Citation: López-González, Mónica. (May 2019). *An Argument and Proposal for Integrating Human Cognitive Intelligence into Autonomous Vehicle Perception*. IS&T Electronic Imaging Symposium: Autonomous Vehicles and Machines, (IS&T, Springfield, VA, 2019).

https://doi.org/10.2352/ISSN.2470-1173.2019.15.AVM-054

Today Is To See and Know: An Argument and Proposal for Integrating Human Cognitive Intelligence into Autonomous Vehicle Perception

Mónica López-González; La Petite Noiseuse Productions; Baltimore, Maryland, U.S.A.

Abstract

The race to commercialize self-driving vehicles is in high gear. As carmakers and tech companies focus on creating cameras and sensors with more nuanced capabilities to achieve maximal effectiveness, efficiency, and safety, an interesting paradox has arisen: the human factor has been dismissed. If fleets of autonomous vehicles are to enter our roadways they must overcome the challenges of scene perception and cognition and be able to understand and interact with us humans. This entails a capacity to deal with the spontaneous, rule breaking, emotional, and improvisatory characteristics of our behaviors. Essentially, machine intelligence must integrate content identification with context understanding. Bridging the gap between engineering and cognitive science, I argue for the importance of translating insights from human perception and cognition to autonomous vehicle perception R&D.

Introduction

We are at a veritable turning point with autonomous vehicle perception technology. Machine intelligence is able to process enormous amounts of complex data simultaneously from cameras, lidar, and radar in a more accurate way than ever before. From a bird's-eye view, autonomous self-driving vehicles have sufficient technical components to be deployed on our roads. Taking stock of major auto companies' predictions regarding the expected year of self-driving vehicle deployment, we can see in Figure 1 that deployment is around the corner with year 2020 seeing significant promise [1], [2]. Furthermore, IEEE community members estimate 75% of all vehicles on the road will be autonomous by 2040 [3].

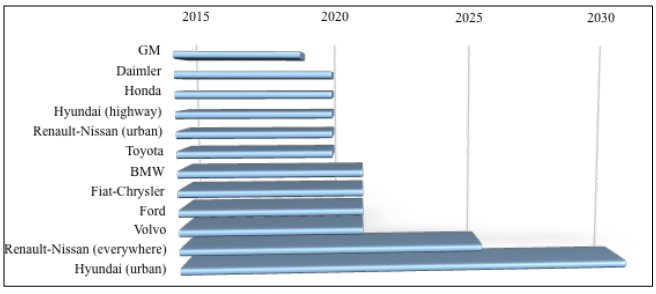


Figure 1. Top ten major global auto companies' predictions for the year of deployment of their autonomous self-driving vehicles on public roads. Data are by no means exhaustive. The word 'predictions' should be read with caution given the rapidly changing state of technologies and the complexity of process from announcement to actuality.

The autonomy automakers forecast to produce as early as 2019 are SAE International levels 4 (high automation whereby the vehicle can drive itself within a limited area and under certain conditions with minimal human input) and 5 (full automation whereby the vehicle can drive itself in all roadway and environmental conditions without any human input) [4]. From a cognitive science perspective, these levels entail a sophistication in higher level reasoning not yet possible by machine intelligence: a capacity to perceive an ever-changing environment with meaning and purpose, to make spontaneous, new predictions based on learned and hypothesized expectations, and to take consequential actions for a future outcome. Despite such vital capacities for an autonomous machine to successfully maneuver from point A to point B amidst dynamic and unpredictable environments common to roadways, Uber, for example, is back on Pittsburgh, Pennsylvania's roadways to continue its on-road testing after halting operations in the wake of a fatality [5]. Moreover, this comes at a time when self-driving vehicle accidents within the State of California, for example, are a very real problem. Figure 2 illustrates every company who has reported at least one accident between their autonomous vehicle and a conventional vehicle, some significantly more than others, from 2014 through the end of 2018.

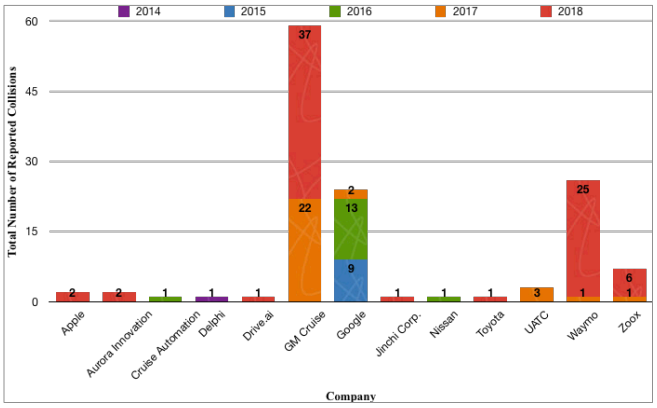


Figure 2. Companies who reported collisions on public roads in the State of California of their autonomous vehicle with a conventional vehicle by year. Data made publicly available by the Department of Motor Vehicles, State of California, USA. A total of 129 reports have been received as of December 21, 2018. Reports are from the dates October 14, 2014 to December 11, 2018 [6].

While these reported accidents (129) across several years are small in number in comparison to the number of fatalities (37,133) from conventional motor vehicle traffic crashes on U.S. roadways alone during 2017 [7], they require attention in this new era of ensuing human-machine interaction.

Conventional vehicles and other unpredictable animate and inanimate elements common to the real world are far from exiting our roadways anytime soon. Success of a fully automated system naturally implies a solid foundation for what constitutes as safe and unsafe in an environment full of action, distractors, and unfamiliarities and a machine capable of causing damage and/or

death. An important element to consider in the building of relevant intelligent technology is the resulting human-machine interaction on the roadway. To dismiss it as a future problem, or not even consider it a factor, is to compromise safety for profitable gains. Damaged property and/or a death/deaths should not be acceptable in the name of possible life-saving technology of the future. As suggested by a few, clear and agreed upon criteria for global standards of safety validation beyond a brute-force approach of road-test driven miles must be determined now at all levels of the testing process, be they closed-course, simulation, on-road, and human-tended systems [8].

