On DeLancey's *The Passionate Engines*

Ho Manh Tung

Ritsumeikan Asia Pacific University,

Beppu, Oita, Japan,

November 8th, 2020

Craig Delaney's *The Passionate Engines* presents a comprehensive account of “what basic emotions reveal about central problems of the philosophy of mind” (2001, p.vii). The book discusses five major issues: The affect program theory, intentionality, phenomenal consciousness, and artificial intelligence (AI). Since its first edition's publication in 2001, the book has received multiple reviews, such as Graham (2002), Radden (2003), and Scarantino (2004). All of them have praised the book for its contribution to the philosophical literature on emotion, and its clear and measured writing style. I would like to briefly review the major tenets in the book and then focus on its discussion of AI, which has not been reviewed in detail. It must be stated that the author could not benefit from witnessing the recent development of the neuroscience of emotions (Barrett, 2017) and the growth of affective computing, i.e., the AI research on how to detect human emotions (McStay, 2018; Picard, 2000). With that in mind, I believe many of the claims in the book are still relevant and can benefit our understanding of emotional AI.

First, DeLancey's overall strategy is that although there are many emotions, and each with its varied expressions, one can still make a case of the existence of the basic emotions, such

as fear and anger. He defines a basic emotion as that which appears in different cultures, has a similar behavioral aperture, has a motivational and action-oriented quality, and appears to be adaptive. Here, he aligns his view on emotions, which he calls the *affect program theory*, with *the essentialist school of thoughts* on emotions (Barrett, 2017), championed by scholars such as Ekman (1999). Using that theory, DeLancey proceeds to criticize several philosophies of emotions: cognitivism, the doctrine that considers emotion to be a form of value judgment, “a propositional attitude like belief or judgement” (DeLancey, 2001, p.31); interpretationism, which states that “some mental states are dependent upon the stance of perspective of an observer” (p.50); social constructionism, which state emotions are a product of our cultural norms and have no or little references to our biology.

After dealing with the contemporary philosophies of emotion, DeLancey focuses on “*affective engineering*,” the effort to engineer emotions into AI systems. He views that programming AI to express and feel emotions is a very rigorous test of any given theory of emotions. He makes a distinction between “shallow affective engineering” and “deep affective engineering.” Shallow affective engineering does not try to *instantiate* affects in AI, i.e., merely giving AI the appearance of an ability to read or display emotions without actually understanding them. Deep affective engineering, however, takes instantiating affects in AI systems as an engineering strategy (p.204). DeLancey argues undertaking deep affective engineering can be very beneficial for the field of AI research.

He advances *the biomorphic argument*, which states that biological evolution has created much better autonomous beings than AI labs, and affects play an important role in autonomous behaviors. Thus, turning to biology can improve the engineering of autonomous AI systems. DeLancey advances six lessons to build *a passionate engine* and contrasting this engine with the

current focus of AI research on the symbol-manipulating, number-crunching engine. First, AI research should focus on the motion before emotion, action before abstraction. Second, in biological systems, affects are an important component of the decision-making process. Third, our affective sub-cognitive systems can be more accurate than the higher cognitive ones. Fourth, embodiment should be taken more seriously in AI research. The fifth and sixth point out the importance of parallel processing of affective sub-cognitive systems and vertical integration of these processes for producing intelligent, autonomous behavior. These lessons stem directly from DeLancey's intuition that basic emotions have motivational and action-oriented qualities that evolved and adapted in our biological history.

I generally agree with DeLancey's viewpoint on the potential of taking inspiration from biology for AI research, and affects are an important foundation for our intellect. However, there are several issues with this view. First, DeLancey did not define what it means to *instantiate* emotions and affects. Does that entail machines have feelings? Feelings are very important for humans and other animals to make goal-oriented behaviors. Still, without an account of machine consciousness, one cannot simply assume feelings are as important for non-organic, mechanical systems. And who to say feelings in a machine, if any, resemble our feelings?

Moreover, it seems to me there is too much focus on motion and embodiment, which can render his passionate engine impractical in the foreseeable future. This issue is connected with the previous issue of instantiation. As DeLancey admits, a test of a good theory of emotions is whether it is programmable in a machine if it is already programmable, why we need to give our machines the ability of motion before emotion, like DeLancey suggests in his first lesson.

