# Critique of the Conceptual Framework of Autonomous Vehicles

Jamey Blackman[1[3183495]]

University of Newcastle, Callaghan NSW, 2308, AUS
c3183495@uon.edu.au

**Abstract.** Autonomous vehicles are one of the fore runners of AI consumer technologies in the circulated media, with companies like Tesla, Waymo, Uber, as well as traditional car manufacturers competing to successfully develop a car that will not require any human attention while driving. This paper attempts to perform clarification of the conceptual framework taken in by researchers and engineers when designing these AI pipelines for autonomous driving, specifically perception mechanisms. One cause of the conceptual incoherence is postulated to be the rapidly demanding research and development driven by the speculative business model GAFA companies occupy.

**Keywords:** autonomous vehicles · self-driving · Tesla · Waymo · phenomenology · deep learning · machine intelligence · platform capitalism · financialisation

## 1 Introduction

Several reasons can be stated that provide the impetus for this paper. There is the recent few years of press that companies such as Tesla, Waymo, and Uber have been generating with promises in delivering fully autonomous vehicles at self-imposed deadlines, most of which have failed [1] [2]. Additionally, recent crashes [3] and the success of general adversarial networks (GAN's) in exploiting the fragility of the vehicles' computer vision systems aided with trained deep learned networks raises concerns about their technical integrity. Conversation sparked by political candidates around automation of the trucking industry puts into question the societal worth of introducing these autonomous vehicles when considering the mass unemployment effect. When these problems arise it can be of use to suspend the head-down approach to development and commercial release, even to go as far as to re-examine core design principles and ethics.

Philosophy can be utilised to clarify the concepts that the realm of science either holds with contradictions or disregards entirely. The fact that empirical scientific work is theory-laden has been stated numerous times by numerous scientific practitioners and philosophers of science since at least turn of the 20th century [4] [5]. When there is a contradiction between the predicted outcomes of empirical testings and the actual, it is just a valid to question the constellation of a priori givens as it is to scrutinise the direct hypothesis being tested. Conceptual clarification need not be brought about upon failed tests, but may also give a basis for *understanding* fruitful results or, to double back, discarding possible empirical investigations from the multitude of pathways. Godel's incompleteness theorems, while not scientific in inquiry, can be considered an analogue to the latter as it obliterated Hilbert's project of formalisation on logical grounds. Of the former, Hacker cites Netwon's evaluations in his *Optics* as an example of conceptual confusion leading to misunderstanding of colours as "sensations in the sensorium"; a tautological idea with no explanatory power [6]. The empirical work however remains a permanent and useful contribution to vision science and colour theory.

## 2 Conceptual Critique of Autonomous Vehicles

If the work on autonomous vehicles is considered as "scientific" work, within the context of the computer science of artificial intelligence, it is not guaranteed its conceptual coherence, regardless of its flourishing development. There is however a subtle dissonance when work on autonomous vehicles is grouped in with the common conception of "science", or the science that is continuous with the history of science. However if we consider the USA National Science Foundation's categorisation [7] we would type work on autonomous vehicles to be mostly Development, Applied Research to a lesser degree, and Basic Research remaining outside the purview. This split of applied research and development (R&D), and basic research in scientific work would commonly be considered the split between "technology" and "science", with another defining characteristic being the source of funding; the prior being mainly privately sourced and the latter public sourced [8].

A comprehensive conceptual analysis and critique has been undertaken for the neighboring field of neuroscience by M.R. Bennett and P.M.S. Hacker. It is arguable that artificial intelligence borrows a good portion of its conceptual framework from the fields of neuroscience and cognitive science, and this paper will try not presume guiding ideas about mind, sensation, perception, will, volition, that have been expounded upon in neuroscience are synonymous or even present in the guidance of autonomous vehicle R&D.

### 2.1 Current autonomous vehicles paradigm

With technology the truly state-of-the-art will never truly be known by those outside the confines of the organisation that creates it, so levelling any critique at a continuously articulating process such as the R&D of autonomous vehicles can result in a fruitless straw-manning. With a lack of theoretical literature the conceptual framework will be mainly intimated from what R&D is focused is on with regards to technical design. Let us proceed to a discussion of deep learning techniques employed in tandem with mechanical components to produce "self-driving" [9], with some specific attention to Waymo's wager on LIDAR sensor reliance as opposed to Tesla's camera-based methods for perception and localisation.

