This manuscript has been accepted for publication in *Cognitive Systems Research*. Please check the published version at: http://dx.doi.org/10.1016/j.cogsys.2016.11.009

Augmented Reality: An Ecoloigcal Blend

Vicente Raja[1], Paco Calvo[2,3]

[1]McMicken College of Arts and Sciences, University of Cincinnati, 7148 Edwards One, Cincinnati, OH 45221-0037, USA

[2]Minimal Intelligence Lab (MINTLab), University of Murcia, Edificio Luis Vives, Campus de Espinardo, 30100 Murcia, Spain

[3]EIDYN Research Centre, University of Edinburgh, UK

Abstract

In this article we present Ecological Augmented Reality (E-AR), an approach that questions the theoretical assumptions of main-stream Augmented Reality (AR). The development of AR systems to date presupposes an information-processing theory of perceptionthat hinders the potential of the field.Generally, in AR devices, virtual symbolic information is superimposed upon the environment in such a way that the real and thevirtual may be processed, informationally speaking, in tandem. Thus, we find information in reality itself, as well as virtual symbolicinformation. But by increasing the burden of symbolic crunching, AR devices run the risk of saturating the user of the technology.

AR systems developed under the principles of an ecological psychology may contribute to new and better levels of performance andadaptation to the user's perceptual abilities. Our proposal is to develop AR devices such that reality itself is augmentednon-symbolically by blending real and virtual layers/information. Although there are seldom AR devices in the market that are designedecologically, two fields of research may well bring inspiration to AR developers. These are the design and manipulation of real objects,and ecological research in the field of sensory substitution. We consider them both in turn with an eye to putting forward a frameworkthat eschews any type of information-processing regarding the nature of our psychological processes. Ultimately, our aim is to providesome guidelines for the exploration of an ecological trend in AR applications.

Keywords: Augmented reality; Ecological psychology; Affordance; Ecological design; Sensory substitution

Augmented Reality: An Ecological Blend

1. Introduction

Augmented Reality (hereafter, AR) makes reference to the real-time perception of an environmental setting that has been enhanced by means of computer-generated virtual components. Google Glass™ or Pokemon Go! are probably the examples that first comes to mind when one thinks of technologies that blend aspects of the real world with virtual ingredients. The field is moving fast, and today's applications, especially mobile AR, promise to revolutionize the field even further. In terms of economic investment, applications, commercial expectations, and scientific development, AR technology is one of the leading research fields of our days. From the humblest applications in mobile or television environments to the future commercialization of ambitious projects, innovations and possibilities introduced by AR start to permeate all areas of human life. Navigation, urban design, education or medicine, to name but a few, will, no doubt, be modified by this new technology (for a survey of AR technologies and applications, see Furht 2011).

AR is a fertile field that allows for scientific and humanistic interdisciplinary reflection. Its relationship with different users' abilities and its impact on society make AR an unparalleled space for creation and analysis. But current AR has endorsed, we suspect rather acritically, an understanding of the way agents perceive and act upon their environment—inspired by mainstream information-processing artificial intelligence and cognitive science—that hinders the potential of the field. Simply, in mainstream contemporary cognitive science (highly influenced by artificial intelligence) perception is thought to be a kind of processing of poor stimulation to construct symbol-like representations (see Marr 1983, Rock 1983). The details of such a processing may vary across different theories, but the idea of internal symbol-like representations as the objects of perception is common to all of them. Under this lens, AR has been mostly developed by superimposing a virtual layer over the real world that requests to be processed along with the latter. In other words, the virtual layer adds more information to be processed and integrated with the real world's one. This tendency plus the use of symbols-to-be-processed as chief features of AR are what hinders its potential. Instead, our aim is to provide some guidelines for the exploration of an ecological trend in AR applications that we shall dub Ecological-Augmented Reality (E-AR).[1] As we shall see, E-AR is not about enhancing our perception of the real world courtesy of some virtual enrichment. It is rather about augmenting reality itself in a novel fashion. In E-AR, the virtual layer does not have meaning on its own, just in its *juxtaposition* with the real environment. So, E-AR augments the meaningful features of the reality itself and no processing or integration is needed. Then, our aim is to bypass the information-processing postulates that underlie current AR devices, and to submit for the reader's consideration ways to augment reality ecologically; an approach that will hopefully allow current technology to spread its full potential.

2. A bit of recent history of AR

The term "Augmented Reality", as such, was coined in the early 90s by Tom Caudell and David Mizell at Boeing.[2] These researchers replaced customized designed wiring instructions on boards with a head-mounted display that would effectively allow for the computer-based recycling of instructions for multiple flight-related purposes. However, it was not until 1997 when Ronald Azuma published the first study that relates such incipient processes of combining the virtual and the real as augmented reality, putting forward what has been endorsed as the canonical definition of AR:

[1] Our usage of the expression "Ecological AR" is not to be confounded with AR Ecology: the type of AR applied to the protection of the environment, the effective use of resources, and the like.

[2] Although the boost in AR only came in the late 90s, strictly speaking AR was born in the 50s, incidentally, in an era that predated digital computing. Later on, courtesy of new computer technology based on fast processing computers and graphics rendering techniques, combined with new techniques for real-time positioning, preceding mechanical devices such as Sensorama soon became obsolete (Carmigniani and Furht 2011).

> ... a variation of *Virtual Environments* (VE)... VE technologies completely immerse a user inside a synthetic environment. While immersed, the user cannot see the real world around him. In contrast, AR allows the user to see the real world, with virtual objects superimposed upon or composited with the real world. (Azuma 1997: 356)

In this first take, Azuma states quite clearly what will eventually become the essence of AR. Namely, the combination of real and virtual elements occurring in the user's senses. But such a definition still fails to identify accurately the differential characteristics of AR with respect to other technologies that, while exploiting combinations of the virtual and the real, would not fit the augmented reality bill. Think, for example, of special optical effects of the sort that the film industry regularly uses. In order to solve this demarcation problem, and to provide a comprehensive characterization of AR, Azuma proposes a number of core properties:

> ... we define an AR system to have the following properties: combines real and virtual objects in a real environment; runs interactively and in real time; and registers (aligns) real and virtual objects with each other. (Azuma et al. 2001: 34)

In this way, AR is characterized as a technology that enables the overlay of virtual information onto the real world in real time and that allows for user-based interaction. But, how does this combination of the virtual and the real take place? And above all, what sort of applications can we envisage?

