How to Build Consciousness into a Robot: The Sensorimotor Approach

J. Kevin O'Regan

Laboratoire Psychologie de la Perception, CNRS Université Paris Descartes, Centre Universitaire des Saints Pères, 45 rue des Saints Pères, Paris 75270 Cedex 06, France
kevin.oregan@psycho.univ-paris5.fr

Abstract. The problem of consciousness has been divided by philosophers into the problem of Access Consciousness and the problem of Phenomenal Consciousness or "raw feel". In this chapter it is suggested that Access Consciousness is something that we can logically envisage building into a robot because it is a cognitive capacity giving rise to behaviors or behavioral tendencies or potentials. A few examples are given of how this is being done in current research. On the other hand, Phenomenal Consciousness or "raw feel" is problematic, since we do not know what we really mean by "feel". It is suggested that three main properties are what characterize feel: the fact that feels are different from each other, that there is structure in these differences, and that feels have sensory presence. It is then shown how, by taking the sensorimotor approach [24], [27] it is possible to account for these properties in a natural way and furthermore to make counter-intuitive empirical predictions which have recently been confirmed. In conclusion it is claimed that when we take the sensorimotor approach to feel, building raw feel into a robot becomes a theoretical possibility, even if we are a long way from actually attaining it.

1 Introduction

Consider a robot programmed so that it *acts* in every way as though it is conscious. For example when injured, it screams and shows avoidance behavior, imitating in all respects what a human would do when in pain. The robot is able to talk about its pain, and it reasons and acts like it has the pain. The philosopher Ned Block would say that the robot has *Access Consciousness* to the pain [6].

However all this would not guarantee that to the robot, there was actually *something it was like* to have the pain. The robot might simply be going through the motions of manifesting its pain: perhaps it actually feels nothing at all. Something extra might be required for the robot to *actually experience* the pain, and that extra thing is *raw feel*, or what Ned Block calls *Phenomenal Consciousness*.

2 Access Consciousness

From a theoretical standpoint (although currently no one has actually done it), there would appear to be no logical obstacle to implementing Access Consciousness in a

M. Lungarella et al. (Eds.): 50 Years of AI, Festschrift, LNAI 4850, pp. 333–347, 2007.

robot: the reason is that Access Consciousness ultimately corresponds to a behavioral capacity. What we mean when we say someone has Access Consciousness to something is that the person currently knows that he (considered as a person with a self) is poised to make use of that thing in his ongoing rational decisions, in his planning, intentions and linguistic behavior [6]. Agreed, the notions of "self", "rational", "decision", "planning", "intention", and "language" required to have access consciousness are all difficult notions. We are far from understanding these notions, and once we do, building them into a robot may require as yet undiscovered principles. But the important point is that there is no logical impossibility preventing this from being done: it has been termed the "easy" problem of consciousness [8]. Indeed, as the following illustrations show, cognitive scientists and artificial intelligence researchers are busy analyzing the components and prerequisites necessary to achieve this goal.

2.1 The Self, Intentions, and Theory of Mind

One critical aspect of Access Consciousness that has to be understood is the notion of self. Studies in cognitive science reveal that the notion is not a unitary notion, but is an umbrella term, covering capacities going from the individual to the social, and going from knowledge about only the organism itself, to knowledge about other organisms and their motivations [13], [22], [32], [33], [47]. Different aspects of the self become established at different times as humans grow up, with social pressures and individual experience contributing to their development in complicated ways. The notion of self is related to "intentions" and to "Theory of Mind", that is, the ability to understand other agents' thoughts and goals. The following are just a few illustrations where current robotics research is attempting to implement some very simple aspects of the self.

Self-discrimination has been investigated with Domo, a robot constructed at the Humanoid Robotics Group at the MIT Computer Science and Artificial Intelligence Laboratory (CSAIL). The robot consists of an upper torso equipped with moveable eyes, head, arms and grippers. It uses vision-based movement detection algorithms to determine whether something is moving in its visual field. It checks whether by commanding movements of its own body, the movements it sees are correlated with the movements it commands. If such a correlation occurs, it assumes that what it is seeing is part of its own body. In this way it is able to figure out what its own hand looks like, and later, what its own fingers look like [11].

Work on the higher notions of self, namely *self-knowledge* and *knowledge of self-knowledge* is being done using the COG platform, also developed at CSAIL. COG is actually one of the first robotic platforms that was built at the Humanoid Robotics group, and one might say it is approaching "retirement". It is an upper-torso humanoid robot equiped with visual, auditory, tactile, vestibular, and kinesthetic sensors, and which can move its waist, spine, eye, head, arms and primitive hands. COG has been used by a variety of groups at MIT to do experiments in object recognition, tactile manipulation, and human-robot interaction. Since the development of COG, many groups throughout the world have been constructing similar devices to study embodied cognition.

