# Enactive Perceptual Supplementation

Francis Jeanson

December 14th 2009

## Introduction

Aspiring to the development of human potential can be seen as a virtuous endeavor. From the development of early tools to the sophistication of modern technologies such as computers and cell phones, artifacts have led to the creation of modern and nearly global societies. By enabling the potential of each individual modern technologies are transforming the way we interact with our environment, transfer our knowledge and communicate with people. In parallel to these technological advancements, theories of cognition have emerged which stand firmly against the traditional cognitivist view. Embodied embedded theories uphold to various degree that intelligence cannot be understood as solely caused by processes internal to the body and brain but must be understood with respect to the environmental context in which it occurs. Practically, these views can not only help situate the significance of technological artifacts in their participatory role for cognition, but also help emit new hypothetical designs for them. In particular, the enactive approach is becoming increasingly accepted as a powerful theoretical framework for the analysis of human interface interactions that are not a priori representational such as iconic or symbolic displays (Varela et al., 1991; Grespan et al., 2008). However, the potential of the enactive approach as a guiding theory for the design of devices that can help us complete tasks through gained situation awareness with minimal disruption is still poorly understood.

In the following, I aim to explore this potential. I will start by introducing contending views on the nature of cognition and favour those which provide support for a workable understanding of how awareness of the world can be gained using artifacts that do not afford a priori representational content. In particular, I will present the enactive approach as a viable contender for this role with support from empirical evidence. I will then discuss existing perceptual supplementation device that best exemplify the enactive approach to device design. Because perceptual supplementation is most useful in high cognitive workload scenarios, I will explain how the enactive approach can account for situation awareness in contrast to classical information processing theories. I will finally approach enactive theory with respect to its practical application in the development of a device aimed at assisting car drivers placed in risky road situations.

## Cognitivist Inadequacies

Since the late 50's Artificial Intelligence has inspired the view that cognition is a mechanistic process that gives rise to rational thought based on the manipulation of symbolic structures (Newell and Simon, 1972; Minsky, 1974; McCarthy and Hayes 1969; Pylyshyn, 1984). In accordance with the computer paradigm, the cognitivist view understands the mind as formed by the interaction between clearly delineated parts, such as memory systems, perceptual systems, inference system, etc., that essentially manipulate physical representational structures (symbols) to give rise to claimed hallmarks of intelligence such as categorization, reasoning, and problem solving. According to this approach, the challenge lies in formulating the algorithms by which intentional or representational states can cause intelligent action. This view, however, has been brought under great scrutiny over the past 30 years. In particular it has suffered in two important ways:

1 - Cognitivism presupposes that representations can be meaningful independently of what's outside of the head (Harnad, 1990).

2 - Cognitivism assumes that perception, thought, and action is a sequential cycle with well defined boundaries (Clancey, 1997).

The first criticism has most generally been up-help by a broad group of cognitive scientists who have formulated the problem within what has come to be known as the embodied embedded paradigm. In particular, the failure of symbolic manipulation systems to satisfy the conditions required to ground these symbols with meaning has led many to more carefully consider the processes that give rise to perception: the transformation of sensory stimulation into representational content. This oversight has even provided some with fuel to dismiss the physical symbol system perspective altogether (Brooks, 1991; Varela et al. 1991). What theses views share is the notion that cognition cannot be studied as a process independent of the body and environment in which it takes place. Thought shouldn't be seen as the product of isolated brain activity but instead as the result of a complex coupling between the world, body and mind (Kelso, 1995; Clancey, 1997).

The second criticism also marks a staple of the embodied embedded approach. The importance of coupling between the individual's actions, perceptions, and mental processing suggests a strong parallelism amongst them. Indeed it is unclear that perception should always come first. Van Dijk, for example, warns that action may come before perception: for instance the act of running changes the optic flow on the retina thereby altering our sensation of the external world (Gibson, 1979; van Dijk, 2009). Also, action may precede planning: individuals make use of dynamic organizations of the environment via action to scaffold planning (Suchman, 2007; van Dijk, 2009). We can also organize our world so to offload knowledge using artifacts such as pen and paper, a computer, etc. (Hollan et al., 2000). Under theoretical scrutiny, the notion of parallelism in embodied coupling bares further weight when we come to understand that coupling is not simply about how the precedence of perception, thought, or action can alternate but about how cognition is the result of time varying self-organizing processes (Varela et al. 1991; Kelso, 1995; Buzsaki, 2006). Take for instance the complex interaction of multi-sensory integration. Sight and audition address qualitatively disjunct properties of the environment with events happening at different time scales. Because of this, a driver can stop a car at a red light while talking to a friend. Perception, thought, and action then self-organize independently for these separate tasks. This aptitude for divided attention isn't fool proof however. If we increase the degree of attention or cognitive load required by the driver for one of the tasks he or she may fail to adequately attend to the other (Broadbent, 1958). I will return to the notion of cognitive load when discussing perception supplementation devices. The point here is that complex behaviours cannot be justified by a clear chronological and functional ordering of perception, thought, and action. Instead, we must consider alternative perspectives that can offer comprehensive accounts of situated human performance.

Now that we have seen how two fundamental criteria of cognition are generally addressed by embodied embedded approaches we can work towards refining them in a way that motivates the use of the enactive approach for perceptually supplementing device design.