There are several ways to create AR, although there are some differences that make them deliver different results and returns. The most commonly used AR displays are varying head devices. These devices are generally referred to as being head-mounted (HMD). An example of HMD is Minolta eyeglass, a holographic based display (figure 1). Within this category we can find from head-up devices (HUP; e.g. Microsoft HoloLens™; see also Newman 1995), to different kinds of glasses (e.g., EyeTap, see Mann and Fung 2002, Mann et al. 2005) or even contact lenses (as the iOptik system developed by Innovega). But AR can also be handheld, such as flat-panel LCD displays. All the mobile AR apps, for example, are of this kind (see Pokemon GO!, AR EdiBear, or AR Tower Defense). In addition, AR may be displayed spatially, as in the case of Spatial Augmented Reality (SAR; see Bimber et al. 2001), where virtual information is directly video-projected onto the object of interest with no need for eye-wear, or even displayed directly onto the retina as in virtual retinal display (VRD) technology (see Tidwell et al. 1995, Lin et al. 2003, McQuaide et al. 2003).

INSERT FIGURE 1 ABOUT HERE

The applications and devices out of these various implementations are multiple. As we noted before, from the world of entertainment to medicine, education, manufacturing and repair, navigation or military environment, you can find the most varied applications of AR systems. Consider the following for the sake of illustration. Blair MacIntyre and his team have developed games and applications that exploit the different capabilities of AR. Its aim is "[to] build games and study them, focusing on everything from interactivity and visualization techniques to the feel of game mechanics to the social experiences they foster." (Kroeker 2010: 20) Thus, they try to open up a new field of entertainment based on the combination of the real and the virtual, and try to assess its impact within the environments in which it operates.

The application of AR also exhibits great potential in fields such as medicine, or education, something that makes of AR a technology with the capacity to transverse social practices. An example of the application of AR in medicine is mARble (figure 2), a mobile AR LCD display app used by the PLRI MedAppLab group at Hannover Medical School that aims to achieve a realistic environment to instruct learners in topics such as forensic medicine (see Albertch et al. 2013):

INSERT FIGURE 2 ABOUT HERE

Another example, this time in the classroom, is Construct3D (figure 3; see Kaufmann & Schmalstieg 2002). It is a HMD device based on the Studierstube system (see Schmalstieg et al. 2002) and designed for mathematics and geometry education:

INSERT FIGURE 3 ABOUT HERE

We can see that the application of AR to the most diverse activities is anything but a pipe dream. There are rapid and continuing developments in order to achieve a profitable integration in such activities. But to appropriately understand how the combination of the virtual and the real does take place and, above all, the type of novel applications to be envisaged, we need to appraise more thoroughly the theoretical commitments of AR. In our view the development of AR systems to date presupposes, mostly, an information-processing theory of perception inherited from mainstream artificial intelligence and cognitive science. But AR systems developed under the principles of ecological psychology (Gibson 1979) may contribute to new and better levels of performance and adaptation to the user's perceptual abilities. Thus, in what follows, we consider, first, the theoretical underpinnings of mainstream AR, and, second, a promising way out of the pitfalls yet to be identified, by the hand of a full-fledged Ecological Augmented Reality (E-AR). Ultimately, we would like to convey the absolute relevance and interest that the evolution of this technology has for the cognitive sciences, writ large.

3. *Constructing* Augmented Reality and its shortcomings

One way to augment reality—effectively, the one and only way that has been pursued in a systematic manner thus far—is under the principles of the cognitive revolution of the 50s of last century; a paradigm marked by the explanation of perception and cognition in information-processing terms. Information-processing relies upon the poverty of environmental stimulation. It is precisely because the environment by itself does not contain sufficient information that it has got to be processed further before an agent can solve a cognitively demanding task or behave adaptively in the face of environmental contingencies. Bluntly, a Helmholtzian constructivist (Rock 1983) process of enriching an otherwise impoverished source of information is called for. A canonical example is the 3D reconstruction of 2D retinal images. In *Vision* (1982), David Marr proposes a quite complex way to explain how this reconstruction is possible. Technicalities aside, the process consists in applying some rules to the raw data given in the retina such that the 2D inputs of light on it are converted into 3D representations. The 3D images are a product of the processing of the luminal input on the retina. They are not the input itself, but mental constructions that have been built up by following a specific set of rules and that are symbolic-like entities insofar as they are inner-body entities standing for environmental objects and properties.

In this way, we can see that symbol manipulation and formal rule-following provide a straightforward way to augment reality. The agent perceives a blend of real and virtual components that requires to be elaborated upon, informationally speaking, before appropriate action or interpretation can take place. Just for the sake of illustration, consider a handful of examples—figures 4-6—of AR symbolic virtual overlaying.

INSERT FIGURES 4, 5 AND 6 ABOUT HERE

In all these devices, symbolic information is composited inferentially with the surrounding environment. In the first example, among the many features that the device offers, all kinds of information can be found (weather, social networking, dating, and navigation services, etc.). Information offered by *Google Glass*^{TM} such as the walking distance to the nearest Tourist Office, or which movie is being shown at the theatre, is often symbolic; from the symbols used to report the weather for the weekend to commands like "Turn right" or "Straight 100 meters" of the navigation system. We find a functionally analogous situation in *Townwear System*. It is a device that points open environments in order to serve as a compass to various means of transport via particular landscape markers (e.g., the Phillips Tower). Trademarks used are symbols such as names, locations, etc. In the sport-broadcasting illustration, symbolic moving labels are simultaneously broadcast in real time with their corresponding race cars. All of these are, again, symbolic information packages that the agent is supposed to be able to make sense of inferentially.

Generally, in AR devices, virtual symbolic information is superimposed upon the environment in such a way that the real and the virtual may be processed, informationally speaking, in tandem. Thus, we find both information in reality itself (the type of information that would still be there were we not to use an AR device), and virtual information (the type of information generated by the AR device and aligned in real time with the data pool from reality). We confront, therefore, two independent sources of information that need to be processed in parallel. This processing will likely involve building up mental representations that stand for the real data pool, doing the same for the virtual one, and, on top of that, connecting the two sources of information in a meaningful manner. It is easy to see how dealing with AR devices requires the use of a large amount of cognitive resources such as memory storage and retrieval, among others.

