Please cite this article in press as: Mobbs et al., Promises and challenges of human computational ethology, Neuron (2021), https://doi.org/10.1016/j.neuron.2021.05.021

Neuron

CellPress

Perspective

Promises and challenges of human computational ethology

Dean Mobbs,[1,2,10,*] Toby Wise,[1,3,4,10] Nanthia Suthana,[5,6] Noah Guzmán,[2] Nikolaus Kriegeskorte,[7,8] and Joel Z. Leibo[9]
[1]Department of Humanities and Social Sciences, 1200 E. California Blvd., HSS 228–77, Pasadena, CA 91125, USA
[2]Computation and Neural Systems Program at the California Institute of Technology, 1200 E. California Blvd., HSS 228–77, Pasadena, CA 91125, USA
[3]Wellcome Centre for Human Neuroimaging, University College London, London, UK
[4]Max Planck UCL Centre for Computational Psychiatry and Ageing Research, University College London, London, UK
[5]Department of Psychiatry and Biobehavioral Sciences, Jane and Terry Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, Los Angeles, CA, USA
[6]Departments of Neurosurgery, Psychology, and Bioengineering, University of California, Los Angeles, Los Angeles, CA, USA
[7]Department of Psychology, Columbia University, New York, NY, USA
[8]Department of Neuroscience, Zuckerman Mind Brain Behavior Institute, Columbia University, New York, NY, USA
[9]DeepMind, London, UK
[10]These authors contributed equally
*Correspondence: dmobbs@caltech.edu
https://doi.org/10.1016/j.neuron.2021.05.021

SUMMARY

The movements an organism makes provide insights into its internal states and motives. This principle is the foundation of the new field of computational ethology, which links rich automatic measurements of natural behaviors to motivational states and neural activity. Computational ethology has proven transformative for animal behavioral neuroscience. This success raises the question of whether rich automatic measurements of behavior can similarly drive progress in human neuroscience and psychology. New technologies for capturing and analyzing complex behaviors in real and virtual environments enable us to probe the human brain during naturalistic dynamic interactions with the environment that so far were beyond experimental investigation. Inspired by nonhuman computational ethology, we explore how these new tools can be used to test important questions in human neuroscience. We argue that application of this methodology will help human neuroscience and psychology extend limited behavioral measurements such as reaction time and accuracy, permit novel insights into how the human brain produces behavior, and ultimately reduce the growing measurement gap between human and animal neuroscience.

INTRODUCTION

In the natural world, animals modify their behavior in response to changes in their environment, such as predation and competition, as well as changes in their internal metabolic drives (e.g., hunger and thirst; LeDoux, 2012; Mobbs et al., 2018). These observable behaviors can range from deliberately controlled to impulsive or reactive but are consistent in that they provide information about the animal's latent motivational state and reflect strategic responses to a variety of survival demands. Measuring these unconstrained and seemingly random behaviors in extensive datasets has been a major challenge for behavioral neuroscientists. However, in non-human studies, machine learning methods that automate the registration and analysis of locomotor activity on the basis of video have given us much richer measurements of behavior. By better characterizing the behavioral motifs (see Box 1) of animals, a deeper understanding of the neural circuits involved in a rich variety of survival behaviors can be achieved (Anderson and Perona, 2014; Datta et al., 2019). This computational enhancement of behavioral observation has created the new field of computational ethology. Its methods appear to be equally pertinent to humans and animals. However, these methods have yet to develop their full effect on human psychology and human neuroscience.

Here, we explore the promises and challenges of applying the methods of computational ethology to human neuroscience. We focus on the fields of human decision and social and affective neuroscience and discuss ways in which experimental paradigms can be designed to evoke a wide range of natural defensive, appetitive, and social behaviors. We detail current and potential future methods, focusing on increased use of virtual ecologies to probe behavioral motifs and their underlying computations. This potential shift in approach to human neuroscience parallels recent calls from the field of comparative neuroscience, where there is a need for more effective probing of behavior (Anderson and Perona, 2014; Datta et al., 2019). In turn, this will provide better models of how the brain produces behaviors (Babayan and Konen, 2019; Niv, 2020; Krakauer et al., 2017). As we argue throughout this review, if our goal is ultimately to explain real-world behavior, we will need to study naturalistic behavior and its neural underpinnings. Such an approach will (1) allow detection of naturalistic behavioral patterns that are hidden in traditional, constrained

Please cite this article in press as: Mobbs et al., Promises and challenges of human computational ethology, Neuron (2021), https://doi.org/10.1016/j.neuron.2021.05.021

Box 1. Glossary

Behavioral motif. A unit of organized movement often used interchangeably with the terms "motif," "movement," "module," "primitive," and "syllable" (Datta et al., 2019; Anderson and Perona, 2014).
Dimensionality. The number of variables in a dataset (also see "Dimensionality reduction"; Datta et al., 2019).
Ethogram. A repeatable and predefined set of movements that are learned or hard-wired. These include such things as thigmotaxis (see below), approach, and pauses.
Machine learning (ML). Where computers are programmed to learn without explicit instructions, providing accurate predictions based on labeled or unlabeled training data. Using a variety of mathematical models, including support vector machines or deep neural networks, a dataset is first used to train the algorithms. When the algorithm is trained, it is then tested on a test dataset. In the case of human behavior, ML can be used to detect and categorize human and animal behavioral motifs using what has been called an "action classifier."
Model-based (MB) inference and decision-making. Inferences and decisions based on an internal model of the world. MB methods exploit a (possibly learned) model of the environment to prospectively calculate the likely consequences of actions, for instance, by simulating possible future states.
Model-free (MF) inference and decision-making. The agent learns what to do to maximize long-run return or learns value estimates of those long-run returns. MF methods acquire values by a bootstrapping process of enforcing consistency between successive estimates.
Protean escape. Unpredictable escape trajectories, such as zigzagging, that prevent a predator anticipating the future position or actions of its prey.
Temporal dynamics. How behavior features change over time.
Trajectory. The movement of the agent through time and space.
Thigmotaxis. A measure of anxiety where animals stay close to walls rather than maneuvering in open spaces.
Virtual ecology. Self-contained virtual environments where subjects can move freely throughout the environment.
For more definitions, see Anderson and Perona (2014) and Datta et al. (2019).

experimental paradigms; (2) characterize neural systems supporting spontaneous, naturalistic behavior; and (3) determine how cognitive processes (e.g., decision-making) unfold in complex, naturalistic scenarios. So far, computational ethology has led to a range of novel computationally identified behaviors and their associated neural basis being cataloged in rodents and *Drosophila*. These include behaviors that reflect appetitive, social, and defensive behaviors (Box 1; Anderson and Perona 2014) that, up until now, were frequently passed off as noise, being too fast and too stochastic to measure. But neither these challenges nor their emerging solutions are unique to animal research. We contend that computational ethology methods will also prove critical for the progress of human neuroscience.

Non-human computational ethology

Description and analysis of animal behavior has traditionally relied on human observation and recording. In recent years, however, the development of modern recording and analysis methods has facilitated the emergence of computational ethology (Anderson and Perona, 2014; Datta et al., 2019), which uses machine learning methods to automatically identify and quantify behavior, obviating the need for human observers. In lab settings, these approaches typically take data acquired from video cameras positioned around one or more animals and put out a continuous representation of the animal's location or pose, for example, by recording limb or head position. This circumvents the subjectivity of human observations and promises observations higher in precision and quality, being unaffected by human visual and attentional capacity. Perhaps most importantly, computational ethology provides a dramatic increase in throughput, the benefits of which have been felt most strongly in fields that depend on high-frequency observations from large numbers of animals, such as those examining the roles of specific neural circuits in *Drosophila* (Hoopfer et al., 2015).

Although a new field, computational ethology has already demonstrated its utility across a range of studies. In *Drosophila*, these methods have allowed identification of neural circuits underlying distinct sensorimotor states (Calhoun et al., 2019), and combining computational ethology with optogenetics has enabled causal links between neural circuits and specific behaviors to be tested (Jovanic et al., 2016). Recently, the combination of automated classification of behavior with high-throughput neural recordings from rodents has revealed distributed patterns of neural activity regarded previously as noise to be associated with specific behavioral patterns (Musall et al., 2019; Stringer et al., 2019), a discovery with obvious relevance to "noisy" human neuroimaging.

Computational ethology has been a major beneficiary of developments in machine vision, which allows streams of video data to be mined automatically for behaviors of interest. This is not a trivial task. First, it is essential to continuously detect and monitor unique animals without confusing separate individuals. Second, animal features that constitute specific behaviors must be extracted and tracked accurately. Although simply tracking an animal's location and direction of movement will be sufficient to answer many questions, a behavioral phenotype often depends on more complex features of the animal's actions. These may be subtle, for example, based on pose or specific limb movements. Finally, classification of behaviors based on these features must be accurate. This process is approached in a supervised way (Graving et al., 2019; Mathis

et al., 2018; Pereira et al., 2019), where an experimenter manually identifies the features of the animal that should be tracked, or an unsupervised way (Berman et al., 2014; Wiltschko et al., 2015), where features of interest are identified without human intervention.

The number of tools developed for this purpose has increased dramatically over the past decade, boosted by the growth of deep learning. Experimental setups used in computational ethology represent the kind of nonlinear many-to-many classification problem (where many features must be mapped to many categories) at which neural networks excel, and their use has enabled automatic identification and classification of behavioral features previously limited to human observation. These methods have evolved from algorithms designed to detect human poses (Insafutdinov et al., 2016) and have been made possible through transfer learning (Donahue et al., 2013), which takes advantage of pre-trained neural networks to facilitate performance without the need for extensive training data from the task at hand. Notable examples here are those that have been able to identify animal body and limb positions (Graving et al., 2019; Mathis et al., 2018), enabling automatic evaluation of movements and pose and classification of behaviors based on these. The result of this is that continuous video recordings of naturalistic animal behavior can be processed automatically, producing detailed ethograms representing how behavior unfolds over time.

In both human and non-human research, we face the challenge of understanding how complex high-dimensional behaviors are generated by neural systems. Computational ethology naturally lends itself to this problem. First, computational methods allow fine-grained assessment of behavior that accounts for its temporal dynamics. Importantly, this permits the dynamics of behavior to be linked to unfolding neural activity, potentially elucidating time-dependent neural processes underlying behavior. Second, computational ethology allows detailed quantification of free, naturalistic movements, making it possible to link neural processes to behavior without the need for highly controlled, unrealistic tasks.

