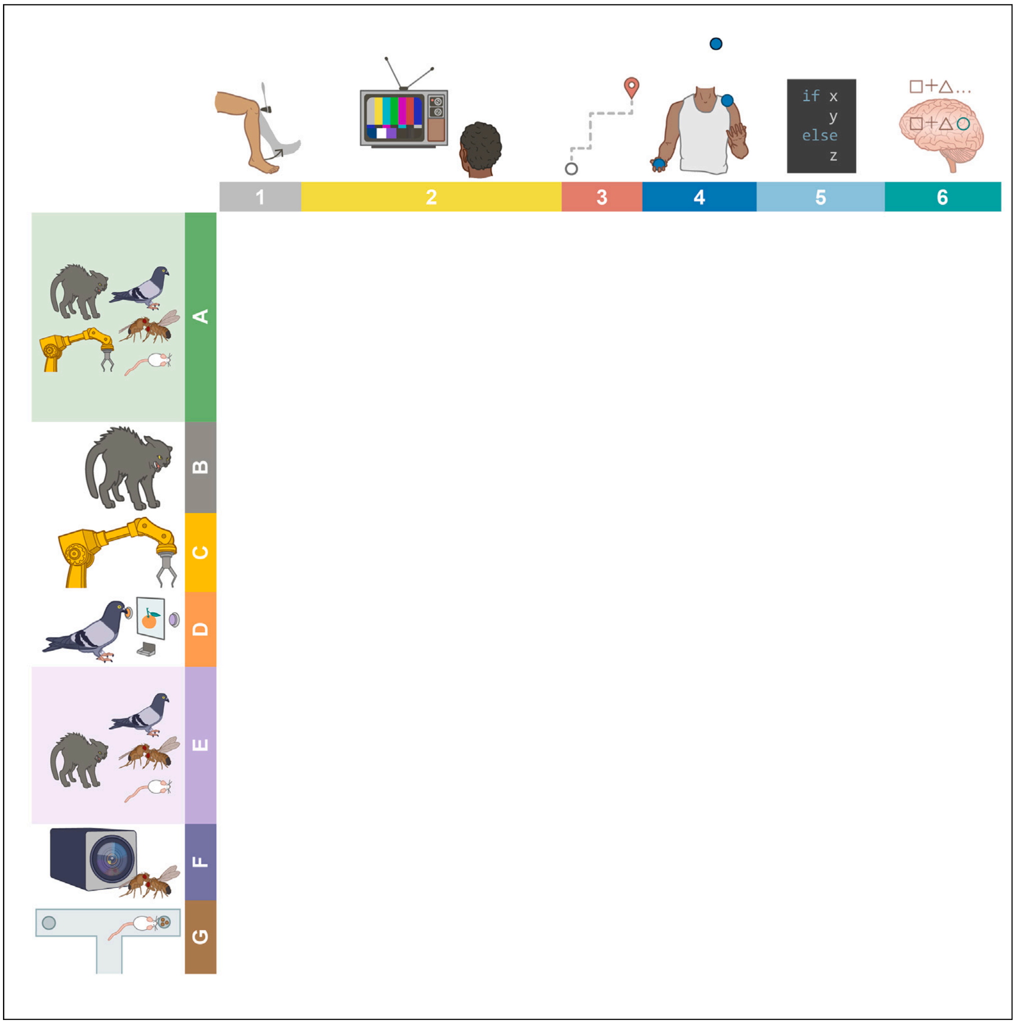
iScience

CellPress
OPEN ACCESS

Article

Everyone knows what behavior is but they just don't agree on it



Adam J. Calhoun, Ahmed El Hady

adam.calhoun@gmail.com (A.J.C.)
ahady@ab.mpg.de (A.E.H.)

Highlights

Large-scale quantitative survey about the meaning of behavior across academic disciplines

Identified discrete definitions of behavior

Individuals have consistent definitions

Different disciplines use different definitions for the word "behavior"

Calhoun & El Hady, iScience *26*, 108210
November 17, 2023
https://doi.org/10.1016/j.isci.2023.108210

iScience

CellPress
OPEN ACCESS

Article

Everyone knows what behavior is but they just don't agree on it

Adam J. Calhoun[1,*] and Ahmed El Hady[1,2,3,4,*]

SUMMARY

Studying "behavior" lies at the heart of many disciplines. Nevertheless, academics rarely provide an explicit definition of what "behavior" actually is. What range of definitions do people use, and how does that vary across disciplines? To answer these questions, we have developed a survey to probe what constitutes "behavior." We find that academics adopt different definitions of behavior according to their academic discipline, animal model that they work with, and level of academic seniority. Using hierarchical clustering, we identify at least six distinct types of "behavior" which are used in seven distinct operational archetypes of "behavior." Individual respondents have clear consistent definitions of behavior, but these definitions are not consistent across the population. Our study is a call for academics to clarify what they mean by "behavior" wherever they study it, with the hope that this will foster interdisciplinary studies that will improve our understanding of behavioral phenomena.

INTRODUCTION

The scientific definition of "behavior" has tended toward the maxim "I know it when I see it." But does everyone know and see the same thing? Much ink has been spilled coming up with precise definitions of behavior, stretching from Aristotle[1] through Wiener's cybernetics (Wiener[2]) (but see also Rosenblueth et al.,[3] Von Neumann,[4] Lorenzian ethology (Lorenz[5]), Tinbergen behavioral theory (Tinbergen[6]), and Von Uexkuell's umwelt (Uexküll[7])). More recent work has operationalized such definitions to create machine learning algorithms capable of "recognizing" when an animal is performing some "behavior" (Berman et al.[8]; Wiltschko et al.[9]; Calhoun et al.[10]). Instead of trying to think up a personal definition of behavior, we could simply ask how experts in a field actually use the word. Within behavioral biology, Levitis et al.[11] surveyed the literature and found 25 distinct operational definitions of behavior. These range from the vague ("what an animal does," Davis[12]) to the specific (A response to external and internal stimuli, following integration of sensory, neural, endocrine, and effector components). Surveys on biological concepts such as "genes" and "consciousness" (JC et al.[13]; Stotz et al.[14]) have similarly identified highly divergent meanings of words that form the basis of their respective fields. However, "behavior" is a concept not only used by behavioral biologists but across many fields. To date, there are no surveys that cover a wide variety of academic fields when it comes to the definition of behavior.

Implicit and unstated definitions of behavior have resulted in scientific results that are overturned or narrowed when the definition is made explicit. For instance, chemotaxis (climbing a chemical gradient) has canonically been assumed to be produced by individuals but instead may be a collective action (Ramdya et al.[15]); fertilization was canonically viewed as a sperm acting upon an egg (insemination behavior), when the ovum actively participates (Kekalainen and Evans[16]); pathways in the brain classically thought to be involved in biasing motor output may instead be biasing cognitive strategies (Bolkan et al.[17]); and so on. In all of these cases, a particular definition of behavior was assumed which precluded the study of the actual underlying phenomena. This is why we need clear definitions of behavior.