If anything, cases such as those depicted in figures 4-6 serve to illustrate that reality is being symbolically augmented at the expense of augmenting the information-processing load likewise. Superimposed symbolic information increases the processing demands upon the agent. Unfortunately, in our opinion, by increasing the burden of symbolic crunching, AR devices run the risk of saturating the user of the technology. Moreover, an AR user might encounter time constraints in response to highly sophisticated environments for which she has not been trained. This may result in massive overload because of the processing of information in tandem: inferential perception of the surrounding environment plus perception of the virtual layer superimposed upon it. In fact, the problem is greater since we must not only interpret each layer separately, but construct both in composition, experiencing a slowdown in troubleshooting. In the case of tasks with great requirements of attention and concentration by the subject (e.g., driving in a variegated environment) this may even prove fatal.

The problem that appears when AR is grounded on cognitivist foundations is ultimately a saturation of the cognitive resources available. By assuming an overall constructivist framework, it is assumed that the agent holds these capabilities ideally or without limit, and AR is based on increasing virtually superimposed symbolic information. But what happens if you put into question for various reasons such capacity?

The problem of symbolic information-processing in fact runs deeper. We can go further and ask what happens in the case of agents who are not in possession of full-fledged cognitive faculties. For example, what about infants or children who have not yet reached cognitive-inferential maturity? What about the elderly who due to aging or disease lack full-cognitive powers? How could these users have access to AR?[3] Thus, cognitive immaturity and degradation of cognitive faculties are two domains where the limitations of mainstream AR become more evident.

[3] The first surveys of volunteer testers of the early prototype of Google GlassTM reported headaches, difficulty in concentrating, or some degree of discomfort, generally speaking, that the use of the device produces: “It's disorienting. You're unable to focus on people or things around you ... Glass is headache-inducing too; you're more or less cross-eyed when focusing on something so close to your face." Hedge fund manager Eric Jackson also tweeted that he heard the same thing: "VC told me this week — who'd tried it and knows many people who have — Google Glass actually is not very good at the moment, gives big headaches.” (Yarrow 2013).

In the case of early ontogenetic stages where biolinguistic development is minimal, the access to AR devices is severely diminished. While it is true that so-called "digital Natives" acquire skills related to digital environments at increasingly early ages, we must remember that such settings do not constitute cases of AR, that is, alignment of virtual and real information in real time does not take place. A baby or toddler who is "fluent" in such environments is no better "equipped" than a *non-Native* adult when dealing with an AR environment. The reason is that usual devices are designed from the perspective of enriching reality by superimposing a virtual symbolic layer. Thus, pre-linguistic agents will have limited capabilities in these environments. Elderly or ill people also present a counterexample to orthodoxy in AR, since the access to these devices presupposes that they keep their information-processing skills and cognitive competence above a minimum threshold, below which their ability to navigate in an AR environment will also be severely diminished. However, these presupposed skills are not met in well-known processes of cognitive aging (see Hofer and Alwin 2008, Gariépy and Ménard 2010). Most of them entail, for example, some problems in terms of attention (McDowd and Hoffman 2008) and request for perceptual saliency in order to successfully interact with the environment (Passow et al. 2012, Passow et al. 2014, Lindenberger and Mayr 2014). It is easy to see how orthodox AR poses problems to these aspects of aging. The perceptual field is more crowded than usual when an orthodox AR device is used. There is a whole new layer of information added to the usual one. This leads to much more stimuli and elements to pay attention to, perceive, integrate, and so on. The consequent supersaturation negatively affects perceptual saliency—things are more difficult to detect,[4] attention—it is more difficult to focus in just the relevant information, and the task performance in general.

In view of all this, information-processing, we contend, cannot be the basis for the revolution in augmented reality that looms in the horizon. In what follows, we put forward a framework that

[4] At the first glance it may be unclear why a crowded environment negatively affects perceptual saliency. To illustrate the fact, we suggest the reader remember when she (likely) played Where's Wally-like games. It was really difficult to find him, among other factors (in that specific game the colors, for example, were determinant as well), because the images were really crowded. Wally was not perceptually salient. The problem of orthodox AR in terms of perceptual saliency, although less extreme, is isomorphic to the one in Where's Wally-like games.

eschews any type of information-processing regarding the nature of our psychological processes. Our working hypothesis is that an ecological paradigm can help circumvent the aforementioned problems.

4. Principles of Ecological Augmented Reality

4.1 Ecological Psychology

According to ecological psychology (Gibson 1979), we can find information in the environment that suffices to specify its features (Michaels & Carello 1981) and that is perceived in terms of affordances, i.e., as opportunities of interaction with the environment (see Turvey 1992, Chemero 2003). These two principles are orthogonal to the theoretical commitments of the information-processing paradigm. In particular, the concept of specification stands against one of the main assumptions of constructivism: the poverty of environmental stimulation. As we noted earlier, a basic tenet of constructivism is that the environmental information that a perceiver is able to access is inherently poor and ambiguous. This is the reason for postulating that information needs to be processed as a fundamental aspect of cognition. Insofar as the stimulus is poor and ambiguous with respect to the complex and rich environmental features, the perceiver needs to process it in order to make it relevant to her behavior. By contrast, Gibsonian environmental information is specificational. In his classical example (Gibson 1979), the luminal information present in the ambient optic array specifies the features of the environment to which it belongs. In this way, all those features of the environment that are relevant for our behavior are specified in this luminal information. In other words, the available environmental information, far from being poor, is rich enough and carries the specification of all the complexity we can find in the environment and everything we need to interact with it. For this reason, no further inferential processing is called for. Perception is not matter a matter of processing poor stimuli but of detecting rich environmental information.

Key to the specificational nature of environmental information are perceptual invariants. As a mathematical concept an invariant is a feature of an object or a system that remains the same when transformations are applied to the object or system. When Gibson (1966) refers to this concept in proposing perceptual invariants, he is talking about environmental features that remain the same when an organism is moving around and exploring her environment:

> Besides the changes in stimuli from place to place and from time to time, it can also be shown that certain higher-order variables of stimulus energy—ratios and proportions, for example—do *not* change. They remain invariant with movements of the observer and with changes in the intensity of stimulation… They constitute, therefore, information about permanent environment. The active observer gets invariant perceptions despite varying sensations. (p. 3)

Perceptual invariants are the base of specification. They are the stable features of the environment that are informationally relevant to the behavior of a given organism. Three main features of perceptual invariants are noteworthy. First, they are high-order variables. Namely, they do not correspond to basic variables of physics such as length, weight, shape, etc. On the contrary, they are such things as ratios, relations, degrees of texture, etc., as figure 7 illustrates, where the cobblestone space that the two bricks occupy respectively remains relationally invariant.