## Embodied Embedded Cognition Reconciled

Recent developments in embodied embedded cognition emphasize the existence of a divide within situated approaches. In particular, van Dijk suggests that phenomenological approaches should be seen as distinct from behaviour-oriented ones especially with respect to object design (van Dijk, 2009). A phenomenological approach favours the view that developing an appreciation for first person experience is central to understanding cognition (Heidegger, 1927; Merleau-Ponty, 1962; Varela et al., 1991; Noe, 2000; Dourish, 2001). Behaviour-oriented approaches in contrast are primarily concerned with cognition as emerging from ongoing interaction between the environment, body, and brain (Brooks, 1991; Hutchins, 1995; Clancey, 1997; Clark, 1997, van Dijk, 2009). Van Dijk is worried that speaking of experiences commits us to the view that consciousness is always present. Because of this, he warns that design approaches may seek to build devices that provide experiences when in fact no consciousness needs to be elicited. As an example, he imagines a haptic device that controls the user's stress level without the user noticing it. Furthermore, he is concerned that couplings between user and the environment may exist but are undetectable from a first person experiential perspective. For instance, the impact of a keyboard on the user's writing compared to a pen may be significant but not necessarily apparent from the user's point of view. But first person reports would keep this effect undetected. Instead of experiences he suggests that we conduct objective studies through the observation of actions or behavioural patterns. By approaching embodied embedded design from a behaviour-oriented perspective, van Dijk hopes to not only better approach questions of device design but also provide a conceptual tool set that can adequately detect and explain the causes of a particular human-device interaction.

Arguably, van Dijk's original unease with phenomenological approaches stems from the influence Paul Dourish has exercised in the field (Dourish, 2001; van Dijk, 2009). For van Dijk, Dourish's emphasis on 'user experience' in guiding design is phenomenological and thus ill fated for the reasons given above. However, van Dijk himself ultimately acknowledges that behaviour-oriented design can benefit from taking into account user experience in the particular case where the state of experience of the device is elevated to the state of experience of the action: "many tools operate best when ready at hand"; borrowing Heidegger's notion (Heidegger, 1927; van Dijk, 2009). This notion of ready at handedness or transparency of use was particularly well illustrated by Merleau-Ponty in his description of the blind man's cane (Merleau-Ponty, 1962; Clark, 2003). The blind man doesn't experience the cane when walking about, but experiences the environment through the cane. This transparency of use may very well constitute the 'holy grail' of interface design and so van Dijk must implicitly acknowledge the importance of person accounts.

What is needed then is a framework that will take cognition to be the outcome of a tight coupling between agent and the environment based on action, admit experience as a method for evaluation, and satisfy the criteria of parallelism and grounding of representations. What I aim to show in the following is that the enactive approach from its early development to its latest applications can meet this challenge.

## The Enactive Approach

The enactive approach was first introduced by Francisco Varela, Evan Thompson and Eleanor Rosch in their book The Embodied Mind in 1991. It is based on two central principles: (1) ”perception consists in perceptually guided action” and (2) ”cognitive structures emerge from the recurrent sensorimotor patterns that enable action to be perceptually guided” (Varela et al., 1991, p. 173). The first notion (1) emphasizes the central role that action plays in perception. By acting in the world we change its situation which in turn changes our perception of it. Hence the world is no longer independent of the agent but coupled to it. But what kind of leverage does it give us as a cognitive theory? At a minimum, we may want this principle to satisfy the criteria for cognition as the result of a concomitant operation of perception, thought, and action that was developed earlier. However, the interwoven dependence of perception and action may leave out thought altogether. Action and perception could regulate one another automatically in a reflexive way. For example, it is relatively easy to build a tracking device that maintains its heading with respect to a moving object (Smit and Tilden, 1991). In this case action is perceptually guided but any sense of thought is arguably absent: the device has no means to obtain meaningful conceptual structures because reflexive behaviour doesn’t require a sufficiently complex internal sensorimotor network. Ross Ashby precisely articulated this in his principle of homeostasis for agent adaptation (Ashby, 1960, p. 80). In his view, a first layer feedback loop between sensation and motor action forms nothing more than a reflexive layer. But a secondary feedback loop which governs the degree to which essential variables (those required for survival) are maintained within viable bounds must involve a complex structural organization that takes over motor actions in case of threat. Addressing representational grounding more directly, Lakoff and Johnson describe how meaningful conceptual structures arise from ”the structured nature of bodily and social experience” and from ”our innate capacity to imaginatively project from certain well-structured aspects of bodily and interactional experience to abstract conceptual structures" (Lakoff, 1988). But reflexive perception guided action is clearly not sufficient to account for abstract conceptual structures, nor can it address for how structures are formed from historical experiences. By introducing the second principle (2), Varela et al. aim at completing their account of cognition. Indeed, by stating that cognitive structures emerge from recurrent sensorimotor interactions they address what is missing. First, the notion of emergent structures suggests that coupling between the agent and environment can not only occur dynamically but can give rise to highly complex processes which may indeed include thought. Hence the parallelism criterion is met. Yet because these structures arise from recurrent sensorimotor interactions their account not only grounds cognitive structures into the physical world but also takes into account historical events. But the claim here is stronger: these structures emerge in order to enable actions to be perceptually guided. The semantic loop is thus closed with no open end. The agent generates meaning itself *from* perceptually guided action *for* perceptually guided action. Hence, meaningful conceptual structures can be obtained in Lakoff's sense (Lakoff, 1988).