The ideas being used to study the emergence of the higher notions of self in COG and similar robots are based on analyses of what psychologists consider to be the most basic capacities postulated to underlie this notion in humans [39]. One such basic capacity is the ability to locate and follow an agent's gaze direction. For this, skin, face and eye detection algorithms in the robot's visual system allow it to locate eyes and infer where a human is looking. By extending this gaze following capacity, the researchers hope to implement algorithms for joint attention, that is, algorithms that allow the robot to attend to an object that another agent is also attending to.

A further example of a capacity that might be involved in the genesis of the self is the ability to distinguish mechanical motion due to inanimate objects and animate motion due to living agents like animals and humans. This is perhaps the basis of the notion of the ability to ascribe intentions and goals to other agents. To test this idea with COG, an algorithm has been used that estimates the variability in the velocity of moving objects. Presumably an object whose velocity does not follow simple laws of physics probably has a "will of its own" and is likely to be animate.

Another robotic platform that is being used to investigate the emergence of the self is Sony's domestic dog robot, the AIBO. At Sony Computer Science Laboratory in Paris the AIBO has for example been used to study joint attention and pointing, except this time in a "social" context, that is, with another AIBO or with a human [18].

The Domo, COG and AIBO projects are just three samples of work in progress. They are only painfully preliminary steps towards implementation of different self notions in a robot. But such studies in developmental robotics are a growing research field in which researchers attempt to show how from a few basic capacities, robots can acquire the social skills that we know humans acquire over the first few years of life, skills that are at the root of humans' notions of intention and "Theory of Mind" [1], [12], [19], [21]. Much work needs to be done, but the vitality of this and related research projects shows that researchers are confident that providing robots with a realistic notion of self and accompanying Theory of Mind is an achievable goal. Even if it takes many more years, the problem of building a robot with a self seems *in principle* solvable.

2.2 Language

Whether beings without language are conscious will probably have to remain a matter of debate. But it is obvious to us humans that insofar as we possess the faculty of language, it is an important component for Access Consciousness: after all among the things we *mean* by having conscious access to something is being able to talk about it.

However the goal of providing artificial agents with human-like natural language understanding is still far from being attained [10], [17]. The main problem seems to be in anchoring the symbols used by machines in the real world. One attempt to do this is progressing by accumulating vast amounts of "common sense" knowledge from large natural language databases like the web [20] (see http://www.cyc.com). More recently researchers are trying to physically embed artificial agents in the real world in order to it facilitate proper human-like use of language [34], [35]. As was the case for the notion of the self, work at MIT and Sony CSL is also illustrative of this.

An example of how immersion in the real world can help solve problems is Ripley, a kind of robot dog from the MIT Media Lab. Ripley can move its neck and pick

things up with its mouth. Because Ripley is embedded in the real world, it does not need to do any very complicated reasoning concerning how it is physically placed with respect to the objects it is dealing with, and how they are placed with respect to the person it is talking to: this kind of information is available at any moment in front of its eyes, so when someone says "pick up the one on your left", it can just look over on the left and find what is being referred to. Furthermore, when it learns words like push, pull, move, shove, light, heavy, red, hard, soft, it can make use of information it obtains from interacting with objects in order to ground the meaning of the words in physical reality, imitating what probably happens when real infants interact with their caretakers [36], [37].

A similar project is being undertaken at Sony CSL, where Sony's robot dog AIBO learns the meanings of simple words by interacting with a human [44], [45] (see also http://www.csl.sony.fr and http://playground.csl.sony.fr/). Other work at Sony CSL is investigating how word meaning and syntax can emerge when humans or robotic agents play language-oriented games together in order to achieve common purposes [42], [43]. Of course in these examples the interactions between robots and humans is much more focussed and the number of utterances involved is much more limited than in normal human interactions. But this work suggests that we may be starting to model human language acquisition in a plausible way.

2.3 Conclusion on Access Consciousness

Though the illustrations in the preceding paragraphs are obviously ridiculously simple, and are clearly only the very first steps towards implementation of Access Consciousness, they nonetheless suggest that Access Consciousness, though a difficult problem, can be decomposed into a collection of simpler problems which are logically not beyond the bounds of robotic implementation. Access Consciousness is ultimately an aggregate of behavioral capacities, and the necessary ingredients, while out of reach today, can conceivably be achieved in the future. Perhaps the most tricky problems are on the one hand the notion of "self", with its accompanying concepts of intention and Theory of Mind, and on the other hand natural language understanding. The hope today is that these problems will be successfully dealt with when researchers start working more with actual physically embodied agents in real world settings. Indeed, one cannot neglect the fact that humans live, move, and interact in particular ways with the objects that they use and with other humans in the real world. Human language and thought are not just raw symbol manipulation: the concepts that are manipulated are constrained by the physical world and the particular way humans interact with it. People have bodies and interact with other people who themselves also have bodies. People live in a shared social environment and have desires, emotions and motivations that play an important role in conditioning communication. It may be that only machines that have human-like immersion in the world will be able to have notions of self and use language like humans [10], [29].

3 Phenomenal Consciousness or "Feel"

We have seen that Access Consciousness, though clearly difficult, is not a logically insoluble problem. We can hope that we will gradually progress towards its

implementation in robots. With *Phenomenal Consciousness* however, we are in quite a different situation. People are convinced that they *feel* things, but it is hard to say exactly what feel is.