Human computational ethology

It has been argued recently (e.g., Babayan and Konen, 2019; Balleine, 2019; Niv, 2020) that behavior is essential for understanding the animal brain, including the human brain. The continuous measurement of behavior that characterizes computational ethology may add to existing measures of behavior, including categorical decisions that are common to most experiments in human cognitive neuroscience (Figures 1A–1J). It is not, however, obvious how to best integrate computational ethological approaches into traditional paradigms in human neuroscience. Consider, for example, the study of human fear and anxiety. Traditional approaches include fear conditioning and presentation of visually aversive stimuli (e.g., fearful faces). These studies provide no clear path to the approaches advocated by computational ethology because these paradigms minimize behavioral dynamics (e.g., binary button presses), in contrast with the rich behavioral outputs that are central to animal computational ethology. However, new experimental paradigms have begun to engage subjects in dynamic interactions, typically in virtual environments. One of the first studies to move beyond classic fear conditioning paradigms used virtual predators to create an active escape task. Although simple, the task involved subjects actively escaping from an attacking virtual predator in a 2D maze that allowed the subjects to visually keep track of the distance to the predator and providing richer behavior than typical for these tasks (Mobbs et al., 2007). Distance, therefore, could be parametrically coupled with blood-oxygen-level-dependent (BOLD) signal measurements taken from fMRI, permitting identification of neural circuits involved in processing proximal and distal threats. This approach was later used and extended by showing that panic-related motor errors correlated with brain areas commonly implicated in human and animal models of panic (Figure 1B; Box 1; wrong button presses resulting in collisions with the virtual walls of the maze; Mobbs et al., 2009). Despite the promise of these early studies, more causal research is needed, and computational ethology may be one direction that can address the shortfalls of previous empirical work.

These experiments were followed by several studies using similar paradigms (Bach et al., 2014; Gold et al., 2015; Meyer et al., 2019). For example, Bach et al. (2014) examined the role of the human hippocampus in arbitrating approach-avoidance conflict under different levels of potential threat. Subjects were instructed to move a green triangle around a 2D gridded environment to collect tokens (exchanged for money), where one of three different dangerous predators was located in a corner of the grid. At any time, the predator could begin to chase the subject, but the subject could also choose to hide in a safe place (a black box in the corner of the grid). When caught, the subject lost all tokens, and the epoch was over. This task was able to measure several variables, including time spent in the safe place, time spent close to the walls, and distance from the threat (Bach et al., 2014; Figure 1E). In another study, Gold et al. (2015) created a task in which subjects were asked to capture prey and evade predators in a 2D maze, similar to previous studies (Bach et al., 2014; Mobbs et al., 2007, 2009). The main result showed that, when a threat was unpredictable, there was increased connectivity between the amygdala and ventromedial prefrontal cortex (vmPFC) (Gold et al., 2015).

Although these studies did not take full advantage of the movement trajectories of the virtual environment or apply unsupervised machine learning, they do provide a template for how to apply computational ethology to human neuroscience. They also highlight the advantages of using less restricted behavioral measures than common in human neuroscience to reveal behavioral and neural patterns that would not be observable otherwise. Creation of such virtual ecologies enables experimentalists to measure less constrained types of behavior (Figure 1).

Examples of novel behavioral assays in humans

The standard behavioral measures used in laboratory tasks are decision accuracy and reaction time (RT). RT is often used as a marker of decision confidence, deliberation, and learning. However, although it does not undermine the importance of RT, it can be problematic because latencies in RT can arise because of many factors, including those of little interest to the experimenter, such as tiredness and distraction. The additional measures of behavioral motifs and sequences help minimize these

Please cite this article in press as: Mobbs et al., Promises and challenges of human computational ethology, Neuron (2021), https://doi.org/10.1016/j.neuron.2021.05.021

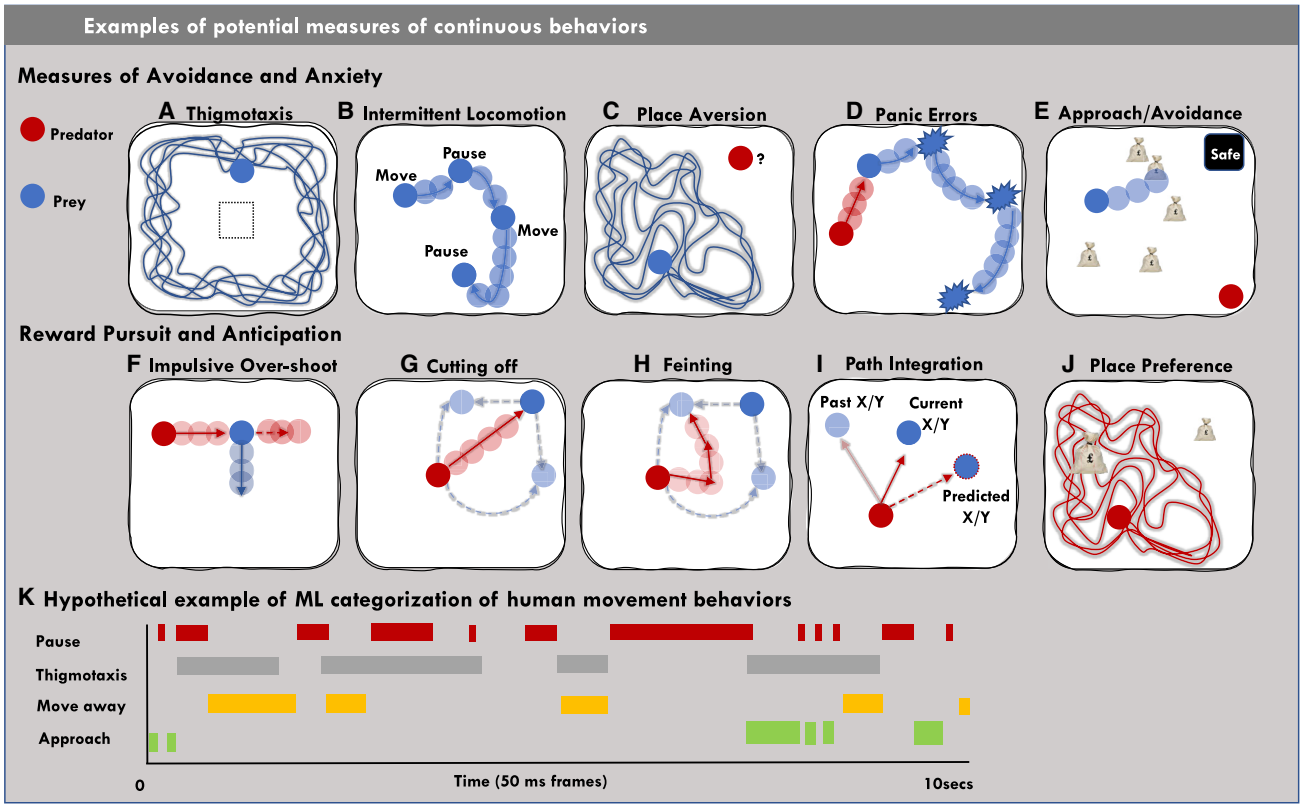

Figure 1. Examples of defensive, avoidance, and appetitive behaviors

(A) Thigmotaxis as a measure of avoidance.
(B) Intermittent locomotion as a measure of cautiousness.
(C) Place aversion.
(D) Escape measures include panic-related errors that reflect motor errors when being pursued by a virtual predator (Mobbs et al., 2009).
(E) Example of an approach/avoidance task used by Bach et al. (2014).
(F) Impulsive behaviors such as overshooting the reward (Mobbs et al., 2009).
(G) Cutting off behaviors, where the area under the curve can be measured to examine attack success.
(H) Feinting is a way of tricking the prey into moving into the direction the predator wishes them to move.
(I) Example of path integration and prediction of prey movement during an appetitive movement task (Yoo et al., 2020).
(J) Place preference can be measured by measuring the subject's time in specific positions of the virtual ecology.
(K) A hypothetical example of how software tracks behavior over time to produce detailed ethograms.

potentially confounding effects. In many tasks, decision accuracy is also used as a measure of learning (see *Neuron*'s special issue on behavior from October 2019), which encourages decision-making paradigms to depend on infrequent decisions between a limited set of options to provide clearly delineated opportunities for learning to occur. Dynamic measurement of complex behaviors replaces the RTs and accuracies of a discrete sequence of actions by an essentially continuous stream of motor control signals and performance measures. This, in turn, provides an excellent way to examine between-subject differences in behavior. Moving beyond the standard protocols, paradigms have been developed based on simple 2D environments or virtual ecologies that can capture multiple measures of threat anticipation, escape, and conflict (Figure 1A). Virtual ecologies can move one step closer to the real world by including levels of threat imminence, visually clear versus opaque environments (e.g., forest versus open field), and changes in competition density (Mobbs et al., 2013; Silston et al., 2020). This approach provides the experimentalist not only with tools to question how the environment changes decision processes but also how it affects locomotor activity. Indeed, in the real world, behaviors can be fluid, stilted, fleeting, and subtle. Below, we give examples of several types of behavior that can be measured using virtual ecologies.

Anticipation of danger

Anticipation of danger is a critical part of anxiety, which is typically defined as a future-oriented emotional state associated with "potential" and "uncertain" threats (Grupe and Nitschke, 2013). One classic measure observed by ethologists and behavioral neuroscientists is thigmotaxis (Box 1), an index of anxiety typically associated with the animal moving to the peripheral area of an open field. Thigmotaxis is observed in rodents, fish, and humans (Walz et al., 2016). Other anxiety-like behaviors include intermittent locomotion (i.e., movement pauses) when a threat is anticipated. Place aversion, a form of Pavlovian conditioning, is also possible in virtual ecologies, where avoidance of

particular areas of the environment demonstrates aversion (e.g., a predator was encountered in the location). Another approach taken from the field of behavioral ecology is the concept of margin of safety, when prey adopts choices that prevent deadly outcomes from occurring by keeping close proximity to a safety refuge and increasing the success of escape (Cooper et al., 2015; Qi et al., 2020). Finally, potential threats lead to vigilance behaviors, including orienting toward and attending to threat, both in non-human species and in humans (Mobbs and Kim, 2015; Wise et al., 2019).

We do not wish to argue that the results of traditional, non-naturalistic tasks are incorrect or entirely invalid. Their ability to break down complex behaviors into their constituent parts has provided substantial insight into the basic processes governing behavior. However, we do argue that a reductionist approach limits their ability to explain naturalistic behavior as a whole. Taking fear conditioning as an example, we now have a rich understanding of the specifics of how humans acquire and lose fears of stimuli linked to aversive outcomes in the lab. However, typical fear conditioning experiments involve repeated, contiguous pairing of unconditioned and conditioned stimuli to engender learning. In the real world, such clean learning experiences are the exception rather than the norm. For example, a student may become fearful of exams after receiving a poor grade on a single exam taken weeks previously without needing repeated experiences of taking exams and receiving immediate negative feedback. In terms of behavioral output, traditional tasks fail to capture the complexity of human behavior. Fear conditioning studies may require subjects to provide an expectancy rating or, in some tasks, may require binary stimulus selection. In the real world, our behaviors in response to feared stimuli are far more varied and complex, but traditional fear-conditioning paradigms tell us little about how acquired fears influence these behaviors. This gap between lab-based fear conditioning and real-world fear has been described previously as a barrier to successful treatment of pathological fears (Scheveneels et al., 2016).