The appeal for solid safety measures comes at a critical moment in this human-machine reality and in the most recent acknowledgement here at the Autonomous Vehicles and Machines 2019 conference in Burlingame, California that the near future is more about deploying SAE International levels 2+ (or 3-) with a driver having some sort of role –most likely remotely– than levels 4 or 5. As initially put forward in [9], human-machine interaction is a relationship to be significantly brought to the forefront if the threefold gold standard of efficiency, effectiveness, and safety is to be met with these new highly intelligent vehicles on our roadways. To put it bluntly, these self-driving vehicles will need to be able to deal with humans' rule-breaking and/or erratic behavior as well as other environmental unknowns at all times and under all sorts of conditions on the spot, be they urban, suburban, and/or weather related. The self-driving vehicle cannot just simply stop operating or operate but incompetently towards a negative outcome because programmed rules in the self-driving vehicle were breached by the human driver in their conventional vehicle and left unchecked in the self-driving vehicle's algorithms for unpredicted if-then cases [10] that did not happen to occur, for example, during the millions of miles it was test-driven.

In [9] I made the theoretical argument that collaborative musical improvisation within live theatre is cognitively analogous to driving a vehicle in a densely populated and chaotic area like a big city. Moreover, I advanced the claim that cognitive research in such area of human creativity [see [11], [12], and [13] for experimental details] is useful for translation in other areas. My main point –drawn from the assumption of an inevitable collaborative relationship between humans and SAE International levels 4 and 5 machines-to-be on roadways and elsewhere– was a call for multidisciplinary action to merge insights from human cognition and behavior like online decision-making, emotion perception, risk prediction and management, and spontaneous collaborative negotiation with vehicle perception R&D to improve machine intelligence outcomes. Continuing in that vein, I begin my argument for this paper with the following: we must make it a R&D priority to understand and integrate into machine intelligence what we currently know –and could know more of– in regards to our ability to recognize, identify, anticipate, and utilize that of *meaning* to us in a sea of multi-sensory information. Unraveling the full potential of human perception and cognition is perhaps our closest bet to dealing with so-called edge cases and overall "what if" scenarios resulting from human behavior and/or in-the-moment landscape/environmental changes. It is by no means an easy feat to answer the biggest conundrums of human intelligence. But if we unify our intellectual efforts and financial resources, my bets are all on for answering them sooner and more effectively.

I will not attempt to fully cover and answer such aspects of human intelligence and why, for example, the human brain is so efficient at what it does –there are more than sixty years of dedicated ink to such vastly loaded questions since Cognitive Science's birth as a discipline in 1956 [14]. Instead, I will emphasize what I believe are fundamental questions and open challenges we must commit to resolving if we want to create fully functioning SAE International level 5 autonomous vehicles and to delineate a more defined path for safety metrics.

What We Have and What Is Missing

> *Sometimes it seems as though each new step towards AI, rather than producing something which everyone agrees is real intelligence, merely reveals what real intelligence is not* [15].

The advent of the Internet and consequent exponential rise of data and data types has afforded impressive advancements for artificial intelligence (AI) since the "AI winter" of the 1990s. Current core machine intelligent technologies classify and identify things. For example, they can distinguish between cat faces and human bodies [16] and label dogs by their breed [17]. Moreover, they can beat humans at quite complex strategic games as chess [18], Jeopardy [19], and Go [20] –all very clear and well-defined environments and good examples of narrow intelligence. The colored schematic boxes in Figure 3 summarize where we are generally. Starting with machine learning, AI learns from past behaviors and events to predict what might happen in the future. Robotics has most recently been focusing on reinforcement learning that uses rewards as feedback with robots learning tasks by trial and error to avoid resets between task episodes and to introduce a more efficient approach to task performance [21] – techniques similar to our own behaviors. Classification is quite sophisticated now as enormous amounts of collected sample data with any number of characteristics productively lead to an outcome. Natural language processing has also greatly improved. Translation is no longer purely dependent on word by word matching as it was done in the past. Rather, it is progressively moving towards the use of large corpora of real-life conversational human-translated data and automated data collection, annotation, and analysis [22].

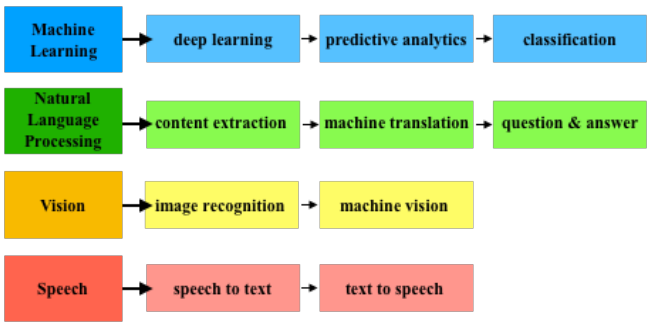


Figure 3. Schematic summary of core AI technologies available today.

As mentioned, machine vision is quite good with image recognition. However, egregious errors persist and discriminating between basic comparative characteristics like same-or-different between patterns and objects is still exceedingly difficult for the algorithms [23]. Furthermore, there are glaring cases where deep neural networks can easily be fooled into identifying a turtle as a rifle [24] and misclassifying a Stop sign as another target like Speed Limit 45 [25], for example, with systematized transformations of pixels in the digital image or with physical perturbations to real-world objects, respectively. The success of

such adversarial examples in causing significant targeted misclassification raises serious questions regarding the need for and development of resilient learning algorithms for real-world self-driving vehicle situations. Moreover, they highlight the weakness of an argument like quantity of test-driven miles as a measure of confidence for safety when an unforeseen scenario of this kind –with perhaps a very low probability of likeliness but with an existent probability– can easily mislead the system not for better but for worse. If not obvious before, these errors are clear reminders that vision, whether object recognition or scene navigation more broadly, is much more than just the detection of low-level visual features. In fact, advancing further my main argument for this paper, we are at the cusp of a much-needed change of approach to automated vehicle perception R&D if full automation is to be achieved.