To try to understand better what is meant by feel, consider what happens when I look at a red patch of color: I see red. What exactly is this feel of red? What do I *experience* when I feel the feel of red?

One aspect of the feel of red is the **mental associations** that I have with red: Redness is associated in my mind with, among other things: the word "red", with roses, ketchup, blood, red traffic lights, stopping, anger, and certain red cough-drops... But these mental associations are *in addition to* the raw sensation: they are added over and above the experience of red itself. Having these mental associations might of course produce *additional* experiences: for example the association with anger may make me more likely to get angry. But such effects, if we want to call them experiences, would appear to be *in addition* to the basic, core, raw experience of red itself.

Another aspect of the feel of red may be the automatic **physiological states or tendencies** it creates. For example, red may be a color that has the direct effect on my nervous system of making me more excited, whereas blue may calm me down. The existence of such effects is controversial, but if they do exist, they are surely *over and above* the actual raw feel of the redness of red: in this example they correspond to excitement, not to red.

Yet another aspect of the feel of red is the **learnt bodily reactions** that redness may engender, caused for example by habits that are associated with red: for example pressing on the brake at red traffic lights. But again, such bodily reactions are add-ons to the actual raw experience of red.

To summarize: experiencing red may be accompanied by various mental associations, physiological tendencies and bodily reactions. These are extra behaviors that *come with* the feel of red, and they may in turn produce their own, additional experiences. But at the root of the feel of red there surely must more than simply mental associations, physiological tendencies and bodily reactions. This extra component, which we could call the "raw feel" of red itself, is presumably what makes red quite different from green, or from the sound of a bell, or any other sensation.

Now it seems clear that if we wanted to build a robot that experienced sensations as humans do, then at least conceptually, building in the *additional components* accompanying the sensations poses no particular problem. This is because these components are behaviors or behavioral tendencies or capacities. One could fairly easily build into the robot a higher probability of saying "red" when it sees red; one could have the robot be more active or agressive in red rooms, and even have it make subliminal brake-pressing movements.

But what could be done to provide the robot with the "core" component of the red sensation, namely the raw red feel? This seems to be a much harder problem. What "circuit" should be added into the robot to provide it with the *raw feel* of red?

The robot played by Arnold Schwarzenegger in the film "Terminator I" is a good example of this problem. The designers of the robot could incorporate circuits that make the robot wince, say "ouch", and otherwise manifest its disapproval in cases where for example it gets its arm chopped off. The robot could know that this kind of

injury is a *bad thing* for it, and it could be programmed to avoid getting into such nasty situations in the future. But what extra circuit would the designers have to build into the robot so that it actually felt the *raw pain itself*, instead of just going through the motions of feeling the pain?

3.1 Three Properties of Raw Feel

Let us look more closely at the raw, core component that is at the basis of feel. In this exercise we are purposefully leaving aside all the "extra" components like the mental associations and bodily manifestations that might come with feel.

There are three important aspects to note about raw feel.

First, *raw feels are different from each other*. For example there is red, green, pink, black. There is the sound of a tractor, of a violin, of middle C, of the wind in the willows. There is the smell of lemon, the taste of onion, the touch of a feather, the cold of ice, among innumerable others.

Second, *there is structure in the differences*. Sensations can be grouped together according to their similarity. For example sensations of light form a collection which is separate from sensations of sound, which are in turn different from sensations of touch, etc. Within each such collection or "modality" there may be further structure. Tones for example can be compared and contrasted, and they form a linear order going from low pitched to high pitched. Color is more complex, since one can distinguish the hue or tint of a color, and its "saturation", that is, the intenseness of the color it contains (a color is less saturated when it contains a lot of grey). Furthermore the dimension of hue is circular rather than linear: you can arrange colors in a closed circle of similarity going from red to orange to yellow to green to blue to purple and back to red again.

Sounds are another complicated case. Clearly loudness is something that can be defined along a linear dimension, but then sounds also have "timbre", which seems not to be describable in terms of dimensions that can easily be agreed upon. Smells also are complicated, and no consensus has been reached on a set of dimensions to describe them. A recent study suggests that a minimum of 30 independent dimensions are needed to account for smell judgments.

But whereas sensations form sensory orders [9], [48] or "modalities" within which they can be compared and contrasted in this way, across such modalities they cannot. For example, how is red different from middle C? Or how is cold different from onion flavor? It seems to make no sense to try and compare Red and middle C, since they have nothing to do with each other, and can't really be compared at all. Perhaps there is one attribute which is common even across different modalities, namely intensity: we can talk about intense colors, intense sounds. But apart from that sounds and colors are incommensurate (except perhaps to synesthetes!). The same goes for cold and onion flavor.

In summary, sensations form more or less separate modalities across which making comparisons is impossible. But within each such modality there may be a structure, which may be of varying complexity, depending on the modality.