Escape behaviors

Escape is associated with fear and is elicited during predatory attack (Mobbs et al., 2020). Escape differs from avoidance in that escape is driven by the moment-to-moment adjusted movements of the attacking predator. Behaviorally, escape is associated with ballistic movements and increased vigor and is less coordinated than avoidance. This also results in protean escape, which is driven by the trajectory of the predator's attack and often results in unpredictable flight (e.g., zigzagging, spinning) (Humphries and Driver, 1970). The first human studies of escape and its neural correlates used a virtual predator that chased subjects in a 2D maze and examined the shift of activity in the brain as the threat came closer or went farther away (Mobbs et al., 2007, 2009). More recent work has used flight initiation distance (FID), or the distance at which the subject flees from the approaching threat. FID is a spatiotemporal measure of threat sensitivity and economic decision-making (Ydenberg and Dill, 1986). In a recent study, Qi et al. (2018) measured a subject's volitional fleeing distance when they encounter a virtual predator. This study was the first to examine escape decisions in humans and, importantly, showed that different parts of the defensive circuits were engaged for fast- and slow-attacking threats (Fung et al., 2019; Qi et al., 2018).

Appetitive behaviors: Pursuit and hunting

Several experiments have used virtual ecologies to measure reward activity. These include foraging for rewards (Bach et al., 2014; Gold et al., 2015), and chasing prey for reward (Mobbs et al., 2009). In nature, appetitive behaviors take several forms, including approach, increased vigor, exploration, stealth, and sit and wait behaviors associated with surprise attack. Other examples include movement strategies when pursuing virtual prey, including angle of attack (e.g., cutting off corners to reduce escape time and feinting; Figure 1), place preference, and impulsive errors, such as overshooting the prey's anticipated movements (Mobbs et al., 2009). Using similar 2D environments as in human studies (Mobbs et al., 2007; Bach et al., 2014), a recent study with non-human primates showed how the dorsal anterior cingulate cortex is involved in pursuit predictions, including velocity, prey position, and acceleration (Yoo et al., 2020). Finally, some studies have taken advantage of virtual reality (VR) to explore human place preference in 3D environments. This has been demonstrated across primary (Astur et al., 2014) and secondary (Astur et al., 2016; Molet et al., 2013) reinforcers, allowing simple conditioning paradigms to be extended to more realistic environments.

Dynamic switching, arbitration, and conflict between circuits

Virtual ecologies allow dynamic switches in behavior. This, of course, can be measured in conventional task designs (e.g., task switching); however, in virtual ecologies, the switches can be reactive, volitional, or ramped up, providing a unique way to study the human brain. One example is the active escape task, where results showed that, when the artificial predator is distant, increased activity is observed in the vmPFC. However, as the artificial predator moves closer, a switch to enhanced activation in the midbrain periaqueductal gray (PAG) is observed (Mobbs et al., 2007). Arbitration between approach and avoidance has also been measured using more dynamic paradigms. Bach et al. (2014) aimed to support well-established animal models of how approach-avoidance conflict drives anxiety by showing that subjects exhibited passive avoidance behavior to threats when foraging for money in a 2D maze. fMRI results implicated the ventral hippocampus in this passive avoidance behavior, with lesions resulting in reduced avoidance (Bach et al., 2014). This was later extended by showing that individuals with amygdala lesions (i.e., two individuals with Urbach-Wiethe syndrome) and healthy subjects administered lorazepam showed reduced avoidance of a threat (Korn et al., 2017).

Social behaviors

Human social interaction features rich temporal and spatial dynamics. In the case of cooperative behaviors, most prior experimental and computational research has treated the choice to cooperate or defect as atomic. For instance, Camerer (2003) established a canon of rigorously controlled experimental paradigms where participants make atomic decisions, such as whether to cooperate or defect in a prisoner's dilemma game. However, studies that abstract over the substructure of group behavior obscure its multi-scale dynamics. More recently, there has been a trend toward more complex paradigms using

Please cite this article in press as: Mobbs et al., Promises and challenges of human computational ethology, Neuron (2021), https://doi.org/10.1016/j.neuron.2021.05.021

computer game-like virtual ecologies (Janssen et al., 2010; Mobbs et al., 2013). In this setting, self-interested individuals cooperate or defect through emergent policies that sequence lower-level actions. That is, participants must sequence primitive actions like move forward, turn left, etc. to implement their higher-level strategic decisions, e.g., to cooperate or defect. Thus, the fine structure of an ecologically valid collective action problem is determined by coupled interactions between natural properties of the environment and the actions of other group members.

Automated classification of behavior

Non-human work in computational ethology has availed itself of advances in deep learning to extract ethograms from high-dimensional behavioral data. In humans, similarly, complex behaviors can be captured with wearable behavioral sensors and video (Carreira and Zisserman, 2017; Topalovic et al., 2020) combined with machine learning, and this technology looks set to transform this field. More dynamic virtual ecologies provide a middle ground, with joystick or mouse input providing relatively rich behavioral data. Such data would lend themselves to analyses with machine learning methods, which can discover useful latent variables that more concisely capture consistent structure in dynamic behaviors. For example, detailed measures of a subject's position and velocity relative to rewarding stimuli could be used to identify particular reward-guided behaviors (Figure 1). Moving beyond purely visual worlds, virtual environments enabling full bodily movement could allow use of video-based techniques similar to those used in animal work. Use of unsupervised methods could also be particularly interesting when applied to human data, allowing identification of behavioral patterns that are not easily detectable by human observers. Furthermore, methods linking behavioral data to other variables of interest (for example, physiological measures or subjective state) could facilitate identification of particular behaviors that have relevance to broader constructs, such as anxiety.

A major disadvantage with collection of large datasets is that much of the data are meaningless. However, when used in combination with dimensionality reduction methods (for example clustering approaches, as discussed elsewhere in this article), signal can be separated from the noise to an extent. The approach we advocate provides two advantages over previous behavioral measures. First, measurement of targeted behaviors such as thigmotaxis or pauses and second, the ability to discover new behaviors that might be indicative of a decision or emotional state. The former uses naturalistic environments and rich data to confirm theory-driven predictions, whereas the latter uses a data-driven approach that can be used to inform new theory.

However, although automated dimensionality reduction can help greatly in the face of high-dimensional, unconstrained data, this does not entirely eliminate the need for theory. First, the choice of dimensionality reduction technique will be guided by theoretically motivated questions (e.g., what is the dimensionality of the data? Are we seeking to cluster brief behavioral motifs or trajectories through an environment?). Second, it will be necessary to validate extracted behavioral patterns based on existing theory (e.g., do these newly detected behaviors have meaningful neural correlates?). Finally, theory can be used to constrain the inferences that can be drawn from new observations and help identify areas where new theory is needed (e.g., are automatically identified threat-related behavioral patterns in line with theories about avoidance behavior?).

There are three examples that we believe illustrate the value of human computational ethology. First, Rosenberg et al. (2021) showed that more naturalistic environments can produce surprising insights into behavior even without introducing complex, multidimensional behavioral measures. In this study, mice were allowed to freely roam through a complex maze in search of reward, and the authors found that learning about the location of rewards was approximately 1,000 times faster than in a standard, two-alternative forced choice task as typically used to study learning and decision-making. This clearly shows that behavior in a standard, constrained task is not necessarily reflective of real-world behavior, which is ultimately what we are trying to explain. Second, Calhoun et al. (2019) used a data-driven modeling approach to identify three discrete behavioral states in *Drosophila* during courtship and were able to identify neural systems supporting behavioral state switching. These behavioral states were not hypothesized *a priori* and would not have been visible under more constrained conditions. This demonstrates that using naturalistic behavior can identify behaviors that would simply not be identified otherwise. Finally, Stringer et al. (2019) combined data-driven parsing of high-dimensional natural behaviors and neural activity in mice to demonstrate that a large proportion of neural activity across the cortex is linked to behavioral patterns. This shows that neural signals that may otherwise be considered noise can, in fact, be linked clearly to behavior, a finding that emerges by virtue of a data-driven approach using multidimensional, naturalistic behavior.

We believe these studies show that (1) naturalistic behavior can be qualitatively different from that seen in constrained tasks, (2) data-driven analysis of naturalistic behavior can identify novel behavioral states, and (3) combining measures of behavior and neural activity in naturalistic environments can provide insights into the role of neural systems that would not be seen otherwise. Drawing parallels with human avoidance, (1) it is possible that naturalistic avoidance decisions are qualitatively different from those seen in constrained avoidance tasks, (2) escape may involve key behavioral states that would only be visible through data-driven analysis of high-dimensional behavioral data, and (3) variability in neural activity may be linked to these behavioral states during escape.

Hybrid approaches: Preprogrammed and automatic classification

Validating machine learning methods

Despite the upsides of automated methods, even in non-human computational ethology, some analyses, such as syllable identification and segmentation in bird song, remain difficult to fully automate, and often hand coding and simple preprogrammed approaches remain the gold standard (Mets and Brainard, 2018). The results of automated machine learning models can be validated using traditional methods, but automated methods, such as unsupervised learning, can also help investigators identify features of behavior they may have missed using

Please cite this article in press as: Mobbs et al., Promises and challenges of human computational ethology, Neuron (2021), https://doi.org/10.1016/j.neuron.2021.05.021

traditional methods. In cases such as these, a fruitful approach is to open a dialog between traditional and automated methods, called "human-in-the-loop" or "interactive" machine learning (Holzinger et al., 2019). In this framework, input from a human user is utilized to select or provide feedback to aspects of a model or learning algorithm, resulting in performance that adheres better to domain-specific expertise. The results produced by these models can give the human user clues regarding features they were not aware of, expanding their domain-specific expertise, which can then be fed back into the model or algorithm. Notably, this approach has been advocated as a training method for human grandmasters in games such as chess and Go (Kasparov, 2018).

Using deep neural networks to develop better virtual ecologies

A limitation of the current virtual ecologies is that they consist of preprogrammed environments whose realism is questionable. For example, some virtual ecologies require subjects to interact with virtual agents whose behavior does not necessarily reflect known strategies used by real agents (Yoo et al., 2020). Hybrid environments are possible, where real agents interact with each other, but the environments in which they interact are impoverished with respect to the information they would encounter and used to guide behavior in more naturalistic scenarios (Tsutsui et al., 2019).

One issue with more naturalistic task environments is the lack of control over properties of the stimuli and information in these environments. Indeed, in naturalistic settings, it is not always clear what the relevant properties are in terms of guiding behavior (Hamilton and Huth, 2018; Patterson, 1974; Sonkusare et al., 2019) because naturalistic stimuli are nonparametric and complex (Geisler, 2008). By pairing insights from ecological psychology with the relatively new tools provided by deep neural networks, investigators can discover the properties of naturalistic stimuli relevant for behavior as well as parameterize them, enabling construction of more realistic virtual ecologies whose properties can be controlled precisely. Recent work on automated, unsupervised environment design for reinforcement learning agents may allow virtual ecologies to be defined without experimenter input (Dennis et al., 2020).