INSERT FIGURE 7 ABOUT HERE

A well-known example of a perceptual invariant is the moment of inertia (Solomon & Turvey 1988), a relational invariant that puts in correspondence muscular activity and the rotational motion of hand-held objects. As an object is manipulated, and despite muscular activity being in constant flux, the muscular torque needed to rotate the object (the moment of inertia) remains invariant (see Richardson et al. 2008, for the details). Second, detecting perceptual invariants requires action. Although they are features of the environment, the way in which organisms are able to detect them is by interacting with the environment. For instance, the way to detect the moment of inertia of a rod is to move it describing circles. Finally, the specificational character of perceptual invariants allows ecological psychologists to develop a naturalistic framework for perception, but also requires them to understand perception and action as two sides of the same coin. For this reason, perception is of affordances from the ecological stance.

Affordances are probably the most famous entities of ecological psychology. They have received much attention—and sometimes an equal amount of misunderstanding—from different fields and disciplines (cognitive science, design, architecture, etc.). But what does Gibson mean when he claims that perception is of affordances? Two dimensions must be considered to answer this question. First, as we have noted earlier, that perception and action are two sides of the same coin. Put bluntly, perception is the system by which organisms control their actions and, at the same time, through action organisms generate the environmental information needed for perception. In this sense, one process cannot be understood without the other. And second, that given this relation between perception and action, what is perceived is nothing but opportunities of interaction in a given environment. Perceiving consists in detecting the possibilities of behavioral interaction for the organism given the environment and the action loop in which it is involved. It could be interpreted that, according to Gibson and ecological psychology, perceiving consists of two different things: the detection of specificational environmental information, on the one hand, and the perception of affordances, on the other. However, these seemingly different processes are the same one. The features of the environment specified by specificational environmental information, i.e., perceptual invariants, are affordances. These features are not exclusively of the environment, but they are environmental features with respect to the organism, its biological features, and its actions—for example, climbability, walkability, eatability, etc. In this sense, the detection of specificational information is, in other words, the perception of affordances. Figure 8 illustrates a number of affordances as perceived by different organisms. Whereas a human adult may perceive, say, the throwability of the stone, a mouse might perceive, for instance, its climbability. As the figure shows, several different affordances are in principle perceivable as a function of the ecological scale of interaction of the organism in question.

INSERT FIGURE 8 ABOUT HERE

4.2 Ecological Augmented Reality (E-AR)

Orthodox AR only considers the semantic context of the real world so as to put the virtual component in semantic agreement with it. None of this makes sense from an E-AR perspective. First, it is not a matter of agreement between real and virtual semantics. Meaning takes the form of affordances, and (symbolic) semantics plays against the true immersion of the AR user and ecological interaction with its enhanced surroundings. But more importantly, meaning under E-AR only makes sense as the type of affordances that can only be perceived out of the reality being augmented, that is, affordances that the user would not be able to perceive neither in the real-world component, nor in the virtual one, considered in isolation.

Now, can the principles of ecological psychology be applied in a straightforward manner to AR? It is not easy to develop E-AR. To the best of our knowledge, there are no forthcoming examples of AR based on the principles of ecological psychology. The reader familiar with the literature may have thought of AR literature that somehow already points towards E-AR. But more work is needed before the two are brought together. Nishizaki (2015), for instance, reports development on an alleged ecological AR application for iPads and iPhones. Nishizaki considers the perception of affordances, contributing therefore to the enhancement of interaction, something also stressed by Carmigniani & Furht (2011). On close inspection, however, we can see that despite the affordance-jargon being employed, these approaches are not truly ecological. Nishizaki claims that his study aims:

> 1) to present an analysis of the affordances of infants' home environments and 2) to propose a new way of visualizing infants' affordances by creating an AR mobile application prototype for iPhones and iPads to provide parents and surrounding adults with better understanding. (2015: 582)

Figure 9 shows an example of AR of an infant's environment, as the infant interacts with cushions on a coach.

INSERT FIGURE 9 ABOUT HERE

AR of this kind aims to make explicit the type of affordances that infants may perceive in their home environment. But as it becomes clear from Nishizaki's discussion, this use of AR has nothing to do with E-AR as herewith understood. His aim is to make use of iPhones and iPads to help parents visualizes the type of affordances their infants may be perceiving. In this way, it is an aiding device for parents, not for the perceiver of the affordance herself. We understand the aim of E-AR to consist in providing *the user* of the technology with information that is specificational in the AR overlay itself. But here all we get is the *simulation* of affordances, providing parents with a visual rather than a verbal description of the way infants interact with objects in their surroundings. In this way, it simply amounts to saying: "Infants may perceive the throwability, or the jumpability-upon, of a cushion".

Ultimately, we are after the development of an AR paradigm that does not rely upon making more explicit the virtual information to be overlaid, but rather upon making the blend more ecological insofar as new opportunities that did not exist, new affordances, can be perceived, or existing affordances become more salient, in the blend itself. As noted above, according to the principles of ecological psychology, *perception is direct and of affordances*. In this sense, what we need from a AR device is to pose new affordances, or to enhance the perceptual saliency of existing affordances, without the need for information processing. In other words, E-AR displays should aid modify direct perception itself by generating new or highlighting existing affordances through the juxtaposition of the virtual and the real layers. Otherwise, we are only complementing environmental information with virtual cues, with the risk of contributing to the collapse of E-AR into mainstream AR.

Let us show the difference between orthodox AR and E-AR by means of an example. Suppose we want to develop an app to navigate towards our destination through relatively crowded environments. How would this app be implemented in terms of orthodox AR and E-AR? In the first case, an example of such an app is Google Glass' navigation app. Bluntly, it is a Google Maps-like app implemented in an AR device. In the virtual layer the user finds a zenithal-view map of the path, arrows marking the direction to aim, names of the streets, and other symbolic information (e.g., miles to go, cardinal direction, and so on). Namely, the user finds a mostly symbolic virtual layer added to the real environment, being forced to process both sources of information at the same time—e.g., by reading the name of the next street she has to walk through while trying to avoid hitting other people and objects in the environment. This is, thus, a clear example of orthodox AR and its high (and problematic) cognitive demands.