Lastly, improvements have also been seen with speech recognition. Though still restricted to specific tasks and with lags, free-form speech translation capabilities are moving forward [26]. Still a long way to go, however, is when the computer will be able to perfectly decode a dictated talk irrespective of a speaker's regional accent or to not auto-correct incorrectly amidst the code switching common to multi-linguals' discourse in multiple shared languages; another example where content identification *and* context understanding will be paramount for reaching human-like intelligence for speech recognition and free-form collaborative narrative discourse. From these successes, failures, and open questions, a pattern arises of what is fundamentally missing: Despite using hundreds or thousands of available training data and evermore powerful deep-learning systems with up to 152 layers [27] requiring yet greater hardware capacity, AI systems continue to lack knowledge, awareness, common sense, and the ability to form and extract meaning and value from context, let alone to pursue a changing goal with intent. In essence, machine learning is still incapable of dealing with new and unknown situations because previously acquired knowledge cannot be generalized and/or transferred from one domain to another to adapt to an ever-changing environment.

To See and Know Is Human

As a result, the ultimate fundamental question is: how do we humans integrate bottom-up sensory processing with top-down knowledge processing to perceive what we perceive in the world and move about our business. For illustrative purposes let us dissect the image in Figure 4 to underscore how important top-down knowledge information is to our perceptual experience of the photograph and any preceding hypothetical insights processed, were behavioral action to ensue in the real world. Why this particular image? The reversed, reflected duplication of the blue-yellow doors/panels on the horizontal axis suggests a mirrored floor. Does the room have a mirrored floor? Further inspection of the still-visible titled floor near the farthest vanishing point –odd to have such a break on a smooth mirrored surface– suggests an encroaching waterline to an almost completely flooded room. Is the room flooded? Lack of a single ripple of movement (of the supposed water), however, a slightly airborne-looking divan against the left panel, a seemingly endless wall height, and shifted doors/panels' height between the actual and the reflected images create a cognitive dissonance between low-level visual feature perception, situational reality, and real-world expectations. The reflection seems off… Is this a real room? Was the reflection artificially created with Photoshop? Would one enter or not enter? Et cetera, et cetera.

Figure 4. Color film photograph of a seemingly flooded room. Shot with a Seagull Twin-Lens Reflex camera using a single exposure and a mirror. Scanned negative was not digitally altered, only dust spots were removed using Photoshop. Image is part of a photographic series by the author on visual illusions titled 'E/Il-lusive Spaces.'

A single static scene (and dynamic scene by consequence) engenders a host of parallel visual low-level and high-level feature and contextual reasoning (in no one particular order): specific and general identifications and classifications of local and global scene features, juggling of expectations, hypotheses tested, focused attention shifted accordingly to local scene features, expectations revised, and so forth. In other words, human vision amounts to a highly complex bidirectional feedforward-feedback process significantly influenced by contextual elements like focused attention, world knowledge expectations, and perceptual tasks as surrounding scene elements affect the perceptual quality of local features and global scene characteristics affect neurons' responses to local features [28]. All of this to call attention to what computer vision algorithms must be able to mimic in regards to melding mental representations of individual objects and evaluating their role within the scene as a whole to make meaning out of an image –let alone a dynamic stream of images– and not be thrown off course due to added physical perturbations on objects in the environment. By further implication, this processing becomes critical to successfully steer a vehicle through a sea of new environments and human agents with their own sets of hypotheses, expectations, implications, etc. and ability to process multiple pieces of information, focus attention appropriately, and produce the relevant behavioral action. I second Sebastian Thrun's remark –without the 'almost' and with an emphasis on 'is'– that "…artificial intelligence is almost a humanities discipline. It's really an attempt to understand human intelligence and human cognition" [29].

Accidents: Insights from A Glaring Problem

To continue the discussion on higher-order mental processing, it is imperative to address the topic of human error. Although expansively complex and multi-layered with many running definitions [30], human error can be generally defined as "…a generic term to encompass all those occasions in which a planned sequence of mental or physical activities fails to achieve its intended outcome, and when these failures cannot be attributed to the intervention of some chance agency" [31]. Developing taxonomies of error is an ever-evolving task and one I will not

discuss here. However, a particular type of error of significant relevance to understand within the conventional vehicle-fully autonomous vehicle (i.e. human-machine) relationship is that of 'mistakes.' Mistakes are particularly interesting in regards to their characterization of occurring at earlier stages of information processing and determined to be the result of either incorrectly assessing a stimulus or inadequately selecting an appropriate response [30]. I bring this topic to the forefront under the assumption that identifying and understanding the cognitive processes that lead to a mistake or an error more generally can provide valuable insights towards building appropriate measures for future prevention. As such, looking at available statistics for the attributed reasons in conventional motor vehicle crash data and the types of traffic accidents involving autonomous vehicles and conventional vehicles can support both the need for highly desired autonomous vehicle intervention and the need for clearly defined metrics for autonomous vehicle safety as well as underscore the computational requirements for achieving such safety.

According to the National Highway Traffic Safety Administration (NHTSA)'s National Motor Vehicle Crash Causation Survey conducted from July 2005 to December 2007 and released in 2018, 94% of an estimated 2,046,000 crashes throughout the US were attributed to driver-related errors [32]. Moreover, 41% of those crashes were attributed to a recognition error and 33% to a decision error. "Driver's inattention, internal and external distractions, and inadequate surveillance" were classified as recognition errors and "driving too fast for conditions, too fast for the curve, false assumption of others' actions, illegal maneuver and misjudgment of gap or others' speed" were classified as decision errors [32, p.2]. All other driver-related errors had much lower rates with 11% due to performance (e.g. overcompensation, poor directional control), 7% to non-performance (e.g. sleeping on the wheel), and 8% to other kinds not specified [32]. As the data lack specification regarding the number of kinds of errors within an attributed category, the attributed error type and resulting crash type, and the factors leading to the error type in the first place, a cognitive analysis of the situation is anyone's best guess. One can extrapolate, however, that the two most significant types of errors observed reveal three important elements that directly inform the driving experience: (i) the driver's level of focused attention to the task at hand of driving, (ii) the driver's perceived characterization of self in relation to her environment, and (iii) the driver's active role as a spontaneous decision-maker as hypotheses, beliefs, and assumptions of others' behaviors are assessed.

