“Should Robots Feel Pain?” – Towards a Computational Theory of Pain in Autonomous Systems

Trevor Richardson, Indranil Sur, and Heni Ben Amor

Abstract We argue that investigating the biological mechanisms underlying the sensation of pain in humans and animals may lead to fundamental new insights about robot cognition, motor skill acquisition, autonomy, memory, and system integration. Despite the fact that pain plays a central role in the life of humans and animals, it has received only peripheral attention in the field of robotics. In this paper, we discuss the complex web of mechanisms and functions underlying biological pain sensation and anticipation. Next, we examine the opportunities and challenges that arise when studying computational frameworks that mimic nociceptive pathways. Further, we propose two initial benchmark tasks that may be leveraged to accelerate such research. Our main objectives are to highlight a critical knowledge gap in our understanding of intelligent physical systems and to identify a new and promising avenue for further research.

1 Introduction

Pain plays a central role in our lives and is of paramount importance to many brain and body mechanisms such as cognition, social interaction, motor control, memory, learning, autonomy and, most importantly, self-preservation. It acts as a critical signal, which guides our decision-making processes and shapes the choices we make. A long-standing theory, as articulated by Descartes [15], describes pain as bodily perturbations that are detected by nerve fibers and communicated to the brain.

Trevor Richardson
Arizona State University, Tempe, AZ 85281 e-mail: trevor.richardson@asu.edu

Indranil Sur
SRI International, Princeton, NJ 08540 e-mail: indranil.sur@sri.com

Heni Ben Amor
Arizona State University, Tempe, AZ 85281 e-mail: hbenamor@asu.edu

Cartesian theory limits the role of pain to the sensation of bodily harm, failing to acknowledge the many other functions involved in complex biological pain systems. More recent scientific evidence suggests that pain is generated through a complex interplay of a variety of signals and predictions involving multiple areas of the brain [1, 14, 23].

Despite its central role in many functions of the human brain, to date, pain has attracted relatively little interest in the robotics community. Pain and its relationship to robotics, however, has not been completely overlooked [8, 10]. Researchers have attempted to formalize pain for robotics and one promising result has been the development of an "artificial Robot Nervous System" that can react to multi-modal stimuli much like a biological organism's pain-reflex [9]. This system is similar in nature to the nociceptive Cartesian view and does not encompass the various other roles and interactions of complex biological pain systems. In contrast to the Cartesian approach, Sur and Ben Amor have shown that perturbations and their causes can be learned and anticipated [10]. Pain is also prominently featured in reinforcement learning (RL) algorithms [11] as a negative reward. RL algorithms, however, require a human expert to specify how this negative reward is calculated, typically resulting in extremely task-specific definitions of pain.

In this paper, we argue in favor of a general, computational theory of pain in autonomous systems that goes beyond the simple Cartesian model typically employed today. Translating the concept of pain into a bio-inspired computational framework allows for resilient machines and systems that can learn to anticipate and avoid harmful sensations, with a concomitant increase in longevity and autonomy. We will first discuss the evolutionary origin of pain, as well as the complex web of underlying mechanisms and functions of biological pain systems. Next, we discuss the opportunities and challenges that arise from studying computational frameworks that mimic nociceptive pathways. Further, we propose two benchmark tasks that can be leveraged to accelerate such research. Our main objectives are to highlight a critical knowledge gap in our understanding of intelligent, physical systems and, to identify a new, promising avenue for further research by the robotics community.

2 The Evolutionary Role Of Pain

Pain is a sensation that many species experience. It is not unique to humans and has been observed in vertebrates as well as invertebrates such as cephalopods. Pain is a dominant neurobiological process that is essential to the survival of our species; its influence is felt in almost all functional areas of the human brain [20]. The central nervous system (CNS) generates pain signals that influence our behavior and guide our learning within the contexts of self-preservation and reproduction.