There are good reasons to believe that different fields use different definitions and that these definitions in turn influence what is studied. For instance, consider the study of a risky choice. The behavior that an economist might be interested in are the biases and heuristics that people employ while performing real-world decisions, while a neuroscientist seeks to understand how the "behavior" is mediated via neural mechanisms. On the other hand, behavioral ecologists seek to study how "behavior" is shaped by the environment, while sociologists seek to understand society through the "behavior" of its individuals. While each of these researchers seeks to understand "behavior," they are attempting to understand fundamentally different things by virtue of these disciplinary differences in mechanism.

This can reflect deeper, more fundamental disagreement about what "behavior" is. For instance, sensorimotor behavior is behavior that relates sensory input to motor output. But is this "behavior" when it is performed by a rat? How about *C. elegans*, with only three or four

[1]Princeton Neuroscience Institute, Princeton University, Princeton, NJ, USA
[2]Cluster for Advanced Study of Collective Behavior, University of Konstanz, Konstanz, Germany
[3]Max Planck Institute of Animal Behavior, Konstanz, Germany
[4]Lead contact
*Correspondence: adam.calhoun@gmail.com (A.J.C.), ahady@ab.mpg.de (A.E.H.)
https://doi.org/10.1016/j.isci.2023.108210

synapses between sensory neurons and motor output? What about a bacterium or a computer program? Are the neurons themselves "behaving"? Furthermore, not only is the animal moving but there are cognitive variables and needs that influence the animal's action. Which of these constitute the "behavior"?

Instead of creating our own definition of behavior, we aimed to identify how the word "behavior" is actually used by the academic community. We designed a survey aimed at determining the answer to this question. Through a quantitative analysis of survey data, we show that what constitutes "behavior" differs between academic disciplines, level of academic seniority, and type of animal model organism studied. We further show that there are at least six "types" of behavior that are used to construct seven archetypes of behaviors.

The study pinpoints the lack of a common interdisciplinary definition of behavior. We argue that we do not need an agreed-upon singular definition of behavior, but we do have to be clear on what the definitions that we are using are in any given study. Those clear working definitions will open up the opportunity for effective interdisciplinary studies that tackle multiple facets of "behavior." We conclude that there is no "*the* behavior" animals produce, but rather they are producing different facets of the many concepts that encompass the singular word "behavior." We offer clear ways that explicitly identifying these archetypes can improve our ability to understand behavior.

RESULTS

Widespread disagreement of definitions of behavior

We developed a set of questions designed to probe how people define "behavior." For instance, "A behavior is always the output of motor activity", "Is learning a behavior?", and "Is a dog that follows a scent trail behaving?". We included questions that were designed to elicit inconsistencies such as, "Is a reflex a behavior?" and "Is sweating a behavior?". The order of questions was randomized and the survey was distributed online and elicited responses from a range of academic disciplines and seniorities (see STAR Methods, Tables S1 and S2). Analysis revealed that respondents did not provide a consistent answer for the majority of questions: 43/48 answers had fewer than 80% of respondents answering in the same manner (Table S1; Figure S1A). While there were some questions that showed widespread agreement in either a "Yes" ("Is a dog that follows a scent trail behaving") or a "No" ("Does a behavior need to be intentional") answer (Figures S1B–S1D), most questions showed disagreement. The questions with the most disagreement (defined as the highest uncertainty, or entropy, in their responses) included "Can computer programs behave?" and "A behavior always involves interaction with the environment?" (Figure 1A). We remain agnostic as to the source of disagreement between responses: it is possible that different subjects are responding to different aspects of each question. This does not impact the interpretation of our results as we believe this itself provides a consistent signal about underlying beliefs (see Discussion).

A previous survey of professionals in animal behavior (Levitis et al.[11]) had observed that respondents were inconsistent and self-contradictory. Was this true for our survey? We used a regression model to ask whether we could predict responses from the answers to previous questions and found that we could (Figure 1B). This was true both on average and across all questions. This suggests that disagreement on behavioral questions represents disagreement on the underlying definition of behavior itself.

We then asked whether this variation in responses could be due to differences in academic disciplines—perhaps those trained in the Humanities would have a fundamentally different idea of "behavior" from those trained in Engineering, for instance. We embedded responses in a lower dimensional space using multiple correspondence analysis, a factor analysis technique for categorical variables (Le Roux and Rouanet[18]). While the factors can be difficult to interpret, examination of the questions most responsible for the first two factors suggest they are related to whether reflexes are behaviors (Factor 1) and cognition (Factor 2) (Figure S2). For visualization purposes, we grouped subjects into broad academic categories: Biological Sciences, Math/Engineering, Ecology, and Humanities (see STAR Methods for definitions, but see Figure S3 for all disciplines). We found differences between each field (Figures 1C, S3A, and S4A). We hypothesized that even within a field there might be substantial differences; a Cognitive Neuroscientist gets different training from a Systems Neuroscientist. Again, plotting the mean responses for each sub-field showed significant differences (Figures 1D and S4B). We had expected molecular neuroscience to be relatively more similar to systems+circuits or computational neuroscience because they investigate similar physical substrates. However, we found that molecular neuroscience was more similar to cognitive neuroscience, which investigates abstract processes.

In behavioral research, scientists can ask fundamentally different questions about behavior when they work with humans than when they work with invertebrates. For instance, *C. elegans* researchers typically investigate chemotaxis and thermotaxis, which are activities that involve sensorimotor transformations, while researchers who study humans may be more likely to study cognitive activities like abstract decision-making. Additionally, those who study invertebrates can map behavior to a single-neuron level, something not typically done in humans. We thus divided subjects by the model organism that they work with. We again find differences between model organisms, where researchers who work with invertebrates have a more similar definition of behavior to those who work with other non-mammalian organisms compared to those who work with mammals or humans (Figures 1E, S3B, and S4C). Those who don't work with animals but work with computer models instead (*in silico*) are the most distinct in their definition. Surprisingly, we also found that academic experience is also associated with differences in behavioral definitions, with a steady cline as experience increases (Figures 1F and S4D).

Hierarchical clustering finds definitions

One possibility is that these differences between fields are not because each has a different single definition of behavior, but because certain behavioral definitions are more common in that field. In order to identify potential definitions of behavior, we performed hierarchical