Finally, an issue that I think could enrich DeLancey's discussion is what role affects and emotions play in our *counterfactual reasoning*. Statistician Judea Pearl, the father of the

Bayesian networks—a common approach in building AI (for more details on the use of Bayesian networks, see Vuong et al. (2020)), argues that AI research has not made much conceptual progress for a failure to take counterfactual reasoning seriously (Bereinboim & Pearl, 2016; Pearl, 2019). And we know a hallmark of how children learn is their ability to use counterfactual thinking (Gopnik, 2012). Indeed, to move beyond the current shallow machine learning model of stimuli-responses association and reinforcement (Deustch, 2020), modeling our counterfactual reasoning seems to be a good place to start.

In DeLancey's word, our affects play an important role in appraisals, which factors in our decision-making process. Thus, it seems reasonable to argue that affects might play a large role in our counterfactual thinking. The straight line that one can draw among appraisals, decision-making, counterfactual thinking makes me believe this is an important contribution I believe DeLancey could discuss, but he ended up missing in his book. Yet, one must acknowledge that his *biomorphic argument* is still an important contribution to the discourse on building intelligent autonomous machines and software.

In an era where machines and algorithms are increasingly delegated the task of decision-making, studying the blueprints of these algorithms—the field of “*transparent AI*”, is not only the matter of improving the quality of our decision, but also the matter of trust in science and technology (Vuong, 2018; Vuong, 2020). On that front, DeLancey's work indeed serves as a bridge between humanities scholars and the technical community.

References

Bereinboim, E., & Pearl, J. (2016). Causal inference and the data-fusion problem. *Proceedings of the National Academy of Sciences*, 113(27), 7345-7352.

- Barrett, L. F. (2017). *How emotions are made: The secret life of the brain*. London: Houghton Mifflin Harcourt.
- DeLancey, C. (2001). *Passionate engines: What emotions reveal about the mind and artificial intelligence*. Oxford: Oxford University Press.
- Deutsch, D. (2020). Beyond reward and punishment. Brockman, J. (Ed.). *Possible Minds: Twenty-Five Ways of Looking at AI*, 113-124. Penguin Books, UK.
- Ekman, P. (1999). Basic Emotions. In T. Dalgleish & M. Power (Eds.), *Handbook of Cognition and Emotion*. Sussex, UK: John Wiley & Sons.
- Gopnik, A. (2012). Scientific thinking in young children: Theoretical advances, empirical research, and policy implications. *Science*, 337(6102), 1623-1627.
- Graham, G. (2002). Review of *Passionate Engines: What Emotions Reveal about Mind and Artificial Intelligence* by Craig DeLancey. *Notre Dame Philosophical Reviews*. Retrieved from <https://ndpr.nd.edu/news/passionate-engines-what-emotions-reveal-about-mind-and-artificial-intelligence/>
- McStay, A. (2018). *Emotional AI: The rise of empathic media*. London: Sage.
- Pearl, J. (2019). The seven tools of causal inference, with reflections on machine learning. *Communications of the ACM*, 62(3), 54-60.
- Picard, R. W. (2000). *Affective computing*. Cambridge, Massachusetts: MIT Press.
- Radden, J. (2003). Review of "Passionate engines: what emotions reveal about mind and artificial intelligence" by Craig DeLancey. *Consciousness & Emotion*, 4(1), 143-148. doi:<https://doi.org/10.1075/ce.4.1.13rad>

- Scarantino, A. (2004). Craig DeLancey: Passionate Engines: What Emotions Reveal about the Mind and Artificial Intelligence. *Philosophy of Science*, 71(2), 227-230. doi:10.1086/381422
- Schneider, S. (2020). *Artificial You: AI and the future of your mind*. Princeton: Princeton University Press.
- Vuong, Q.-H. (2018). The (ir)rational consideration of the cost of science in transition economies. *Nature Human Behaviour*, 2(1), 5-5. doi:10.1038/s41562-017-0281-4
- Vuong, Q. H. (2020). Reform retractions to make them more transparent. *Nature*, 582(149). doi:<https://doi.org/10.1038/d41586-020-01694-x>
- Vuong, Q.-H., Ho, M.-T., Nguyen, H.-K. T., et al. (2020). On how religions could accidentally incite lies and violence: folktales as a cultural transmitter. *Palgrave Communications*, 6(1), 82. doi:10.1057/s41599-020-0442-3