The CNS has proven to be evolutionarily advantageous, having arisen as a result of natural selection. Accordingly, pain pathways are heritable traits that promote the fitness of biological populations. Individuals with congenital insensitivity to pain frequently die at a relatively young age due to tissue damage, infections, or both

[18]. Dawkins [13] offers a persuasive thought experiment to illustrate the importance of pain's role in evolution. He asks his audience to consider the potential fitness of a gazelle population with genes that cause analgesic states when fleeing predators. He concludes that this gene pool of gazelles would not be favored by natural selection unless tranquilizing a gazelle evading predation improved that gazelle's probability of reproducing [13]. He asserts that one must infer that gazelles experience extreme agony before death, because this system promotes self-preservation and the likelihood of reproduction. His thought experiment reinforces the principle that pain is a product of natural selection.

In addition to its evolutionary advantage, pain functions as a guiding signal by which a biological organism learns to navigate its environment safely, interact with living beings and inanimate objects, and promote its own well-being. It is therefore central to the behaviors learned and exhibited by a species within a lifetime. According to Craig [7], pain is not only a sensation, but also a motivation that is rooted in an emotional drive that results in homeostatic behavior. Not only does pain impose evolutionarily significant influences on behavior, it also serves as an educational feedback signal. The use and maintenance of a CNS of sufficient complexity to experience pain requires the expenditure of considerable energy. Such a CNS would be wasteful unless it served an evolutionarily advantageous purpose. Current research indicates that most insects do not experience pain [24]. A prevalent scientific theory suggests that this is evidence that pain is more advantageous to organisms with longer life spans because learning complex relationships is more important to organisms that live longer. Pain is beneficial to complex learning within the scope of self-preservation. Even more important may be the relationship between pain and emotional learning. Apkarian found that the representation of acute pain is related to the areas of the brain primarily responsible for emotional learning, memory and reward/addictive behavior [2]. Scientific evidence indicates that as we navigate our lives, pain consistently influences our behavior; it plays a central role in our ability to safely learn complex relationships while engaging with our environment [1].

3 Learning, Empathy, Memory and Fear

The neurobiological processes that produce pain significantly influence other human and animal functions. In particular, our ability to learn, empathize, remember and fear are all mechanisms that are affected by pain pathways in the central nervous system [4, 5, 6, 14, 16, 17, 23]. With regard to learning, this impact is realized in two distinct ways. First, fear-based conditioning leads to associative and avoidance learning [5, 22, 25, 26, 27]. Second, pain impacts learning through the bidirectional relationship between the formation of an individual's motivations and the pain that is experienced when those motivational goals are pursued [5, 22, 25, 26, 27]. Contemporary research has shown that personal experiences of pain are altered based on an individual's motivations and conditioned fear-based associations, as well as social factors that are unique to that individual [2, 26, 27]. Because the impact of experiencing pain is bidirectional, it functions as a fear-based conditioner. Fear-based conditioning that results from pain directs an organism's motivations and, as a result, affects how that organism learns [22, 25, 26] and how quickly it learns [14].

Another function that bears a close relationship to pain is *empathy*. There is a considerable overlap of brain activation between individuals experiencing pain and those experiencing empathy [6, 17, 23]. Research has shown that the brain's signature for empathy overlaps specifically with the brain's signature for pain in areas that are associated with pain's affective as opposed to sensory qualities [23]. Some results suggest that empathy is not exclusively a human emotive state, but instead is one that also exists in other living organisms such as rats [6]. The attentional resource needs of pain systems are considerable and parts of the brain outside the pain matrix can be altered by CNS pain networks. For example, an individual's memories can be altered by a painful experience [4, 12, 16, 19, 21]. In many cases, experiencing intense pain results in an enhanced ability to accurately recall a memory or to recall the emotions experienced during the painful event [4, 16, 21]. In contrast, due to pain's attentional requirements, painful experiences can also limit an individual's capability to remember information about his or her environment, especially when the information is not directly related to the cause of the painful experience [4]. In one study, subjects were found to remember their emotions with high accuracy after experiencing intense, acute pain. The same subjects, however, were much less accurate in recalling an unrelated stimulus present during the painful event [4]. Pain's functional role for an organism impacts that organism's ability to learn, empathize, experience, fear and remember.

4 Pain Maps to Robotics and the Benefit of Robotic Pain Systems

The development of robotic systems with the capacity to perceive pain would further the dream of creating a fully autonomous robot that could safely explore dangerous environments. With advances in hardware and software, the aspects of pain described in sections 2 and 3 have the potential to be realized in robotic systems. Using biology as inspiration, robotic software and hardware can incorporate systems that, to a considerable extent, mirror the functionality, complexity and interplay of biological pain systems. Such developments would provide immense benefits to functioning robots.