The main deep learning neural networks utilised in autonomous vehicles are convolutional neural networks (CNN), deep networks utilising reinforcement learning algorithms (DRL) and to a lesser extent, recurrent neural networks (RNN). We will gauge their conceptual coherency and present an alternative NN structure that appears to model itself on the phenomenological view of consciousness in section 2.2.

For vehicle perception and localisation CNN's have become the standard. The efficacy of CNN's for processing spatial information has been attributed to it's analogy with the neurologically understood dual-flux theory of the visual cortex, where the dorsal stream establishes spatial relations of objects and visual guidance of action, and the ventral stream establishes the recognition of objects and perceptual representation. The CNN analogue of these dedicated streams are the pooling nodes and convolutional nodes, respectively. This inspiration from the biological to the computational should be encouraged however the dual-flux theory has it's assumptions: the dissociation of the two streams that the theory inculcates seems to entail that there exists aspects of perceptual representation that are disconnected from the visual guidance of action. Alva Noe illustrates the folly of this entailed possibility with a reference to traditional approaches to the problem of vision [10]. In this view it is proposed that vision is the process of producing a 3D representation of the external world with what is given in the 2D retinal image. This proposition extricates embodiment and volitional movement from the perceiver which are required for 3D representation, an example of the view's deficiency being when a perceiver is presented with a large object at a far distance away and a small object closer up they would be perceived as the same size. Spatial-depth orientation crumbles in this instance. These instances should not be possible which the dissociation of streams allows for. Furthermore the series structure of a CNN formalises this problematic dissociation of perceptual streams.

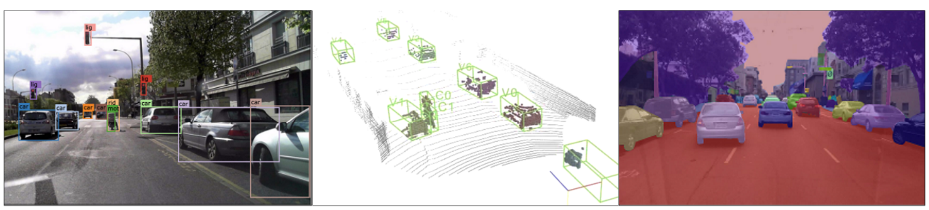
The advent of adversarial examples has sullied the naive view that current CNN's perceive like humans do. With a slight perturbation to a CNN classifiable object (a physical adversarial example would be a sticker on a stop sign [11], with images, the introduction of a slight noise vector) the classifier will mis-classify the object, sometimes more wildly then others. The exact reasoning behind this phenomenon is debated.

The modules for vehicle path planning and behavior arbitration are mainly DRL based. At any time the vehicle is required to consider all possible obstacles within given states and determine a collision-free path by acting the specific trajectory required. Path Planning has two deep learning approaches: Imitation Learning, which records actual human driving experience, and Deep Reinforcement Learning Path Planning, which records trajectories within a simulated virtual environment. The former gains real world-coping trajectories within safe and comfortable limits of manoeuvrability. The latter gains all possible trajectories including during long-tail scenarios and doesn't end up trying to formalise the expertise of a human driver. With regards to last module, motion controllers have not seen a big leap in deep learning integration, only to improve accuracy on traditional controller methods.

Up until now the modular approach to pipeline architecture has been elucidated and is the standard for any currently commercially aspiring autonomous vehicle R&D. Ultimately the modularity of the pipeline architectures does not imitate the concurrency of perception, planning and action that takes place in humans. The alternative pipeline approach, End2End, has achieved popularity with NVIDIA's PilotNet project. It invites much heterogeneity with its deep learning methodologies, with RNN models fairing well due to the shared closed continuous circularity with the overarching pipeline architecture.

With all main deep learning methodologies addressed, we can continue onto the somewhat tired debate about sensing hardware between big tech forerunners Tesla and Waymo but with a conceptually critical approach as opposed to arguments of cost and bad-weather performance.

Waymo employs a "more is better" approach with its reliance on LIDAR sensors for perception and localisation. This expensive method for driving-scene-input loudly sirens the concepts of representationalist view and the problems that follow. LIDAR sensors provide a 3D representation of the surrounding environment in the form 3D point clouds (see Fig. 1).



