Could a machine ever experience emotions like we can?

Eden Zan

January 6th 2025

There are two approaches which could be taken to respond to this question. The first is whether we could discover a reliable method to producing machines with experiential properties, and the second whether there is simply some arrangement of material to form a machine with experiential properties. However, it is only the second approach that should be taken, as affirming the former implies the truth of the latter, and if the latter is true, then it is at least theoretically possible, with an arbitrarily large amount of resources, to brute force a successful arrangement of material.

There exists a successful arrangement of material iff experience is purely physical, as we will only ever have access to physical matter, which we may call physicalism[1]. Unlike the first approach, it is irrelevant whether humans can understand the full nature of experiential phenomena. Some, including myself, hold that the origin of consciousness may never be explained, a view given the title of 'mysterianism' [1], however as long as it is purely physical, then there is the guarantee that some arrangement of matter gives rise to experience.

Likewise, there doesn't exist a successful arrangement of material iff experiential phenomena is not solely caused by physical matter, and is caused, at least partly, by non-physical matter, a view promoted by substance dualism, or simply dualism.

The question then is, whether dualism or physicalism is more likely.

There are good reasons both for believing in some form of dualism and of physicalism. For example, John M. DePoe writes that it is plain to all that 'they are not identical with a material body or bundle of mental events, but that they are "a seat of consciousness" that possesses a body and experiences mental events"' [2]. He claims that this 'seat of consciousness' is immaterial and irreducible, akin to Descartes' conception of the mind as baring no parts [3]. Many dualists are also motivated by a belief in a God who gave us souls, which is to be resurrected into the afterlife. On the other hand, physicalists may appeal to the principle of ontological parsimony, or Occam's razor, which here states that entities are not to be multiplied beyond necessity. If a physicalist worldview is sufficient to explain the phenomena we observe, then we should not believe in anything beyond it.

I feel that the arguments for each perspective carry equal weight, so it is therefore it is more productive to consider the criticisms of each worldview.

For dualists, one traditional objection arises from mind-body causation. For instance, if the body is wounded, then the mind will experience a sensation of pain, and if the mind wills that the body moves in a certain way, then the body will proceed to do so. One naturally questions how these two substances, if they are of wholly heterogeneous natures, may interact with each other.

Intuitively, it feels that the mind and body, if they are truly of fundamentally different natures, should not be able to interact, and it is demonstrably problematic if they do. For if

[1] True physicalism would hold that everything that exists is purely physical, so this is not a precise definition. For example, you can believe in a God beyond the material world and still hold that the human mind is purely physical. However, the definition I am using serves the same purpose.

observable physical processes can cause effects on mental substance, and unobservable mental processes had effects on physical substance, then energy would be flowing in and out of physical space. This contradicts the law of the conversation of energy, a fundamental scientific principle.

The dualist is not without response. For instance, 'parallelism' denies the interaction between the body and mind, and instead claims they run on their own separate courses with corresponding events [4]. As Liebniz writes, "bodies act as if there were no souls, and souls act as if there were no bodies, and both act as if each influenced the other" [5]. To parallelists, there is likely an external harmonising force, classically considered to be God. Yet while God, of course in his omnipotence, would be able to harmonise the effects on the body and mind, this hardly seems like a reasonable response to the question. Now, while the physicalist only posits one substance to explain what we observe, now the dualist posits 3: the body, the soul, and now a deity. In addition, to me it seems that if God were harmonising the effects of the body and mind, then he could and would do it better than it is. There is an interval, for example, between the experience of a sensation and the cause of that sensation. If you are wounded, you will not instantaneously feel pain. Why would God create this interval in the harmonisation? In addition, our minds can be fooled by our senses, for example by optical illusions. The body does not perceive this illusion, for the body does not perceive at all, yet the mind does. Is God deceiving us in this case? The parallelist now has a greater explanatory burden.