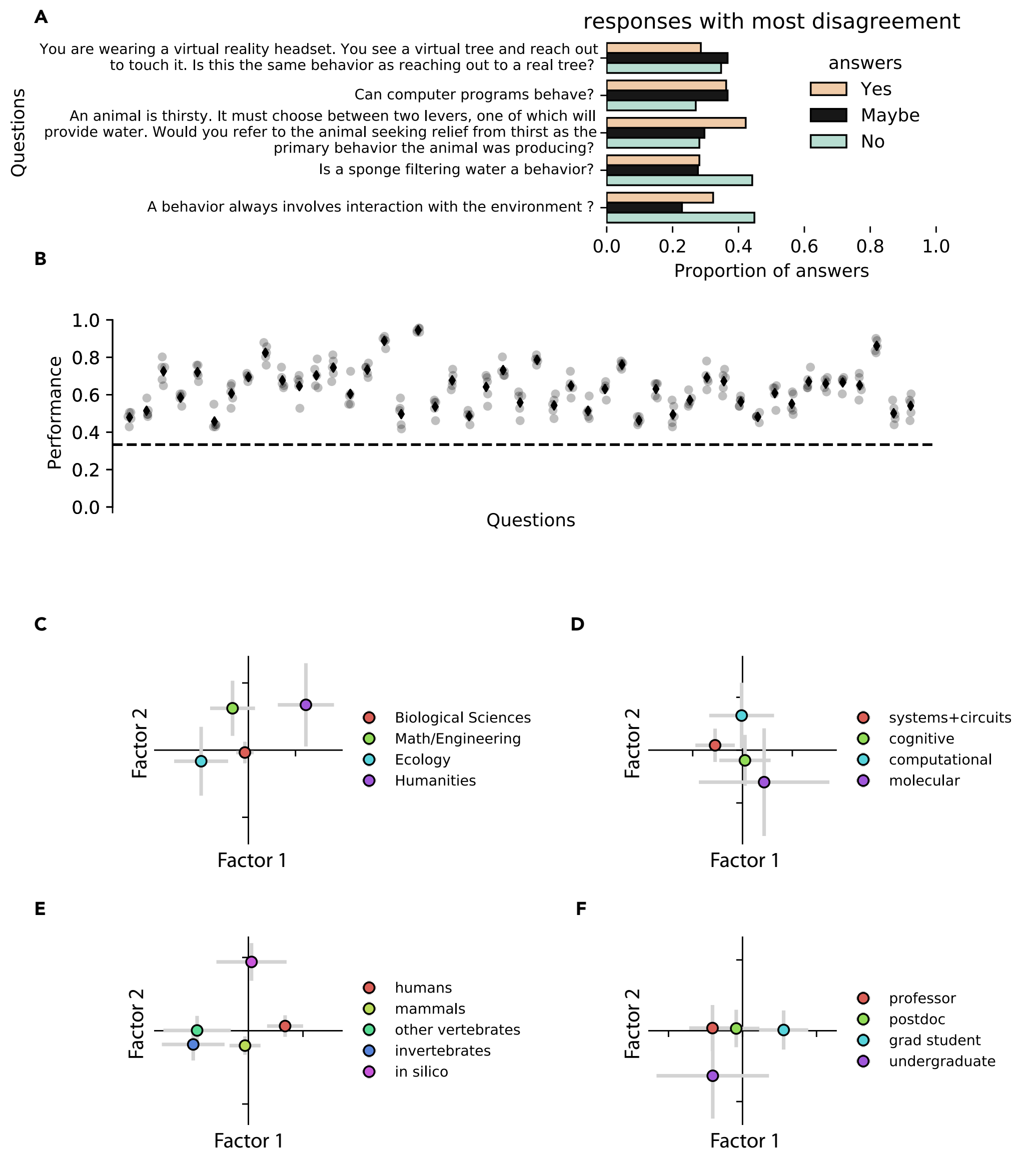


Figure 1. Definitions of behavior are varied but internally consistent

(A) The five questions with the most disagreement between respondents as defined by the response entropy (n = 455).

(B) Responses are predictable with a regression model. Gray dots represent the performance at predicting answers on held-out data using 5-fold cross-validation, black diamond is mean. Mean prediction accuracy across responses is 63%. The dashed line represents chance performance.

(C–F) We used a factor analysis (multiple correspondence analysis (MCA)) to reveal how groups responded on similar questions. Plotting responses on the two largest factors, colored by (C) scientific field, (D) neuroscience specialty, (E) model organism used, and (F) academic seniority. All data shown are the mean, and error bars are SEM.

clustering on both the responses and respondents (Figure 2). We found six potential classes of questions and seven potential classes of responses.

We began by identifying whether the different clusters were meaningful. Examining the questions in the response clusters revealed that similar questions were in each cluster (Figure 3). For instance, Cluster 1 was filled with questions about whether reflexes count as behaviors (“Is a baby urinating a behavior?”, “Is a reflex a behavior?”, “Is the sucking reflex a behavior?”, and so on). Based on this, we term these “behavior definition clusters” and labeled them as the following.

1. Reflex (Figure 3A)
2. Actions (Figure 3B)
3. Understanding the mind (Figure 3C)

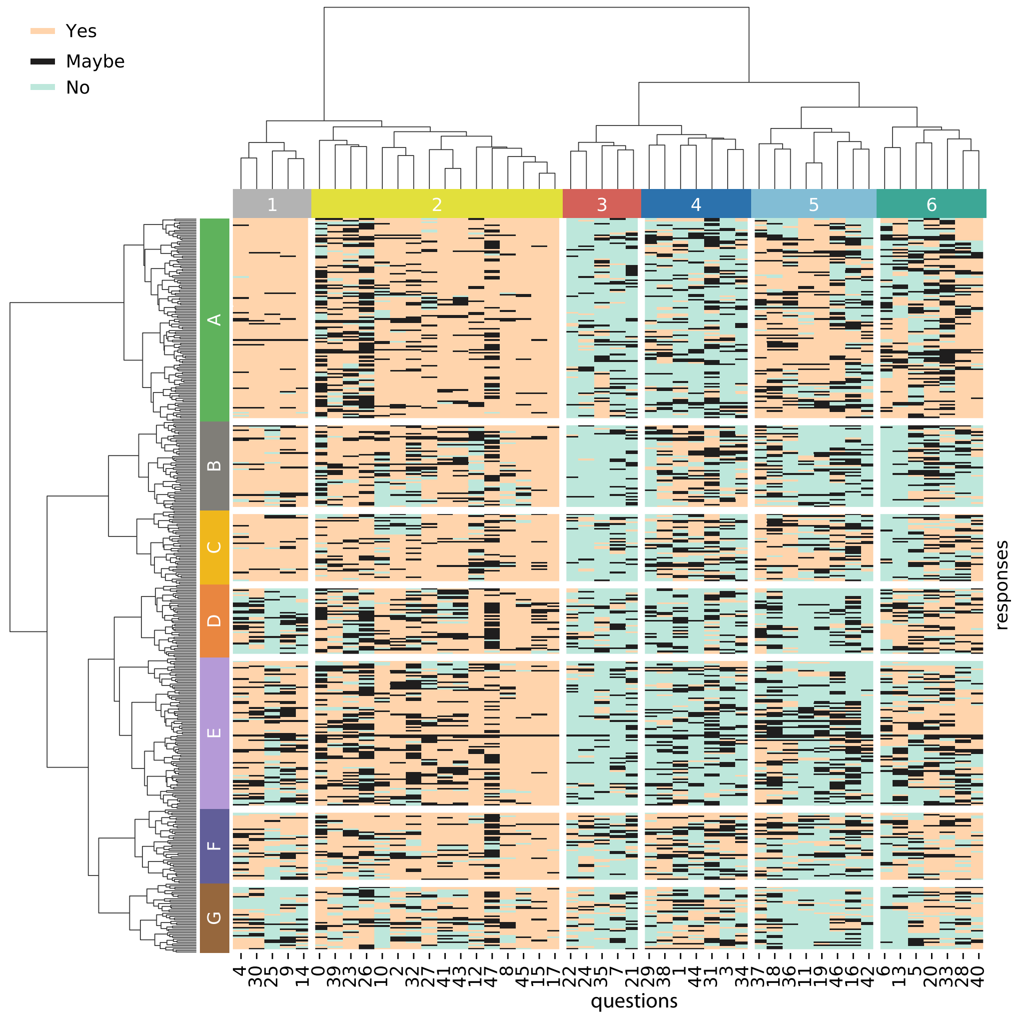


Figure 2. Survey responses reveal clustered definitions of behavior

Hierarchical clustering on the 455 responses reveals six clusters in the questions and seven clusters in the responses. The six categories of questions, labeled 1–6, are unique definitions of behavior, and the seven categories of responses, labeled A–G, are behavioral archetypes. See Figures 3 and 4.

