Q6. How can AI improve the reliability of operant, reflex, or observational nociceptive assays?
Authors: Kayla M. Moehn and Joshua J. Emrick

Response: Traditionally, scoring of nociceptive assays has relied on human judgement, which can lead to subjectivity, variability, and bias. Additionally, scoring nociceptive assays can be very time-consuming. Thus, machine learning-based tools may address this by enabling greater objective throughput and standardization. While nascent machine learning tools have been aimed at pain measurement, these methods still require careful internal validation and evaluation across multiple pain models. Thus, researchers should establish an understanding of the appropriate application of machine learning independent of its ease of use.

Level of Evidence: Moderate

Delegate Vote: Agree: Disagree: Abstain:

Rationale:

Commonly used nociceptive assays and current limitations: Nociceptive behavioral assays include operant, reflex, and observational assays. Each of these assays attempts to measure a different aspect of the pain experience. We provide current limitations that can be addressed using machine learning.

Operant nociceptive assays are designed to model the affective and motivational components of the pain experience observed in humans[1]. Operant tests use a conflict paradigm in which animals must decide whether to endure an aversive mechanical or thermal stimulus to receive a reward[2–4]. The outputs of these tests are simple and include measurements like latency to obtain a reward or the number of times a reward is obtained throughout a trial. A benefit of these assays is that they may capture the tolerability of pain[5] and are relatively easy to score. However, a drawback is that affect and motivation can be influenced by factors unrelated to pain, such as the effect of a pharmacological treatment on motivation itself[6]. The integration of videography and machine learning to analyze facial expressions and other behaviors during operant tests may provide validation that an experimental manipulation affects both pain-related motivation and sensory processing.

Reflex assays examine changes in hypersensitivity to either innocuous or noxious stimuli. In humans, it is known that allodynia and hyperalgesia can accompany tissue damage or inflammation[7,8], so reflex assays attempt to recapitulate this aspect of the pain experience in animal models. Reflex assays are often scored by measuring the latency to withdrawal or the minimum intensity of a stimulus needed to evoke withdrawal. The most popular reflex assay[5] measures mechanical sensitivity through the application of various intensities of mechanical stimuli to the bottom of the hind paw or cheek. Researchers may also evaluate changes in thermal sensitivity[9–11]. Like operant nociceptive assays, reflexive tests are attractive because they are relatively easy to perform and the outputs are straightforward. However, a major drawback is that these tests can be largely variable even when performed within the same laboratory[12]. Advances in machine learning have the potential to improve standardization of these tests through automation and remove potential confounds, such as the presence of the researcher in the room[13–15].

Observational assays are designed to capture changes in animal facial expressions or behavioral repertoire that may indicate pain. The grimace scale uses a cumulative pain score to quantify changes in facial features, such as orbital tightening, nose and cheek bulge, and ear and whisker position to approximate pain[16]. Traditionally, grimace has been manually scored by human observers, which can make it subject to bias and variability between researchers. It is also common for researchers to measure the frequency or total time an animal spends engaging in nocifensive behavior, such as paw guarding, licking or shaking[17,18] or changes in weight-bearing or gait[19]. Manual scoring of these behaviors can be time-consuming, and like the grimace scale, the subjectivity and bias of the researcher can lead to variable results. Machine learning tools that automatically track body position or behavior for each frame of a video are a promising approach to address these limitations.

Machine learning can improve throughput and standardization between researchers: Recently, computational neuroscientists have developed user-friendly machine learning algorithms that automate behavioral analysis. These approaches can be broadly divided into two categories: body-part position tracking and behavior classification. Both types typically rely on supervised machine learning, in which a researcher manually labels a subset of video frames to train a model. Once trained, the model can identify user-defined body parts or behaviors in entire videos.