With this formulation at hand the enactive approach can provide a strong theoretical grounding for existing experimental work. Take for instance a psychological experiment performed by Held and Hein in the early 1960's where kittens were raised in the dark and exposed to light in specific control scenarios (Held and Hein, 1963). Kittens were separated in two groups. Anytime the light was turned on kittens of the first group continued to move freely whereas kittens of the second group were placed in a basket harnessed to one of the free roaming kittens. Each group was thus exposed to the same visual experience. But kittens of the first group would actively perceive their environment whereas those of the second group would remain passive. When released a few weeks later kittens of the first group had normal behaviour whereas kittens of the second group would fall over edges and hit obstacles. From this we can conclude that visual perception alone is not sufficient to gain meaningful concepts of objects with which an agent interacts, but that, in this case, visual guidance of action is necessary. These results fit quite nicely within the enactive view. Since 1991 the enactive approach has also emerged as a novel theoretical contender for the backing of a growing trend in computer vision which, in refutation to cognitivist strategies, saw vision as an active process (Bajcsy, 1988; Aloimonos, 1993). More recently, research in evolutionary robotics has unveiled principles of structural coupling through active perception in neural network controlled robots. Scheier et al. demonstrated how sensorimotor coupling can allow robots with a limited sensory capacity to reduce hard sensory problems into easy ones (Scheier et al., 1998). Hard problems, known as type-2 problems, are present in the case where regularities are scarce or hidden in a complex sensory pattern (Clark and Thornton, 1997). In easy problems, type-1 problems, regularities are readily available. Scheier et al. hand designed a feedforward neural controller that would adjust its weights using backpropagation. When placed in a world filled with small or large cylinders no regularity was immediately available to the robot because of its mediocre sensory apparatus, namely eight infrared proximity sensors positioned around its circular body. However, by actively exploring the space the neural network controlled robot was able to distinguish large cylinders from smaller ones in virtue of a coupling between its odometric wheel sensors and proximity sensors. By circling cylinders the robot revealed regularities through sensory coupling by detecting the presence of an object and the difference in speed at which its inside and outside wheels where turning. Because this speed difference is greater around small cylinders the robot could learn to categorize cylinders as either large or small. These results first suggest that enaction can transform perceptually noisy environments into more clearly discernible ones using very simple sensory cues. Furthermore, enaction also enables higher cognitive feet to occur, such as categorization in this case.

Buy improving our formal understanding of active perception, is it possible to adopt the enactive approach as a productive framework for human interface design? In the next section I will discuss existing work in device design that successfully developed enactive enabling devices.

## Enaction Enabling Devices

Pioneering work by Bach-y-Rita et al. in perceptual supplementation initiated four decades of research in the field (Bach-y-Rita et al., 1969). They developed a tactile vision supplementation system (TVSS) that could help visually impaired people perceive remote objects in their surrounding space. This device consisted of a 20x20 array of vibrating pins placed in a chair to make contact with the subject's back. The tactile array would respond to input from a directional video camera. Subjects could manipulate the orientation of the camera and adjust the zoom level. Remarkably, after about 10 hours of training in detecting a simple set of objects through active exploration, subjects where not only able to recognize these objects quickly (5 to 20 seconds) but they would report that the sensation was felt as coming from the front of the camera rather than from the vibrotactile array on their back.

Hence, by exploiting the principle of perception guided action[1] subjects, in virtue of their sensory and cognitive apparatus, became the structural couplers[2] and organized their actions to extract meaningful representations. Intuitively it is not obvious that rudimentary vibrating pins can invoke, by stimulating the skin, a perception so strongly associated with vision. But for Bach-y-Rita et al. the two-dimensional spatial distribution of the pins on the subject's back could serve as a rough global corresponding description of what was seen through the naked eye. What is unclear, however, is the degree to which the spatial distribution of the pins would play a role in the detection and recognition of objects versus the role played by action perception. The sophistication of the device and lack of control measure in Bach-y-Rita et al.'s work prevents us from answering this question from these results alone. Similarly, more recent work in perceptual substitution or supplementation has generally been aimed at improving the portability, cost, and sensory sophistication of these devices for the sensory impaired. Perceptual analysis has focussed on the kinds of substitutions that could be obtained from various inner modal configurations such as touch-to-touch, or cross modal arrangements such as touch-to-sight (Bach-y-Rita and Kercel, 2003; Visell, 2009). These developments have elicited a significant debate regarding the physiological implications of these devices. Opinions vary as to whether substituted perceptions are extensions of the substituted modality, constitutions of percepts in the substituted modality, or formed by an entirely new sensorimotor profile (Auvray and Myin, in press)[3]. But because these devices are manipulated, carried, or worn, their enactive role is implicit while not controlled for. Controlling for the role of enaction with such devices should help formulate a better understanding of the nature of perceptual substitution at a physiological level (Grespan et al., 2008). Not having a clear understanding of this role however, is especially crucial for the exploration of the applicability and design of perceptual supplementation devices. In particular, it would be helpful to understand the limitations of such devices in terms of situation awareness.

By controlling the degrees of freedom of a minimal perceptually supplementing device, it should be possible to elucidate what these limits are and how to improve on them for enhanced perception in diving situations. In 2004, Adam Spiers developed a minimal haptic device originally intended

---

[1] the first principle of enactive cognition.

[2] the second principle of enactive cognition.