Third*, raw feels have a quality rather than no quality, and are perceptually present.*

To understand this statement, consider the fact that the brain continually monitors blood oxygen and carbon dioxide, keeps the heartbeat steady and controls a variety of other bodily functions. All these activities involve sensors signalling their measurements via neural circuits and are processed by the brain. Yet one does not feel them, whereas one does feel the redness of the light or the prick of the needle.

Why should brain processes involved in processing input from certain sensors (namely the eyes, the ears, etc.), give rise to a felt sensation, whereas other brain processes, deriving from other senses (namely those measuring blood oxygen levels etc.) do not give rise to a felt sensation?

A related, but not identical case is thinking. Thinking obviously involves brain processing like controlling the oxygen level in the blood, and like analyzing inputs from the sensory systems. But does thinking have a feel?

Clearly, like the situation for sensory inputs, one is *aware* of one's thoughts. One knows what one is thinking about, and one can, to a large degree, control one's thoughts. But being *aware* of something is not the same as feeling something. Indeed, thoughts are more like blood oxygen levels than like sensory inputs: thoughts are not associated with any kind of sensory presence. Thoughts may be *about* things like blood and red traffic lights and red cough drops, or even about the raw feel of red, but such thoughts do not *themselves* have a red quality or indeed any sensory quality at all. Thoughts may of course be accompanied by feels: the thought of an injection makes me almost feel the pain and almost makes me pass out. But the pain I feel is the sensory pain normally associated with the injection, not the sensory quality of thinking. The thought ***itself*** has no sensory quality.

3.2 Neurophysiological Explanations for Feel

We have concluded that there are three important aspects of raw feel: feels are different from each other, there is structure in the differences, and feels have a quality and sensory presence, rather than no quality.

One's first impulse in seeking for an explanation for these facts is to look in the brain.

Neuroscientists have certainly localised different brain areas which seem to be involved in consciousness, but to date no explanation of how any such areas contribute is in view. Many hypotheses are entertained and discussed in the literature, such as the possibility that consciousness is generated by recurrent activation in corticothalamic networks, or by widespread synchrony of oscillations in the gamma band, or even that consciousness could be linked to quantum gravity effects in neuron microtubules. Such mechanisms might account for the behavioral capacities involved in Access Consciousness, but how any such mechanisms could explain why feels have the properties that they do is never addressed.

It would seem that there is a logical problem: whatever mechanism is invoked to generate consciousness, additional "linking hypotheses" will always have to be made: a linking hypothesis is a hypothesis that establishes a link that justifies, for example, why different neurons or neural mechanisms or firing patterns or quantum mechanisms should produce the particular different sensory qualities, with the particular structure of similarities and differences that is found, and with the

experienced sensory presence. The problem is that there would appear logically to be no non-arbitrary way of making such links between the neural states or physical characteristics of the firing patterns of neurons, and the experienced sensations.

4 The Sensorimotor Approach

A possible alternative way to understand the problem of Phenomenal Consciousness of feel is the sensorimotor approach [24], [25], [27]. This starts from the postulate that looking for a circuit or mechanism that generates Phenomenal Consciousness is to make what the philosopher Gilbert Ryle called a "category mistake" [38]: Phenomenal Consciousness is simply not the kind of thing that can be generated at all. Just as it makes no sense to search for the meaning of a word in the shapes of the particular letters that compose it, it makes no sense to search for a circuit that generates Phenomenal Consciousness in the brain.

Instead, the sensorimotor approach suggests that what it really means for a person to have Phenomenal Consciousness or feel is that the person:

1. is currently engaged in exercising a certain sensorimotor skill, and
2. is attending to this engagement and the skill's qualities.

Under this approach, the *quality* of a feel is constituted by the particular laws that govern an individual's sensorimotor interaction when he is experiencing the feel.

To understand the idea, one can take as analogy the feel of driving a Porsche as compared to driving a Volkswagen. Where lies the essential difference between the feel of Porsche driving and the feel of Volkswagen driving? It comes from the mode of sensorimotor interaction you have with the cars. It comes from *the things you can do* and the *way the car reacts when you do them*. When you press on the accelerator, the Porsche *whooshes rapidly forward*, whereas nothing very much happens in a Volkswagen. When you so much as slightly touch the steering wheel the Porsche *swerves immediately* whereas the Volkswagen only lumbers slightly to the side.

Thus:

1. The Porsche driving feel comes from being engaged in exercising a certain sensorimotor skill, namely the Porsche driving skill. What provides the Porsche driving feel with its distinctive quality is the different mode of interaction you have with the Porsche as compared to other cars.

2. Furthermore to actually feel the Porsche driving feel, you have to be paying attention to the fact that you are doing Porsche driving things. If while you drive you get very involved in a discussion with a friend, you might no longer be noticing that you were driving the Porsche, and you would rather be experiencing the fact that you are conversing with your friend even though your body was actually doing the same Porsche driving things as before.