The most popular tool used to track user-defined body parts is DeepLabCut[20]. Additional tools have been developed to address specific needs like SLEAP[21] for tracking multiple animals and Facemap[22] for tracking facial features. Pain researchers have begun to utilize these tools with promising results. For example, DeepLabCut has been used to quantify detailed paw withdrawal mechanics such as paw velocity, maximum paw height, and guarding duration after the application of a stimulus[23,24]. DeepLabCut has also been used to measure non-reflexive behaviors associated with pain, like orbital tightening, after optogenetic activation of nociceptors or capsaicin adminstration[25]. In

osteoarthritis research, DeepLabCut has been successfully applied to gait analysis in both rodents[26–28] and dogs[29]. Position tracking has enabled researchers to measure the angle of the knee joint during walking bouts in models of osteoarthritis which was previously impossible with other tools[27]. Altogether, position tracking may reduce variability in reflex and observational assays by replacing qualitative pain assessments with standardized quantitative measures that can be compared across laboratories. Additionally, these tools only require a small subset of labeled frames for initial training and refinement, greatly reducing the total workload of behavioral analysis and scoring.

One limitation of using body part position tracking is that further analysis of the outputs is necessary to obtain meaningful information about behavioral or kinematic changes. In response to this, user-friendly software packages have been developed to analyze these outputs[13–15,30–32] and even automate behavioral acquisition itself[13–15]. These tools also allow the researcher to remotely control data acquisition[13–15,31] to avoid any observer-induced confounds of accurate preclinical pain measurement[5]. Automated processing of DeepLabCut outputs allows researchers to obtain a quantitative analysis of reflexive and observational assays quickly without computer programming knowledge. Altogether, these software packages and automated behavior apparatuses make collecting and analyzing both reflexive and observational position-tracking data efficient and more reliable.

A complementary approach to body-part position tracking is holistic behavioral classification. Unlike position-tracking, behavioral classification tools do not require downstream analysis to determine meaningful information about behavioral changes. DeepEthogram[33] and LabGym[34] are two supervised machine learning tools that can be trained to identify an animal's behavior by frame throughout an entire video. The researcher trains the model to identify a set of predefined behaviors. When the model analyzes a video, it outputs an ethogram, which is a plot of which predefined behavior an animal performs at each moment. The application of these tools to pain research is relatively new, but LabGym has been successfully trained to detect mouse grimace, resting, grooming, and rearing in a model of tooth pain[35]. It is possible for behavioral classification tools to be trained to identify key behaviors that might change in models of osteoarthritis, such as grimace and rearing. Behavioral classification tools allow researchers to analyze observational data with less bias and with higher throughput. Importantly, researchers should ensure that they are blinded to experimental conditions when labeling behaviors for model training, so that they do not bias the model to detecting a behavior of interest under a certain experimental condition.

In addition to position-tracking and behavioral classification, PainFace is a new machine learning tool that has been developed by pain researchers to score mouse grimace[36]. Unlike DeepEthogram and LabGym, PainFace only quantifies grimace, not other behaviors. The automation of mouse grimace scoring can remove researcher bias, decrease variability, and increase throughput. One limitation of PainFace is that it is less flexible than the previously described machine learning tools, and users must mimic the acquisition parameters used by the developers.

Machine learning may enable the discovery of novel behavioral repertoires that indicate pain: In addition to the application of supervised machine learning, some pain researchers have started to use unsupervised machine learning tools, like MoSeq[37] and B-SOiD[38], to identify novel behavioral signatures of pain and its relief. MoSeq and B-SOiD apply unsupervised machine learning to uncover changes in the sequence and frequency of distinct behavioral subcomponents, or "syllables," in observational and reflex assays. MoSeq has been used to identify changes in the behavioral repertoire of a diabetic mouse model during an open field test, and these changes may indicate neuropathy[39]. Both B-SOiD and MoSeq have been used to identify novel behavioral signatures in reflex and observational assays that may indicate paw or knee pain and its relief[23]. Similarly, a study probing the function of spinal circuits in allodynia applied B-SOiD to identify behavioral subcomponents of paw withdrawal after spared nerve injury[40]. Altogether, unsupervised machine learning has the potential to uncover and/or correlate behavioral signatures of pain that have been previously unnoticed by human observation and scoring alone.

Important considerations and limitations of machine learning methods: While machine learning enables the detection of minute details that may have previously been undetectable by human observers, it is important to understand that not every detail may have a biological meaning or relevance to pain. For example, one study found that paw height and velocity during withdrawal after a mechanical stimulus were not indicative of inflammatory injury and produced variable results that were explained by unknown variables[41]. Nevertheless, machine learning remains a powerful tool that can replace qualitative assessments of nociceptive behavior with standardized quantitative measurements. To achieve this, we recommend that researchers using machine learning for behavioral analysis carefully label training data, remain blinded to conditions, and ensure that their experiments are well-controlled. Additionally, trained models and videos should be widely accessible to evaluate reproducibility and application across a variety of models of pain.