[3] Neural pathways of sensory activity have supported the view that substitution of vision, for instance, triggers activation in the visual cortex of well trained device users (Kupers et al., 2003).

to aid the visually impaired (Spiers and Harwin, 2004; Froese and Spiers, 2007). The enactive torch is a hand held device that vibrates when an object is located in front of it within a meter. Vibration intensity can be set to vary with respect to the distance of objects or stay constant, the latter being the most basic mode (the binary mode). In order to explore the perceptual affordances that the enactive torch can give rise to Grespan et al. have performed controlled experiments where the user had limited manipulative range of the device (Grespan et al., 2008). In the first task the device was set to its simple binary mode and fixed to a cart that could translate on a horizontal rail. With this single degree of freedom novice blindfolded subjects where able to accurately detect the width and centre of objects in the scene. In the second task, the rotation of the device fixed to the cart was made possible. By sliding and rotating the enactive torch, subjects where able to reliably detect whether objects of different sizes where near or far. Although subjects used different strategies[4] their ability to determine position, size and distance using only two degrees of freedom and a simple binary signal indicates that enactive perception plays a substantial role in elementary object detection and recognition. Indeed, in accordance with the first principle of enactive cognition these results support the prediction that simple reflexive devices may be developed in such a way that the user 'fills in' the thinking gap where structural couplings can emerge and allow the user to gain meaningful conceptual structures.

Although these two examples provide a general idea of the ways in which enaction enabling devices (EED) can be implemented, it is important to more carefully consider what sorts of device fall under this label. What the above devices demonstrate is that they can enable enaction if they allow the user to pick up regularities in the environment through bodily motion that would otherwise be unavailable. From this we can derive the conditions that make a perceptual supplementation device an EED. First, a perceptual supplementation device can be called an EED if it can allow the user to perceive regularities in the environment without exploiting iconic or symbolic signs that may already be available thought the modality that the device aims to interface with. In the case of the eye, a device is not enactive if it visually presents an image with stimuli that are already detectable through histories of coupling. For instance a video monitor that presents an image taken by a standard video camera is not an EED because it merely exploits existing perceptual networks of the user. This is because such devices overlook the

---

[4] In most cases subjects reported using a cognitive or intuitive strategy (Grespan et al., 2008).

first principle of perceptual guided action altogether. Second, an EED can increase in complexity as long as it doesn't violate the previous condition. There are two ways in which this complexity can increase: (1) by augmenting the number the reflexive responses in either time or space. Bach-y-Rita's et al.'s device precisely exploits spatial complexity by arranging reflexive vibrotactile units in an array. (2) An EED can itself go beyond the reflexive and become cognitively implicated through self-organizing sub-networks as long as the coupling with the individual exploits novel means, i.e. that it respects the first condition. Finally, the enactive approach predicts that, in fact, even if a perceptual supplementation device is regarded as complex, it will only successfully give rise to meaningful regularities if it respects the first principle of enactive cognition.

From this understanding, it should become possible to devise an adequate strategy for the design of enactive devices that could allow users placed in complex task environments to enhance their awareness of the present situation. To do so, it is useful to first overview some fundamental psychological developments pertaining to situation awareness and how the conceptual framework offered by the enactive approach can address these developments.

## Enaction in Situation Awareness

The phenomenal technological progress of the past century has led to the design of increasingly sophisticated work environments for systems control personnel and pilots. The technical sophistication of airplane cockpits has elicited particular interest and desire to gain a better understanding of human factors. Estimated to be a central psychological aspects in successful piloting, situated awareness was motivated in the mid 80's as a central concern for high cognitive demand environments (Hamilton, 1987; Emerson et al., 1987; Endsley, 1995). Endsley formulated the most comprehensive and accepted account of situation awareness to date. According to Endsley, situation awareness is "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future" (Endsley, 1988). Hence according to this view, a first level is involved with the perception of the status, attributes and dynamics of salient objects in the environment. This level is thus concerned with the cognitive processes required for the identification and aggregation of elements that play a significant role in task completion. The second level is concerned with how an agent gains an understanding of the global state of the situation given the elements obtained at the first level. Thus, a cognitive mechanism for integration and sense-making is required at this stage. The third level involves the ability for agents to predict the future state of the perceptual objects. This requires knowledge of their individual and global dynamics. From these three levels we can infer that situated awareness is understood as the result of a process that is operationally closed: the result of a process in a system is another process in the system (Varela, 1979, p. 55). In this case, knowledge is gained from the coupling of the agent with the environment where histories are involved (past predictions) and new projections emerge so to scaffold future couplings. This level of interpretation fits well with the enactive approach because taken that relevant elements in the environment are obtained from perception guided actions, the projections obtained from them give rise, in turn, to new perceptually guided actions. But it is important that we not overly conflate the traditional sense of situated awareness with enaction. By more carefully considering the content of Endsley's proposal it should become apparent that both approach are in reality quite distinct.

The classic sense of situated awareness, as the one proposed by Endsley, is encompassed in a theory of information processing. In his words an agent "perceives information" (Endsley, 1995). But to what degree are our senses receiving readily identifiable regularities? Information processing theories provide no account of the mechanisms involved in how information is gained or grounded, but instead rely on the assumption that information is representational a priori: that, for instance, the object is a speed dial, or that it is a pedestrian, or a car. As Clancey notes: "in information processing psychology [...] the idea of *information* is often conflated with the ideas of *data, representation, model,* and *knowledge*, so the only distinctions are what is input and what is output in a given situation" (Clancey, 1997, p. 77, emphasis original). Information processing theories thus fall typically fall under the cognitivist umbrella. Indeed, this can lead to a convenient systematization of the cognitive processes involved in situation awareness. Not only does it allow models to easily integrate psychological findings that use the same vocabulary, but it can also draw categorical distinctions between variably contingent processes. Although this is also the case with respect to Endsley's theory of situation awareness, it is important to note that even in this information processing account, he appeals to pioneering concepts of embodied embedded cognition that come from ecological psychology.