4. Motor or sensorimotor (Figure 3D)
5. Non-animal (Figure 3E)
6. Cognition (Figure 3F)

Behavioral archetypes

We next examined what each response cluster corresponds to. For each respondent cluster—which we call a "behavioral archetype," answers were relatively consistent in each behavioral definition but answers varied widely between clusters (Figures 4A and S5). For instance, in the "reflex" behavioral definition, the mean across all responses was 67% "Yes". However, the "A" archetype had 94% of all responses as "Yes" while the "D" and "G" clusters had 19% and 25% of responses as Yes, respectively. To better understand how each archetype corresponds to the respondent's answers, we plotted the change in responses relative to the overall mean for "Yes", "No", and "Maybe" responses (Figures 4B and S6).

Based on these, we named these archetypes by examining which behavioral definitions were being used.

A Broad
B Motor outputs (animals only)
C Broad (but mostly motor)
D Cognition (and anthropomorphizable animal behaviors)

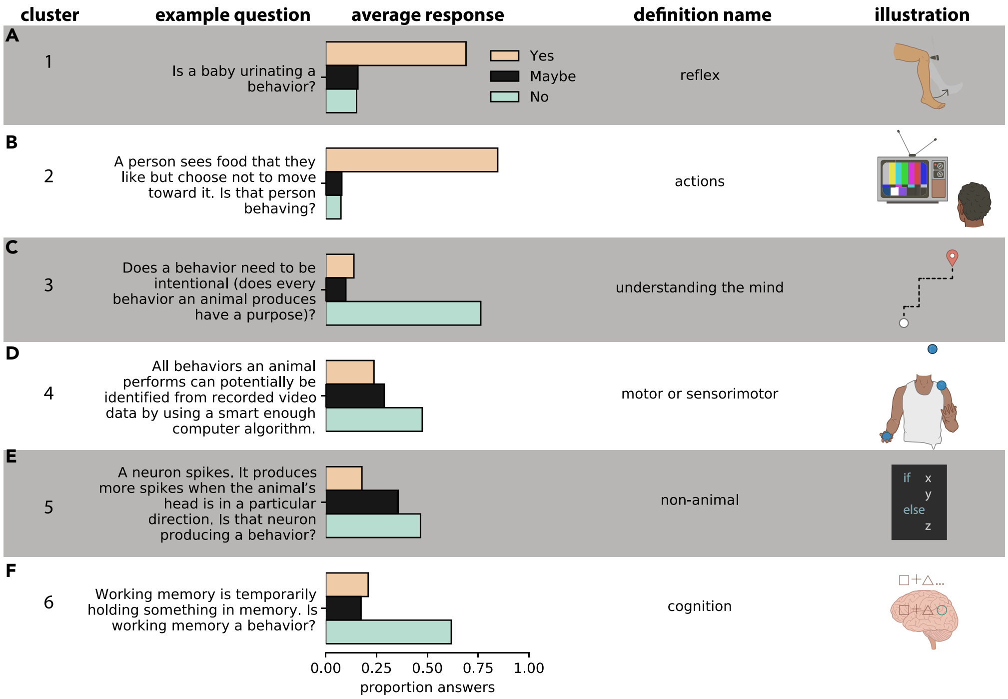


Figure 3. Behavioral definitions revealed by clustering

(A–F) Example questions for each type of behavior (left). Illustration representing that type of behavior (right). These definitions are (A) reflex, (B) actions are behavior, (C) we must understand the mind of an animal to identify its behavior, (D) motor or sensorimotor, (E) non-animal behaviors, (F) learning and memory/cognition.

E Broad (animals only)
F Well-defined/understood animal behaviors
G Animals acting with intentions

Within each archetype, there were unique patterns of beliefs for each definition (Figure 4). For instance, whereas "Broad" (A) archetypes accept any reflex as constituting a behavior, "Cognition" (D) and "Animal intentions" (G) rejected almost all reflexes as constituting "behavior." "Cognition" was more uncertain, being much more likely to respond with "Maybe" to whether a baby urinating constitutes behavior (Q4) or whether a reflex counts as behavior at all (Q9), even while both flatly reject the knee-jerk reflex as a behavior (Q25). These responses do not exist in a vacuum but must be interpreted in relation to each other; in this example, for instance, it is possible that respondents are uncertain whether a baby urinating may have cognitive motives that go beyond simple reflex.

Prior to conducting this study, we had expected two prominent archetypes to appear in responses: behaviorism, in which behavior can be explained through conditioned motor patterns without regard to cognitive states (Skinner[19,20]), and cognitivism, in which cognitive processes are thought to be driving behavior and are central to it (Haugeland[21]). More specifically, behaviorism (Skinner[19,20]) considers behavior as either a reflex to a certain stimulus (Dewey[22]) or one that is shaped by past reinforcement or punishment contingencies.

However, it is unclear if the colloquial definitions of behavior used by the experts surveyed here actually relate to these definitions. In order to answer these questions, we assigned each of our survey questions to supporting behaviorism, supporting cognitivism, or being unrelated to either (see STAR Methods). Individual responses were assigned a score from −1 to 1 relating to the number of "Yes"es or "No"s that were answered for the relevant questions (Figure 5). We find that very few respondents are strongly Behaviorist or Cognitivist, on either an individual level (Figure 5, histograms) or a group level. We thus conclude that if we were to define "behavior" in a purely Behaviorist or Cognitivist fashion, we would not be using the term as it is actually used, and would not be studying what is actually being studied.

Taking this all together, we show how behavioral archetypes are made up of different behavioral definitions (Figure 6).