Fig. 1 displays specific interrelated and essential properties of pain-based systems. Hardware and software research has the potential to map these properties to the robotic domain. The construction of systems that improve a robot's self-awareness and promote self-preservation have far reaching applications. Such systems would benefit robotics in general by enhancing various learning capabilities, reducing the financial cost of replacing robotic parts and ensuring the longevity of robotic systems as a whole. Self-preservation systems for robots have the potential to map to empathy, enhanced learning, augmented memory and emotional fear-based conditioning. A robotic system that learns to associate certain scenes with negative internal states has the potential to understand the behaviors of other robots and organisms in the same environment. Empathy involves the comprehension of another organism's internal states. If a robot can predict a potentially precarious situation facing another robot, then not only can that robot learn from the other robot's experience, but also potentially provide assistance to the endangered robot. A robot's ability to understand another damaged robot's behavior promotes the well being of both the observing and damaged robots. A robot that understands negative internal harm and events that may cause harm also has the potential to help protect human beings and other organisms in dangerous situations. It is infeasible to preprogram all possible noxious states. In many experiments, robots only learn how to interact with their environment over a small period of time within fixed boundaries. In the future, robots will need to learn how to accomplish significantly more complex tasks that require them to reason about and interact with their environment over much longer periods of time and in multiple and varying locations. One could imagine giving a robot the tasks of shopping at a grocery store that it has never been to before and then cooking a meal for you and your family. These tasks present many perilous possibilities for the robot as it interacts with new situations. Learning how to accomplish these tasks and overcome the obstacles it will face (such as navigating new doorways in high traffic areas, transporting the required ingredients and overcoming food preparation dangers) can best be addressed through a system that promotes learning that involves concerns about internal safety and sustainability. Such a system would need to be able to reevaluate its own goals and motivations based on new data coming from nociceptive pain systems. Any learned information about noxious experiences with the environment would need to be stored, reused and, most importantly, generalized to new contexts.

Fig. 1 Critical functions that are interconnected to pain.

It is crucial that intelligent, autonomous robots with finite capacity for storing memories retain the most relevant and important experiences. A fully integrated pain system can assist in grading the relative importance of specific memories. For a fully autonomous robot to be realized, these types of Bayesian conditioning associations must be learned in real time. A pain-based robotics system would provide important feedback that would enable probabilistic conditioning. Nociceptive feedback would provide valuable insight regarding whether a given action or state would be beneficial or harmful to the robot. In general, robotic pain systems will promote robotic autonomy, system lifespan and robotic altruism – a robot's ability to assist other robots, humans and other living organisms.

5 Experimental Frameworks to Test Robotic Pain Systems

The creation of robotic pain systems requires experimental platforms that enable scientists to collaborate in this promising new area of research. We propose two scenarios to encourage research on robot pain perception – the Rotating Bar task and the Dodge Ball task. These scenarios allow the exploration of fundamental questions regarding computational theories of pain.

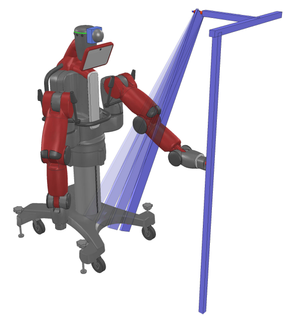

Fig. 2 Rotating Bar Benchmark.

Rotating Bar: In the Rotating Bar task displayed in Fig. 2, a fixed-position robot must learn how to extend its arms and end effectors, while at the same time avoiding damage from a rotating bar. In this scenario, the robot uses an RGB vision camera to observe its environment. The robot can query about its own internal states such as the position and orientation of its end effectors and arms. Parameters that can be varied include the angular velocity of the bar, angular acceleration of the bar, size of the bar, length of the robotic arms, and complexity of the task assigned. The robot's position, however, is fixed. A proper pain solution requires the robot to rapidly learn to avoid painful negative noxious states and to balance these nociceptive stimuli against its desire to complete its assigned task. The robot must learn to balance its goals and motivations with its own well-being. Note that any solution should not include any hard-coded reward function, e.g., `if(arm_torque > thresh) pain = 1`. This scenario tests the bi-directionality of competing interests and avoidance learning and provides the ability to analyze various nociceptive software models. The robot must learn to understand where negative feedback is occurring and how to respond to possible harm such as a damaged arm or end effector.