The example of such a navigation app in a E-AR device would be radically different. No single element of orthodox AR's virtual layer would appear in E-AR. No arrows, maps, names, or any other kind of symbol would be found in the E-AR app. Instead, a way to really augment the reality to lead the user towards the right place is to modify the affordances of the environment. In this specific case, the augmentation would consist in enhancing the *walkability* of some routes (the ones the user has to use to reach its objective) and blocking others (the ones the user has to avoid to reach its objective). It may be achieved by different virtual layers, but a simple one would be a layer which darkens or defocuses the routes to avoid. This would make the right route perceptually salient. Some features of a solution like this one must be highlighted. First, that the newfangled perceptually salient affordance is such due to the blend of the virtual and the real layers. The virtual layer itself would consist in some textures that either occlude or defocus the routes to avoid, leaving non-occluded or focused just the path the user has to follow. However, these textures would have no meaning by themselves; namely, they would just seem arbitrary textures if they were combined with a different real environment. They would be meaningful, in the sense of making an affordance (walkability) perceptually salient, by their superimposition with some specific real layer. And second, that such an augmentation will avoid the above mentioned problems posed by orthodox AR to elderly users or users in early ontogenetic stages (section 3). For example, as far as no information-to-process is added to reality, problems related to attention or perceptual saliency would be avoided. Actually, in a case like the one proposed, E-AR would pose a blending of virtual and real layers where the attention to perceptually salient affordances would be easier—only one perceptually salient affordance would be available to the user!

Our proposal is to develop AR devices ecologically in such a way that reality itself is literally augmented. It is noteworthy that we are not calling for a mere change in the technology, but rather in the way of making AR. We must rethink the kind of elements that are virtually superimposed upon the external environment. As a result, we contend, the user will be able to perceive brand new sets of affordances, or existing affordances in a new perceptually salient fashion, courtesy of information of a different sort altogether that remains specificational. Although, at the moment, there are no AR devices in the market that are designed ecologically, two fields of research may well bring inspiration to AR developers. These are the design and manipulation of real objects, and ecological research in the field of sensory substitution. To these we now turn.

5. Ecological design

In *The Design of Everyday Things* (1989/2002), and more recently in his *Living with Complexity* (2011), Donald A. Norman proposes some requirements for the intuitive usability of everyday products. Norman laments that technology and designs at our disposal were not more useful and simple to use. The key resides not in simplicity but in the very design of objects and in our environment; design that ultimately hinders rather than enables optimal interaction with our immediate environment. A classic example is doors that provide conflicting information (what has become known in the literature as *Norman doors*—figures 10 and 11, below).

INSERT FIGURES 10 AND 11 ABOUT HERE

How are Norman doors meant to be opened? Pulling or pushing? This kind of design with contradictory opportunities for interaction is found in hundreds of everyday objects (doors, taps, switches, kitchenware, automobile dashboards, and so on) and it is counterproductive to our daily activity.

Norman proposes to rethink the design of these objects in such a way that the information rendered is unequivocally specificational. In the case of the humorous vignette in figure 12, the aim is to design the button so that the physical properties of the object specify what to do with it, without colliding with the level of the symbolic information printed on its surface. In the case of Norman doors, the affordance perceived ought to invite us either to push or to pull. In an exclusively real environment, with no virtual involvement, this can be achieved, for example, by simply changing the shooter for a metal plate, by drawing the outline of a hand, etc. (figure 13).

INSERT FIGURES 12 AND 13 ABOUT HERE

Another example is the design of user interfaces based on affordances for interacting with computing devices (Sheridan & Kourtem 2006). The idea is similar to Norman's. It is illustrative that Norman speaks about ecological design that allows redirecting the highly technical complexity of the environment in which we find ourselves ("Good design can tame complexity") (Norman 2011). However, as the following comment evidences, Norman's proposal is not directed to AR, but only to the (re)design of physical objects:

> affordances, both real and perceived, play very different roles in physical products than they do in the world of screen-based products. In the latter case, affordances play a relatively minor role: cultural conventions are much more important. (Norman 1999: 39)

Actually, this comment does not even suggest potential applications of E-AR. By contrasting physical products to digital ones, Norman emphasizes the actual design of both products and not the new reality resulting from the juxtaposition of a virtual environment and the real world itself. Thus, our proposal can be seen as a continuation of this methodology, but raised to the level of AR.

Norman's ideas (see also Sheridan & Kourtem 2006), are productive insofar as the design of new objects is concerned, but what about already designed objects with which opportunities for interacting are opaque? Here, the potential of AR devices comes into play.

The information layer that AR devices superimpose upon reality should, in its very combination with reality, bring to light new affordances. Hence, physical elements that are not designed in such a way as to deliver specificational information could do so when appropriately combined with ecological virtual layers. It is in the combination where new affordances crop up. For example, a Norman door may become an ecological door by superimposing on it a texture that composites with the door itself, generating thus a new opportunity for behavioral interaction. In this case, by superimposing, say, a virtual image of a steel plate on the handle and the sign. A cursory reading of the ecological literature shows that the possibilities are limitless. Consider the amount of poorly designed everyday objects about which we don't have the slightest idea how they should be handled.[5]

The advantage of this proposal is twofold: on the one hand, in terms of the ability to operate a *customized cognitive science*; on the other hand, in terms of learning opportunities: to operate in a hybrid environment (real-virtual) allows us to customize opportunities for interaction. From a Gibsonian perspective, what matters is the ecological scale that we describe agent-environment interactions from. Thus, various agents may well perceive different opportunities for interaction. As we saw earlier, a human, a cat, and a mouse perceive different opportunities for interaction as they all confront a stone laying on the road. But taking the value of the relative ecological scale into account, the physical (re)design advocated by Norman, however good it was, may muddle the perception of affordances by different agents. This is evident from the previously discussed case of infants and the elderly. However, an ecological AR allows for the exploration of a bigger set of interaction possibilities than the one provided by physical design alone; design that although optimal is static and especially focused on an average, ideal, user.

It is in this sense that our proposed application allows different customizations on the same hardware to overlay an unlimited set of virtual layers, focused, not in themselves, but in their juxtaposition with the physical product, which may result is the direct perception of new affordances on a one-to-one basis.

[5] According to Norman (2011), affordances can be *signifiers* (i.e, perceivable affordances), but even if they are not signifiers they are fixed in environment, so, even if the agent does not realize them, affordances are there. The same happens with environmental constrains. When Norman talks about the re-design of everyday things, he has in mind the re-design of things in order to make more explicit their signifiers or to create new ones. The distinctive point of AR technology and our thesis is that with the juxtaposition of a virtual environment on a real world we can not only create new signifiers, but we can create new affordances or constrains and even delete some affordances or constrains we do not need anymore. For this reason, the scope of our thesis, led by the specificity of AR technology, is wider that Norman's.