(a) (b) (c)

**Fig. 1.** Examples of 2D object recognition in (a), the bonding box detector mapped onto 3D Lidar point cloud in (b), Semantic segmentation results (c)

This "objective" birds-eye view of perception is counter to how humans actually perceive. But if for a moment we were to accept that to perceive we had to access an internal representational model a few things would follow. Firstly, the correct mapping of our internal representation and the external world would not be guaranteed. Additionally how would it be known that any objects delineated before a representational update were the same delineated objects afterwards. A internal interpreter maybe posited that would confirm the similarity of two representations, however this relationship could entail an infinite regress of internal interpreters which only relocates the problem of perception and has no explanatory power. The world as it's own representation is just as viable, as we are *in* the world, and arguably evolutionary favourable considering the theoretical information processing burden consulting an internal representation would bare instead of off-loading representations to the world. Secondly, normal representations have non-representational properties that channel the semantic quality. The idea of an internal mental representation suggests it has no medium, no way of signalling the sign; they are posited as pure meaning which excludes them from the category of representations. It is the same with scientific models and instruments that are affected by the causality of the external world. They have the model/instrument property that is able to be affected and the interpretative property.

To return to the sensory mechanism of LIDAR, another problem arises when considering any visual paramaterisation of the driving scene that is of an extremely large order of magnitude with this method of sensing.

Now, facts are *relevant* to particular moments; they are concretely instantiated events in time. What results from this extreme bloat of representational facts is the high probability of the inability to choose from a comparatively small amount of relevant representational frames for relevant action [12]. Neural networks are prone to this classic problem of frames as well [13]. Optimistic proponents assert NN's automatically have relevance-sensitivity, that is, when relevant information is inputted it is then proliferated throughout the architecture via weight revision. But this is a causal explanation; that the relevance is within there cannot be verified.

While Tesla has correctly avoided the crutch of LIDAR with the relative thrift of camera-reliant computer vision, there are general problems not yet endemic to the deep learning neural network techniques employed in the perception/localisation module or throughout any autonomous vehicle pipeline architecture.

Calls for accountability and intelligibility have surrounded the recent crashes of autonomous vehicles. The black box nature of the deep learning networks that act as the intelligence of the vehicles occludes the reasoning behind these crashes. One response to this vexation of inscrutability is to clarify that these systems are ecological; they latch onto the environmental cues with which they are trained on, their efficacy lies in the correlation to targets. Environments are inherently specific so the act of misapplication of a NN system should be the thing held to scrutiny.

Another response would be to imbue these systems with a capacity for "understanding" or "common sense", capacities that are lambasted for their ambiguity and possible appeal to a naive humanist exclusivity. This paper has opted to defend the essence of this missing capacity in the upcoming section 2.2, however we will subsume these terms into the concept of compositionality which will borrow from linguistics and phenomenological concepts for its precise language and conceptual coherence.

### 2.2 Alternative directions for the research programme

The conceptual criticisms of the deep learning methods laid out above have always utilised a different, more coherent understanding of perception. Phenomenology is the study of necessary connection between human experience and the appearing world, which is traditionally undergone through the suspension of the naive assumption of a physical world so that one can be more conscious of oneself in the act of perceiving. These first questions should be the ones studied in a field that wishes to artificially recreate the questioning phenomena. That there is a structure of perception first off is something AV design seems to neglect with its modular pipeline design or the standard choice of CNN's for perception of a scene and localisation of the vehicle, which are unidirectional state-act flows and thus do not account for opposite-heading anticipations from the subject.

Jun Tani has produced a comprehensive account of his theoretical work and empirical investigations in robotics [15], with studies on self-organising closed circuit RNN variations and emergent behaviours, work that most of the embodied cognition school would exalt as approaching the extent to what *being* is. However Tani realises the one-sidedness of this school when he observes these robots emergent behaviours are merely reactive generated by perception-to-action mapping. Thus he returns to the coherent anthropology of phenomenology, specifically the founder Edmund Husserl, and bases a model of the mind that while retaining the systematicity of the most self-organisational NN's, accounts for subjectivity: or the faculty with which compositionality is conducted.

In linguistics [16], to generalise compostionally there must be systemacity; the current ability to understand/produce a set of sentences is intrinsically connected with the understanding/production of others (e.g. "I am going for a walk" and "I am going" are systematic). They must also be productive; able to infinitely generate new compounds with a finite set of sentences at any given time. Substitutivity is premised on the synonymity of two constituents not affecting the meaning of an overall expression when one constituent is swapped out for the other inside the expression. These meaning-retaining rules are ways in which given parts are composed into new wholes. Husserl, whose work has had influence in linguistics via Jakobson and his structuralism, considered two intentionalities of time: traversal (retention of what has recently passed and protention of what is anticipated e.g. ray in the sung "do-ray-mi" protends mi and retains do) and longitudinal (where the pre-empirical continuous flow of time can be objectified after the fact and recalled e.g. thinking can be recalled as discrete thoughts). These objectified recalls can act as components to be recombined to anticipate new continuous incoming direct perceptions, hence intentionally is compositional; anticipations are composed of past experience obviously.