Feature learning methods can help identify such variables by extracting higher-order features of environments that reliably differ between task conditions using artificial neural networks (ANNs; Dosovitskiy et al., 2015). Recent work has argued that ANNs, rather than explicitly representing features in their environment, implicitly learn the structure of their environment that corresponds to task-appropriate actions (Hasson et al., 2020); this work lends further credibility to use of ANNs for identifying higher-order latent variables in naturalistic stimuli. Deep generative models, such as generative adversarial networks (GANs; Goodfellow et al., 2014) and variational autoencoders (VAEs; Doersch, 2016) can be used to construct generative models of naturalistic stimuli, such as images, audio, videos, and even task-specific video game environments (Li et al., 2019; Yan et al., 2016), which can then be sampled. These tools are already being used for generation of naturalistic audio stimuli in the animal vocalization community (Sainburg et al., 2019). New methods for feature-specific guidance of the output of these deep generative models can provide investigators with precise control over the statistics and dynamics of the higher-order latent variables relevant for behavior (Brookes et al., 2020; Lee and Seok, 2019).

Linking behavioral ethograms to neural circuits

Computational ethology provides tools to generate ethograms (representations of different types of behavior over time; Figure 2) with great accuracy and ease. Ethograms provide a detailed representation of the frequency of different behaviors over time; for example, in an avoidance task, we may wish to identify patterns of danger anticipation and escape behavior in response to environmental threat. With computational methods, we can automatically generate an ethogram describing how different behaviors emerge over the course of the task. The temporal information embedded within ethograms makes them ideal candidates for linking to unfolding neural events. This presents a new challenge, however: how can we optimally map these detailed behavioral observations onto high-dimensional data provided by neuroimaging?

Multivariate decoders (Haxby et al., 2001; Kriegeskorte and Douglas, 2019) have great potential when we wish to identify distributed patterns of brain activity or connectivity associated with specific behaviors. Given behavioral labels taken from ethograms, multivariate classifiers may be trained to identify neural patterns associated with distinct behaviors. This would then permit identification of distinct, distributed patterns of activity associated with specific behavioral patterns emerging from naturalistic behavior in a similar way to previous work relating activation measured using fMRI to naturalistic movie stimuli (Spiers and Maguire, 2007) or continuous speech (Willems et al., 2016). Alternatively, for more hypothesis-driven work, joint brain-behavior modeling approaches may be effective (Turner et al., 2019). These methods rely on a single pre-specified model that accounts for behavioral and neural data, accounting for their covariance through a hierarchical parameter structure, with shared parameters at the top level constraining the two modalities.

Encoding models (Kay et al., 2008) predict each channel of measured brain activity from external variables. They, too, could provide an effective method for identifying how brain activity corresponds to complex behavior. Alternatively, given a rich characterization of behavioral patterns, decoders could be trained on behavior to predict associated neural states, in line with work in non-human animals (Clemens et al., 2015).

A complementary multivariate method that could link ethograms to neural data is representational similarity analysis (RSA; Kriegeskorte et al., 2008). RSA characterizes the representation in each brain region by means of a representational dissimilarity matrix that reveals how dissimilar the activity patterns are for each pair of experimental conditions. RSA could help establish relationships between complex behavioral descriptions and high-dimensional response patterns with minimal need for fitting of parameters that define the relationship between each channel of measured brain activity and each behavior as needed when using encoding and decoding models (Kriegeskorte and Douglas, 2019).

A promising avenue toward reducing the number of states to be considered is clustering of behavioral and neural data. Tools

Please cite this article in press as: Mobbs et al., Promises and challenges of human computational ethology, Neuron (2021), https://doi.org/10.1016/j.neuron.2021.05.021

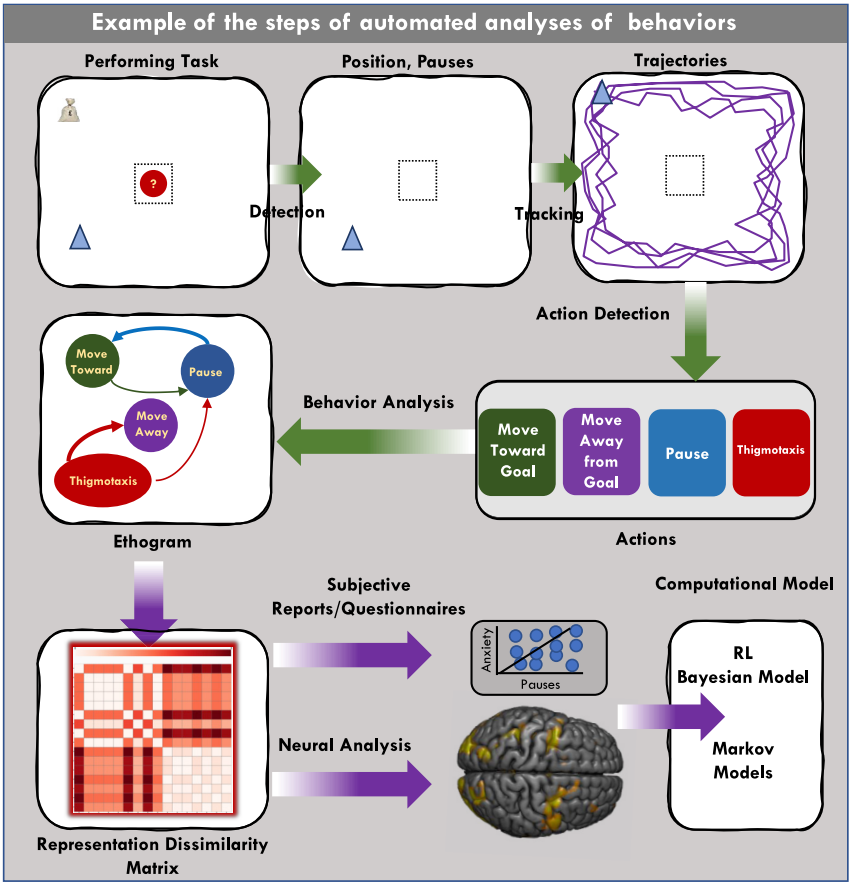

Figure 2. Steps in automated analysis and modeling of natural behavior

Shown is an example of how software can measure human behavior in a 2D or 3D environment. This occurs in several stages, including tracking the movements, action classification, and behavior analysis (Anderson and Perona, 2014). Starting from the top left: the subject performs a task where they learn about safe patches and where rewards of high or low value will appear. Such a task should result in place preference or aversion. The software occurs in several stages, including detection, tracking of the movements, action classification, and behavior analysis (Anderson and Perona, 2014; Dankert et al., 2009). This will result in an ethogram that will illustrate the different behaviors of pausing, thigmotaxis, movement away, and so forth. The behaviors can then be used for correlation with neural activity or subjective reports and questionnaire data. Finally, these data can be used to inform or create computational accounts of the behavior.

from computational ethology already constitute a form of clustering, typically identifying combinations of actions that constitute specific behaviors, resulting in detailed ethograms. In a similar vein, clustering methods have been applied to fMRI (Allen et al., 2014) and magnetoencephalography (MEG) data (Baker et al., 2014) to identify recurrent patterns of functional connectivity or activity, demonstrating that the brain cycles through a limited number of neural states. Exploring relationships between recurrent behavioral and neural states could aid with identification of robust associations between brain and behavior.

Linking ethograms and neural circuits to computational models

Ultimately, we would like to have computational models that explain the information processing performed by brains and predict neural and behavioral activity (Kriegeskorte and Douglas, 2018). Reinforcement learning algorithms are commonly divided into two categories: model-free (MF) and model-based (MB). MF learning gradually updates cached value estimates retrospectively from experience, and MF control uses those value estimates for decision-making. MF learning is associated with algorithms like temporal-difference learning (Sutton and Barto, 1998) and Q-learning (Watkins and Dayan, 1992). In contrast, MB algorithms calculate prospectively, for instance, by simulating possible future states (as in replay and preplay; Mattar and Daw, 2018; Pfeiffer and Foster, 2013; Wise et al., 2020).

Applying computational models to dynamic and unconstrained behaviors presents a new challenge: in terms of decision-making, we are now faced with a series of complex decisions unconstrained in time and reflected in behavior more elaborate than a button press. For example, human tasks that depend on multi-step decision trees typically focus on a single initial decision point at the start of the tree, with a decision made once per trial (Daw et al., 2011; Momennejad et al., 2017). When generating richer behavioral datasets, virtual ecologies make such models more difficult to test in relation to behavior. However, recent advances in deep reinforcement learning (RL) algorithms have created the opportunity for modeling complex behavior in virtual ecologies (Mnih et al., 2015). This is especially true in the MF case but now shows promise for incorporating MB algorithms as well (Schrittwieser et al., 2020).

In one notable example, artificial agents learned through reinforcement learning to play a first-person shooter computer game with realistic physics and complex objectives involving competition and teamwork (capture the flag) (Jaderberg et al., 2019). This was a computer game played by simulated agents, so in principle, the researcher could have full experimental access to any internal variable of the system. However, in practice, the long timescale and high dimensionality of the generated dataset meant that computational ethology methods were still necessary to analyze the resulting agent behavior. In particular, the authors employed an unsupervised computational ethology analysis inspired by Wiltschko et al. (2015). The results showed that internal representations of important game events like teammate following and home-base defense emerged as a result of reinforcement learning in this environment, suggesting the significant extent to which

Please cite this article in press as: Mobbs et al., Promises and challenges of human computational ethology, Neuron (2021), https://doi.org/10.1016/j.neuron.2021.05.021

the agents come to "understand" these game-related concepts.

Multi-agent deep reinforcement learning algorithms have also been applied to model the cooperation behaviors of groups in mixed-motivation settings (Foerster et al., 2018; Leibo et al., 2017; Lerer and Peysakhovich, 2018; Perolat et al., 2017). This line of work extends classical game-theoretic models based on matrix game formulations (Camerer, 2003) to capture complex spatiotemporally extended virtual ecologies. "Rational" (selfish) agent models are no better at cooperating in virtual ecologies than they are in matrix games. For example, Perolat et al. (2017) studied a virtual ecology designed to model common-pool resource appropriation (Janssen et al., 2010). As in the human studies, where individuals could not communicate (Janssen et al., 2014), they found that failures of cooperation yielded a tragedy of the commons, where overuse of resources degraded the environment to mutual detriment. In other circumstances, they found spontaneous emergence of exclusion behaviors and inequality (Perolat et al., 2017). Another study measured proxemics (i.e., distance to conspecifics) and preferences for different individuals that emerged from reinforcement learning in mixed-motivation scenarios (McKee et al., 2020). This approach, underpinned by models derived from multi-agent reinforcement learning research, holds promise to uncover interactions between the fine spatiotemporal structure of behavior and its strategic content that are not easily seen in traditional paradigms.