In the case of agents with diminished cognitive abilities (e.g., an Alzheimer's patient), affordances offered by the composition of the virtual and the real would ease things up. It is common practice, for example, to find reminders in the domestic environment of an Alzheimer's patient. But those sticky notes must be processed inferentially, hence ultimately they run the risk of becoming likewise indecipherable to the gracefully, but continually degrading agent. However, an AR device that allows access to our daily chores without resource to complex inferential routines, may be very beneficial, and it can be customized for patients with various diseases that may require different approaching strategies.

A second advantage relates to the ability to *learn* new interaction possibilities (see Raja 2016). Agents are not only able to perceive different affordances already perceived by other agents, but to learn new routines. This may be due to the fact that the environment is constantly changing, or simply as a result of the greater degree of expertise we reach as we become familiar with certain tasks to the point of automatizing them. In the case of severe pathologies, the patient often must relearn skills partially or completely lost. Interestingly, although learning is sometimes presented as an insurmountable barrier to the ecological psychologist, in recent years ecological psychology has moved from *direct perception* to *direct learning* (Jacobs & Michaels 2007). According to the principles of direct learning, when an agent changes an informational variable for another the shift itself is guided *directly* (Jacobs & Michaels 2007). However, again, we have the possibility not to exploit a simulation environment,[6] but to ecologically augment the type of reality where such learning shifts take place. In short, the potential is to learn AR informational invariants. In the case of children, agents with immature inferential skills, superimposing an environment of explicit affordances over reality can have benefits on learning rate, and novel application possibilities in the educational environment.

[6] Jacobs & Michaels (2007) tested their hypothesis experimentally in flight simulation environments.

To recap, we propose the development of AR devices that combine the virtual with the real, generating a brand new collection of affordances that make information specificational in a way absent when considering either the real or the virtual in isolation.

6. Sensory Substitution as E-AR

Researchers are aware of the potential of AR enhancement of missing senses. Carmigniani and Furht (2011), for example,[7] point out that:

> AR could be used as a sensory substitution device. Hearing-impaired users could receive visual cues informing them of missed audio signals and sightless users could receive audio cues notifying them of unknown visual events. (2011: 39)

In effect, there is a literature in sensory substitution showing that these developments could be possible, but Carmigniani and Furht do not elaborate further beyond some preliminary remarks. They consider, for example, the use of an iPhone as a hearing device for impaired people by displaying visual information on the screen, or alternatively, as a seeing device for blind people by displaying auditory information. Hugues et al. claim:

> augmented reality could in this way augment some people's view in an unconventional manner for AR not by imposing virtual tags in the real environment, but by the use of audio tags. (2011: 42)

But Carmigniani and Furht's sensory substitution solutions do not go far enough. In fact, we are not sure in what sense their understanding of AR can be said to be ecological. The insertion of visual

[7] See also Hugues et al. 2011.

and auditory cues for deaf and blind people respectively is pretty much compatible with the orthodox type of AR we are trying to move away from here.

In what follows, we consider sensory substitution under a full-fledged E-AR. Sensory substitution was introduced in the 60s by Paul Bach-y-Rita (Baha-y-Rita et al. 1969) as a means of using one sensory modality to supply information usually perceived through another sensory modality. For instance, using touch or hearing to gather information usually perceived through vision. This process is carried out by using devices (Renier & de Volder 2013; see figure 14 for an example) that transform one kind of information (e.g., light) into another one (e.g., sound). In this sense, through a sensory substitution device, the 'substituent' sensory modality has access to an augmented array of information constituted by the usual array of information gathered by this modality plus—at least a part of—the informational array of the 'substituted' modality transformed by the device.

Given this framework, it is not difficult to see the resemblance between sensory substitution and augmented reality. On the one hand, in both cases we find a combination between a layer of natural information and a layer of virtual information. Natural information is the real world in the case of augmented reality and the usual array of information accessible by the substituent sensory modality in the case of sensory substitution. Otherwise, the virtual information is the layer generated by the augmented reality device in the first case and the transformed array of information of the substituted sensory modality generated by the sensory substitution device in the second case. On the other hand, sensory substitution meets the other core properties of augmented reality proposed by Azuma as well: the combination between both layers happens in real time and both layers are aligned with each other. The only difference between them is that virtual information is a transformation of natural information in sensory substitution, while in augmented reality any other kind of information may be transformed. However, there are studies under the principles of sensory substitution that use non-natural information as virtual layers. This sort of process has been coined, curiously, sensory augmentation (see Nagel et al. 2005, Kärcher et al. 2012), making even more explicit the relation between the processes of substitution and augmentation.

The resemblance between sensory substitution and augmented reality, we claim, is not trivial. Our point here is not just that the former may be seen as a subset of the latter, but that sensory substitution is already a kind of E-AR. Namely, that the information we perceive while using a sensory substitution device only makes sense insofar as it is a blend of the two layers. The virtual array of information does not make sense in or by itself. Let us examine two vision-to-haptics devices, the enactive torch (Froese et al. 2012; see figure 14) and lower leg device (Lobo et al. 2014; see figure 15), to illustrate our claim. Before, though, a caveat is needed. These two devices are cases of special-purpose sensory substitution and not general-purpose sensory substitution (for a review on the differences between these two ways to develop sensory substitution devices, see Loomis et al. 2012). Thus, our claim is that special-purpose sensory substitution as a kind of E-AR. It is not impossible, however, to claim that general-purpose sensory substitution also meets the requirements to be understood as a kind of E-AR (the blending of virtual and real layers is present as well in devices of this kind). Nevertheless, given the current state of the research on sensory substitution and the doubts concerning the feasibility of general-purpose devices, the former claim seems more adequate to us.

INSERT FIGURES 14 AND 15 ABOUT HERE

The enactive torch is a special-purpose sensory substitution device that can be handled as a flashlight. The torch is equipped with distance sensors and a vibrator that can be strapped to the wrist. The device is really easy to use: a close environment may be explored by steering the torch towards any direction and once it faces some obstacle—i.e., an object in the direction of exploration—the vibrator activates. The features of the vibrations depend of the object explored and the pattern of exploration. The device vibrates while the distance sensors are facing the object, so by exploration it is possible to know its width, length, and so on. Also, the closer the object is, the more intensely the device vibrates. The virtual layer generated by the enactive torch consists in these vibrations that are vision-to-haptic transformed information about distance and other features of the obstacle. The enactive torch is then conceived as a 'virtual cane' which could serve blind or blindfolded people to navigate through a crowed environment thanks to the new information gathered by the haptic system. But why we do claim that it is an E-AR device?