With respect to focused task attention, studies on multitasking during driving, for example, reveal both positive and negative effects on driving performance with such effects significantly depending on driving circumstances (e.g. no traffic vs. substantial traffic) and secondary task types (e.g. listening to the radio vs. using a tablet) [33]. In regards to the (mis)perceived characterization of self in relation to one's environment and the good or bad decisions made in the moment, understanding what went cognitively wrong, so-to-speak, is murky territory as it would at least first depend on the driver's self-reporting directly after an accident, if at all even possible. Any number of miscalculations on the part of the driver can occur as the result of either an individual mental event or a conjunction of mental events of the following kind: distance from the vehicle in front; distance of turn from origin point; reaction time to braking or speeding; control of vehicle's speed and arc of road curvature; length of road curvature; familiarity with the roadway; trust in vehicle's response to one's actions; meaning of the acceleration/deceleration, maneuvering patterns, light signals, honking, facial/hand/finger movements, etc. of other drivers; others' intentions; others' reaction abilities; others' driving habits; etc. While not all encompassing, considering such possible mental events emphasizes just how complex every driver's cognitive environment is during driving and suggests a complexity to be expected with the introduction of autonomous driverless vehicles on roadways made for humans and their conventional vehicles.

Turning a critical eye to the success rate of current autonomous vehicles leads us to accident reports. As shown in Figure 2 above, driverless vehicle accidents within the State of California in the United States are not to be disregarded when every company listed has reported at least one accident since 2014. Accident statistics on testing data from September 2014 to March 2017 reveal a very significant kind of reported collision [34]: fender-benders with a conventional vehicle hitting the autonomous vehicle from behind. Not surprisingly, most of the damage to the self-driving vehicle was 62% of the time to the rear, followed by 23% to the side, and 15% to the front. Consequently, as the conventional vehicle hit the self-driving vehicle, damage was significantly less. Moreover, 89% of reported accidents happened at an intersection. Additionally, a review of all 75 reported collisions for the year of 2018 (January 2 to December 11) reveals that out of the 46 vehicles in autonomous mode 52% were rear-ended by a conventional vehicle and out of the 29 vehicles in manual mode 28% were rear-ended by a conventional vehicle; whereby conventional vehicle refers to car, motorcycle, or bicycle. While not addressed in [34] but making a connection between the conventional motor vehicle crash causation data in [32] mentioned above and the possible mental events involved also described above, the following picture results: an intersection –much like being on the yield-and-merge lane to enter the highway or on the highway itself– gives rise to a vastly complex cognitive situation of a human driver in a conventional vehicle with a host of expectations about the vehicle in front and behind of them and an adverse consequence resulting from the poor interaction between the human in their conventional vehicle and the autonomous self-driving vehicle due, in part, to the lack of strategic, decision-making intelligence from the part of the autonomous self-driving vehicle. If these relatively innocuous fender-benders are occurring at intersections, one can just imagine the kind of paralysis possible in highly dense, changing, and unprogrammable urban environments.

A misconstrued argument to this dissection of human error is that it suggests the very opposite of what fully autonomous vehicle driving hopes to deliver one day: that making autonomous self-driving vehicles to drive like humans is to create equally faulty results and propagate erroneous and dangerous behaviors. I by no means advocate creating machines equally as fallible, chaotic, and/or unpredictable as humans when it comes to driving and human lives and environmental awareness (in the domain of AI and artistic creativity that is another argument). What I am highlighting with this discussion is that we must understand better from a cognitive perspective the when, where, and why of human driving behaviors and integrate the various resulting insights to create autonomous systems capable of dealing with such behaviors until the day when fully autonomous vehicles are the only machines whatsoever operating on our roadways and such human matters become inconsequential.

Cognitive Principles To Consider

In the meantime, there are still various human perception and cognition factors to understand and mimic in autonomous driverless vehicle software. As mentioned earlier in regards to our breakdown of the image in Figure 4, for every feedforward connection there is a reciprocal feedback connection that carries information about the behavioral context. This reinforces the actuality that human perception and cognition is a cycle of interaction. Perceptual information directly guides our decisions and actions and shapes our beliefs just as internal knowledge, expectations, attention, and working memory influence the way we perceive the world. Perception and cognition are interdependent and failure to integrate the two results in an incomplete system for true human-machine interaction to succeed because one biological system, us, remains intelligently superior to the other synthetic system, the machine. Currently, this is what we have if we make a computational analogy to our perception and cognition. Computer vision systems follow a linear feedforward process whereby information has a one-way flow through the layers of the network and even with supervised and unsupervised deep learning we do not know exactly how the system is learning. Rule extraction algorithms are being proposed for neural networks to explain their decision-making, but they are still in their infancy [35]. The point being that our visual system does not follow a visual cortical hierarchy in which information is conveyed in a single, feedforward manner to progressively higher levels in the hierarchy. How then can we expect a new digital system with linear feedforward flow to successfully see and know similar to an organic cognitive system like ourselves that took millions of years to evolve as such? We cannot. And brute-force training of systems on hundreds and thousands of training data until the systems eventually learn –somehow– is not the answer.

When testing computer vision algorithms to compare randomly generated black and white shapes as same or different within the same visual space, the algorithms were no better than chance at recognizing the appropriate relationship [23]. Such a result underscores the fact that computers only identify a collection of pixels that have similar patterns to collections of pixels they have "learned" to associate with particular labels and are incapable of discriminating where one object in the image stops and the background, or another object begins. Current computer vision systems, additionally, cannot deal with the following placement of unusual objects within a scene. As shown in a recent study where the image of an elephant was inserted within a typical-looking living room scene, the system started misidentifying and misclassifying all the other objects it had previously correctly identified as well as the elephant itself [36]. This has significant consequences, for example, for the adversarial inputs mentioned earlier which are successful precisely because the vision system does not have world knowledge of object and categorical constancy across time and context and does not engage in a constant loop of increasingly complex feedforward-feedback cognitive processing. How then do we build resilient learning algorithms for real-world autonomous driverless vehicle situations?