Fig. 2 tries to convey a way of modelling this perceptual structure. Once initialised, we have two movements through the system happening concurrently. The first movement starts from the "intentions" top-level with a chaotic attractor (trajectories exhibit infinite periodicity and thereby form fractal structures) then state values are sampled sequentially when they cross the Poincare section and fed into a parameterised dynamic system in the lower level. The parameters in the figure depict a dimensionally-reduced visuo-proprioceptive space for the intended acts of approaching an object and shaking it. The comparison of the predicted state and the actual state produce an error that is monitored and modifies the higher level to minimize this error.

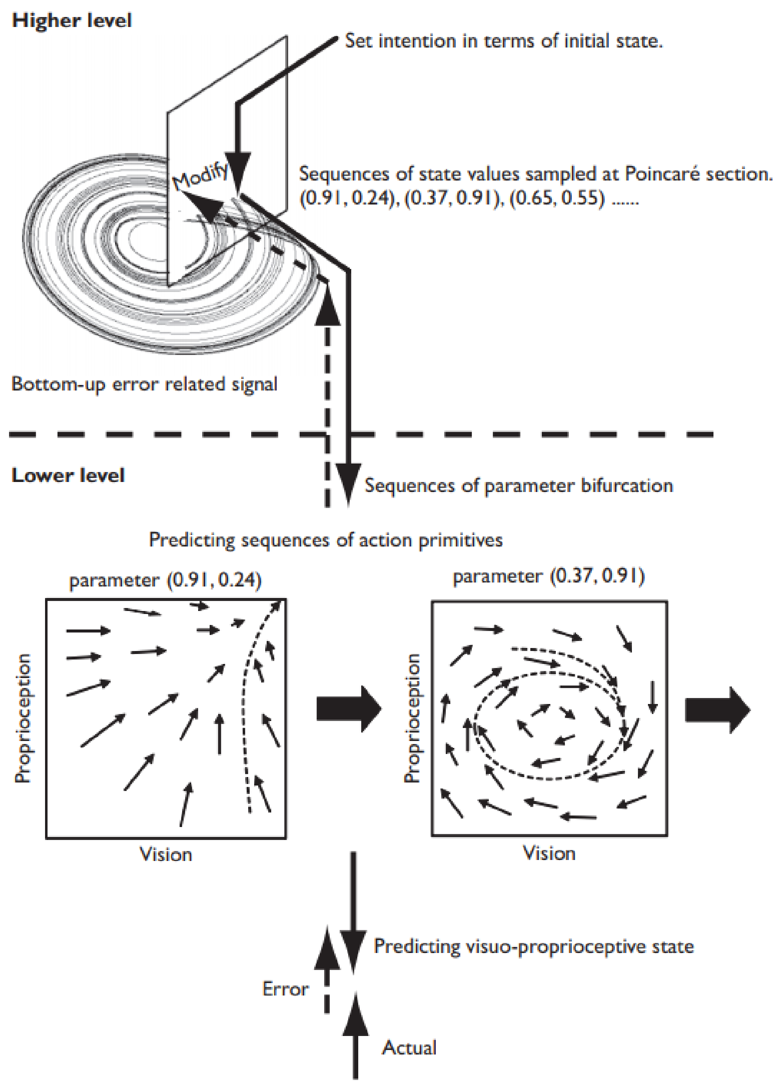



**Fig. 2.** Tani's top-down discrete intentionality, bottom-up continuous coping model of the mind

This structuring proves to capture the outgoing compositional intentionally, not by renewing symbolic AI methods, but through nonlinear neurodynamics which allows also for the continuous incoming perception of embodied cognition without any strife amongst the two. The bifurcation of subject-object isn't retained, neither is it dissolved by one or the other, but reconciled as an inseparable entity.

Tani's conceptualisation and designed model can be just one example of how innovative and useful basic research can be. This paper prescribes the return of basic research and study of relevant theoretical literature back into equal footing with R&D. Also, the compounding use of terms in their old customary sense but with the intention of a novel technical meaning in the context of scientific discourse is a folly that the field has been prone too which leads to conceptual incoherence. The presence phenomena in scientific discourse is illustrated profoundly if one were to consider the two-way terms ”code”, ”message”, ”signal”, introduced by information theorists and molecular biologists of the 1950's [14].