Dodge Ball: In the Dodge Ball scenario, see Fig. 3, N balls with Gaussian distributed initial positions are sent into projectile motion with randomly distributed initial velocities in the x, y, z planes toward a robot that can move along one dimension only. The robot is confined to a limited space. It accesses information about its environment using an RGB camera. It also can obtain information about its own internal state such as its velocity, orientation, and position. If the robot chooses not to move, or moves randomly, it will eventually be hit by some of the projectiles. Movement guided by intelligent anticipation is necessary to minimize the number of collisions with incoming projectiles. The robot needs to learn from experience what visual information is predictive of impending harm. This platform provides multiple parameters of interest, such as the rate and speed at which balls are fired, the distance from the robot to the balls, and the damage incurred to the robot from each collision.

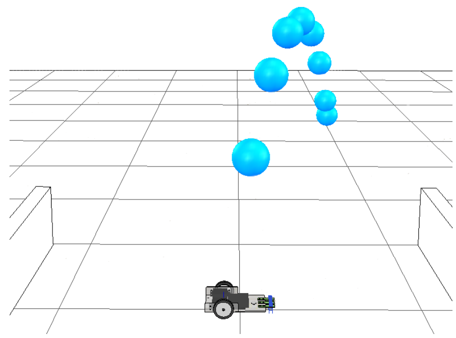

Fig. 3 Dodge Ball Scenario

The collisions provided by this simulation offer a way to test and build software that learns to predict negative future internal states. The robot must prioritize some noxious states over others. In these situations, the robot will need to endure some painful stimuli in order to develop an effective long term self-preservation strategy. The robot needs to reason about future states in order to maximize its longevity by incurring minimal damage over time.

In order to approach the development of a fully integrated pain system that promotes robotic well-being and autonomy, scenarios such as the two described above are required for comparison and testing purposes.

6 Conclusion

Pain is an adaptive trait produced by natural selection that promotes homeostatic behavior by influencing the way organisms learn, empathize, remember, and fear. Pain's feedback is essential to the biological fitness of many organisms. The biological motivation and the two scenarios outlined in this paper illuminate what we propose as a new direction of research that aspires to develop a computational theory of pain. Any such theory and mechanism has the potential to unify and integrate many critical functions of an intelligent system. In turn, this may lead to fundamental breakthroughs in smart, autonomous and resilient robots that require minimal if any human intervention.

References

1. Valry Legrain, Gian Domenico Iannetti, Lon Plaghki, and Andr Mouraux. The pain matrix reloaded: A salience detection system for the body. *Progress in Neurobiology*, 93(1): 111 – 124, 2011. ISSN 0301-0082. https://doi.org/10.1016/j.pneurobio.2010.10.005. URL http://www.sciencedirect.com/science/article/pii/S0301008210001759.
2. A Vania Apkarian. Pain perception in relation to emotional learning. *Current opinion in neurobiology*, 18(4):464–468, 2008.
3. V Bonavita and R De Simone. Pain as an evolutionary necessity. *Neurological Sciences*, 32 (1):61–66, 2011.
4. Angela Beese and Stephen Morley. Memory for acute pain experience is specifically inaccurate but generally reliable. *Pain*, 53(2):183–189, 1993.