I believe the answer lies in taking a multidisciplinary approach, one espoused by the field of cognitive science whereby insights and methods from neuroscience, psychology, philosophy, anthropology, linguistics, and computer science are integrated in its fundamental goal to reverse engineer the mind/brain. Furthermore, it means significantly moving away from reductionist methods common to machine learning research and asking the bigger question of humanity: general not narrow intelligence. Consider the hallmarks of human intelligence. Overcome by computers in speed, accuracy, and precision, we continue to surpass computers in our capacity to generalize, learn, and manipulate and integrate multiple streams of known and novel information. Zooming into the issue of data efficiency, one of the biggest challenges in Cognitive Science is modeling how our minds do so much cognitively despite minimal amounts of information. More specifically, what allows for a human infant to only need a single example or very few examples of a cat, for example, to correctly deduce the next animal or object she encounters is a cat or not a cat? In comparison, a deep learning algorithm needs hundreds if not thousands of labeled training examples to correctly identify a cat and even still, be susceptible to radical misidentification upon minute pixel changes or real-world added perturbations. Part of the answer lies in our ability to generalize or efficiently extract key fundamental attributes within a category and compare and contrast those attributes against those of a novel object to determine its status within the pre-identified category. We can see this quite easily in Figure 5 where the new chair object at the bottom center is identified as a type of chair because it shares many core categorical chair properties with the other chairs shown. Classic studies with adults and children have revealed that when shown a particular nonce object called a "tufa" amidst a set of nonce objects, both adults and children correctly categorize and identify all the "tufa" objects in the set [37].

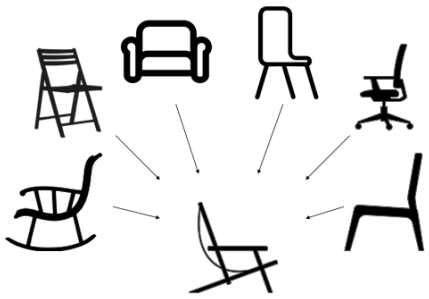


Figure 5. Six different kinds of chairs that all fall under the category of a 'chair' because of shared attributes such as a backside, a seat area, and a set of legs, irrespective of the shape and size of those attributes. Sharing such properties, the chair object bottom center is thus categorized as a 'chair.'

Scene representation much like we analyzed in Figure 4 is a wonderful example to illustrate our ability to dynamically integrate many different computations in real-time and flexibly transform, alter, and change mental representations to deal with incoming information. Accurate scene perception is a highly complex state of multi-level processing that includes low-level processing of simple environmental properties, mid-level processing of object identification and extraction, and most crucially, high-level processing of meaning whereby expectations of the world as determined by knowledge and past experience directly influence the what and how of low and mid-level processing [38]. As we have already seen, visual processing is so critical to how we understand and navigate our world that any intelligent system that is detecting and categorizing objects on the street must be able to do the same so not to hallucinate a squirrel crossing a road when it was in fact a human. Again, this is all importantly tied to our knowledge about the world and how inanimate things and animate beings work, or common sense. In an image of a standing elephant and a basketball on the floor next to an open doorway used to illustrate the power of common sense in deducing truths about

objects, it is absurdly obvious to us that the elephant, no matter how much it wiggles around, simply cannot fit through [39]. The consequent question arises: how is this abstract knowledge acquired and how does it guide learning and inference from infancy to adulthood? The answer is unsatisfactory and an open challenge within the field of Cognitive Science. Nativists who stand in contrast to a tabula rasa viewpoint postulate innate structures that facilitate model building. Connectionists who view mental phenomena as interconnected networks postulate learned representations. More recent modelers postulate hierarchical Bayesian models that rely on multiple levels of hypotheses with priors on priors.

To end, I will briefly mention the fruits of three interesting examples of cross-disciplinary R&D experimentation which I believe set up a very constructive path towards achieving models of general intelligence. The first pertains to CAPTCHAs or completely automated public Turing tests to tell computers and humans apart. The example in Figure 6 spells out CAPTCHA! and the symbols are effortlessly readable despite differing textures and the incomplete and cluttered symbolic representations. On closer inspection, the symbols illustrate the various principles of grouping known as Gestalt laws and are easily identifiable (closure, similarity, continuation, figure and ground, and proximity). Or perhaps they are not easily identifiable. Without knowledge of the philosophy of mind from experimental psychology to understand the laws behind the meaning we make from perception, the symbols are just a string of letters and represent nothing more. The main point emphasizing that context is key.

Figure 6. An example of a CAPTCHA that consists of the symbols C A P T C H A !. The symbols are composed in such a way to illustrate the following Gestalt principles of perception: the incomplete C to represent closure, the first and second appearance of A to represent similarity, the connected P to represent graphic continuation from A, the 3D H to represent figure and ground with the appearance of windowed arches, and the overall lack of spacing between CAPTCHA! to represent proximity of symbols and thus, using top-down knowledge of the meaning of CAPTCHAs, represent an actual English language acronym.

We seamlessly integrate the bottom-up sensory information of lines and curves with our top-down knowledge of written letter representations and flexibly and dynamically identify and categorize transformed letters as we individuate one letter from another and its background. Incorporating theoretical principles from cognitive science about the mind/brain to build software, researchers have successfully moved many steps ahead of traditional deep-learning approaches. Using an object-based recursive cortical network (RCN) model, they have created a human-like generative model that successfully breaks text-based CAPTCHAs and uses 300-fold less training data [40].