Use of complex behavioral measures will require a move away from models of simple choice likelihood based on the value of individual options, as is common in standard decision-making tasks, toward models that make predictions about more complex ongoing aspects of locomotor activity. Modeling approaches to more ecologically realistic behaviors have been developed, for example, using computational models of reward-guided place preference-like behavior (Wu et al., 2018) or threat-guided place aversion (Wise and Dolan, 2020) in 2D environments; however, these have focused on relatively coarse-grained trial-by-trial behavioral measures of location. Fully explaining behavior in these environments will require modeling not only position on a 2D grid but also motion measures, such as velocity and acceleration. In essence, the action space for modeling becomes larger and more complex. Additionally, it will be necessary to determine the optimal level of granularity for behavioral outcome measures. However, despite this added complexity, these rich measures may greatly improve our behavioral models. Additionally, unconstrained virtual ecologies naturally result in richer and more varied behaviors, a characteristic that provides more flexibility when designing experiments to elicit behavioral patterns that will differentiate candidate models (Palminteri et al., 2017).

The uses and future of VR

2D environments provide simple and clear ways to provide the subjects with task-relevant information that may not be visible from a first-person point of view. On the other hand, in some circumstances, the perceptual uncertainty provided by a 3D environment could be useful. For example, given that 3D environments most accurately represent our perception of the real world, they increase the ecological validity of the task while also allowing identification of behavioral dynamics, including intermittent locomotion and eye movements. To strengthen connections between behavioral research with human participants and its counterpart with artificial agent participants, it is sometimes even helpful to simulate, within a 3D environment, a scenario where the agent stands in front of a flat screen to perform a task. This allows virtual "eye movements" on the 2D environment projected within the simulated 3D environment (Leibo et al., 2018). Furthermore, 3D environments where aspects of the world are obscured from view encourage the subjects to build and use an internal model of their environment rather than relying on what is directly in front of them (Wayne et al., 2018).

Immersive VR technology has moved forward significantly in the last decade. Its use in human neuroimaging and psychological experiments has been revolutionary because it provides a more enriched and naturalistic approach to computerized environments (Bohil et al., 2011; Reggente et al., 2018; Figure 3). As with simple 2D environments, defensive behaviors can be measured, including thigmotaxis, place aversion, and escape. As VR becomes more realistic and feasible (for example, with smaller headsets equipped with improved eye tracking and pupillometry capability) and integrated with mobile intracranial electroencephalogram (iEEG; Topalovic et al., 2020), MEG, and fMRI hardware (Figure 3), there will be a need to understand how to best use this technology. Limitations of the immersive experience in the MRI scanner, however, are a major challenge, given the absence of body-based cues related to vestibular, motor, and somatosensory input. Several creative ways around these limitations have begun to emerge, including use of 3D glasses and VR training outside of the scanner. For example, Huffman and Ekstrom (2019) used VR outside of the MRI scanner to train people in enriched (on a treadmill), limited (using a joystick and head-mounted display), and impoverished (joystick only) environments, where their goal was to spatially navigate a virtual large-scale environment. After training, subjects were placed in an MRI scanner, where they performed a "judgement of relative direction" task, showing that body-based cues influenced spatial navigation. Full immersion in a virtual environment did not result in any behavioral or neural differences between conditions and proved the hypothesis that body-based cues are not necessary for retrieval of spatial information related to large-scale environments, a hypothesis that would be difficult to test without VR. Numerous other studies have shown that VR using desktop computers, head-mounted displays, and other technologies can be used to study several aspects of human cognition and related neural mechanisms (Ekstrom et al., 2003; Jacobs et al., 2013; Chrastil et al., 2015; Diersch and Wolbers, 2019; Hartley et al., 2003). Although several aspects of real-world human behavior seem to be modeled effectively in VR (Huffman and Ekstrom, 2021; Chrastil and Warren, 2015), there may be certain cognitive abilities that do not transfer as well in VR. For example, the type of learning strategy used to accomplish a spatial navigation task and transfer of this knowledge to novel situations may be altered in VR compared with the real world (Clemenson et al.,

Please cite this article in press as: Mobbs et al., Promises and challenges of human computational ethology, Neuron (2021), https://doi.org/10.1016/j.neuron.2021.05.021

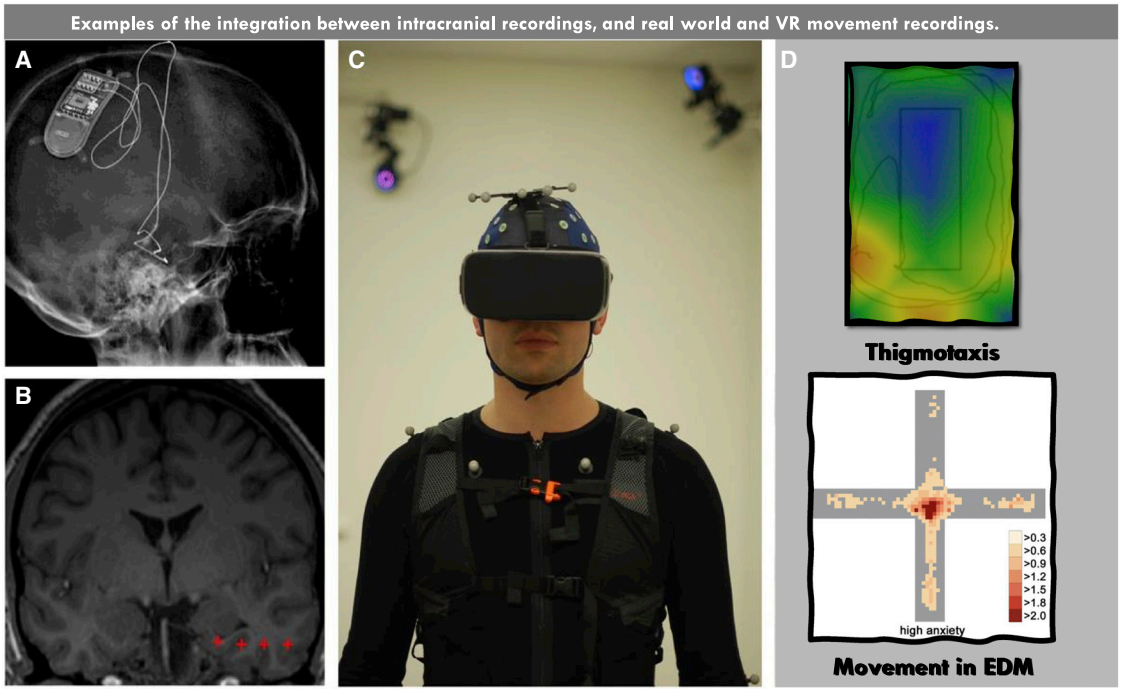

Figure 3. Example of how VR can be used to create virtual ecologies that measure naturalistic behaviors in participants with chronically implanted electrodes

(A–C) X-ray image (A) and MRI (B) of an example participant with a chronically implanted electrode with four contacts (red crosses) in the temporal lobe for iEEG recording, during which ambulatory VR and full-body motion capture (C) can be integrated (Stangl et al., 2021).

(D) Example of how VR can be used to create virtual ecologies that measure behaviors similar to those observed in rodents, e.g., thigmotaxis (Walz et al., 2016) and movement in the EDM (Biedermann et al., 2017).

2020; Hejtmanek et al., 2020). It will thus be important for future studies to determine the boundary conditions where reality can and cannot be modeled effectively with immersive VR technologies.

Evidence suggests that some of these challenges may become less prominent as the level of immersion continues to improve and full vestibular, motor, and somatosensory inputs are present (Huffman and Ekstrom, 2021; Hejtmanek et al., 2020). Recently, Topalovic et al. (2020) combined fully immersive VR technologies with full body and eye tracking as well as biometrics (e.g., heart rate, respiration, and galvanic skin response) in participants implanted with chronic deep brain devices capable of recording iEEG activity that is unsusceptible to motion-related artifacts. One recent study used combined immersive VR and iEEG recordings in moving subjects to understand the neural representations of actual physical space during memory formation and retrieval (Aghajan et al., 2019). Future research studies of a similar nature can use this technology to record synchronized behavioral and neural data from a wide range of brain structures (e.g., amygdala, hippocampus, and vmPFC) in naturally behaving humans.

This technology can also be integrated with augmented reality (AR) headsets that allow for objects/events/agents to be superimposed onto the real-world (Topalovic et al., 2020). One recent study used full-body motion capture combined with on-body world-view cameras and eye tracking in an environment shared with others to investigate social neural mechanisms of location encoding (Stangl et al., 2021; Figures 3A–3C). These studies open up exciting opportunities for applying computational ethological methods to high-resolution behavioral data captured during naturalistic experiences in social scenarios within real, virtual, or augmented environments. Use of iEEG in participants with temporary deep brain electrodes (e.g., in the epilepsy monitoring unit) combined with biometric recordings is also primed for VR use (Yilmaz Balban et al., 2021). Further, with creation of moveable optically pumped magnetometer (OPM)-MEG (Boto et al., 2018), providing electrophysiological measurements at millisecond resolution, there is promise in combining VR with spatially and temporally high-resolution brain imaging on a wider population of subjects not limited to only those who have implanted brain electrodes.

A small but accumulating set of VR studies are beginning to demonstrate how VR can be used to study fear and anxiety. Traditionally, it has been difficult to expose subjects to realistic threats in the lab, and instead studies have typically relied on painful stimuli, such as electric shocks. In contrast, VR has permitted assessment of common fears, such as height and public speaking, in the lab (Gromer et al., 2019; Stupar-Rutenfrans et al., 2017). Virtual versions of the elevated plus maze (EPM) have been used in humans, and, like rodents, high-anxiety individuals show increased avoidance of open arms (Biedermann et al., 2017). Others have shown that VR can be used to elicit anxiety in flight phobics (Mühlberger et al., 2001). Interestingly, VR can have therapeutic effects in flight

Please cite this article in press as: Mobbs et al., Promises and challenges of human computational ethology, Neuron (2021), https://doi.org/10.1016/j.neuron.2021.05.021

Box 2. Implications for psychiatric populations and research domain criteria (RDoCs)

Movement in psychiatric disorders. The emergence of ML techniques provides a new avenue from which to study a variety of complex movements that capture behaviors that, until now, have been difficult to measure. This, in turn, could present unique opportunities for identifying novel markers of mental health problems. Similar approaches have already shown promise in prior studies, demonstrating subtle behavioral signatures of mental health problems. For example, inspired by animal models, researchers have shown that thigmotaxis is higher in individuals with social phobia compared with healthy control individuals (Walz et al., 2016). Movement kinetics have also been used to detect altered movement patterns in autism (Cook et al., 2013) and may be used to detect prodromal markers of psychiatric conditions. These early studies demonstrate that taking a more naturalistic approach to the study of behavior, with rich indices of the movements human subjects make, can facilitate identification of markers of disorder that could not be seen otherwise.