Taking the Porsche driving analogy seriously and applying it to sensory feels in general provides a way of accounting for feel in which feel is not something which can be located in some circuit, or which is generated by some mechanism. Instead, feel is a way of doing things. Taking this stance allows one to escape from many of the conundrums connected with phenomenal consciousness, and provides a principled way of explaining the three main properties of feel:

Thus, *why are feels different from one another*, and how exactly are they different? If we were to take the neurophysiologist's view that feels are different because they correspond to different input channels or different areas or mechanisms in the brain, we would be left with the question: What makes this channel or brain area or mechanism generate an experience of seeing, and that channel or brain area or mechanism generate an experience of hearing?

But the sensorimotor approach suggests that we should look for the differences between sensations in the *things that we can potentially do* when we have sensations: Thus, take the example of seeing and hearing. Seeing is a form of interaction in which blinks, movements of the eyes, of the body and of outside objects provoke very particular types of change in sensory input. The laws governing these changes are quite different from the laws governing sensory input in the auditory modality. For example, when one sees, moving forward potentially produces an expanding flow-field on the retina, whereas when one hears, the change in sensory input is now mainly an increase in amplitude of the signal. The claim is now that the sum total of these differences constitute precisely what differentiates the sensations of seeing and hearing. The same would be true for differences between experiences across other sensory modalities. This explanation for differences in sensory modalities escapes from the arbitrariness inherent in neurophysiological explanations appealing to brain channels, areas or mechanisms. No "linking hypothesis" need be made, because the quality of feel is considered to be constituted by what one does when one engages in a particular sensorimotor interaction.

The second main question one can ask about feel is: *What determines the structure of the differences between feels* within a sensory modality? The sensorimotor approach suggests that these differences correspond to differences in the laws that govern one's interaction with the world when one is experiencing different sensations. Contrary to a neural correlate explanation where we have no natural metric linking neural firing rates or other brain phenomena with differences in sensation, in the sensorimotor approach, there is a natural metric allowing sensations to be compared, namely the same metric used by subjects to compare sensorimotor skills in everyday parlance. Quite naturally, because the laws governing sensorimotor interactions are complex and vary from modality to modality, the structure of the differences between feels will be complex. Across two modalities, the sensorimotor interactions are so different that little comparison is possible. Within a modality, each change in one's mode of interaction determines the change in the quality of the experience involved. Later sections will discuss two examples, namely the sensation of touch and the sensation of color.

And finally, what can be said about the third question we asked about sensations, namely: *Why do they have sensory presence,* that is*, Why do they have a feel at all, rather than having no feel?* We will devote a few paragraphs to this question here.

If having a feel consists in attending to the fact that one is engaged in exercising a sensorimotor skill, and if the quality of the feel is constituted by the laws of sensorimotor interaction that the skill involves, then by the very definition of feel, the feel must have a quality, namely the quality constituted by exercising the particular sensorimotor law involved. Thus feels have a quality rather than no quality.

Then, just as the sensorimotor approach invokes differences in skills to account for differences between sensations, the approach will also invoke differences in skills to

account for the difference between experiences involved in perceptual acts and the experiences associated with other brain activities. In particular, two facts about perceptual skills distinguish them from other brain activities.

First, whereas perceptual acts invariably involve, at least potentially, changes caused by motor behavior, this is not true either of internal physiological states or "mental" activities. What we call sensory experience can always potentially be modified by a voluntary motion of the body: Sensory input to the eyes, ears, or any other sensory system is immediately changed in a systematic and lawful way by body motions. On the other hand, people cannot reliably control their internal physiological states by moving their bodies (although of course states like heartbeat and blood oxygen will be affected indirectly by body motions). Likewise, "mental" activities like thoughts, memories, and decisions, to the extent that these can be considered as skills, are not skills that intrinsically involve voluntary body motions. This then is one thing that makes the skills constituting sensory experiences special as compared to other brain processes: they are by nature sensorimotor. Even if at any particular moment there need be no motion, they have what the sensorimotor approach calls "corporality" or "bodiliness" [24],[25],[26]. This strong potential effect of body motions on sensory input is what distinguishes sensor states deriving from the world from sensor states internal to the body or brain.

A second characteristic that distinguishes the skills involved in sensory experience from those of mental functions is what is termed "alerting capacity" or "grabbiness" [24],[25],[26]: this is the fact that sensory systems are genetically endowed with the capacity to deflect our cognitive processing. A loud noise or bright flash will automatically, incontrovertibly, attract our attention to the locus of the event. We are thus, in some sense, "cognitively at the mercy" of sensory input. This is not generally the case for either internal states or mental activities. If a change occurs in the visual field, like a mouse flitting across the floor, one's attention will immediately be caught by it. But variations in heart beat, for example, generally provoke no attentional orienting. Only exceptionnally, and then only indirectly through the pounding it produces on the chest, for example, is one aware that one's heart is beating very fast. Like visceral states, memory, and in general other mental activities, possess no alerting capacity or grabbiness: If one forgets a fact, one only discovers this if one actively tries to recover the fact from memory (an exception might however be, for example, obsessive thoughts).