The second example pertains to the feature-binding problem in vision. In this particular domain the question is: how does the visual system represent hierarchical relationships between and within features (e.g. edges, objects). In other words, how does the visual system represent which low-level features belong to the doors/panels, divan, floor, and walls in Figure 4 we so fluidly segmented from one another in order to make sense of the scene? Incorporating acknowledged neural dynamics from neuroscience, researchers have significantly moved forward our understanding of how we make sense of our visuospatial world that relies on the emergence of polychronization, or the phenomenon in which a subpopulation of neurons fire in regularly repeating spatio-temporal patterns in response to specific visual stimuli. Significantly, neurons embedded within these polychronous neuronal groups receive convergent inputs from neurons representing lower- and higher-level visual features and they appear to encode the hierarchical binding relationship between features [41]. The crucial point here being that such semantically rich, hierarchal visuospatial representation is key for the brain to make sense of its sensory world and behave intelligently within it.

The third and final example is facial recognition. How does the brain process and recognize the myriad array of faces we see every day? Combining functional magnetic resonance imaging, single-cell recording techniques in macaques, and twenty years of research on face recognition, researchers have recently cracked the neural code for how faces are identified in the brain [42]. They identified about 200 neurons that are functionally specialized to distinguish facial features along specific axes in face space like the distance between eyes, hairline shape, face width, skin tone, texture, among others, and that neuronal response is proportional to the strength of the features. In other words, a strong response results from a large inter-eye distance but a minimal response to a small inter-eye distance. Essentially, these 200 neurons can combine in different ways to encode every possible face. The resulting model was used to accurately re-create faces monkeys were viewing. The implications for now modeling how the brain processes non-facial shapes and object recognition at large is an avenue ripe for investigation.

Conclusions for Yesterday

Once again, the critical question underlying this discussion resurfaces: what do we want (and need) to create at the end of the day? A machine that is like us? A machine that is like us but better (and better in what ways)? An entirely other kind of thing meant to enter our world for us to passively accept? An entirely other kind of thing meant to be part of our world and complement us?

As a cognitive scientist, business executive, and consumer, I propose the following if the auto industry wants a real competitive advantage in the business of AI: (1) It is time to restructure the challenges of creating efficient, effective, and safe autonomous systems and answer the following questions: What exactly does the industry want these autonomous vehicles to do? How does the current technology fit, in cognitive terms, with its ability to interact with us humans? What elements of human cognition do we need to integrate now within these systems to improve the outcomes of their intervention and eventual integration within our lives? To change goals ad-hoc as an industry from robo-cars for all to robo-service in closed-course, sunny areas simply because the reality of computer vision intelligence looks bleak for fast profits, at least in the near future, is to undermine the potential ahead.

(2) It is time to re-evaluate priorities within the industry. I agree with decreasing the thousands of deaths due to traffic accidents. I agree with decreasing traffic congestion to provide greener traffic options. I agree with democratizing mobility for a more equitable economy. And I agree with maximizing efforts towards building the software and hardware necessary for the growth of such complex systems-of-systems interaction required for autonomous vehicles. But the ultimate goal is human benefit for human survival, for environmental health. Human benefit will

not result from prioritized commercialization and advertised false hope if the industry does not move ahead in a judicious manner. I have argued in this paper that we need to significantly integrate the various strengths in our understanding of human cognition, behavior, and neurology with computer engineering and consequent software and hardware development. I am appealing specifically to a change in how the industry deals with the limitations, and in some cases hazards, of current computer vision systems.

(3) Lastly, to move forward with changed priorities it is time to promote a paradigm shift in automated vehicle R&D and overall start-up culture within the industry. Cognitive Science is, in simple terms, aiming to reverse-engineer the human mind/brain. Today is to see and know, not just to see. The foundation to move toward new, much more advanced intelligent systems is in place. But the industry needs to delve deeper and faster into the biggest questions of human cognition and merge old, new, and yet-to-be-discovered insights regarding how we categorize, learn, think, problem-solve, and make decisions to move beyond the acquisition of thousands of training data points, tweaking of deep learning approaches, and rise of hardware problems to keep up with the massive amounts of unlabeled, accumulated data. Cognition is as much a part of the computational, mechanistic, ethical, legal, and infrastructural aspects of making autonomous vehicles a scaled, integrated reality within society. Cognitive scientists need to have a seat at the table alongside engineers, ethicists, lawyers, and policymakers.

Ponder the problem this way. We stand before a bifurcated path of our own design. Are our most recent developments in machine intelligence a boon to humanity or a threat? Know the future we cannot. But imagine it we can. Machine intelligence, about sixty plus years old, is completely incapable of hoping, of dreaming, of transforming the known into the unknown –2.5 millions years of human evolution have helped to achieve these unique cognitive abilities on which our very existence as a species so greatly depends. So consider your duty as a citizen of humanity resolved and know that the time is more than ripe to participate in, dialogue and engage with others towards a more sustainable and equitable path ahead.

Acknowledgements

The author would like to thank La Petite Noiseuse Productions for supporting this research and the Chief Business Officer, Dr. Irene Gonzalez, for her comments and review of this paper.

References

[1] J. Walker, “The Self-Driving Car Timeline – Predictions from The Top 11 Global Automakers,” Emerj, 21 December, 2018 <https://emerj.com/ai-adoption-timelines/self-driving-car-timeline-themselves-top-11-automakers/>

[2] N. Naughton, “GM Moves To Deploy Driverless Car Fleet in 2019,” The Detroit News, 12 January, 2018 <https://www.detroitnews.com/story/business/autos/general-motors/2018/01/12/gm-driverless-car-fleet-cruise-av/109381232/>

[3] News Releases, IEEE, 5 September, 2012 <https://www.ieee.org/about/news/2012/5september-2-2012.html>

[4] SAE On-road Automated Vehicle Standards Committee, “Taxonomy and Definitions for Terms Related to On-road Motor Vehicle Automated Driving Systems,” 2014.

[5] K. Korosec, “Uber Reboots Its Self-Driving Car Program,” TechCrunch, 20 December, 2018 <https://techcrunch.com/2018/12/20/uber-self-driving-car-testing-resumes-pittsburgh/?utm_source=ActiveCampaign&utm_medium=email&utm_content=Uber+gets+back+on+the+road&utm_campaign=76+1+AV HW+Week+13+-+Main+Mail+%28Unopens%29>

[6] Report of Traffic Collision Involving An Autonomous Vehicle (OL 316), Department of Motor Vehicles, State of California, USA, as of 21 December, 2018 <https://www.dmv.ca.gov/portal/dmv/detail/vr/autonomous/autonomousveh_ol316+>

[7] 2017 Fatal Motor Vehicle Crashes: Overview, NHTSA’s National Center for Statistics and Analysis, Traffic Safety Facts – Research Note, DOT HS 812 603, October 2018.