In Endsley's more careful development of situated awareness, he introduces seven information processing mechanisms and structures that participate in it: pre-attentive processing, attention, perception, working memory, long-term memory, automaticity, and goals. Especially relevant to the purpose of this paper, pre-attentive processing and attention are here seen as part of the perception process. First, pre-attentive processing refers to the parallel detection through pre-attentive sensory stores of environmental properties such as proximity, color, simple properties of shape, and movement, which cue further attention (Endsley, 1995). This view has the merit of not taking objects as being perceivable a priori. Instead, it acknowledges the requirement for a detection mechanism. This view appeals to an ecological perspective in so far as it suggests that invariance's in the organization of the environment, such as ambient light, can directly specify the properties of this environment (Gibson, 1979). But this cannot the case, as I have illustrated with Scheier et al.'s robot, regularities must be extracted via active perception where stimuli are coupled over time (Scheier et al., 1998). In enactive terms, the environment is enacted through histories of coupling (Varela et al., 1991, p. 204). Next, attention enforces constraints on this parallelism but it is also a capacity shared with other processes such as decision-making and action. Attention can be a limited resource and can suffer from over solicitation. Attention is thus seen here as a mechanism for focusing and filtering with limited resources. But how is attention instantiated? How can one explain the mechanism by which an agent attends to a particular scene, event, or object? Two basic principles in attention theory are proposed for this: automatic response to an alarm (bottom-up), and deliberate monitoring (top-down) (Wickens and McCarley, 2008). The enactive approach is only supercifially compatible with this interpretation. In the enactive sense, automatic response to an alarm can be understood in terms of a reflex response to a perturbation in sensorimotor coupling (Varela et al.,1991, p. 151). Because sensorimotor coupling takes the form of dynamic process in which some form of attractive state is reached, a sufficiently strong disruption to this coupling will induce a transition of the coupled network dynamics[5]. Deliberate monitoring can also be satisfied by the enactive approach. Enaction is precisely a supervenient state by which an agent is coupled to objects, states or events via perceptual guided action. However, parallel filtering of sensory 'inputs' is instead understood in terms of parallel self-organized sub-networks. Instead of seeing the agent as an input output machine detached from the environment, enactive theory adopts the view that an agent is in dynamic consonance with the environment: information is not extracted from the world but generated by sensorimotor coordination. Finally, for Endsley, perception is constituted by pre-attentive and atten-

---

[5] Varela et al. illustrate this in the case of stable cellular automata which undergo state transitions upon the introduction (input) of disruptive states in a few cells.

tional processes, but perception also encompasses the notion that experience in the world allows one to develop expectations about future world states. According to him, these expectations are made available from memory. As illustrated earlier, histories of sensorimotor coupling are formed by past experiences. This supports enaction as a viable contender as a complete theory of situation awareness with respect to perception and the formation of appropriate responses. However, contrary to information psychology perspective, the enactive view doesn't approach situation awareness as a complex amalgam of distinct sequential causal procedures that accomplish sophisticated parallel processing feats. Instead, situated awareness is achieved via bodily coupling with the environment. Complex tasks are achieved through the appropriate asynchronous interaction of sensorimotor sub-networks obtained from perceptually guided actions. But with respect to situation awareness and the design of helpful interfaces one might ask: what does enaction give us that the classic approach to situation awareness doesn't? To address this question I propose to look at a concrete example which can motivate the use of enaction as a powerful framework for device design aimed at improving situation awareness during automobile operation in dense traffic.

## Enactive Supplemented Perception for Driving Assistance

Although to many it may feel as second nature, driving is an activity that demands significant attention to perform safely. Varying road conditions such as traffic density, traveling speed, weather, along with driver space distractions like passengers, music, telephones, etc. can place significant stress on the driver's ability to keep the vehicle in control and perform the right actions when needed. But beyond distractions, ergonomic factors also play a role in road accidents. Because driving requires significant visual awareness obstructions such as blind spots can be significant impediments. Wang and Knipling found that 4% of all police reported traffic accidents in the USA in 1991 were caused by lane changes or merges (Wang and Knipling, 1993). These accidents cause more than 10% of all crash caused congestions leading to significant negative economic impacts (Chovan et al., 1994). Causal analysis by Chovan et al. indicates that most of these accidents are due to drivers being unaware of other vehicles in the adjacent lane. To help remedy this they suggest that driver warning systems could be developed (Chovan et al., 1994). Various complexities for such systems are proposed in their report. A basic system would serve as a presence indicator that would take the form of a visual, auditory or haptic display. It would indicate when a vehicle is present in an adjacent lane when performing a lane-change or merge. The second, more sophisticated system would allow the driver to gain knowledge of the position of surrounding vehicles, their speed, and perhaps even acceleration. This is what they call a situation display. As Chovan et al. note, however, these systems can either be passive or overt/intrusive. A situation display would typically be considered passive, because it would most commonly be implemented as a visual display that would inform the driver via iconic or symbolic representations of the situation. Because the driver can optionally access this information, the system is passive. Overt/intrusive systems are those which impose on the driver a signal about the situation. For instance sound signals or skin bound haptic devices. Because the driver cannot optionally ignore them they are overt. Importantly there are significant tradeoffs between both system types. Whereas visual situation displays afford more situation awareness to the user than do alert systems, they also demand greater visual and cognitive resources. The driver must not only take his or her gaze away from the road to attend the display, but also employ specialized cognitive resources to extract knowledge of the display. This sort of intermodal interference (visual-visual) has been identified as a significant causal factor of impairment during driving because of the significant visual demands required (Parkes and Coleman, 1990; Srinivasan and Jovanis, 1997). Wickens and McCarley suggest instead that cross-modal systems can significantly alleviate this cognitive load because no interference takes place (Wickens and McCarley, 2008, p. 134). In the case of driving, a situation display would most reliably be auditory or haptic. Although auditory displays can easily provide symbolic stimuli via speech, they are recognized as often irritating and can interrupt higher priority tasks (Ho et al. 2004). Alternatively, haptic systems have not shown to be obstructive. However, it isn't clear that such systems can go beyond the overt alert system and serve as situation displays. By approaching this problem from an enactive standpoint I hope to motivate the idea that this is indeed possible.