This is not the only possible response. In a paper by José Gusmão Rodrigues, he claims that "It is possible to concede that there is no good model of psychophysical causal interaction without giving up dualism" [6]. I must ask, then: What problem does the existence of a soul solve that is posed by a physicalism which claims that there is no good scientific model for sentience, i.e. mysterianism? He later offers a more plausible response though, that causation does not entail energy transference. For example, the interruption of blood flow to the brain causes fainting, and the turning off a switch to turn the light off. These are sufficient counterexamples, yet one can accept this and still hold that interaction between mind and body may still break the law of conservation of energy. Both of these examples are at least related to the transference of energy: cells in the brain use oxygen from the blood to produce energy, and turning off a switch ceases the continuous transfer of energy, so it could still be the case that interactions between the mind and body are related to energy transference in some way. The final point it makes that I will discuss is the possibility of 'spatial dualism', under which the soul bounded by space and can have energy, meaning the law of the conservation of energy is conserved. But to me this proposition is almost meaningless. Is a soul penetrable: could another object occupy the same space? If it is, then what does it mean for the soul to occupy this space? And if it isn't, then can we push the soul out of someone's body e.g. by filling them with fluid? This conception of the soul simply seems unreasonable.

On the other hand, the physicalist view of mind is not without issues, arguably the most famous one being the knowledge argument proposed by Frank Jackson. We are asked to imagine Mary confined to a monochromatic room, who learns everything about the physical world, and then is introduced to the outside world or even just colour. Upon seeing these new colours, Jackson argues that she has learnt something new: she learns what the experience of seeing red *feels like* [7]. This is incompatible with physicalism, because then she would already know everything about the colour and should not learn anything.

This idea that there is a gap between objective facts about the world and an individual's subjective experience was further developed by David Chalmers in his 'hard problem of consciousness'. He writes, "It is widely agreed that experience arises from a physical basis, but we have no good explanation of why and how it so arises. Why should physical processing give rise to a rich inner life at all?" [8]

For reductive physicalists, this is a major challenge. According to Philip Goff[2], it seems also the case that in the field of neuroscience research has come no closer to explaining sentience [9], so it may seem increasingly unlikely that we will ever have an answer. In addition, the scientific method can only give us quantitative information, yet one's subjective experience is qualitative - another indicator that the question might never be solved.

However, all that the hard problem of consciousness *can* prove is that we will not discover sufficient physicalist explanation for subjective experience. Even granting Chalmer's point to the maximum, this does not necessarily make it more rational to posit a non-physicalist explanation, such as a soul as do substance dualists. As I have claimed earlier, this dualism has no advantage over a non-reductive physicalism such as mysterianism. In addition, the problems here for physicalism do not at all address the claims made by mysterianism, and can in fact support it, and the criticisms for this view mainly come from other physicalists, such as Daniel Dennett [10][3], who claims that this view presents a pessimistic view of science. Although there are certainly physical observations that we may never be able to explain apart from consciousness, such as the fine-tuning of the universe - if this is indeed physically caused - so why not also consider consciousness physical yet unexplainable?

Thus I feel that a mysterian view of the mind is the most reasonable, and therefore it follows that there is some arrangement of physical matter that can have a conscious, and even human experience. Then we can conclude that it is at least theoretically possible that humans can arrange matter in this way, even perhaps by shear luck or brute force, and that machines *can* ever experience emotions like we can.

References

[1] Nicholas G. Carr. Mysterianism. `https://www.edge.org/response-detail/27017`, 2017. Accessed: 2025-1-6.

[2] John M. DePoe. A defense of dualism. `https://www.newdualism.org/papers/J.DePoe/dualism.htm`. Accessed: 2025-1-6.

[3] René Descartes. *Meditations on First Philosophy*. Oxford University Press, 2008.

[4] Leslie Joseph Walker. Psycho-physical parallelism. *Catholic Encyclopedia*, 11, 1911.

[5] Gottfried Wilhelm Liebniz. Monadology. 1714.

[6] José Gusmão Rodrigues. There are no good objecions to substance dualism. *Philosophy*, 89:199–222, 2014.

[7] Frank Jackson. What mary didn't know. *The Journal of Philosophy*, 93(5):291–295, 1986.

[8] David J. Chalmers. Facing up to the problem of consciousness. *Journal of Consciousness Studies*, 2(3):200–219, 1995.

[9] Philip Goff. *Galileo's Error*. Pantheon Books, New York, 2019.

[10] Daniel C. Dennett. The brain and its boundaries. *Times Literary Supplement*, 1991.

[2]It is worth noting that Goff is a panpsychist, meaning that consciousness is a fundamental property of matter. However plausible this view is, it would follow that machines could experience emotions like we do, a conclusion I agree with for different reasons.

[3]It is also worth noting that Dennett is an illusionist, and argued directly against qualia. This view seems absurd to me - it seems intuitive that I experience what I experience.