Academic fields use different definitions

We next used these behavioral archetypes to ask whether different academic fields are consistently using different archetypes when they refer to "behavior." Respondents who come from the humanities were dramatically less likely to be part of the "broad" (A) archetype and much more likely to be a part of the "acting with intentions" (G) archetype. Ecologists were relatively more likely to be part of the "broad (mostly motor)" (C) and "broad (animals only)" (E) archetypes (Figure 7A). Within neuroscience specialties, molecular neuroscientists were more likely

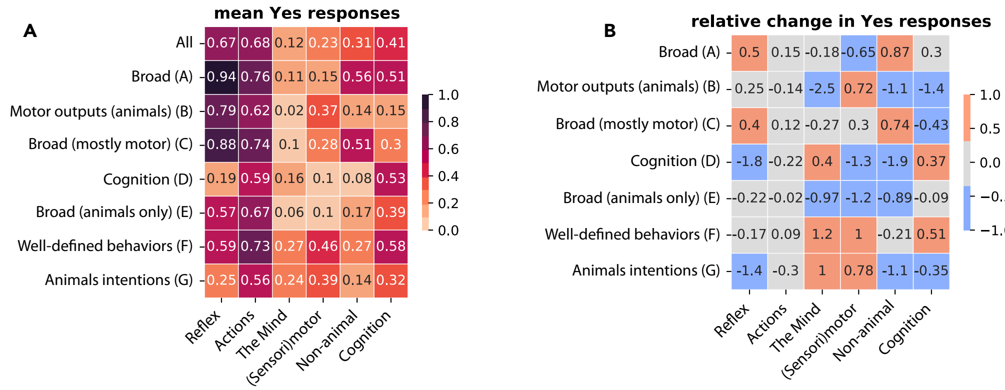


Figure 4. Validation and consistency of behavioral archetypes

(A) The mean "Yes" response in each category and (B) its relative (log-valued) difference from the overall average allow us to understand what the definitions are composed of.

to be a part of the "broad (mostly motor)" (C) archetype and cognitive to be part of the "well-defined/understood behaviors" (F) archetype. The "broad (animals only)" (E) archetype was almost entirely Systems and Computational neuroscientists (Figure 7B).

When broken down by the organism used for the work, those who worked with humans were the least likely to use "broad" (A) and "broad (animals only)" (E) archetypes but the most likely to use the "cognition" (D) and "well-defined/understood behaviors" (F) definitions. "Other vertebrates" dominated "broad" (A) and "broad (animals only)" (E) and invertebrates were 50% more likely to be a part of "broad (but mostly motor)" (C) (Figure 7C).

Undergraduates were predominantly a part of archetypes "broad" (A) and "acting with intentions" (G), whereas professors were the most likely to be part of cluster "motor outputs (animals only)" (B) (Figure 7D).

DISCUSSION

In this study, we have shown that "behavior" is not universally defined by academics. We have developed and analyzed a survey that has allowed us to identify how the term "behavior" is used, and what it means to people belonging to different academic communities. Using this survey, we have shown that important differences exist in what an academic community would consider to be "behavior," and additionally identified six behavior clusters and seven archetype clusters. These behavioral archetypes correspond roughly to the following labels "broad," "animal motor outputs are behavior," "broad but motor," "animal+cognitive," "animal motor+cognitive," "intentional motor behaviors," and "intentional motor outputs by animals." We speculate that other sub-categories exist—we could, for instance, have cut the tree at other branches or used even more questions and gotten more responses.

Overall, there are three concrete concepts to take away from the results of the survey and resulting analysis. First, that what we call "behavior" is highly heterogeneous—and that is okay. Contrary to hierarchical theories of behavior (Rosenblueth et al.[3]), animals are producing many different things that constitute "behavior" in parallel. Each of these likely have their own relationship to the animals' internal world (or nervous system) and external world. Second, each of these will inform us differently about the relationship between the animal and its internal/external world, resulting in different scientific questions and needing distinct experimental paradigms. This could be the relationship between motor output and kinematics, the production of repeated behavioral motifs, behavioral states, cognitive strategies, or collective behavior, among many other things. Finally, that clear definitions of "behavior" will have an electrifying effect on our ability to analyze and understand it, as already seen by the recent work quantifying certain aspects of animal "behavior" (Berman et al.[8]; Wiltschko et al.[9]; Mathis et al.[23]; Pereira et al.[24]).

Our study is the first to report the heterogeneous definitions of "behaviors" across academic communities, as previous studies have focused on a particular sub-field. For example, Levitis et al.[11] looked at how behavioral biologists define "behavior" and found them to be self-contradictory. However, we have shown that we are able to predict the subjects' definition of behavior on an individual basis and the definitions are to a great extent self-consistent relative to some underlying set of beliefs. We believe that a follow-up survey focusing on particular terms in behavior, such as "decision-making," "foraging," or "computation" could lead to similar classification of terms that are widely used in literature. It is important to note that we have taken our utmost care in formulating the questions using precise language but this does not prevent that some of the words might be read in a more complex manner than what we intended to. This is especially true with words that have ambiguous boundaries such as whether a "computer behaves." It might be understood as "it is useful sometimes to consider a computer's output as a behavior." We instead remain agnostic to how people interpret these questions, and primarily interpret the answers as reflecting some common underlying set of beliefs.

Despite all this, these inconsistent and unclear definitions may obscure otherwise clear differences. For instance, one of the most controversial questions in our survey was Q31: "You are wearing a virtual reality headset. You see a virtual tree and reach out to touch it. Is this the same behavior as reaching out to a real tree?", with roughly equal numbers of Yes, No, and Maybe answers. This has direct relevance to many

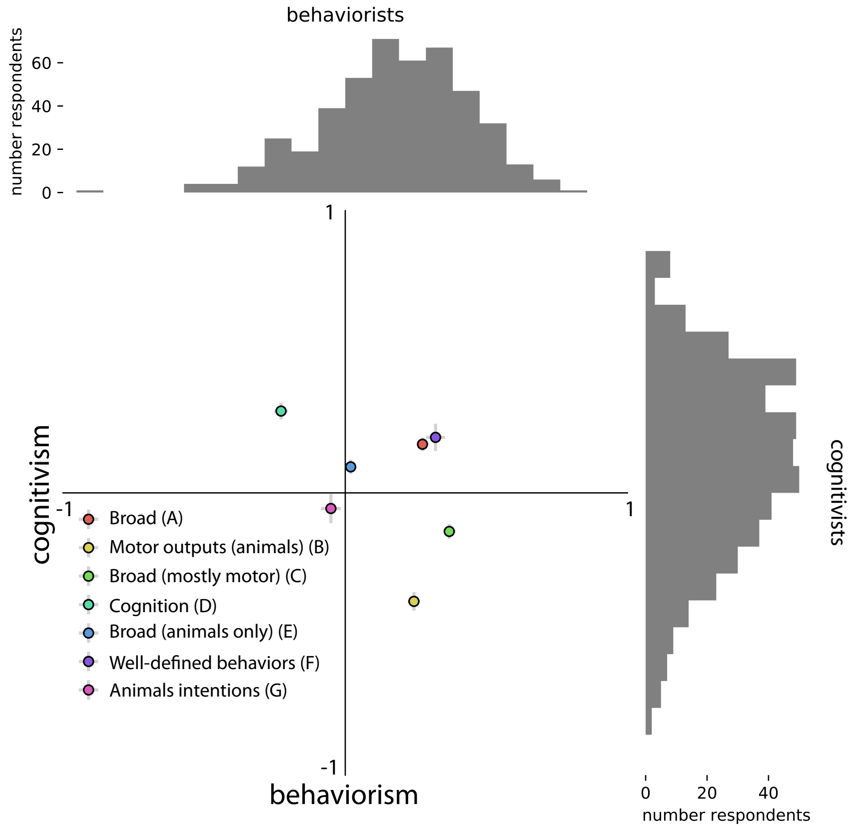


Figure 5. No one cares about behaviorism and cognitivism

Questions were labeled "behaviorist," "cognitivist," or "neither." Responses were scored to either totally agree (+1) or totally disagree (−1) with each category. Respondents did not agree with behaviorism or cognitivism, but used definitions that were a mix of the two. Circles represent mean and error bars are +/− SEM.

experiments in neuroscience, where behaviors are "simplified" for experimental tractability: a rodent may be fixed in place and make a "decision" by licking from one of two reward ports. Would this "behavior" be the same in a different context in which the animal was free to explore and move about? It is increasingly apparent that movement itself influences many aspects of the decision process, and animals solve decisions differently when they are head fixed and when they are freely moving (Eisenreich et al.[25])—which would not be obviously apparent if we consider them the "same behavior."