References:

1. Chen, Z. S. & Wang, J. Pain, from perception to action: A computational perspective. *iScience* **26**, 105707 (2023).

2. Reker, A. N. *et al.* The Operant Plantar Thermal Assay: A Novel Device for Assessing Thermal Pain Tolerance in Mice. *eNeuro* **7**, ENEURO.0210-19.2020 (2020).

3. Harte, S. E., Meyers, J. B., Donahue, R. R., Taylor, B. K. & Morrow, T. J. Mechanical Conflict System: A Novel Operant Method for the Assessment of Nociceptive Behavior. *PLoS ONE* **11**, e0150164 (2016).

4. Anderson, E. M. *et al.* Use of the Operant Orofacial Pain Assessment Device (OPAD) to Measure Changes in Nociceptive Behavior. *JoVE* 50336 (2013) doi:10.3791/50336.

5. Sadler, K. E., Mogil, J. S. & Stucky, C. L. Innovations and advances in modelling and measuring pain in animals. *Nat Rev Neurosci* **23**, 70–85 (2022).

6. Tappe-Theodor, A., King, T. & Morgan, M. M. Pros and Cons of Clinically Relevant Methods to Assess Pain in Rodents. *Neuroscience & Biobehavioral Reviews* **100**, 335–343 (2019).

7. Coutaux, A., Adam, F., Willer, J.-C. & Le Bars, D. Hyperalgesia and allodynia: peripheral mechanisms. *Joint Bone Spine* **72**, 359–371 (2005).

8. Jensen, T. S. & Finnerup, N. B. Allodynia and hyperalgesia in neuropathic pain: clinical manifestations and mechanisms. *The Lancet Neurology* **13**, 924–935 (2014).

9. Woolfe, G. & Macdonald, A. D. THE EVALUATION OF THE ANALGESIC ACTION OF PETHIDINE HYDROCHLORIDE (DEMEROL). *The Journal of Pharmacology and Experimental Therapeutics* **80**, 300–307 (1944).

10. Hargreaves, K., Dubner, R., Brown, F., Flores, C. & Joris, J. A new and sensitive method for measuring thermal nociception in cutaneous hyperalgesia. *Pain* **32**, 77–88 (1988).

11. Brenner, D. S., Golden, J. P. & Gereau, R. W. A Novel Behavioral Assay for Measuring Cold Sensation in Mice. *PLoS ONE* **7**, e39765 (2012).

12. Zumbusch, A. S., McEachern, E. L. F., Morgan, O. B., Nickner, E. & Mogil, J. S. Normative Preclinical Algesiometry Data on the von Frey and Radiant Heat Paw-Withdrawal Tests: An Analysis of Data from More Than 8,000 Mice Over 20 Years. *The Journal of Pain* **25**, 104468 (2024).

13. Dedek, C., Azadgoleh, M. A. & Prescott, S. A. Reproducible and fully automated testing of nocifensive behavior in mice. *Cell Reports Methods* **3**, 100650 (2023).

14. Burdge, J. *et al.* Remote automated delivery of mechanical stimuli coupled to brain recordings in behaving mice. *eLife* **13**, RP99614 (2025).

15. Zhang, Z. *et al.* Automated preclinical detection of mechanical pain hypersensitivity and analgesia. *Pain* **163**, 2326–2336 (2022).

16. Langford, D. J. *et al.* Coding of facial expressions of pain in the laboratory mouse. *Nat Methods* **7**, 447–449 (2010).

17. Klein, A. H. *et al.* A tingling sanshool derivative excites primary sensory neurons and elicits nocifensive behavior in rats. *Journal of Neurophysiology* **105**, 1701–1710 (2011).

18. Barik, A., Thompson, J. H., Seltzer, M., Ghitani, N. & Chesler, A. T. A Brainstem-Spinal Circuit Controlling Nocifensive Behavior. *Neuron* **100**, 1491-1503.e3 (2018).