In order to count as a kind of E-AR the meaningful information accessible to the perceiver while using a device of this kind must be the resulting blend of the real with the virtual layer. This means, as we have noted, that the information of the virtual layer must not be meaningful by itself and, actually, this is exactly what happens with the enactive torch. Vibrations (the virtual layer) depend on the pattern of exploration, that is, they depend on how the perceiver moves around the environment, on how she moves her hand, on how she approaches the obstacle, etc. But the pattern of exploration depends on all the other haptic information as well. Exploration is possible just if we are aware of the position and movements of our body—i.e., it is possible, among other things, because of haptic proprioception. We may say that vibrations depend on proprioception. Moreover, in order for specific patterns of vibrations to make sense, they have to take place over specific patterns of proprioception. For instance, the height of an obstacle is perceived because the torch vibrates while the user moves her hand and arm from the bottom to the top of the obstacle. But, at the same time, she is perceiving her own movement and, crucially, the blend between the virtual and the real information. Indeed, the same pattern of vibration would make no sense to her were she not to move her arm—probably, she would take this vibration as an error in the device. Moreover, the same pattern of vibration could stand for the width of the object if the movement (and the perception of the movement itself) were different. These facts lead to a conclusion: vibrations are not meaningful by themselves, but only in combination with all the other haptic information. The blend of both layers of information is what is meaningful. Thus, under our definition, the enactive torch happens to be an E-AR device.

A similar case is the lower leg device (Lobo et al. 2014), a vision-to-haptic special-purpose sensory substitution device that can be strapped to the lower leg. Like the enactive torch, it consists of sensors and vibrators. In this case, a light sensor and an array of vibrators. While using it, as the perceiver approaches a step a pattern of vibration is expanding through the array of vibrators. This pattern specifies different features of the step (mainly height), so the perceiver is able to step on it. Again, it could be used to help blind or blindfolded people.

The lower leg device is likewise an E-AR device. The pattern of vibration generated by the array of vibrators, although more complex than the one generated by the enactive torch, remains meaningless unless blended with haptic proprioception. The lower leg device has some technical advantages insofar as it is able to generate more complex patterns of information—especially, thanks to the array of vibrators, as the perceiver approaches an object the optical expansion that happens in vision can be transformed in haptic expansion. However, these patterns still depend on the pattern of exploration and the movement of the perceiver. The rate of expansion of the step in the lower leg will depend on the approaching velocity towards the step, the feeling of the top of the step will depend of the relative position of our lower leg with respect to the step, and so on. And, again, our proprioception is crucial for these patterns of vibrations to be meaningful to us.

As a last remark, it is worth to note that the underlying principles in these two devices are already present in some commercial ones. Lechal Shoes™ implement a system of navigation by producing vibrations on the user's feet. They are able to guide us towards our destination by changing the pattern of vibration in our feet. Like in the two cases above, the meaning of these vibrations is depending on our movement—whether we are walking forward or backward, etc. So, perhaps we could take Lechal Shoes™ as the first commercial E-AR device ever.

6. Conclusion

In this article we have introduced Ecological Augmented Reality (E-AR), an approach that questions the tacit theoretical assumptions of mainstream AR. In the past, effort has been put into making AR technology “socially acceptable, natural to interact with, and fashionably acceptable” (Carmigniani and Furht 2011: 21). However, narrowing AR to interfacing challenges to be solved by smart engineering prevent us from appreciating the potential that lies ahead, were we to reconsider what exactly is after all being augmented in AR under ecological principles. In our view more effort should be put into mulling over the whole framework root and branch. However, as computer vision, object recognition, and the like advance, high-tech overlays will still be an add-on extra unless the very notion of perception is understood, ecologically, as the detection, and not the processing, of information.

A further word of caution is needed as we should not be carried away by the visual metaphor. It is true that the easiest way to conceptualize AR is from visual devices that overlay 3D graphics and renderings on our view of the real world. It is also true that visual perception is the field with the best AR development. But we must not forget that perceptual changes introduced by AR can be applied to any other sensory modality. Within AR quarters, it is commonly claimed that the role of AR will end up being functionally equivalent to the role customarily played by the senses, especially sight, hearing and touch, but also smell and taste. In a sense, AR furnishes us with yet another additional sensory *organ*. Thus, overlaying, say, virtual smells and tastes to the real world is considered part of the possibilities of AR systems. If AR is, so to speak, a child of the 21st century, E-AR has not even been born yet. As we saw in the previous section, both in the enactive torch and the lower leg device, it is easy to see how the relevant meaningful information for the perceiver is in the blend of the real and the virtual layers, and not in the virtual layer by itself. However, although such special-purpose sensory substitution devices already meet the requirements of E-AR, the technological development of AR devices opens up a wider range of implementational possibilities in the most diverse forms and systems.

Among others, innovation focuses upon research fields such as education, medicine, entertainment, of course, and iPhone applications. The engineering emphasis has to do with things like reducing parallax error, latency, calibration, etc., or resolving problems of occlusion between the virtual and the real, or on display resolution, field of view, and the like, or on marketing worries (lowering prices, making AR technology easier, lighter, to carry or wear, less power consuming, etc.). In our view, despite the fact that research, for example, in mobile AR is boosting (see also Azuma, Billinghurst, and Klinker 2011; Wagner and Schmalstieg 2009) innovation should focus not in the research field itself, but in the very cognitive scientific approach. E-AR makes sense because (i) the surrounding environment is structured; (ii) the user of the technology can acquire knowledge of her surroundings by perceiving-acting in the world (to perceive is nothing other than to detect structure); and (iii) perception can be fleshed out as the pick-up of information (the user detects the relevant invariant structure that affords specific actions by moving around). Agents don't have an experience of objects in the world, but of the relational properties derived from the world. Relational properties themselves are perceived. E-AR is meant to allow the user of the technology to engage her vicinity from different orientations in order to extract the invariant structure itself. The purpose of E-AR is to enable display users to directly perceive possibilities of interaction that couldn't possibly be perceived prior to the real environment having been ecologically augmented. Augmenting reality, we contend, is equivalent to constructing a niche: altering the environment permits the pick-up of new affordances. We alter the environment to better adapt to it. E-AR creates new properties that provide affordances relative to the agent. Developers are welcome to explore these ideas, and, in turn, the ecological blend.