[8] P. Koopman & M. Wagner, “Toward A Framework for Highly Automated Vehicle Safety Validation,” SAE Technical Paper, no. 2018-01-1071, pp. 1-13, 2018.

[9] M. López-González, “Theoretically Automated Conversations: Collaborative artistic creativity for autonomous machines,” in IS&T Electronic Imaging Symposium: Human Vision and Electronic Imaging, IS&T, Springfield, Virginia, 2018. DOI: https://doi.org/10.2352/ISSN.2470-1173.2018.14.HVEI-531.

[10] M. Campbell, M. Egerstedt, J. P. How, & R. M. Murray, “Autonomous Driving in Urban Environments: Approaches, lessons, and challenges,” Phil. Trans. R. Soc. A, vol. 368, no. 1928, pp. 4649-4672, 2010.

[11] M. López-González, “Cognitive Psychology Meets Art: Exploring creativity, language, and emotion through live musical improvisation in film and theatre,” in Proceedings of SPIE 9394, Human Vision and Electronic Imaging, 2015. DOI: https://doi.org/10.1117/12.2083880.

[12] M. López-González, “Minds in the Spotlight: Using live performance art to uncover creative thinking processes,” in IS&T Electronic Imaging Symposium: Human Vision and Electronic Imaging, IS&T, Springfield, Virginia, 2016. DOI: https://doi.org/10.2352/ISSN.2470-1173.2016.16.HVEI-143.

[13] M. López-González, “Trading Conversations Between Science and Art: When musical improvisation enters the dialogue on stage,” in IS&T Electronic Imaging Symposium: Human Vision and Electronic Imaging, IS&T, Springfield, Virginia, 2017. DOI: https://doi.org/10.2352/ISSN.2470-1173.2017.14.HVEI-156.

[14] G. A. Miller, “The Cognitive Revolution: A historical perspective,” Trends Cogn. Sci., vol. 7, no. 3, pp. 141-144, 2003.

[15] D. R. Hofstadter, Gödel, Escher, Bach: An eternal golden braid, London, United Kingdom: Penguin Books Ltd., p. 569, 1994.

[16] Q. V. Le, M.’A. Ranzato, R. Monga, M. Devin, K. Chen, G. S. Corrado, J. Dean, & A. Y. Ng, “Building High-Level Features Using Large Scale Unsupervised Learning,” in 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 8595-8598, 2013.

[17] D. Hsu, “Using Convolutional Neural Networks To Classify Dog Breeds.”

[18] L. Greenemeier, “20 Years After Deep Blue: How AI Has Advanced Since Conquering Chess,” Scientific American, 2 June, 2017, < https://www.scientificamerican.com/article/20-years-after-deep-blue-how-ai-has-advanced-since-conquering-chess/>

[19] T. Reddy, “Why It Matters That AI Is Better Than Humans at Games Like Jeopardy,” IBM AI for the Enterprise, 27 June, 2017

<https://www.ibm.com/blogs/watson/2017/06/why-it-matters-that-ai-is-better-than-humans-at-their-own-games/>

[20] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, & D. Hassabis, "Mastering The Game of Go with Deep Neural Networks and Tree Search," Nature, vol. 529, no. 7587, pp. 484-503, 2016.

[21] K. Chatzilygeroudis, V. Vassiliades, & J.-B. Mouret, "Reset-free Trial-and-Error Learning for Robot Damage Recovery," Rob. Auton. Syst., vol. 100, pp. 236-250, 2018.

[22] I. Zeroual & A. Lakhouaja, "Data Science in Light of Natural Language Processing: An overview," Procedia Comput. Sci., vol. 127, pp. 82-91, 2018.

[23] M. Ricci, J. Kim, & T. Serre, "Same-different Problems Strain Convolutional Neural Networks," ArXiv180203390 Cs Q-Bio Available at: http://arxiv. org/abs/1802.03390 [Accessed May 28, 2018].

[24] A. Athalye, L. Engstrom, A. Ilyas, & K. Kwok, "Synthesizing Robust Adversarial Examples," in Proceedings of the 25th International Conference on Machine Learning, PMLR 80, 2018. arXiv preprint arXiv:1701.07397, 2017.

[25] K. Eykholt, I. Evtimov, E. Fernandes, B. Li, A. Rahmati, C. Xiao, A. Prakash, T. Kohno, & D. Song, "Robust Physical-world Attacks on Deep Learning Visual Classification," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1625-1634, 2018.

[26] D. Khurana, A. Koli, K. Khatter, & S. Singh, "Natural Language Processing: State of the art, current trends and challenges," arXiv preprint arXiv:1708.05148, 2017.

[27] K. He, X. Zhang, S. Ren, & J. Sun, "Deep Residual Learning for Image Recognition," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770-778, 2016.

[28] C. D. Gilbert & W. Li, "Top-down Influences on Visual Processing," Nature Reviews – Neuroscience, vol. 14, no. 5, pp. 350-363, 2013.

[29] M. Chafkin, "Udacity's Sebastian Thrun, Godfather of Free Online Education, Changes Course," Fast Company - Tech Forecast, 14 November, 2013, <https://www.fastcompany.com/3021473/udacity-sebastian-thrun-uphill-climb>

[30] J. R. Fedota & R. Parasuraman, "Neuroergonomics and Human Error," Theor. Issues Ergon. Sci., vol. 11, no. 5, pp. 402-421, 2010.

[31] J. Reason, Human Error, New York: Cambridge University Press, 1990.

[32] S. Singh, "Critical Reasons for Crashes Investigated in The National Motor Vehicle Crash Causation Survey," National Highway Traffic Safety Administration – Traffic Safety Facts Crash-Stats, Report No. DOT HS 812 506, Washington, DC, pp. 1-3, March 2018.