19. Coulthard, P., Pleuvry, B. J., Brewster, M., Wilson, K. L. & Macfarlane, T. V. Gait analysis as an objective measure in a chronic pain model. *Journal of Neuroscience Methods* **116**, 197–213 (2002).

20. Mathis, A. *et al.* DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nat Neurosci* **21**, 1281–1289 (2018).

21. Pereira, T. D. *et al.* SLEAP: A deep learning system for multi-animal pose tracking. *Nat Methods* **19**, 486–495 (2022).

22. Syeda, A. *et al.* Facemap: a framework for modeling neural activity based on orofacial tracking. *Nat Neurosci* **27**, 187–195 (2024).

23. Bohic, M. *et al.* Mapping the neuroethological signatures of pain, analgesia, and recovery in mice. *Neuron* **111**, 2811-2830.e8 (2023).

24. Gonzalez, M. *et al.* Using DeepLabCut-Live to probe state dependent neural circuits of behavior with closed-loop optogenetic stimulation. *Journal of Neuroscience Methods* **422**, (2025).

25. Gupta, S., Yamada, A., Ling, J. & Gu, J. G. Quantitative orbital tightening for pain assessment using machine learning with DeepLabCut. *Neurobiology of Pain* **16**, 100164 (2024).

26. Fiker, R., Kim, L. H., Molina, L. A., Chomiak, T. & Whelan, P. J. Visual Gait Lab: A user-friendly approach to gait analysis. *Journal of Neuroscience Methods* **341**, 108775 (2020).

27. Savannah Dewberry, L., Cruz, C. J., Southern, K. M., Otto, K. J. & Allen, K. D. Pipeline Validation for Rodent Gait Analysis Using DeepLabCut. *Journal of Biomechanical Engineering* **147**, 111004 (2025).

28. Takenouchi, S., Minato, T., Fukuda, M., Kobayashi, K. & Murata, T. Development of a quantitative gait analysis in an osteoarthritis rat model using machine learning. *Journal of Pharmacological Sciences* **160**, 59–63 (2026).

29. Gill, H. *et al.* Gait tracking in dogs using DeepLabCut: A markerless machine learning approach for controlled settings. *Applied Animal Behaviour Science* **292**, 106769 (2025).

30. Jones, J. M. *et al.* A machine-vision approach for automated pain measurement at millisecond timescales. *eLife* **9**, e57258 (2020).

31. Oswell, C. S. *et al.* Mimicking opioid analgesia in cortical pain circuits. *Nature* **649**, 938–947 (2026).

32. Goodwin, N. L. *et al.* Simple Behavioral Analysis (SimBA) as a platform for explainable machine learning in behavioral neuroscience. *Nat Neurosci* **27**, 1411–1424 (2024).

33. Bohnslav, J. P. *et al.* DeepEthogram, a machine learning pipeline for supervised behavior classification from raw pixels. *eLife* **10**, e63377 (2021).

34. Hu, Y. *et al.* LabGym: Quantification of user-defined animal behaviors using learning-based holistic assessment. *Cell Rep Methods* **3**, 100415 (2023).

35. Ronan, E. A. *et al.* Intradental mechano-nociceptors serve as sentinels that prevent tooth damage. *Cell Reports* **44**, 116017 (2025).

36. McCoy, E. S. *et al.* Development of PainFace software to simplify, standardize, and scale up mouse grimace analyses. *Pain* **165**, 1793–1805 (2024).

37. Wiltschko, A. B. *et al.* Mapping Sub-Second Structure in Mouse Behavior. *Neuron* **88**, 1121–1135 (2015).

38. Hsu, A. I. & Yttri, E. A. B-SOiD, an open-source unsupervised algorithm for identification and fast prediction of behaviors. *Nat Commun* **12**, 5188 (2021).

39. Ashiquzzaman, A. *et al.* MoSeq based 3D behavioral profiling uncovers neuropathic behavior changes in diabetic mouse model. *Sci Rep* **15**, 15114 (2025).

40. Upadhyay, A. *et al.* The dorsal column nuclei scale mechanical sensitivity in naive and neuropathic pain states. *Cell Reports* **44**, 115556 (2025).

41. Rodríguez García, D. M. *et al.* High-speed imaging of evoked rodent mechanical behaviors yields variable results that are not predictive of inflammatory injury. *Pain* **165**, 1569–1582 (2024).