Academics often resort to behavioral reductionism by choosing a simplified version of a behavior and then claim that this reflects the entirety of the behavior under study. However, this can be highly misleading for anyone outside of that academic's field, who may not understand the implicit definition of the "behavior," with all of its caveats and consequences. For example, a series of nose pokes in an operant chamber may constitute behavior in a systems neuroscience experiment. On the other hand, an anthropologist or a lay person would focus on other aspects, for example the grooming of the animal, the tail movement, and so on. Even within one's own work, there is the danger of reifying the simplified version of a behavior as "the behavior" the animal is producing, rather than the connected set of other aspects of its behavior which may be vital to the question at hand. While we acknowledge the limitations of studying multiple levels of behavioral descriptions simultaneously, we also stress that we should be cognizant of those limitations, indicating which level we are studying and use more precise language when describing it. This also reflects on the scientific literature that academics publish and if the exact level of description/shortcomings are clarified, interdisciplinary interactions would be far easier and more fruitful.

We believe this survey makes clear that "behavior" constitutes a wide class of concepts (as seen through the different definitions). One clear suggestion that can be taken from this is the need for integrating different types of "behavior" to make further advances on its study. There are many reviews that have provided a critical analysis of how behavior is studied (Gomez-Marin and Ghazanfar[26]; Krakauer et al.[27]). To supplement those critiques, here we focus on quantitative analysis of how a sample of academics define behaviors thus providing evidence for the variety of the epistemological frameworks at play in different academic fields. A challenge here is that the hegemony of particular techniques in an academic field can lead to methodological bias which itself delineates the meaning of "behavior." For example, in systems neuroscience researchers will systematically use brain recording techniques to look for "correlates" of "behavior," leading to a neurocentric view of behavioral phenomenology. Others will use pure video tracking combined with machine learning algorithms to classify states of behavior regarded in this sense as a behavioralist mechanism. Although we also recognize the difficulty of combining different methods, technological

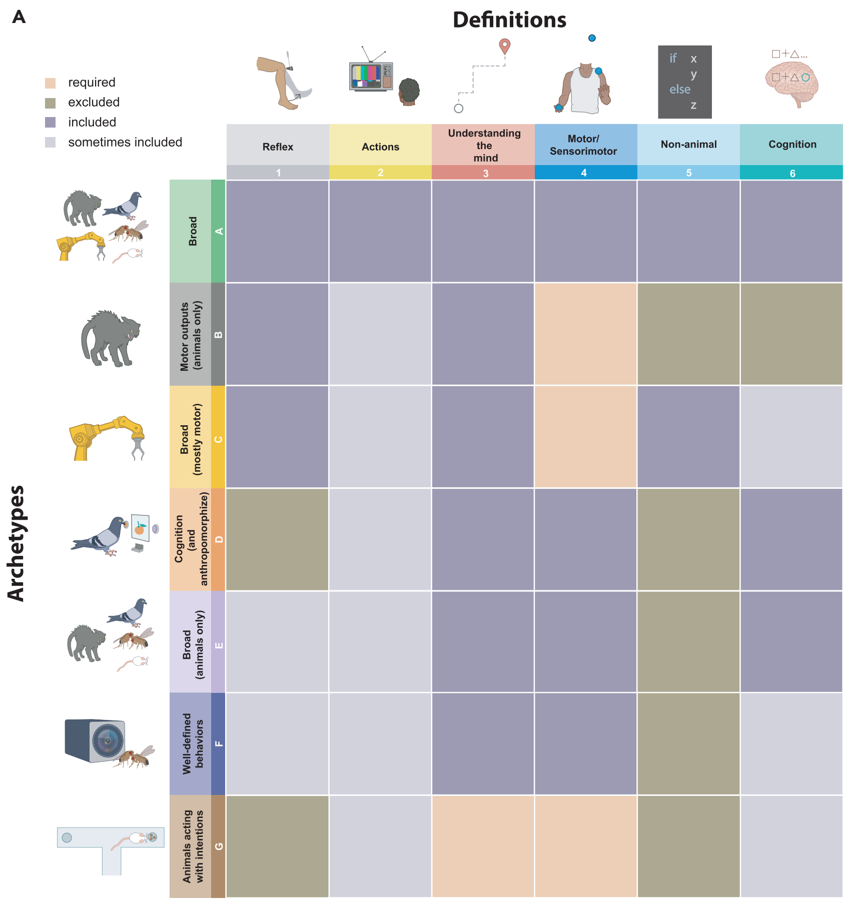


Figure 6. Validation and consistency of behavioral archetypes

(A) Behavioral archetypes are built out of the responses to different definitions. To illustrate the archetypes, we use the relative responses (Figures 4, S5, and S6). For each archetype, we color the definitions by whether they are required (pale pink), excluded (brown), included (purple), or sometimes included (gray).

advances are making the possibility of combining multiple techniques more streamlined (Markowitz et al.[28]). It is possible to go further and take an integrative approach which considers how the ecological environment is shaping the behavior alongside the brain, the body, and the animal physiological state.

Limitations of the study

However, we have shown that we are able to predict the subjects' definition of behavior. The individuals' definitions are to a great extent self-consistent relative to some underlying set of beliefs. We believe that a follow-up survey focusing on particular terms in behavior, such as "decision-making," "foraging," or "computation" could lead to similar classification of terms that are widely used in literature. It is important to note that we have taken our utmost care in formulating the questions using precise language but this does not prevent that some of the words might be read in a more complex manner than what we intended to. This is especially true with words that have ambiguous boundaries such as whether a "computer behaves." It might be understood as "it is useful sometimes to consider a computer's output as a behavior." We instead remain agnostic to how people interpret these questions, and primarily interpret the answers as reflecting some common underlying set of beliefs which might be regarded as a limitation for the study at hand. Another limitation is that a sliding bar from No to Yes with Maybe in the center would have led to more nuanced unbiased results instead of choosing among Yes, No, and Maybe, which might exaggerate the choice of Maybe.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability

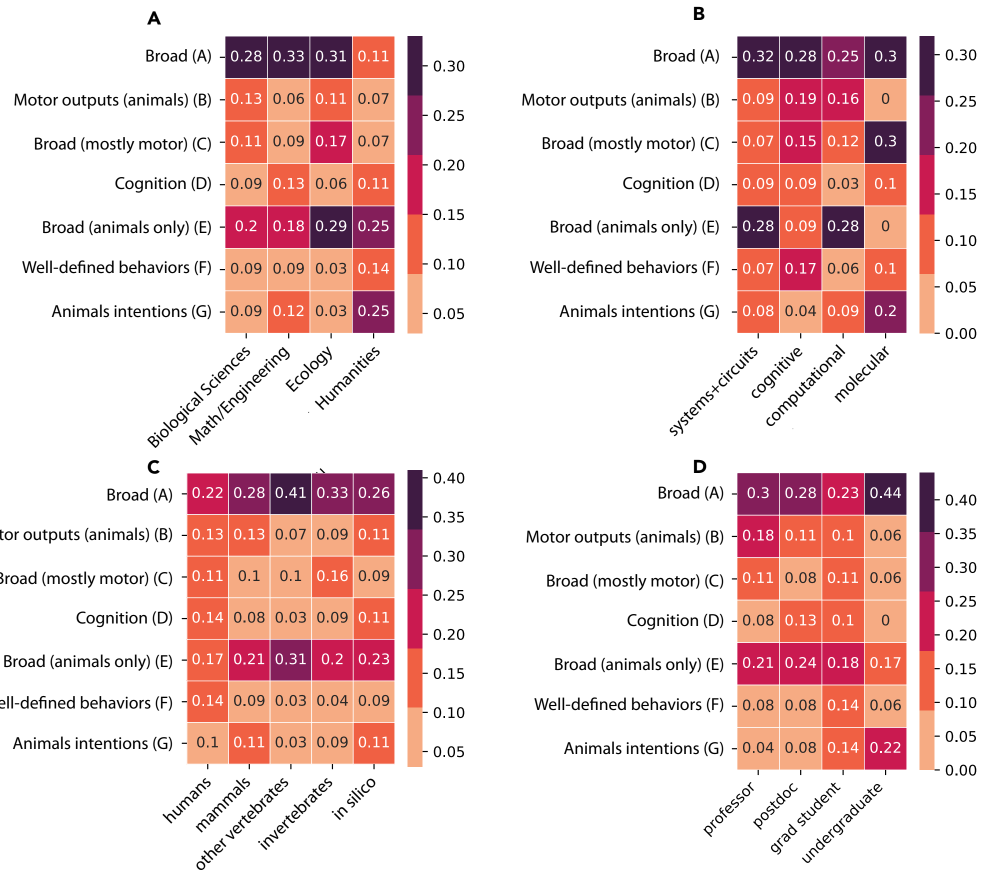


Figure 7. Different groups have distinct distribution of definitions

(A) Academic fields show different propensities to use each category of "behavior." This is also true of (B) neuroscience specialties, (C) model organism used, and (D) academic seniority. See STAR Methods for definitions of each field.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2023.108210.

ACKNOWLEDGMENTS

We would like to thank Emily Dennis, Emily Mackevicius, Asif Ghazanfar, Yael Niv, David Barack, Jakob Voigts, Patrick Dylan Rich, Selmaan Chettih, Jess Breda, Sandeep Robert Datta, Benjamin Hayden, and Alex Gomez-Marin for insightful feedback. We would also like to thank Kathleen Quach for the illustrations used in the manuscript. This work was funded by the Simons Foundation SCGB (AWD494712, A.J.C.) as well as the NIH (1R21MH121889-01, A.E.H.).

AUTHOR CONTRIBUTIONS

A.J.C. and A.E.H. have both designed the survey, analyzed the data, and wrote the manuscript.

DECLARATION OF INTERESTS

Authors have no competing interests.

Received: December 20, 2021
Revised: March 8, 2023
Accepted: October 11, 2023
Published: October 18, 2023

REFERENCES

1. Aristotle (350 BC), 'On the Soul'.
2. Wiener, N. (2019). Cybernetics or Control and Communication in the Animal and the Machine (MIT press).
3. Rosenblueth, A., Wiener, N., and Bigelow, J. (1943). Behavior, purpose and teleology. Philos. Sci. *10*, 18–24.
4. Von Neumann, J. (2012). The Computer and the Brain (Yale university press).
5. Lorenz, K. (2013). The Foundations of Ethology (Springer Science & Business Media).
6. Tinbergen, N. (1972). The Animal in its World: Explorations of an Ethologist, *84* (Harvard University Press), pp. 1932–1972.
7. Uexküll. (1921). Umwelt und innenwelt der tiere (Springer Berlin Heidelberg).
8. Berman, G.J., Choi, D.M., Bialek, W., and Shaevitz, J.W. (2014). Mapping the stereotyped behaviour of freely moving fruit flies. Journal of The Royal Society Interface *11*, 20140672.
9. 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.
10. 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.
11. Levitis, D.A., Lidicker, W.Z., Jr., and Freund, G. (2009). Behavioural biologists do not agree on what constitutes behaviour. Anim. Behav. *78*, 103–110.
12. Davis, D.E. (1966). Integral Animal Behavior.
13. Francken, J.C., Beerendonk, L., Molenaar, D., Fahrenfort, J.J., Kiverstein, J.D., Seth, A.K., and van Gaal, S. (2021). An academic survey on theoretical foundations, common assumptions and the current state of the field of consciousness science. Preprint at PsyArXiv. https://doi.org/10.1093/nc/niac011.
14. Stotz, K., Griffiths, P.E., and Knight, R. (2004). How biologists conceptualize genes: an empirical study. Stud. Hist. Philos. Sci. C Stud. Hist. Philos. Biol. Biomed. Sci. *35* (4), 647–673.
15. Ramdya, P., Lichocki, P., Cruchet, S., Frisch, L., Tse, W., Floreano, D., and Benton, R. (2015). Mechanosensory interactions drive collective behaviour in drosophila. Nature *519*, 233–236.
16. Kekalainen, J., and Evans, J.P. (2018). Gamete-mediated mate choice: towards a more inclusive view of sexual selection. Proc. Roy. Soc. B *285*.
17. Bolkan, S.S., Stone, I.R., Pinto, L., Ashwood, Z.C., Iravedra Garcia, J.M., Herman, A.L., Singh, P., Bandi, A., Cox, J., Zimmerman, C.A., et al. (2021). Strong and opponent contributions of dorsomedial striatal pathways to behavior depends on cognitive demands and task strategy. Preprint at bioRxiv. https://doi.org/10.1101/2021.07.23.453573.
18. Le Roux, B., and Rouanet, H. (2010). Multiple Correspondence Analysis, *163* (Sage).
19. Skinner, B.F. (1986). Is it behaviorism? Behav. Brain Sci. *9*, 716.
20. Skinner, B.F. (2011). About Behaviorism (Vintage).
21. Haugeland, J. (1978). The nature and plausibility of cognitivism. Behav. Brain Sci. *1*, 215–226.
22. Dewey, J. (1896). The reflex arc concept in psychology. Psychol. Rev. *3*, 357–370.
23. 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.
24. Pereira, T.D., Aldarondo, D.E., Willmore, L., Kislin, M., Wang, S.S.H., Murthy, M., and Shaevitz, J.W. (2018). Fast animal pose estimation using deep neural networks. Nat. Methods *16*, 117–125.
25. Eisenreich, B.R., Hayden, B.Y., and Zimmermann, J. (2019). Macaques are risk-averse in a freely moving foraging task. Sci. Rep. *9*, 1–12.
26. Gomez-Marin, A., and Ghazanfar, A.A. (2019). The life of behavior. Neuron *104*, 25–36.
27. 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.
28. Markowitz, J.E., Gillis, W.F., Beron, C.C., Neufeld, S.Q., Robertson, K., Bhagat, N.D., Peterson, R.E., Peterson, E., Hyun, M., Linderman, S.W., et al. (2018). The striatum organizes 3d behavior via moment-to-moment action selection. Cell *174*, 44–58.e17.
29. Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al. (2020). 'SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. Nat. Methods *17*, 261–272.
30. Waskom, M.; the seaborn development team (2020). mwaskom/seaborn. https://doi.org/10.5281/zenodo.592845.
31. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine learning in Python. J. Mach. Learn. Res. *12*, 2825–2830.