Acknowledgments

We are grateful to Frank Faries for his helpful comments and to the audiences of 'iCog 2 (2014): Perspectives on Learning' at the University of Edinburgh and '5th Global Meeting (2015): Virtual, Augmented, and Mixed Reality: Changing the Face of Learning' (Dubrovnik, Croatia), where VR presented some of these ideas, for their helpful comments and suggestions. Research supported by Fundación Séneca-Agencia de Ciencia y Tecnología de la Región de Murcia, through project 11944/PHCS/09 to PC.

Figure 1. Minolta eyeglass display with holographic element. Courtesy of Hiroaki Ueda, Minolta (from Azuma et al. 2001)

Figure 2. AR applied to Medicine (from Albertch et al. 2013)

Figure 3. AR applied to Education (from Kaufmann & Schmalstieg 2002)

Figure 4. Wearable AR, such as Google's Glass, can give users an extra layer of information added onto to their view of the real world (from Edwards 2013).


TOURIST
INFORMATION
20m/0.012miles
Play: The Woman in Black
CYBER CAFE
Cinema: 2 Screen
Showing Scream (18)
Saw XI (PG)

Figure 5. Motion- stabilized labels annotate the Phillips Tower, as seen from two different viewpoints (courtesy HRL Laboratories). From Azuma et al. 2001.


Phillips Twr

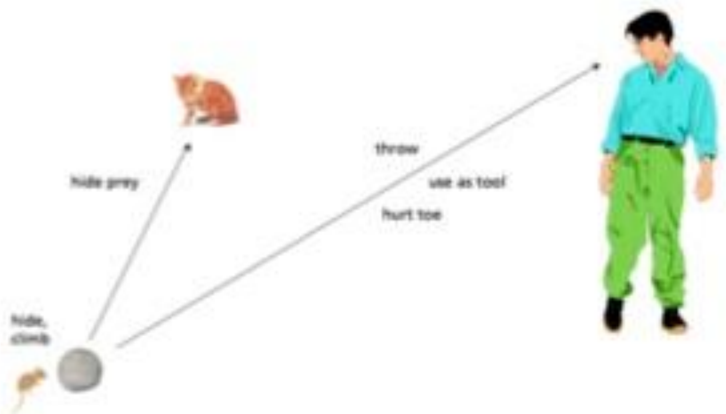

Phillips Twr

Figure 6. AR in sports broadcasting. The annotations on the race cars … are inserted into the broad- cast in real time. (Courtesy of NASCAR and Sportvision, top and bottom, respectively) (from Azuma et al. 2001).


101 To Go
90 - Stricklin
33 - Nemechek
2ND
149 MPH
MARLIN
3RD
153 MPH
10
BENSON

Figure 7. Example of Informational Invariant: The space occupied by the base of both blocks is the same with respect to the cobblestone on which they are. That is an invariant that specifies reality. (from Goldstein 1981).

Figure 8. Example of different affordances that various actors see in a rock (http://epi-thinking.org/AFFORDANCE.html Accessed: 23/05/2013 - 18.57H)



hide prey
throw
use as tool
hurt toe
hide,
climb

Figure 9. Example of AR animation of an infant (age 7 months, 12 days) with a cushion (from Nishizaki 2015).

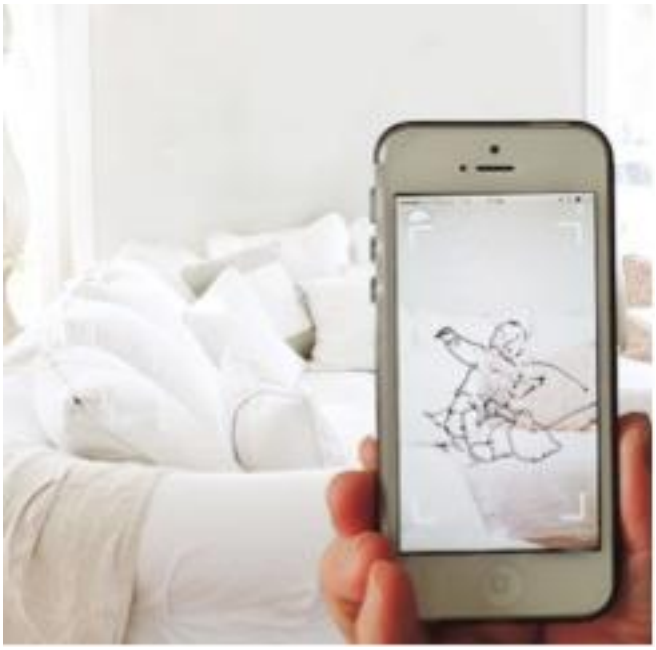


Figure 10. Norman door (Norman 1989/2002)


EXIT
EXIT
PUSH

Figure 11. A poor design, according to Norman (1989/2002)

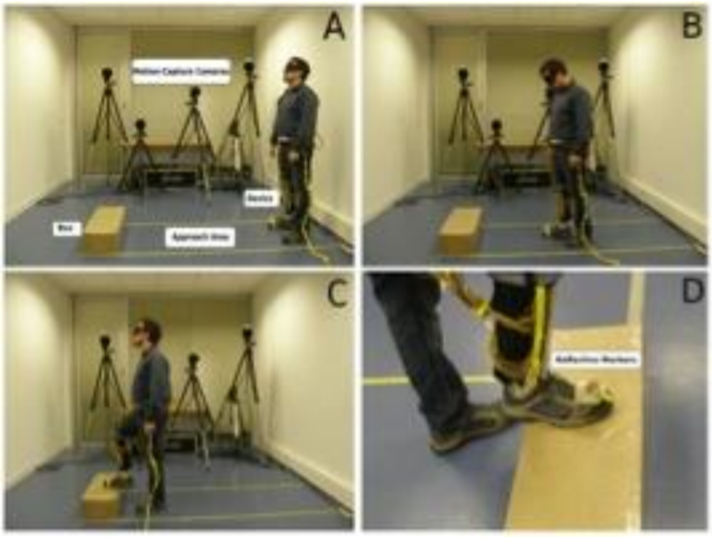
Figure 12. A button should not be pressed while it cries out to be so (in http://www.officeplayground.com/Do-Not-Press-Sound-Button-P3052.aspx) Accessed: 23/05/2013 – 18.54H).


DO NOT
PRESS THE
BIG RED
BUTTON

Figure 13. A door affording 'pushability'.

Figure 14. Enactive Torch (http://techcrunch.com/2014/08/13/the-enactive-torch-is-a-sensor-that-helps-the-blind-see/).

Figure 15. Lower leg device (Lobo et al. 2014).


A
B
C
D