[33] M. Nijboer, J. P. Borst, H. van Rijn, & N. A. Taatgen, "Driving and Multitasking: The good, the bad, and the dangerous," Front. Psychol., vol. 7, 1718, 2016. doi: 10.3389/fpsyg.2016.01718.

[34] F. M. Favarò, N. Nader, S. O. Eurich, M. Tripp, & N. Varadaraju, "Examining Accident Reports Involving Autonomous Vehicles in California," PLoS ONE, vol. 12, no. 9: e0184952, 2017.

[35] T. Hailesilassie, "Rule Extraction Algorithm for Deep Neural Networks: A review," arXiv reprint arXiv: 1610.05267, 2016.

[36] A. Rosenfeld, R. Zemel, & J. K. Tsotsos, "The Elephant in The Room," arXiv preprint arXiv: 1808.03305, 2018.

[37] J. B. Tenenbaum, C. Kemp, T. L. Griffiths, & N. D. Goodman, "How To Grow A Mind: Statistics, structure, and abstraction," Science, vol. 331, no. 6022, pp. 1279-1285, 2011.

[38] R. A. Rensink, "Scene Perception," in A.E. Kazdin (ed.), Encyclopedia of Psychology, vol. 7. New York: Oxford University Press, pp. 151-155, 2000.

[39] Mosaic Project, AI2, Allen Institute for Artificial Intelligence <mosaic.allenai.org>

[40] D. George, W. Lehrach, K. Kansky, M. Lázaro-Gredilla, C. Laan, B. Marthi, X. Lou, Z. Meng, Y. Lu, H. Wang, A. Lavin, & D. S. Phoenix, "A Generative Vision Model That Trains with High Data Efficiency and Breaks Text-based CAPTCHAs," Science, vol. 358, no. 1271, pp. 1-9, 2017.

[41] J. B. Isbister, A. Eguchi, N. Ahmad, J. M. Galeazzi, M. J. Buckley, & S. Stringer, "A New Approach To Solving The Feature-binding Problem in Primate Vision," Interface Focus, vol. 8, no. 4, pp. 1-23, 2018.

[42] L. Chang & D. Y. Tsao, "The Code for Facial Identity in The Primate Brain," Cell, vol. 169, no. 6, pp. 1013-1028, 2017.

Author Biography

Mónica López-González received her BAs (2005) in Psychology and French, MA (2007) and PhD (2010) in Cognitive Science, all from Johns Hopkins University. She has a Certificate of Art in Photography from Maryland Institute College of Art (2009). She held a postdoctoral fellowship at Johns Hopkins University School of Medicine from 2010 to 2013. Since then she has worked as a business executive, cognitive scientist, educator, entrepreneur, multidisciplinary artist, and public speaker as Co-Founder and Chief Science & Art Officer of La Petite Noiseuse Productions. Her work as a thought leader, strategy analyst, and evaluator for building equitable and sustainable artificial intelligent systems has taken her worldwide across different industries. In 2016 she was recognized as a "particularly imaginative polymath" by the Imagination Institute based at the University of Pennsylvania's Positive Psychology Center. She is a committee member of Human Vision & Electronic Imaging.

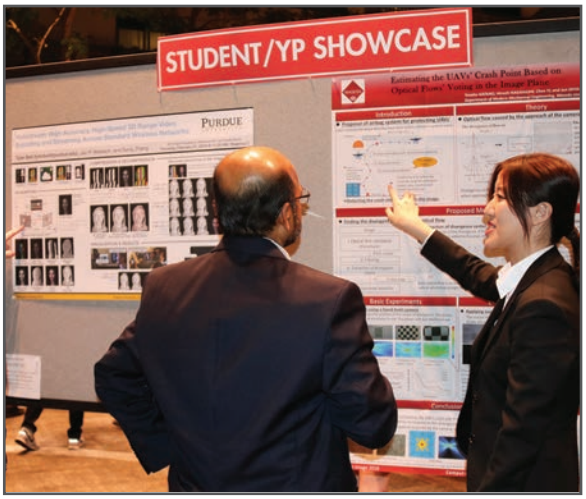
JOIN US AT THE NEXT EI!

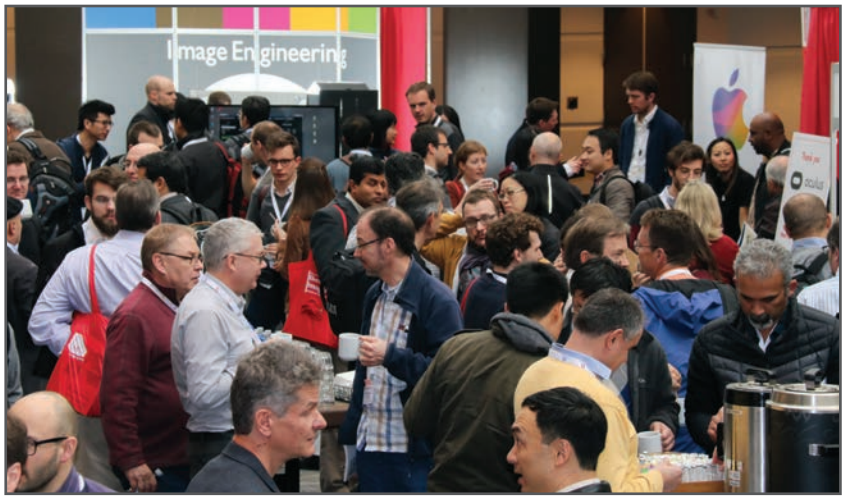
IS&T International Symposium on

Electronic Imaging

SCIENCE AND TECHNOLOGY

Imaging across applications . . . Where industry and academia meet!





• SHORT COURSES • EXHIBITS • DEMONSTRATION SESSION • PLENARY TALKS •
• INTERACTIVE PAPER SESSION • SPECIAL EVENTS • TECHNICAL SESSIONS •

www.electronicimaging.org

IS&T
imaging.org