New transdiagnostic models of psychiatric disorders. The benefits of computational ethology go beyond movement-based markers of psychiatric disorder. In recent years, computational psychiatry has begun to demonstrate how dysfunction in learning and decision-making processes can result in symptoms of mental health problems. However, these studies have relied on highly constrained, artificial tasks with limited behavioral measures. As detailed in other sections, computational ethology has the potential to bring new insights into learning and decision-making through richer and more natural behavioral measures. This deeper understanding of how humans learn about and act within their environment will naturally provide further targets for studies of how these processes go awry in psychiatric disorders. As an example, prior work has considered the importance of MB planning in compulsive symptoms, showing that individuals high in these traits have difficulty in learning a model of the world and in using this learned model to guide behavior (Gillan et al., 2016; Sharp et al., 2020). However, tasks used to assess these processes rely on simplistic task structures with only a few task states. Using methods from computational ethology could encourage development of new models that are able to explain the relationship between MB and MF control in more complex and naturalistic environments, which, in turn, would provide new targets for studies investigating dysfunction in these processes and its association with symptom dimensions such as compulsivity. This may take inspiration from artificial intelligence, for example, where planning in complex environments has received a great deal of attention (Schrittwieser et al., 2020; Silver and Veness, 2010). As a result, new models of dysfunction could be developed that account for behavior that only emerges in these more naturalistic virtual environments.

RDoCs. RDoCs are a framework within which to investigate and classify mental disorders, focusing on systems that span traditional diagnostic categories (Insel et al., 2010). The ability to detect and measure new behaviors will also advance the objectives of the RDoCs, where one goal is to measure a full range of behaviors and link them to health and disorder. For example, although the RDoC negative valance systems matrix suggests that researchers should measure freezing, risk assessment, approach, avoidance, and escape, there are few existing paradigms that can evoke these behaviors in the truest sense. It follows that gaining methods to measure these behaviors in human subjects will aid translation of animal models to humans and identification of human behaviors that can also be evoked in animals, which could have benefits for development of new pharmacological therapies. Drug development in psychiatry has largely stalled (Brady et al., 2019), and there have been numerous examples of drug candidates that were apparently efficacious in animals but failed to show benefits in human trials, likely as a result of animal models of disease not truly representing the conditions they intend to (Grabb et al., 2016). Computational ethology and its focus on naturalistic behaviors will allow proper measurement of key behaviors, highlighted in the RDoCs, across humans and animals and, in turn, could facilitate drug development.

and spider phobics (Mühlberger et al., 2003; Shiban et al., 2015). Similarly, exposure therapy in VR has begun to see use in treating PTSD in war veterans. Rizzo et al. (2010) developed an exposure therapy system for veterans of Iraq and Afghanistan, combining realistic 3D environments in VR with physiological measurements, such as the galvanic skin response. The same team later employed this system in conjunction with fMRI to monitor improvements in cerebral function in veterans undergoing treatment for PTSD (Roy et al., 2010). Finally, Yilmaz Balban et al. (2021) have shown that exposure to virtual threats such as scary heights can elicit increases in autonomic arousal. Further, using iEEG, the authors show higher gamma activity in the insula for virtual heights compared with no-height control conditions. These studies show how VR can elicit autonomic and neural responses to threat and show promise for implementing the behavioral measures advocated by computational ethology.

Conclusions

Introducing methods from computational ethology to human neuroscience promises to help us uncover novel behavioral assays and better understand the dynamic nature of the human brain and how it might function in the real world. Further, through extraction of individualized ethograms from tasks using virtual ecologies, we can link patterns of behavior in naturalistic environments to brain states, potentially revealing links between neural circuits and behavior that are not observable with current methods. Current experimental paradigms are restricted to specific processes thought to be important by the researcher and, therefore, miss behavioral characteristics (in healthy function and psychiatric disorders) that might be essential for understanding brain function in naturalistic environments. The approaches laid out in this paper enable quantification of behavior and its disruption in a variety of psychiatric conditions at a behavioral and neural level. Unsupervised methods of behavior

classification may identify novel patterns of behavior that are diagnostic of particular symptom clusters (Box 2). Although there are challenges to overcome, approaches advocated by computational ethology (i.e., unsupervised quantification of behavior) are an exciting and powerful way to engage the dynamics of natural cognition in human neuroscience.

ACKNOWLEDGMENTS

This work was supported by National Institute of Mental Health grant 2P50MH094258, a Chen Institute award (P2026052), and Templeton Foundation grant TWCF0366 (all to D.M.). T.W. is supported by a Wellcome Trust Sir Henry Wellcome Fellowship (206460/17/Z). This work is also supported by the National Institutes of Health (NIH) National Institute of Neurological Disorders and Stroke (NINDS; NS103802 and NS117838), the McKnight Foundation (Technological Innovations Award in Neuroscience to N.S.), and a Keck Junior Faculty Award (to N.S.). We thank Matthew Botvinick for feedback on an earlier version of this paper.

REFERENCES

Aghajan, Z.M., Villaroman, D., Hiller, S., Wishard, T.J., Topalovic, U., Christov-Moore, L., Shaterian, N., Hasulak, N.R., Knowlton, B., Eliashiv, D., et al. (2019). Modulation of human intracranial theta oscillations during freely moving spatial navigation and memory. bioRxiv. https://doi.org/10.1101/738807.

Allen, E.A., Damaraju, E., Plis, S.M., Erhardt, E.B., Eichele, T., and Calhoun, V.D. (2014). Tracking whole-brain connectivity dynamics in the resting state. Cereb. Cortex *24*, 663–676.

Anderson, D.J., and Perona, P. (2014). Toward a science of computational ethology. Neuron *84*, 18–31.

Astur, R.S., Carew, A.W., and Deaton, B.E. (2014). Conditioned place preferences in humans using virtual reality. Behav. Brain Res. *267*, 173–177.

Astur, R.S., Purton, A.J., Zaniewski, M.J., Cimadevilla, J., and Markus, E.J. (2016). Human sex differences in solving a virtual navigation problem. Behav. Brain Res. *308*, 236–243.

Babayan, B.M., and Konen, C.S. (2019). Behavior Matters. Neuron *104*, 1.

Bach, D.R., Guitart-Masip, M., Packard, P.A., Miró, J., Falip, M., Fuentemilla, L., and Dolan, R.J. (2014). Human hippocampus arbitrates approach-avoidance conflict. Curr. Biol. *24*, 541–547.

Baker, A.P., Brookes, M.J., Rezek, I.A., Smith, S.M., Behrens, T., Probert Smith, P.J., and Woolrich, M. (2014). Fast transient networks in spontaneous human brain activity. eLife *3*, e01867.

Balleine, B.W. (2019). The Meaning of Behavior: Discriminating Reflex and Volition in the Brain. Neuron *104*, 47–62.

Berman, G.J., Choi, D.M., Bialek, W., and Shaevitz, J.W. (2014). Mapping the stereotyped behaviour of freely moving fruit flies. J. R. Soc. Interface *11*, 20140672.

Biedermann, S.V., Biedermann, D.G., Wenzlaff, F., Kurjak, T., Nouri, S., Auer, M.K., Wiedemann, K., Briken, P., Haaker, J., Lonsdorf, T.B., and Fuss, J. (2017). An elevated plus-maze in mixed reality for studying human anxiety-related behavior. BMC Biol. *15*, 125.

Bohil, C.J., Alicea, B., and Biocca, F.A. (2011). Virtual reality in neuroscience research and therapy. Nat. Rev. Neurosci. *12*, 752–762.

Boto, E., Holmes, N., Leggett, J., Roberts, G., Shah, V., Meyer, S.S., Muñoz, L.D., Mullinger, K.J., Tierney, T.M., Bestmann, S., et al. (2018). Moving magnetoencephalography towards real-world applications with a wearable system. Nature *555*, 657–661.

Brady, L.S., Potter, W.Z., and Gordon, J.A. (2019). Redirecting the revolution: new developments in drug development for psychiatry. Expert Opin. Drug Discov. *14*, 1213–1219.

Brookes, D.H., Park, H., and Listgarten, J. (2020). Conditioning by adaptive sampling for robust design. arXiv, 1901.10060 https://arxiv.org/abs/1901.10060.

Calhoun, A.J., Pillow, J.W., and Murthy, M. (2019). Unsupervised identification of the internal states that shape natural behavior. Nat. Neurosci. *22*, 2040–2049.

Camerer, C.F. (2003). Behavioural studies of strategic thinking in games. Trends Cogn. Sci. *7*, 225–231.

Carreira, J., and Zisserman, A. (2017). Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset. IEEE Xplore, 4724–4733.

Chrastil, E.R., and Warren, W.H. (2015). Active and passive spatial learning in human navigation: Acquisition of graph knowledge. J. Exp. Psychol. Learn. Mem. Cogn. *41*, 1162–1178.

Chrastil, E.R., Sherrill, K.R., Hasselmo, M.E., and Stern, C.E. (2015). There and back again: hippocampus and retrosplenial cortex track homing distance during human path integration. J. Neurosci. *35*, 15442–15452.

Clemenson, G.D., Wang, L., Mao, Z., Stark, S.M., and Stark, C.E.L. (2020). Exploring the Spatial Relationships Between Real and Virtual Experiences: What Transfers and What Doesn't. Front. Virtual Real. Published online October 8, 2020. https://doi.org/10.3389/frvir.2020.572122.

Clemens, J., Girardin, C.C., Coen, P., Guan, X.-J., Dickson, B.J., and Murthy, M. (2015). Connecting Neural Codes with Behavior in the Auditory System of Drosophila. Neuron *87*, 1332–1343.

Cook, J.L., Blakemore, S.-J., and Press, C. (2013). Atypical basic movement kinematics in autism spectrum conditions. Brain *136*, 2816–2824.

Cooper, J., William, E., and Blumstein, D.T. (2015). Escaping From Predators: An Integrative View of Escape Decisions (Cambridge University Press).

Dankert, H., Wang, L., Hoopfer, E., Anderson, D.J., and Perona, P. (2009). Automated monitoring and analysis of social behavior in Drosophila. Nat. Methods *6*, 297–303.

Datta, S.R., Anderson, D.J., Branson, K., Perona, P., and Leifer, A. (2019). Computational Neuroethology: A Call to Action. Neuron *104*, 11–24.

Daw, N.D., Gershman, S.J., Seymour, B., Dayan, P., and Dolan, R.J. (2011). Model-based influences on humans' choices and striatal prediction errors. Neuron *69*, 1204–1215.

Dennis, M., Jaques, N., Vinitsky, E., Bayen, A., Russell, S., Critch, A., and Levine, S. (2020). Emergent Complexity and Zero-shot Transfer via Unsupervised Environment Design. arXiv, 2012.02096 https://arxiv.org/abs/2012.02096.

Diersch, N., and Wolbers, T. (2019). The potential of virtual reality for spatial navigation research across the adult lifespan. J. Exp. Biol. *222*, jeb.187252.

Doersch, C. (2016). Tutorial on Variational Autoencoders. arXiv, 1606.05908 https://arxiv.org/abs/1606.05908.

Donahue, J., Jia, Y., Vinyals, O., Hoffman, J., Zhang, N., Tzeng, E., and Darrell, T. (2013). DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition. arXiv, 1310.1531 https://arxiv.org/abs/1310.1531.