Thus: a characterisation of the differences in skills associated with sensory acts, as compared to those involved in internal physiological states and mental acts, reveals differences which naturally account for the difference in felt quality of sensations as compared to other brain processes. What has been called the "presence" of sensations seems precisely to consist in the fact that they are both under our control (in that we can modify sensory input by our bodily actions — they have corporality or bodiliness), and also not under our control (they can cause uncontrollable alerting reactions that interfere with our normal cognitive processing: they have alerting capacity or grabbiness). The sensorimotor approach suggests that this captures the idea that there is "something it is like" to have a sensory experience provoked by an outside sensory event, as opposed to it feeling like nothing to have one's heart beat or to have a thought.

In summary for this section: the sensorimotor approach, unlike appeals to neurophysiological mechanisms that generate consciousness, provides a principled account of the three main properties of raw feel, namely

1. the fact that raw feels are different from each other
2. there is structure in the differences
3. raw feels have a quality rather than no quality, and are perceptually present.

The next sections will consider applications of the sensorimotor approach that provide interesting new predictions and results.

4.1 Sensory Substitution

The sensorimotor approach claims that what determines whether one has the feel of seeing, rather than say, the feel of hearing or the feel of smelling, is not the particular sensory input channel, but the laws that characterize the sensorimotor interactions that are involved when one sees, hears or smells. If this is true, then one ought to be able to create conditions where for example one "sees" through one's ears, or through stimulation of the skin: In order to do this, things would have to be arranged so that the laws that govern the sensory input through the ears or skin corresponded to visual-type laws, rather than the usual auditory-type laws.

Indeed the possibility of such "sensory substitution" has been known since Bach y Rita equipped blind subjects with an array of tactile stimulators positioned on their abdomen or back, connected to a video camera that produced a tactile "image" of the world on the observer's skin. Bach y Rita reports in his book that whereas passive stimulation was inconclusive, active camera manipulation by subjects very rapidly provided them with a sensation that they qualified as "seeing" [5], [14], [15].

Recently, technical advances have facilitated further development of sensory substitution devices, and an active community is investigating different types of substitution, ranging from transposing vision to audition, vision to tongue stimulation, vestibular to tongue, among others. For reviews see [3], [4].

4.2 The Localisation of Touch Sensation

Another prediction of the sensorimotor approach is as follows. If the quality of sensory feel is provided, not by the particular nervous pathways, but by the particular mode of sensorimotor interaction that is involved, then one should predict for example that the perceived localisation of a touch, say, on the arm, is not caused by the activation of a particular brain region, but by the particular sensorimotor laws that are involved when that location is touched. More precisely, a touch is felt to be on one's arm, rather than, say, on one's leg, when the touch can be modified by moving one's arm rather than one's leg; when the touch is accompanied by a visual stimulation in the region of the arm, rather than in the region of the leg.

Conversely, if one were presented with a tactile stimulation on the arm that is systematically accompanied by a visual stimulation on some outside object, like say a fake arm put on the table in front of one, then the prediction is that one should come to feel the sensation on the fake arm.

This is precisely the situation that has been investigated in an extensive literature on the "rubber hand illusion", which confirms this counter-intuitive prediction [7], [46].

The finding is also compatible with a large body of work showing that people can rapidly adapt when their body is "extended" by the use of tools or other artefacts. For example, you "feel" the paper under the tip of your pencil, not in your fingers. When you park your car, you "feel" the curb on the wheels of the car, not on the steering wheel.

4.3 The Structure of Color Sensations

Psychophysicists since Hering and Helmholtz have been trying to understand the structure of color space. Certain colors are considered to be "special" or "unique" in the sense that other colors are perceived to be composed of them. For example red, yellow, blue and green are seen as "pure", and not containing other colors, whereas orange is not pure because it is seen to contain red and yellow. Though some success is obtained by the classic opponent process theory of color vision, color scientists today agree that the finer details of these phenomena have not up till now been adequately accounted for by any neurophysiological findings.

The sensorimotor approach claims that in fact the structure of color space is to be sought not in sensory channels per se, but in the laws of interaction that characterize color perception. Thus when one moves a coloured piece of paper under different illuminants, or when one moves one's eyes on or off the paper, there are precise laws that govern the changes in photon catches made by the three photoreceptor types that humans possess. A recent attempt to apply this idea has come up with surprising success [30]. In the case of red, for example, it is found that the changes in photon catches are confined to a single dimension of variation, suggesting why red is a special colour as compared to, say, orange, where three dimensions of variation are observed. Unique hues are accurately predicted in this way. Furthermore, the approach also accurately correlates with well-known anthropological data concerning the way people name colors [30].

4.4 Change Blindness

An interesting point about the sensorimotor approach is the way it explains humans' experience of a very rich and continually present visual world. Instead of supposing, as does the classic approach to vision, that the perceived richness of visual experience requires continuous activation of a rich internal representation of the world, the theory says that richness and continuity are due to the fact that a perceiver has immediate access, via a flick of attention or an eye movement, to any information about the outside world that the perceiver wishes to investigate. The analogy is made of the light in the refrigerator: every time you open the fridge, the light is on, so you assume it is continually on. Similarly, the reason you feel the visual world as being continually present is that whenever you attend to any portion of it, information is available about that portion [23].