The scientists drive to circumvent the acknowledgement of the conceptual framework and expunging of its incoherence can in part be attributed to speculative way in which funding is accrued [17]. All these GAFA and start-up companies involved in the autonomous vehicle race have yet to return a profit. The failed deadlines are to generate hype, enact market capitalisation, and seduce more outside investment.

## 3 Conclusion

The computer scientist or engineer has what could be described as a technological will animating these efforts which circumvent the acknowledgement and good faith understanding of the structure of perception. This structure is missed when one considers the world as standing reserve [18]; elements external to each other to be imposingly restructured or analytically broken down. There is an efficacy to this mode of seeing that should not be discarded, to separate something from it's prior necessary connections allows one to focus on it exclusively. However a concealment concurrently takes place and this amounts to a confusion especially when it comes to consciousness of the subject itself. If basic research was to be reasserted with public funding and these innovative technologies were not a means for the generation of monopolies in new markets it would foster a more robust scientific community with original epistemic aims set back in sight.

## References

1. Maurer, B.: Tesla: Continued Hype Falls Flat, `https://seekingalpha.com/article/4346298-tesla-continued-hype-falls-flat`
2. Bloomberg: The State of the Self-Driving Car Race 2020, `https://www.bloomberg.com/features/2020-self-driving-car-race/`
3. Colias, M., Higgins, T., Spector, M.: Tesla, Uber Deaths Raise Questions About the Perils of Partly Autonomous Driving,

https://www.wsj.com/articles/tesla-uber-deaths-raise-questions-about-the-perils-of-partly-autonomous-driving-1522661400

4. Duhem, P.: The Aim and Structure of Physical Theory. 2nd ed. Princeton University Press, Princeton, NJ (1991)
5. Kuhn, T.: The Structure of Scientific Revolutions. 4th edn. The University of Chicago Press, Chicago 60637 (2012)
6. Bennett, M. R., Hacker, P. M. S.: Philosophical Foundations of Neuroscience. 1st edn. Blackwell Publishing, 350 Main St, Marlden, MA 02148-5018 (2003)
7. Kennedy, J. V.: The sources and uses of U.S. science funding. The New Atlantis **36**, 3–22 (2012)
8. Avin, S: Breaking the Grant Cycle: On the Rational Allocation of Public Resources to Scientific Research Projects. University of Cambridge (2014)
9. Cocias, T., Grigorescu, S., Macesanu, G., Trasnea, B.: A Survey of Deep Learning Techniques for Autonomous Driving. Journal of Field Robotics, 1556–4967 (2019)
10. Noe, Alva.: Action in Perception. 1st edn. MIT Press, Massachusetts Institute of Technology Cambridge, Massachusetts 02142 (2004)
11. Eykholt, K., Evtimov, I., Fernandes, E., Kohno, T., Li, B., Prakash, A., Rahmati, A., Song, D., Xiao, C.: Robust Physical-World Attacks on Deep Learning Visual Classification. In: Editor, F., Editor, S. (eds.) Conference on Computer Vision and Pattern Recognition (CVPR) 2018, pp. 1625–1634. IEEE (2017). https://doi.org/10.1109/CVPR.2018.00175
12. Thill, S., Windridge, D.: Representational fluidity in embodied (artificial) cognition, Biosystems **172**, 9–17 (2018)
13. Samuels, R.: Classical computationalism and the many problems of cognitive relevance. Studies In History and Philosophy of Science Part A **41**(3), 280–293 (2010)
14. Kay, L., E.: Who Wrote the Book of Life?: A History of the Genetic Code, Stanford University Press (2000)
15. Tani, J.: Exploring Robotic Minds: Actions, Symbols, and Consciousness as Self-Organizing Dynamic Phenomena. 1st edn. Oxford University Press, 198 Madison Avenue, New York, NY 10016 (2017)
16. Bruni, E., Dankers, V., Dieuwke, H., Mathijs, M.: The compositionality of neural networks: integrating symbolism and connectionism. ArXiv abs/1908.08351 (2019)
17. Frigant, V., Jullien, B., Montalban, M.: Platform economy as a new form of capitalism. A Régulationist research program. Cambridge Journal of Economics **43**(4), 805–824 (2019)
18. Heidegger, M.: Basic Writings. 2nd edn. Routledge, 11 New Fetter Lane, London EC4P 4EE (1993)