Dosovitskiy, A., Fischer, P., Springenberg, J.T., Riedmiller, M., and Brox, T. (2015). Discriminative Unsupervised Feature Learning with Exemplar Convolutional Neural Networks. arXiv, 1406.6909 https://arxiv.org/abs/1406.6909.

Ekstrom, A.D., Kahana, M.J., Caplan, J.B., Fields, T.A., Isham, E.A., Newman, E.L., and Fried, I. (2003). Cellular networks underlying human spatial navigation. Nature *425*, 184–188.

Foerster, J.N., Chen, R.Y., Al-Shedivat, M., Whiteson, S., Abbeel, P., and Mordatch, I. (2018). Learning with Opponent-Learning Awareness. Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems, 122–130.

Fung, B.J., Qi, S., Hassabis, D., Daw, N., and Mobbs, D. (2019). Slow escape decisions are swayed by trait anxiety. Nat. Hum. Behav. *3*, 702–708.

Geisler, W.S. (2008). Visual perception and the statistical properties of natural scenes. Annu. Rev. Psychol. *59*, 167–192.

Gillan, C.M., Kosinski, M., Whelan, R., Phelps, E.A., and Daw, N.D. (2016). Characterizing a psychiatric symptom dimension related to deficits in goal-directed control. eLife *5*, e11305.

Gold, A.L., Morey, R.A., and McCarthy, G. (2015). Amygdala-prefrontal cortex functional connectivity during threat-induced anxiety and goal distraction. Biol. Psychiatry *77*, 394–403.

Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative Adversarial Networks. arXiv, 1406.2661 https://arxiv.org/abs/1406.2661.

Grabb, M.C., Cross, A.J., Potter, W.Z., and McCracken, J.T. (2016). Derisking Psychiatric Drug Development: The NIMH's Fast Fail Program, A Novel Precompetitive Model. J. Clin. Psychopharmacol. *36*, 419–421.

Graving, J.M., Chae, D., Naik, H., Li, L., Koger, B., Costelloe, B.R., and Couzin, I.D. (2019). DeepPoseKit, a software toolkit for fast and robust animal pose estimation using deep learning. eLife *8*, e47994.

Gromer, D., Reinke, M., Christner, I., and Pauli, P. (2019). Causal Interactive Links Between Presence and Fear in Virtual Reality Height Exposure. Front. Psychol. *10*, 141.

Grupe, D.W., and Nitschke, J.B. (2013). Uncertainty and anticipation in anxiety: an integrated neurobiological and psychological perspective. Nat. Rev. Neurosci. *14*, 488–501.

Hamilton, L.S., and Huth, A.G. (2018). The revolution will not be controlled: natural stimuli in speech neuroscience. Lang. Cogn. Neurosci. *35*, 573–582.

Hartley, T., Maguire, E.A., Spiers, H.J., and Burgess, N. (2003). The well-worn route and the path less traveled: distinct neural bases of route following and wayfinding in humans. Neuron *37*, 877–888.

Hasson, U., Nastase, S.A., and Goldstein, A. (2020). Direct Fit to Nature: An Evolutionary Perspective on Biological and Artificial Neural Networks. Neuron *105*, 416–434.

Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., and Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science *293*, 2425–2430.

Hejtmanek, L., Starrett, M., Ferrer, E., and Ekstrom, A.D. (2020). How Much of What We Learn in Virtual Reality Transfers to Real-World Navigation? Multisens. Res. *33*, 479–503.

Holzinger, A., Plass, M., Kickmeier-Rust, M., Holzinger, K., Crişan, G.C., Pintea, C.-M., and Palade, V. (2019). Interactive machine learning: experimental evidence for the human in the algorithmic loop. Applied Intelligence *49*, 2401–2414.

Hoopfer, E.D., Jung, Y., Inagaki, H.K., Rubin, G.M., and Anderson, D.J. (2015). P1 interneurons promote a persistent internal state that enhances inter-male aggression in Drosophila. eLife *4*, e11346.

Huffman, D.J., and Ekstrom, A.D. (2019). A Modality-Independent Network Underlies the Retrieval of Large-Scale Spatial Environments in the Human Brain. Neuron *104*, 611–622.e7.

Huffman, D.J., and Ekstrom, A.D. (2021). An Important Step toward Understanding the Role of Body-based Cues on Human Spatial Memory for Large-Scale Environments. J. Cogn. Neurosci. *33*, 167–179.

Humphries, D.A., and Driver, P.M. (1970). Protean defence by prey animals. Oecologia *5*, 285–302.

Insafutdinov, E., Pishchulin, L., Andres, B., Andriluka, M., and Schiele, B. (2016). DeeperCut: A Deeper, Stronger, and Faster Multi-Person Pose Estimation Model. European Conference on Computer Vision, 34–50.

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D.S., Quinn, K., Sanislow, C., and Wang, P. (2010). Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. Am. J. Psychiatry *167*, 748–751.

Jacobs, J., Weidemann, C.T., Miller, J.F., Solway, A., Burke, J.F., Wei, X.X., Suthana, N., Sperling, M.R., Sharan, A.D., Fried, I., and Kahana, M.J. (2013). Direct recordings of grid-like neuronal activity in human spatial navigation. Nat. Neurosci. *16*, 1188–1190.

Jaderberg, M., Czarnecki, W.M., Dunning, I., Marris, L., Lever, G., Castañeda, A.G., Beattie, C., Rabinowitz, N.C., Morcos, A.S., Ruderman, A., et al. (2019). Human-level performance in 3D multiplayer games with population-based reinforcement learning. Science *364*, 859–865.

Janssen, M.A., Holahan, R., Lee, A., and Ostrom, E. (2010). Lab experiments for the study of social-ecological systems. Science *328*, 613–617. https://doi.org/10.1126/science.1183532.

Janssen, M., Tyson, M., and Lee, A. (2014). The effect of constrained communication and limited information in governing a common resource. Int. J. Commons *8*, 617–635.

Jovanic, T., Schneider-Mizell, C.M., Shao, M., Masson, J.-B., Denisov, G., Fetter, R.D., Mensh, B.D., Truman, J.W., Cardona, A., and Zlatic, M. (2016). Competitive Disinhibition Mediates Behavioral Choice and Sequences in Drosophila. Cell *167*, 858–870.e19.

Kasparov, G. (2018). Chess, a *Drosophila* of reasoning. Science *362*, 1087, 1087.

Kay, K.N., Naselaris, T., Prenger, R.J., and Gallant, J.L. (2008). Identifying natural images from human brain activity. Nature *452*, 352–355.

Korn, C.W., Vunder, J., Miró, J., Fuentemilla, L., Hurlemann, R., and Bach, D.R. (2017). Amygdala Lesions Reduce Anxiety-like Behavior in a Human Benzodiazepine-Sensitive Approach-Avoidance Conflict Test. Biol. Psychiatry *82*, 522–531.

Krakauer, J.W., Ghazanfar, A.A., Gomez-Marin, A., MacIver, M.A., and Poeppel, D. (2017). Neuroscience Needs Behavior: Correcting a Reductionist Bias. Neuron *93*, 480–490.

Kriegeskorte, N., and Douglas, P.K. (2018). Cognitive computational neuroscience. Nat. Neurosci. *21*, 1148–1160.

Kriegeskorte, N., and Douglas, P.K. (2019). Interpreting encoding and decoding models. Curr. Opin. Neurobiol. *55*, 167–179.

Kriegeskorte, N., Mur, M., and Bandettini, P. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. Front. Syst. Neurosci. *2*, 4.

LeDoux, J. (2012). Rethinking the emotional brain. Neuron *73*, 653–676.

Lee, M., and Seok, J. (2019). Controllable Generative Adversarial Network. IEEE Access *7*, 28158–28169.

Leibo, J.Z., d'Autume, C. de M., Zoran, D., Amos, D., Beattie, C., Anderson, K., Castañeda, A.G., Sanchez, M., Green, S., Gruslys, A., et al. (2018). Psychlab: A Psychology Laboratory for Deep Reinforcement Learning Agents. arXiv, 1801.08116 https://arxiv.org/abs/1801.08116.

Leibo, J.Z., Zambaldi, V., Lanctot, M., Marecki, J., and Graepel, T. (2017). Multi-agent Reinforcement Learning in Sequential Social Dilemmas. Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems, 464–473.

Lerer, A., and Peysakhovich, A. (2018). Maintaining cooperation in complex social dilemmas using deep reinforcement learning. arXiv, 1707.01068 https://arxiv.org/abs/1707.01068.

Li, J., Ma, H., and Tomizuka, M. (2019). Conditional Generative Neural System for Probabilistic Trajectory Prediction. arXiv, 1905.01631 https://arxiv.org/abs/1905.01631.

Mathis, A., Mamidanna, P., Cury, K.M., Abe, T., Murthy, V.N., Mathis, M.W., and Bethge, M. (2018). DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. Nat. Neurosci. *21*, 1281–1289.

Mattar, M.G., and Daw, N.D. (2018). Prioritized memory access explains planning and hippocampal replay. Nat. Neurosci. *21*, 1609–1617.

McKee, K.R., Gemp, I., McWilliams, B., Duéñez-Guzmán, E.A., Hughes, E., and Leibo, J.Z. (2020). Social diversity and social preferences in mixed-motive reinforcement learning. Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems, 869–877.

Mets, D.G., and Brainard, M.S. (2018). An automated approach to the quantitation of vocalizations and vocal learning in the songbird. PLoS Comput. Biol. *14*, e1006437.

Meyer, C., Padmala, S., and Pessoa, L. (2019). Dynamic Threat Processing. J. Cogn. Neurosci. *31*, 522–542.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A.A., Veness, J., Bellemare, M.G., Graves, A., Riedmiller, M., Fidjeland, A.K., Ostrovski, G., et al. (2015). Human-level control through deep reinforcement learning. Nature *518*, 529–533.

Mobbs, D., Petrovic, P., Marchant, J.L., Hassabis, D., Weiskopf, N., Seymour, B., Dolan, R.J., and Frith, C.D. (2007). When fear is near: threat imminence elicits prefrontal-periaqueductal gray shifts in humans. Science *317*, 1079–1083.

Mobbs, D., Marchant, J.L., Hassabis, D., Seymour, B., Tan, G., Gray, M., Petrovic, P., Dolan, R.J., and Frith, C.D. (2009). From threat to fear: the neural organization of defensive fear systems in humans. J. Neurosci. *29*, 12236–12243.

Mobbs, D., Hassabis, D., Yu, R., Chu, C., Rushworth, M., Boorman, E., and Dalgleish, T. (2013). Foraging under competition: the neural basis of input-matching in humans. J. Neurosci. *33*, 9866–9872.

Mobbs, D., and Kim, J.J. (2015). Neuroethological studies of fear and risky decision-making in rat and humans. Curr. Opin. Behav. Sci. *5*, 8–15.

Mobbs, D., Trimmer, P.C., Blumstein, D.T., and Dayan, P. (2018). Foraging for foundations in decision neuroscience: insights from ethology. Nat. Rev. Neurosci. *19*, 419–427.