According to the sensorimotor approach, a further fact that buttresses the illusion of continual presence lies in the "grabbiness" of visual stimuli. The low level visual system is equipped with "transient-detectors" that register fast changes in luminance or color. These automatically provoke attentional orienting: when a flash of light occurs in the visual periphery, you cannot help moving your eye in that direction. One

therefore has the illusion that one is continually seeing everything, because if anything should suddenly change, one's attention is automatically directed to it, and one sees the change [24], [26].

An exception to this however would occur if the transient detectors were somehow rendered inoperative. This can be done by swamping them with extraneous luminance transients. "Change blindness" is a phenomenon which is coherent with this prediction: large changes in pictures can go unnoticed if the change occurs simultaneously with a global white flash [31] or with several "mudsplashes" [28] distributed all over the visual field. Another way of rendering transient detectors inoperative is to make the changes so slow that they are no longer "grabby". This is what happens in experiments with progressive changes [2], [41], where a large region of a picture changes color or appears or disappears without this being noticed. Although the interpretation has been contested (cf. [40]), because the phenomenon of change blindness is so striking and counter-intuitive, it serves as a convincing confirmation of the sensorimotor approach.

5 Conclusion: Building Consciousness into a Robot

Can we build Access Consciousness into a robot?

The first part of this chapter argued that Access Consciousness is a behavioral capacity, which, though outside the bounds of current work in AI and robotics, presents no fundamental *logical* problem to a robotic implementation. Future work, particularly with embodied systems, bears the hope of gradually approaching the notions of self, intentions, and Theory of Mind, as well as natural language understanding that are undoubtedly prerequisites to Access Consciousness.

Can we build Phenomenal Consciousness or "feel" into a robot?

Although Phenomenal Consciousness or feel is generally considered the "hard" problem of consciousness, this chapter has argued that feel may in fact turn out to be the easier problem (for a similar view see [16]). If we take the stance suggested by the sensorimotor approach, according to which having a feel is, first, engaging in a sensorimotor skill, and second, having access consciousness to the skill, then clearly once a robot has access consciousness, it will suffice for the robot to engage in an embodied interaction with the environment for it to have feel.

References

1. Adolphs, R.: Could a robot have emotions? theoretical perspectives from social cognitive neuroscience. In: Fellous, J.M., Arbib, M. (eds.) Who Needs Emotions: The Brain Meets the Robot, pp. 9–28. Oxford University Press, Oxford (2005)
2. Auvray, M., O'Regan, J.K.: L'Infuence Des Facteurs Sémantiques Sur La Cécité Aux Changements Progressifs Dans Les Scènes Visuelles. Annee Psychologique 103, 9–32 (2003)
3. Auvray, M., O'Regan, J.K.: Voir Avec Les Oreilles: Enjeux De La Substitution Sensorielle. Pour la Science 39, 30–35 (2003)
4. Bach-y-Rita, P., Kercel, S.W.: Sensory Substitution and the Human-Machine Interface. Trends in cognitive sciences 7, 541–546 (2003)

5. Bach-y-Rita, P.: Brain mechanisms in sensory substitution. Academic Press, New York (1972)
6. Block, N.: On a Confusion about a Function of Consciousness. Behav. Brain Sci. 18, 227–247 (1995)
7. Botvinick, M., Cohen, J.: Rubber Hands 'Feel' Touch that Eyes See [Letter]. Nature 391, 756 (1998)
8. Chalmers, D.J.: Facing Up to the Problem of Consciousness. Journal of Consciousness Studies 2, 200–219 (1995)
9. Clark, A.: Sensory qualities. Clarendon Press, Oxford (1993)
10. Dreyfus, H.L.: What computers still can't do: A critique of artificial reason. MIT Press, Cambridge, Mass. (1992)
11. Edsinger, A., Kemp, C.C.: What can I Control? A Framework for Robot Self-Discovery (2006)
12. Fellous, J., Arbib, M.: Who needs emotions?: The brain meets the robot. Oxford University Press, Oxford (2005)
13. Gallagher, S., Shear, J.: Models of the self. Imprint Academic (1999)
14. Guarniero, G.: Tactile Vision: A Personal View. Journal of Visual Impairment and Blindness 71, 125–130 (1977)
15. Guarniero, G.: Experience of Tactile Vision. Perception 3, 101–104 (1974)
16. Harvey, I.: Evolving robot consciousness: The easy problems and the rest. In: Fetzer, J.H. (ed.) Evolving Consciousness. Advances in Consciousness Research Series, pp. 205–219. John Benjamins, Amsterdam (2002)
17. Hutchins, J.: Has Machine Translation Improved? In: MT Summit IX: proceedings of the Ninth Machine Translation Summit, New Orleans, USA, September 23-27, 2003, pp. 181–188 (2003)
18. Kaplan, F., Hafner, V.: The challenges of joint attention. In: Berthouze, L., Kozima, H., Prince, C.G., et al. (eds.) Proceedings of the Fourth International Workshop on Epigenetic Robotics, pp. 67–74 (2004)
19. Lee, M.H., Meng, Q.: Psychologically Insired Sensory-Motor Development in Early Robot Learning. International Journal of Advanced Robotic Systems 2, 325–334 (2005)
20. Lenat, D.B.: CYC: A Large-Scale Investment in Knowledge Infrastructure. Commun. ACM 38, 33–38 (1995)
21. Lungarella, M., Metta, G., Pfeifer, R., et al.: Developmental Robotics: A Survey. Connect. Sci. 15, 151–190 (2003)
22. Martin, R., Barresi, J.: The rise and fall of soul and self: An intellectual history of personal identity. Columbia University Press, Vancouver (2006)
23. O'Regan, J.K.: Solving the "real" Mysteries of Visual Perception: The World as an Outside Memory. Can. J. Psychol. 46, 461–488 (1992)
24. O'Regan, J.K., Myin, E., Noë, A.: Skill, Corporality and Alerting Capacity in an Account of Sensory Consciousness. Prog. Brain Res. 150, 55–68 (2006)
25. O'Regan, J.K., Myin, E., Noë, A.: Phenomenal Consciousness Explained (Better) in Terms of Bodiliness and Grabbiness. Phenomenology and the Cognitive Sciences 4, 369–387 (2005)
26. O'Regan, J.K., Myin, E., Noë, A.: Towards an analytic phenomenology: The concepts of "bodiliness" and "grabbiness". In: Carsetti, A. (ed.) Proceedings of the International Colloquium: Seeing and Thinking. Reflections on Kanizsa's Studies in Visual Cognition, University Tor Vergata, Rome, June 8-9, 2001, pp. 103–114. Kluwer Academic Publishers, Dordrecht (2004)