5. Donald M Broom. Evolution of pain. *Vlaams Diergeneeskundig Tijdschrift*, 70(1):17–21, 2001.
6. J Chen, Z Li, YF Lv, CL Li, Y Wang, RR Wang, KW Geng, and T He. Empathy for pain: A novel bio-psychosocial-behavioral laboratory animal model. *Sheng li xue bao:[Acta physiologica Sinica]*, 67(6):561–570, 2015.
7. AD Craig. A new view of pain as a homeostatic emotion. *Trends in neurosciences*, 26(6): 303–307, 2003.
8. Sami Haddadin. *Towards safe robots: approaching Asimovs 1st law*, volume 90. Springer, 2013.
9. Johannes Kuehn and Sami Haddadin. An artificial robot nervous system to teach robots how to feel pain and reflexively react to potentially damaging contacts. *IEEE Robotics and Automation Letters*, 2(1):72–79, 2017.
10. Indranil Sur and Heni Ben Amor. Robots that anticipate pain: Anticipating physical perturbations from visual cues through deep predictive models. In *Intelligent Robots and Systems (IROS), 2017 IEEE/RSJ International Conference on*, pp. 5541–5548. IEEE, 2017.
11. Richard S. Sutton and Andrew G. Barto. *Introduction to Reinforcement Learning*. MIT Press, Cambridge, MA, USA, 1st edition, 1998.
12. Arthur D Craig. Opinion: How do you feel? interoception: the sense of the physiological condition of the body. *Nature reviews. Neuroscience*, 3(8):655, 2002.
13. Richard Dawkins. *River out of Eden: A Darwinian view of life*. Basic books, 2008.
14. Carolin Gramsch, Joswin Kattoor, Adriane Icenhour, Michael Forsting, Manfred Schedlowski, Elke R Gizewski, and Sigrid Elsenbruch. Learning pain-related fear: neural mechanisms mediating rapid differential conditioning, extinction and reinstatement processes in human visceral pain. *Neurobiology of learning and memory*, 116:36–45, 2014.
15. René Descartes. Treatise on man. *The Nature of Life: Classical and Contemporary Perspectives from Philosophy and Science*, pages 15–20, 2010.
16. Myra Hunter, Clare Philips, and Stanley Rachman. Memory for pain. *Pain*, 6(1):35–46, 1979.
17. Claus Lamm, Jean Decety, and Tania Singer. Meta-analytic evidence for common and distinct neural networks associated with directly experienced pain and empathy for pain. *Neuroimage*, 54(3):2492–2502, 2011.
18. Elna M Nagasako, Anne Louise Oaklander, and Robert H Dworkin. Congenital insensitivity to pain: an update. *Pain*, 101(3):213–219, 2003.
19. SA Pearce, S Isherwood, D Hrouda, PH Richardson, A Erskine, and J Skinner. Memory and pain: tests of mood congruity and state dependent learning in experimentally induced and clinical pain. *Pain*, 43(2):187–193, 1990.
20. Don Ranney. Anatomy of pain. In *Ontario Inter-Urban Pain Conference, Waterloo*, volume 29, 1996.
21. Patricia A Roche and Karel Gijsbers. A comparison of memory for induced ischaemic pain and chronic rheumatoid pain. *Pain*, 25(3):337–343, 1986.
22. Neil Schwartz, Catriona Miller, and Howard L Fields. Cortico-accumbens regulation of approach-avoidance behavior is modified by experience and chronic pain. *Cell Reports*, 19 (8):1522–1531, 2017.
23. Tania Singer, Ben Seymour, John O'doherty, Holger Kaube, Raymond J Dolan, and Chris D Frith. Empathy for pain involves the affective but not sensory components of pain. *Science*, 303(5661):1157–1162, 2004.
24. Christina G Skibinsky. Changes in store for the livestock industry-canada's recurring proposed animal cruelty amendments. *Sask. L. Rev.*, 68:173, 2005.
25. Deborah Talmi, Peter Dayan, Stefan J Kiebel, Chris D Frith, and Raymond J Dolan. How humans integrate the prospects of pain and reward during choice. *Journal of Neuroscience*, 29(46):14617–14626, 2009.
26. Katja Wiech and Irene Tracey. Pain, decisions, and actions: a motivational perspective. *Frontiers in neuroscience*, 7:46, 2013.
27. Jonas Zaman, Victoria J Madden, Julie Iven, Katja Wiech, Nathalie Weltens, Huynh Giao Ly, Johan WS Vlaeyen, Lukas Van Oudenhove, and Ilse Van Diest. Biased visceral perception through fear learning biased intensity judgements of visceral sensations after learning to fear visceral stimuli: a drift diffusion approach. *The Journal of Pain*, 2017.