In the case of lane-change and merging a haptic situation display should help drivers gain awareness of any obstacles in adjacent lanes without requiring a head turn. At the very least a simple haptic device could provide cues as to weather a vehicle is near in the left or right blind zone. With proximity sensors placed at the back left and right of the car adjacent vehicles could be detected. We can imagine a vibrating signal in the lower left and right of the driver's seat. Affecting two distinct sensory locations such a device would serve as an indexical cue with respect to the location of the adjacent vehicle. This spatial organization suggests something interesting for enactive cognition: spatial arrangement of stimuli at the interface between device and user can compensate for time dependent actions by translation of time domain events into spatial domain events. In this case, the driver can gain approximate knowledge of the position of vehicles without action (head turning) in virtue of the spatial quality of the stimulus. If the stimulus was located only at a single location on the driver's back, however, it would have to encode the signal over time. Note that this is only true if the driver knows that a left vibration indicates the presence of a vehicle on the left side or vice versa for the right. Thus the driver must predispose of meaningful couplings with the device. There are two ways the driver could have gained this knowledge. First from prior understanding of the way the device works. Because the system is simple, a verbal explanation could instruct a driver on how to use the device. But it is also possible that a driver would come to understand the utility of the vibrotactile system while driving normally. When he or she would want to change lane, a head turn would coincide with a vibration. Conditional learning would then take place whereby the driver would perceive the vibration as an accurate indicator of the presence of a nearby vehicle. Situation awareness is thus obtained in Endsley's sense (Endsley, 1995). Perception, from histories of coupling, is enabled where vibrations become distinguishable from other percepts. Their meaning is integrated from their spatial organization. Finally, structural couplings emerge to form reinforcing histories of expectations of future events. This compensation for action through spatial organization seems to go against the first principle of enactive theory which states that perception is perceptually guided action. This is not actually the case, although it suggests that spatial organization of senses can compensate for action in detecting sensory regularities, these perceptions direct action but also change with respect to those actions, e.g. if the driver decides to slow down or accelerate the haptic stimuli will change. Perception, in this case, is thus guided by driver/vehicle actions rather than driver actions alone.

Interestingly, this conception allows us to simplify and improve the system. Overt systems, as warned by Chovan et al. can become a nuisance or in-vehicle distraction (Chovan et al., 1994). Furthermore, perceptual supplementation devices can impact the reliance and compliance of the user depending on the device's activation threshold (Wickens and McCarley, 2008, p. 37). On the one hand, a driver's reliance on the system is reduced if the activation of system occurs to infrequently or because of a high threshold setting. If the driver is notified of a danger too late then the system is inefficient and becomes unreliable. On the other hand, a driver's compliance to this presence indicator may wane if the threshold is too low. Compliance is reduced because the event is too frequent and/or dissociated from the driver's intended actions. The driver is less likely to act upon haptic stimulation when it is in fact required. Threshold is low if, for instance, the device is highly sensitive to the presence of objects in the environment because its proximity sensors sense even very distant objects, or if it simply activates when the driver doesn't intend to perform a lane-change or merge. This latter case is especially relevant to the system suggested above. Compliance is likely to diminish if each time a vehicle is passed or is passing the haptic system is activated, even if the driver intents to stay in the same lane. As a solution, Chovan et al. proposed an alert system that would only become active when the driver uses the turn-signal (Chovan et al., 1994). But this system becomes ineffective if the driver forgets to use them. Alternatively, the system could activate only when the driver commits to a turn with the steering wheel. Turning the wheel left for a short but sufficient amount of time would cause the system to activate if a car in the left blind zone is present and vice versa for a right turn. Of course other issues such as reduced reliance may arise from this kind of setup, but a proper activation mechanism can help simplify the haptic system suggested above. Because haptic perception is now guided by the act of turning left or right, the vibrotactile signal does not need to be spatially segregated into a left of right indexical signal. The interface between user and device could be established as a single point on the driver's body. From this arrangement a very simple form of sensorimotor coupling takes place where spatial knowledge of the environment is successfully gained.