Mobbs, D., Headley, D.B., Ding, W., and Dayan, P. (2020). Space, Time, and Fear: Survival Computations along Defensive Circuits. Trends Cogn. Sci. *24*, 228–241.

Molet, M., Billiet, G., and Bardo, M.T. (2013). Conditioned place preference and aversion for music in a virtual reality environment. Behav. Processes *92*, 31–35.

Momennejad, I., Russek, E.M., Cheong, J.H., Botvinick, M.M., Daw, N.D., and Gershman, S.J. (2017). The successor representation in human reinforcement learning. Nat. Hum. Behav. *1*, 680–692.

Mühlberger, A., Herrmann, M.J., Wiedemann, G.C., Ellgring, H., and Pauli, P. (2001). Repeated exposure of flight phobics to flights in virtual reality. Behav. Res. Ther. *39*, 1033–1050.

Mühlberger, A., Wiedemann, G., and Pauli, P. (2003). Efficacy of a one-session virtual reality exposure treatment for fear of flying. Psychother. Res. *13*, 323–336.

Musall, S., Kaufman, M.T., Juavinett, A.L., Gluf, S., and Churchland, A.K. (2019). Single-trial neural dynamics are dominated by richly varied movements. Nat. Neurosci. *22*, 1677–1686.

Niv, Y. (2020). The primacy of behavioral research for understanding the brain. PsyArXiv. https://doi.org/10.31234/osf.io/y8mxe.

Palminteri, S., Wyart, V., and Koechlin, E. (2017). The Importance of Falsification in Computational Cognitive Modeling. Trends Cogn. Sci. *21*, 425–433.

Patterson, G.R. (1974). A basis for identifying stimuli which control behaviors in natural settings. Child Dev. *45*, 900–911.

Pereira, T.D., Aldarondo, D.E., Willmore, L., Kislin, M., Wang, S.S.-H., Murthy, M., and Shaevitz, J.W. (2019). Fast animal pose estimation using deep neural networks. Nat. Methods *16*, 117–125.

Perolat, J., Leibo, J.Z., Zambaldi, V., Beattie, C., Tuyls, K., and Graepel, T. (2017). A multi-agent reinforcement learning model of common-pool resource appropriation. Proceedings of the 31st International Conference on Neural Information Processing Systems, 3646–3655.

Pfeiffer, B.E., and Foster, D.J. (2013). Hippocampal place-cell sequences depict future paths to remembered goals. Nature *497*, 74–79.

Qi, S., Hassabis, D., Sun, J., Guo, F., Daw, N., and Mobbs, D. (2018). How cognitive and reactive fear circuits optimize escape decisions in humans. Proc. Natl. Acad. Sci. USA *115*, 3186–3191.

Qi, S., Cross, L., Wise, T., Sui, X., O'Doherty, J., and Mobbs, D. (2020). The Role of the Medial Prefrontal Cortex in Spatial Margin of Safety Calculations. bioRxiv. https://doi.org/10.1101/2020.06.05.137075.

Reggente, N., Essoe, J.K.-Y., Aghajan, Z.M., Tavakoli, A.V., McGuire, J.F., Suthana, N.A., and Rissman, J. (2018). Enhancing the Ecological Validity of fMRI Memory Research Using Virtual Reality. Front. Neurosci. *12*, 408.

Rizzo, A.S., Difede, J., Rothbaum, B.O., Reger, G., Spitalnick, J., Cukor, J., and McLay, R. (2010). Development and early evaluation of the Virtual Iraq/Afghanistan exposure therapy system for combat-related PTSD. Ann. N Y Acad. Sci. *1208*, 114–125.

Rosenberg, M., Zhang, T., Perona, P., and Meister, M. (2021). Mice in a labyrinth: Rapid learning, sudden insight, and efficient exploration. bioRxiv. https://doi.org/10.1101/2021.01.14.426746.

Roy, M., Harvey, P.-O., Berlim, M.T., Mamdani, F., Beaulieu, M.-M., Turecki, G., and Lepage, M. (2010). Medial prefrontal cortex activity during memory encoding of pictures and its relation to symptomatic improvement after citalopram treatment in patients with major depression. J. Psychiatry Neurosci. *35*, 152–162.

Sainburg, T., Thielk, M., and Gentner, T.Q. (2019). Latent space visualization, characterization, and generation of diverse vocal communication signals. PLOS Comput. Biol. Published online October 15, 2020. https://doi.org/10.1371/journal.pcbi.1008228.

Scheveneels, S., Boddez, Y., Vervliet, B., and Hermans, D. (2016). The validity of laboratory-based treatment research: Bridging the gap between fear extinction and exposure treatment. Behav. Res. Ther. *86*, 87–94.

Schrittwieser, J., Antonoglou, I., Hubert, T., Simonyan, K., Sifre, L., Schmitt, S., Guez, A., Lockhart, E., Hassabis, D., Graepel, T., et al. (2020). Mastering Atari, Go, Chess and Shogi by Planning with a Learned Model. Nature *588*, 604–609.

Sharp, P.B., Dolan, R.J., and Eldar, E. (2020). Cognitive map learning is disrupted in compulsivity and anxious arousal. PsyArXiv. https://doi.org/10.31234/osf.io/x29jq.

Shiban, Y., Brütting, J., Pauli, P., and Mühlberger, A. (2015). Fear reactivation prior to exposure therapy: does it facilitate the effects of VR exposure in a randomized clinical sample? J. Behav. Ther. Exp. Psychiatry *46*, 133–140.

Silston, B., Wise, T., Qi, S., Sui, X., Dayan, P., and Mobbs, D. (2020). Neural encoding of socially adjusted value during competitive and hazardous foraging. bioRxiv. https://doi.org/10.1101/2020.09.11.294058.

Silver, D., and Veness, J. (2010). Monte-Carlo Planning in Large POMDPs. In Advances in Neural Information Processing Systems 23, J.D. Lafferty, C.K.I. Williams, J. Shawe-Taylor, R.S. Zemel, and A. Culotta, eds. (Curran Associates, Inc.), pp. 2164–2172.

Sonkusare, S., Breakspear, M., and Guo, C. (2019). Naturalistic Stimuli in Neuroscience: Critically Acclaimed. Trends Cogn. Sci. *23*, 699–714.

Spiers, H.J., and Maguire, E.A. (2007). Decoding human brain activity during real-world experiences. Trends Cogn. Sci. *11*, 356–365.

Stangl, M., Topalovic, U., Inman, C.S., Hiller, S., Villaroman, D., Aghajan, Z.M., Christov-Moore, L., Hasulak, N.R., Rao, V.R., Halpern, C.H., et al. (2021). Boundary-anchored neural mechanisms of location-encoding for self and others. Nature *589*, 420–425.

Stringer, C., Pachitariu, M., Steinmetz, N., Reddy, C.B., Carandini, M., and Harris, K.D. (2019). Spontaneous behaviors drive multidimensional, brainwide activity. Science *364*, 255.

Stupar-Rutenfrans, S., Ketelaars, L.E.H., and van Gisbergen, M.S. (2017). Beat the Fear of Public Speaking: Mobile 360° Video Virtual Reality Exposure Training in Home Environment Reduces Public Speaking Anxiety. Cyberpsychol. Behav. Soc. Netw. *20*, 624–633.

Sutton, R.S., and Barto, A.G. (1998). Introduction to Reinforcement Learning (MIT Press).

Topalovic, U., Aghajan, Z.M., Villaroman, D., Hiller, S., Christov-Moore, L., Wishard, T.J., Stangl, M., Hasulak, N.R., Inman, C.S., Fields, T.A., et al. (2020). Wireless Programmable Recording and Stimulation of Deep Brain Activity in Freely Moving Humans. Neuron *108*, 322–334.e9.

Tsutsui, K., Shinya, M., and Kudo, K. (2019). Spatiotemporal characteristics of an attacker's strategy to pass a defender effectively in a computer-based one-on-one task. Sci. Rep. *9*, 17260.

Turner, B.M., Palestro, J.J., Miletić, S., and Forstmann, B.U. (2019). Advances in techniques for imposing reciprocity in brain-behavior relations. Neurosci. Biobehav. Rev. *102*, 327–336.

Walz, N., Mühlberger, A., and Pauli, P. (2016). A Human Open Field Test Reveals Thigmotaxis Related to Agoraphobic Fear. Biol. Psychiatry *80*, 390–397.

Watkins, C.J.C.H., and Dayan, P. (1992). Q-learning. Mach. Learn. *8*, 279–292.

Wayne, G., Hung, C.-C., Amos, D., Mirza, M., Ahuja, A., Grabska-Barwinska, A., Rae, J., Mirowski, P., Leibo, J.Z., Santoro, A., et al. (2018). Unsupervised Predictive Memory in a Goal-Directed Agent. arXiv, 1803.10760 https://arxiv.org/abs/1803.10760.

Willems, R.M., Frank, S.L., Nijhof, A.D., Hagoort, P., and van den Bosch, A. (2016). Prediction During Natural Language Comprehension. Cereb. Cortex *26*, 2506–2516.

Wiltschko, A.B., Johnson, M.J., Iurilli, G., Peterson, R.E., Katon, J.M., Pashkovski, S.L., Abraira, V.E., Adams, R.P., and Datta, S.R. (2015). Mapping Sub-Second Structure in Mouse Behavior. Neuron *88*, 1121–1135.

Wise, T., and Dolan, R.J. (2020). Associations between aversive learning processes and transdiagnostic psychiatric symptoms in a general population sample. Nat. Commun. *11*, 4179.

Wise, T., Michely, J., Dayan, P., and Dolan, R.J. (2019). A computational account of threat-related attentional bias. PLoS Comput. Biol. *15*, e1007341.

Wise, T., Liu, Y., Chowdhury, F., and Dolan, R.J. (2020). Model-based aversive learning in humans is supported by preferential task state reactivation. bioRxiv. https://doi.org/10.1101/2020.11.30.404491.

Wu, C.M., Schulz, E., Speekenbrink, M., Nelson, J.D., and Meder, B. (2018). Generalization guides human exploration in vast decision spaces. Nat. Hum. Behav. *2*, 915–924.

Yan, X., Yang, J., Sohn, K., and Lee, H. (2016). Attribute2Image: Conditional Image Generation from Visual Attributes. European Conference on Computer Vision, 776–791.

Ydenberg, R.C., and Dill, L.M. (1986). The Economics of Fleeing from Predators. In Advances in the Study of Behavior, J.S. Rosenblatt, C. Beer, M.-C. Busnel, and P.J.B. Slater, eds. (Academic Press), pp. 229–249.

Yilmaz Balban, M., Cafaro, E., Saue-Fletcher, L., Washington, M.J., Bijanzadeh, M., Lee, A.M., Chang, E.F., and Huberman, A.D. (2021). Human Responses to Visually Evoked Threat. Curr. Biol. *31*, 601–612.e3.

Yoo, S.B.M., Tu, J.C., Piantadosi, S.T., and Hayden, B.Y. (2020). The neural basis of predictive pursuit. Nat. Neurosci. *23*, 252–259.