27. O'Regan, J.K., Noe, A.: A Sensorimotor Account of Vision and Visual Consciousness. Behav. Brain Sci. 24, 939–973 discussion 973-1031 (2001)
28. O'Regan, J.K., Rensink, R.A., Clark, J.J.: Change-Blindness as a Result of Mudsplashes. Nature 398, 34 (1999)
29. Pfeifer, R., Bongard, J.: How the body shapes the way we think: A new view of intelligence. MIT Press, Cambridge (2006)
30. Philipona, D.L., O'Regan, J.K.: Color Naming, Unique Hues, and Hue Cancellation Predicted from Singularities in Reflection Properties. Vis. Neurosci. 23, 331–339 (2006)
31. Rensink, R.A., O'Regan, J.K., Clark, J.J.: On the Failure to Detect Changes in Scenes Across Brief Interruptions. Visual Cognition 7, 127–145 (2000)
32. Rochat, P.: Five Levels of Self-Awareness as they Unfold Early in Life. Conscious. Cogn. 12, 717–731 (2003)
33. Rochat, P.: The self in infancy: Theory and research. Advances in Psychology. Elsevier, Amsterdam, New York (1995)
34. Roy, D.: Grounding Words in Perception and Action: Computational Insights. Trends Cogn. Sci. 9, 389–396 (2005)
35. Roy, D.: Semiotic Schemas: A Framework for Grounding Language in Action and Perception. Artif. Intell. 167, 170–205 (2005)
36. Roy, D.: Grounded Spoken Language Acquisition: Experiments in Word Learning. Multimedia, IEEE Transactions on 5, 197–209 (2003)
37. Roy, D., Pentland, A.: Learning Words from Sights and Sounds: A Computational Model. Cognitive Science 26, 113–146 (2002)
38. Ryle, G.: The concept of mind. Hutchinson, London (1949)
39. Scassellati, B.: Theory of Mind for a Humanoid Robot. Autonomous Robots 12, 13–24 (2002)
40. Simons, D.J., Rensink, R.A.: Change Blindness: Past, Present, and Future. Trends Cogn. Sci. 9, 16–20 (2005)
41. Simons, D.J., Franconeri, S.L., Reimer, R.L.: Change Blindness in the Absence of a Visual Disruption. Perception 29, 1143–1154 (2000)
42. Steels, L.: The Emergence and Evolution of Linguistic Structure: From Lexical to Grammatical Communication Systems. Connect. Sci. 17, 213–230 (2005)
43. Steels, L.: Grounding Symbols through Evolutionary Language Games. Simulating the evolution of language table of contents, 211–226 (2002)
44. Steels, L., Kaplan, F., McIntyre, A., et al.: Crucial Factors in the Origins of Word-Meaning. The Transition to Language, 252–271 (2002)
45. Steels, L., Kaplan, F.: AIBO's First Words: The Social Learning of Language and Meaning. Evol. Commun. 4, 3–32 (2000)
46. Tsakiris, M., Haggard, P.: The Rubber Hand Illusion Revisited: Visuotactile Integration and Self-Attribution. J. Exp. Psychol. Hum. Percept. Perform. 31, 80–91 (2005)
47. Vierkant, T.: Is the self real? LIT Verlag, Münster (2003)
48. von Hayek, F.A.: Routledge & Kegan Paul, London (1952)