Although this is sufficient to detect the presence or absence of an object it does not allow the driver to gain any further knowledge about the environment such as distance, speed, or acceleration of the adjacent vehicle. But by taking the driver/vehicle pair as the actual agent new forms of coupling can be seen to emerge. Critically however, according to enactive theory, as we seek to have the user extract representations of increasing sophistication the user must dispose of more ways to manipulate the device to explore the space and reveal regularities. As seen with the enactive torch, determining the distance of objects required that the torch be manipulated in at least two dimensions: horizontal translation and rotation (Grespan et al. 2008). Arguably, in the case of the driver/vehicle pair forward translation of the vehicle when passing neighboring vehicles may not be sufficient. For such regularities to emerge, the device itself must increase its enactive role. Earlier, I showed how an enactive device could vary in complexity in terms of its spatial-temporal architecture but also in the degree to which it itself autonomously develops sensorimotor couplings. As demonstrated by Bach-y-Rita et al., a translation of a two-dimensional video input into a two-dimensional vibrotactile array is sufficient for subjects to detect and recognize objects (Bach-y-Rita et al., 1968). Situation awareness with representational content can be achieved by improving the device's sensory apparatus and the device user interface. In a similar way Froese and Spiers' enactive torch in an advanced mode could change vibration intensity depending on the proximity of the object (Froese and Spiers, 2007). Encoding environmental conditions through spatial-temporal arrangements could then provide the driver with a number of representational cues. Similar to the enactive torch, drivers could be warned of the proximity of a vehicle from the vibration strength. In addition, speed of a neighboring vehicle could be encoded via vibration frequency. A more sophisticated virbotactile device composed of an array of vibrating pins could also serve to encode other environmental conditions. Much like Bach-y-Rita et al.'s TVSS system, an array could encode acceleration of adjacent vehicles via displacement of vibration from one point in space on the driver's back to another. Recognition of the vehicle type, such as car, truck, motorbike, could also potentially be encoded in such a configuration. What the enactive approach predicts is that the manner in which environmental conditions are transmitted to the agent doesn't matter as long as the modality and cognitive architecture of the agent can self-organize into sensorimotor sub-networks that promote the use of the sensations into actions. If implementation conditions meet these criteria then not only can perception be achieved but sense-making as well.

By reducing our distinction between the driver and the vehicle as independent sensorimotor systems we can achieve a more clear understanding of the kind of experience the driver would obtain from driving a vehicle in which he or she can more completely integrate with. Indeed, the possibility of having drivers experience the vehicle as an extension of their own body could improve their situation awareness or road conditions. The motorcyclist for instance, by leaning with the vehicle in turns, feels a much stronger coupling with the machine and the surrounding environment (personal experience). By developing a phenomenology of such driver/vehicle couplings it may be possible to better guide system design in order to maximize this feeling of ready-at-handedness (Merleau-Ponty, 1962). This is a novel prospect for perceptual supplementation that is well accounted for by the enactive approach.

## Conclusion

In the present paper, I have first shown that a traditional cognitivist approach is insufficient to account for how agents obtain meaningful content of the world and that a serial account of the processes involved in cognition fails to capture the importance of parallel sensorimotor integration for sense-making. Instead I motivated that an embodied embedded approach could satisfy these conditions. I then illustrated how embodied embedded approaches to perceptual device design need to emphasize the importance of action in cognition and that they should not ignore the potential significance of phenomenological accounts which value agent experiences. I then provided an outline of enactive cognition and its viability as an effective framework to account for cognition and perceptual supplementation device design. To better locate this approach with respect to contemporary approaches to situation awareness I discussed how it could account for the mechanisms classically viewed as information processes. Finally, I described how enactive theory could aid in the interpretation of cognitive structures developed by a car driver who improves situation awareness via perceptual supplementation. I also motivate the manner in which existing enaction enabling devices can inspire novel designs for increased driver awareness, and how accounting for a driver's first person experience could improve road safety.

## References

Ashby, W. R. 1960. Design for a Brain, Second Edition, John Wiley and Sons: New York, NY.

Aloimonos, Y. 1993. Active Perception. Lawrence Erlbaum Associates.

Auvray, M. and Myin, E. (in press). Perception with compensatory devices: From sensory substitution to sensorimotor extension. Cognitive Science.

Bach-y-Rita, P. and Kercel, S.W. 2003. Sensory substitution and the human-machine interface. Trends in Cognitive Sciences Vol.7 No.12.

Bach-y-Rita, P., Collins, C. C., Saunders, F. A., White B. and Scadden L. 1969. Vision substitution by tactile image projection. Nature, vol. 221 pp. 963-964.

Bajcsy, R. 1988. Active Perception. Proceedings of the IEEE, (76)8:966-1005.

Broadbent, D.E. 1958. Perception and Communication. London: Pergamon.

Brooks, R. A. 1991. Intelligence without representation. Artificial Intelligence, 47, 139-159.

Buszaki, G. 2006. Rythms of the Brain. Oxford University Press.

Chomsky, N. 1957. Syntactic Structures, The Hague/Paris: Mouton.

Chovan, J.D., Tijerina, L., Alexander, G., and Hendricks, D. 1994. Examination of Lane Change Crashes and Potential IVHS Countermeasures. National Highway Traffic Safety Administration.

Clancey, W. J. 1997. Situated Cognition: on Human Knowledge and Computer Representations. Cambridge University Press.

Clark, A. 1997. Being there: Putting brain, body and world together again. Cambridge, MA, USA: MIT Press.

Clark, A. 2003. Natural-born cyborgs: minds, technologies, and the future of human intelligence. Oxford University Press, NY.

Clark, A. and Thornton, C. 1997. Trading spaces: computation, representation and the limits of uninformed learning. Behavioural and Brain Sciences, 20, 5790.

Dourish, P. 2001. Where the action is: The foundations of embodied interaction. Cambridge, MA, USA: MIT Press.

Emerson, T. J., Reising, J. M., Britten-Austin, H. G. 1987. Workload and situation awareness in future aircraft. SAE Technical Paper (No. 871803). Warrendale, PA: Society of Automotive Engineers.

