitis, D. A., Lidicker, Jr., W. Z., & Freund, G. (2009). Behavioural biologists do not agree on what constitutes behaviour. *Animal Behaviour*, 78(1), 103–110.
- Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., & Zitnick, C. L. (2014). Microsoft coco: Common objects in context. *European Conference on Computer Vision* (pp. 740–755). Springer.
- Lozano, A. M., Mayberg, H. S., Giacobbe, P., Hamani, C., Craddock, R. C., & Kennedy, S. H. (2008). Subcallosal cingulate gyrus deep brain stimulation for treatment-resistant depression. *Biological Psychiatry*, 64(6), 461–467.
- Maisson, D. J. N., Cash-Padgett, T. V., Wang, M. Z., Hayden, B. Y., Heilbronner, S. R., & Zimmermann, J. (2021). Choice-relevant information transformation along a ventrodorsal axis in the medial prefrontal cortex. *Nature Communications*, 12(1), 1–14.
- Marks, M., Qiuhuan, J., Sturman, O., von Ziegler, L., Kollmorgen, S., von der Behrens, W., & Yanik, M. F. (2021). Deep-learning based identification, pose estimation and end-to-end behavior classification for interacting primates and mice in complex environments. *bioRxiv*, 2020-10.
- Marshall, J. D., Aldarondo, D. E., Dunn, T. W., Wang, W. L., Berman, G. J., & Ölveczky, B. P. (2021). Continuous whole-body 3D kinematic recordings across the rodent behavioral repertoire. *Neuron*, 109(3), 420–437.
- Mathis, A., Mamidanna, P., Cury, K. M., Abe, T., Murthy, V. N., Mathis, M. W., & Bethge, M. (2018). DeepLabCut: Markerless pose estimation of user-defined body parts with deep learning. *Nature Neuroscience*, 21(9), 1281–1289.
- Mathis, A., Schneider, S., Lauer, J., & Mathis, M. W. (2020). A primer on motion capture with deep learning: Principles, pitfalls, and perspectives. *Neuron*, 108(1), 44–65.
- Mayberg, H. S., Lozano, A. M., Voon, V., McNeely, H. E., Seminowicz, D., Hamani, C., & Kennedy, S. H. (2005). Deep brain stimulation for treatment-resistant depression. *Neuron*, 45(5), 651–660.
- McGrath, C. L., Kelley, M. E., Holtzheimer, P. E., Dunlop, B. W., Craighead, W. E., Franco, A. R., & Mayberg, H. S. (2013). Toward a neuroimaging treatment selection biomarker for major depressive disorder. *JAMA Psychiatry*, 70(8), 821–829.
- Nath, T., Mathis, A., Chen, A. C., Patel, A., Bethge, M., & Mathis, M. W. (2019). Using DeepLabCut for 3D markerless pose estimation across species and behaviors. *Nature Protocols*, 14(7), 2152–2176.
- Negrete, S. B., Labuguen, R., Matsumoto, J., Go, Y., Inoue, K. I., & Shibata, T. (2021). Multiple Monkey Pose Estimation Using OpenPose. *bioRxiv*.
- Newell, A., Yang, K., & Deng, J. (2016). Stacked hourglass networks for human pose estimation. *European conference on computer vision* (pp. 483–499). Springer.
- Pearson, J. M., Watson, K. K., & Platt, M. L. (2014). Decision making: The neuroethological turn. *Neuron*, 82(5), 950–965.
- Pereira, T. D., Aldarondo, D. E., Willmore, L., Kislin, M., Wang, S. S. H., Murthy, M., & Shaevitz, J. W. (2019). Fast animal pose estimation using deep neural networks. *Nature Methods*, 16(1), 117–125.
- Pereira, T. D., Shaevitz, J. W., & Murthy, M. (2020). Quantifying behavior to understand the brain. *Nature Neuroscience*, 23(12), 1537–1549.
- Roberts, A. C., & Clarke, H. F. (2019). Why we need nonhuman primates to study the role of ventromedial prefrontal cortex in the regulation of threat-and reward-elicited responses. *Proceedings of the National Academy of Sciences of the United States of America*, 116(52), 26297–26304.
- Rudebeck, P. H., Rich, E. L., & Mayberg, H. S. (2019). From bed to bench side: Reverse translation to optimize neuromodulation for mood disorders. *Proceedings of the National Academy of Sciences of the United States of America*, 116(52), 26288–26296.
- Santos, L. R., & Rosati, A. G. (2015). The evolutionary roots of human decision making. *Annual Review of Psychology*, 66, 321–347.
- Schiff, S. J. (2010). Towards model-based control of Parkinson's disease. *Philosophical Transactions of the Royal Society A: Mathematical Physical and Engineering Sciences*, 368(1918), 2269–2308.
- Shively & Willard (2012). Behavioral and neurobiological characteristics of social stress versus depression in nonhuman primates. *Experimental Neurology*, 233(1), 87–94.
- Stevens, J. R., Hallinan, E. V., & Hauser, M. D. (2005). The ecology and evolution of patience in two New World monkeys. *Biology Letters*, 1(2), 223–226.
- Testard, C., Tremblay, S., & Platt, M. (2021). From the field to the lab and back: neuroethology of primate social behavior. *Current Opinion in Neurobiology*, 68, 76–83.
- Tuytens, F. A. M., de Graaf, S., Heerkens, J. L., Jacobs, L., Nalon, E., Ott, S., & Ampe, B. (2014). Observer bias in animal behaviour research: Can we believe what we score, if we score what we believe? *Animal Behaviour*, 90, 273–280.
- Voloh, B., Eisenreich, B. R., Maisson, D. J. -N., Ebitz, R. B., Park, H. S., Hayden, B. Y., & Zimmermann, J. (2021). Hierarchical organization of rhesus macaque behavior. *bioRxiv*.
- Walter, T., & 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.

- Wei, S. E., Ramakrishna, V., Kanade, T., & Sheikh, Y. (2016). Convolutional pose machines. *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 4724–4732.
- Wolfe, J. M., Horowitz, T. S., & Kenner, N. M. (2005). Rare items often missed in visual searches. *Nature*, 435(7041), 439–440.
- Yao, Y., Mohan, A. A., Bliss-Moreau, E., Coleman, K., Freeman, S. M., Machado, C. J., Zimmermann, J., Hayden, B. Y., & Park, H. S. (2021). OpenMonkeyChallenge: Dataset and Benchmark Challenges for Pose Tracking of Non-human Primates. *bioRxiv*.
- Yoo, S. B. M., Hayden, B. Y., & Pearson, J. M. (2021). Continuous decisions. *Philosophical Transactions of the Royal Society B*, 376(1819), 20190664.
- Yoo, S. B. M., Tu, J. C., & Hayden, B. Y. (2021). Multicentric tracking of multiple agents by anterior cingulate cortex during pursuit and evasion. *Nature Communications*, 12(1), 1–14.
- Yoo, S. B. M., Tu, J. C., Piantadosi, S. T., & Hayden, B. Y. (2020). The neural basis of predictive pursuit. *Nature Neuroscience*, 23(2), 252–259.
- von Ziegler, L., Sturman, O., & Bohacek, J. (2021). Big behavior: Challenges and opportunities in a new era of deep behavior profiling. *Neuropsychopharmacology*, 46(1), 33–44.

**How to cite this article:** Hayden, B. Y., Park, H. S., & Zimmermann, J. (2021). Automated pose estimation in primates. *American Journal of Primatology*, e23348. <https://doi.org/10.1002/ajp.23348>