STAR★METHODS

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Ahmed El Hady (ahady@ab.mpg.de).

Materials availability

Code and data are available at https://github.com/adamjcalhoun/WhatIsBehavior. A version of the survey can be taken at http://adamcalhoun.com/WhatIsBehavior.

Data and code availability

The resources available from our study are the participants data and the codes used in the data analysis along with the survey itself:

- Code in python is available https://github.com/adamjcalhoun/WhatIsBehavior.
- All the participants responses to the survey questions along with the metadata can be found at https://github.com/adamjcalhoun/WhatIsBehavior.
- A version of the survey can be taken at http://adamcalhoun.com/WhatIsBehavior.

METHOD DETAILS

In order to assess whether there is a consistent definition of behavior used across academia, we constructed an online survey using Qualtrics. We developed the questions by first identifying categories of behavior whose definition we hypothesized to be disputed among the broader community based on our prior conversations with experts in the field. Additionally, we drew from questions from a previous survey performed by other authors Levitis et al.[11] We then sent those questions to colleagues who were experts in animal behavior and ensured that none of the questions were consistently answered 'Yes', 'No' or 'Maybe' in order to identify whether they represented differences of opinion in the field. We attempted to balance the overall composition of questions to reflect the categories we believed to be most important, and to reflect the categories that were disputed when we sent out the questions to our colleagues. The survey was then disseminated via Twitter, online mailing lists, as well as to colleagues. Subjects were allowed to answer 'Yes', 'Maybe', or 'No' to all questions. Subjects were asked for metadata consisting of their gender, University or Institute affiliation, country and state of residence, level of seniority, research area, and model system used (as detailed in Table S2). Survey questions were randomized and placed in blocks of five questions. Subjects had to answer all the questions before proceeding to the next set of questions. Subjects were free to decline participation in the survey. Subjects were offered the option of leaving their email address to enter a raffle to win a $50 Amazon gift card. The research was approved by the Princeton IRB (IRB# 12895) informed consent was obtained from all subjects.

QUANTIFICATION AND STATISTICAL ANALYSIS

Data pre-processing

Due to a coding error when designing the survey, subjects did not have to answer the first question in order to continue ('An animal is thirsty. It must choose between two levers, one of which will provide water. Would you refer to the animal pressing the lever as the primary behavior it is producing?'). For consistency, this question was excluded from all analyses. Only data where the questionnaire was finished were used for analysis. Due to the methods used for analysis, we only used responses that finished the entire survey which consisted of 42.48% of total responses. The overwhelming majority of these unfinished responses only answered a single question before terminating the survey.

Factor analysis

In order to explore whether answers were similar across fields, data dimensionality was reduced using the python library Prince (available at). Due to the categorical nature of the data, we used Multiple Correspondence Analysis (MCA). Data was transformed into one-hot encodings and then dimensionality was reduced using MCA. Plotting the components (Figure S2A) showed that the first four components each explain at least 5% of the variance, while subsequent components each explained less than 5% of the variance. Because of this, all analyses were done on the first four components. All data was plotted using the mean and standard deviation across metadata groups.

Sub-categories definition

Sub-categories were defined in several ways. First was their academic discipline. Academic disciplines were sorted into several broad categories. 'Biological sciences' included 'neuroscience', 'psychology', 'biology', and 'medicine', 'Math/engineering' included 'engineering', 'statistics', 'mathematics', and 'machine learning'. 'Ecology' included 'ethology' and 'ecology'. Humanities included 'philosophy', 'history', 'sociology', and 'languages and literature'. Subjects were also allowed to enter their 'sub-field' in a text box if they indicated their academic discipline belonged to neuroscience. We categorized these sub-fields into four broad disciplines: 'systems + circuits', 'cognitive', 'computational', and 'molecular'. Subjects were categorized as belonging to 'systems + circuits' if they entered 'systems', 'circuit', or 'circuits' in the sub-field box. They were categorized as 'cognitive' if they entered 'cognitive' or 'cognition'. They were categorized as 'computational' if they entered 'computational', 'theoretical', or 'theory'. They were categorized as 'molecular' if they entered 'molecular' or 'cellular'.

The next set of categories was their model organism. They were categorized as 'mammals' if they entered 'rodents' or 'other mammals', as 'other vertebrate' if they entered 'fish', 'birds', or 'other non-mammalian vertebrates', and as 'invertebrate' if they entered 'drosophila', 'c. elegans', or 'other invertebrates'.

Hierarchical clustering

Hierarchical clustering was performed using Seaborn and SciPy Virtanen et al.,[29] Waskom and the seaborn development team[30] using the Ward linkage and Euclidean distance (Figure 2).

Regression analysis of answers

Regression was performed using multinomial logistic regression in scikit-learn Pedregosa et al.,[31] with an elastic net penalty and an l1 ratio of 0.5. Answers to questions were fit using 5-fold cross-validation. Questions from the second half of the survey were predicted using one-hot encoding of the answers to questions from the first half, and questions from the first half were predicted from one-hot encoding of the answers to the first half.

Analysis identifying behaviorism and cognitivism

After the survey was completed, one author identified questions that a behaviorist or a cognitivist would agree with. For example, Q23 ("A behavior is always potentially measurable") was chosen to be representative of behaviorism and Q22 ("Does a behavior need to be intentional (does every behavior an animal produces have a purpose)?") was chosen to be representative of cognitivism.

For each respondent, a belief index was calculated as $\frac{-1*(\text{"}No\text{" } answers)+1*(\text{"}Yes\text{" } answers)}{number\ of\ questions}$.

The following questions were chosen as being behaviorist: Q1, Q4, Q9, Q14, Q15, Q17, Q21, Q23, Q25, Q26, Q27, Q29, Q30, Q33, Q34, Q38, Q40, Q42, Q44, Q46.

The following questions were chosen as being cognitivist: Q0, Q6, Q7, Q8, Q10, Q12, Q13, Q22, Q31, Q28.

This is also available as a CSV file in Supplementary Files and at the GitHub repository.