Endsley, M.R. 1988. Design and evaluation for situation awareness enhancement. Proceedings of the Human Factors Society 32nd Annual Meeting. pp.

97-101.

Endsley, M. R. 1995. Towards a Theory of Situation Awareness in Dynamic Systems. Human Factors, 37(1), 32- 64.

Froese, T. and Spiers, A. 2007. Toward a Phenomenological Pragmatics of Enactive Perception In Enactive/07: Proc. of the 4th Int. Conf. on Enactive Interfaces, Grenoble, France: Association ACROE, pp. 105-108

Gibson, J.J. 1979. The Ecological Approach to Visual Perception. Boston: Houghton Mifflin.

Grespan, L., Froese, T., Di Paolo, E. A., Seth, A., Spiers, A., and Bigge, W. 2008. Investigating the role of movement in the constitution of spatial perception using the Enactive Torch. in: E. Ruffaldi & M. Fontana (eds.), Enactive/08: Proc. of the 5th Int. Conf. on Enactive Interfaces, Pisa, Italy: Edizioni ETS, pp. 105-110

Hamilton, W. L. 1987. Situation Awareness Metrics Program (SAE Technical Paper Series No. 871767).

Harnad, S. 1990. The Symbol Grounding Problem. Physica D 42: 335-346.

Heidegger, M. 1986. Sein und Zeit [Being and time]. Tbingen: Max Niemeyer Verlag. (Original work published in 1927)

Held, R. and Hein, A. 1963. Movement-produced stimulation in the development of visually guided behavior. Journal of Comparative and Physiological Psychology, 56(5), 872-876.

Ho, C. Nikolic, M.I., Waters, M.J., Sarter, N.B. 2004. Not Now! Supporting Interruption Management by Indicating the Modality and Urgency of Pending Tasks. Human Factors, Vol. 46.

Hollan, J., Hutchins, E., Kirsh, D. 2000. Distributed Cognition: Toward a New Foundation for Human-Computer Interaction Research. ACM Transactions on Computer-Human Interaction, Vol. 7, No. 2, Pages 174-196.

Hutchins, E. 1995. Cognition in the wild. Cambridge, MA, USA: MIT

Press.

Kelso, J.A.S. 1995. Dynamic patterns: the self-organization of brain and behavior. MIT Press, Cambridge.

Kupers, R. et al. 2003. Activation of visual cortex by electrotactile stimulation of the tongue in early-blind subjects. Neuroimage 19, S65.

Lakoff, G. 1988. Cognitive Semantics. In Meaning and Mental Representations, ed. Umberto Eco et al. Bloomington: Indiana University Press.

McCarthy, J. and Hayes, P. 1969. Some philosophical problems from the standpoint of artificial intelligence. In: Meltzer B. & Michie, P. Machine Intelligence Volume 4. Edinburgh: Edinburgh University Press.

Merleau-Ponty, M. 1962. Phenomenology of Perception. C. Smith (translator). Routledge & Kegan Paul.

Minsky, M. 1974. A framework for Representing knowledge. MIT Lab Memo # 306.

Neisser, U. 1967. Cognitive Psychology. New York: Appleton-Century, Crofts.

Newell, A., and Simon, H. A. 1972. Human problem solving. Englewood Cliffs, NJ, USA: Prentice Hall.

Noe, A. 2004. Action in Perception, The MIT Press.

Parkes, A. and Coleman, N. 1990. Route Guidance Systems: A Comparison of Methods of Presentating Directional Information to the Driver. Proceedings of ergonomic society 1990 Annual Conference, E. J. Lovesay (ed.), Taylor and Francis.

Pylyshyn, Z. W. 1984. Computation and cognition. Cambridge MA: MIT/Bradford

Scheier, C. Pfeifer, R., Kunyioshi, Y. 1998. Embedded neural network Exploiting constraints. Neural Networks, (7-8):1551-1569.

Smit, M.C. and Tilden, M. 1991. Beam Robotics. Algorithm, Vol. 2, No. 2, p.15-19

Spiers, A. and Harwin, W. 2004. The Haptic Torch: A Torch for the Blind. Int. Conf. on Disability, Virtual Reality and Assistive Technology, Oxford.

Srinivasan, R. and Jovanis, P. P. 1997. Effect of in-vehicle route guidance systems on driver workload and choice of vehicle speed: findings from a driving simulator experiment. In Ergonomics and Safety of intelligent Driver interfaces, Y. I. Noy, Ed. Lawrence Erlbaum Human Factors In Transportation Series. L. Erlbaum Associates, Hillsdale, NJ, 97-114.

Suchman, L. A. 2007. Human machine reconfigurations: Plans and situated actions (2nd ed.). Cambridge UK: Cambridge University Press.

van Dijk, J. 2009. Cognition Is Not What It Used to Be: Reconsidering Usability From an Embodied Embedded Cognition Perspective. Human Technology. Volume 5 (1), 29-46.

Varela, F.J. 1979. Principles of Biological Autonomy. New York: Elsevier North Holland.

Varela, F.J., Thompson, E., Rosch, E. 1991. The embodied mind: Cognitive science and human experience. MIT Press, Cambridge, MA, USA.

Visell, Y. 2009. Tactile sensory substitution: Models for enaction in HCI. Interacting with Computers 21, p. 38-53.

Wang, J.S., Knipling, R.R. 1993. Lane Change/Merge: Problem Size Assessment and Statistical Description. National Highway Traffic Safety Administration.

Wickens, C.D., McCarley, J.S. 2008. Applied Attention Theory. CRC Press.