Reinventing Ourselves: The Plasticity of Embodiment, Sensing, and Mind*.

Short title: Reinventing Ourselves

Andy Clark

Abstract:

Recent advances in Cognitive Science and Cognitive Neuroscience open up new vistas for human enhancement. Central to much of this work is the idea of new Human-Machine interfaces (in general) and new Brain-Machine interfaces (in particular). But despite the increasing prominence of such ideas, the very idea of such an interface remains surprisingly under-explored. In particular, the notion of human *enhancement* suggests an image of the embodied and reasoning agent as literally extended or augmented, rather than the more conservative image of a standard (non-enhanced) agent *using a tool via some new interface*. In this essay, I explore this difference, and attempt to lay out some of the conditions under which the more radical reading (positing brand new integrated agents or systemic wholes) becomes justified.

1

Introduction: Where the Rubber Meets the Road.

In a short article in the May 2004 edition of WIRED magazine (revealingly subititled 'Fear and Loathing on the Human-Machine Frontier') the futurist and science fiction writer Bruce Sterling sounds an increasingly familiar alarm. After warning us of the imminent dangers of "brain augmentation" he adds:

> "Another troubling frontier is physical, as opposed to mental, augmentation. Japan has a rapidly growing elderly population and a serious shortage of caretakers. So Japanese roboticists…envision walking wheelchairs and mobile arms that manipulate and fetch. But there's ethical hell at the interfaces. The peripherals may be dizzyingly clever gizmos…but the CPU is a human being: old, weak, vulnerable, pitifully limited, possibly senile" (Sterling, 2004 p.116).

This kind of fear is rooted, I shall argue, in a fundamentally misconceived vision of our own humanity: a vision that depicts us as 'locked-in agents': as beings whose minds and physical abilities are fixed quantities, apt (at best) for mere support and scaffolding by their best technologies. In contrast to this view I have argued (Clark 1997, 2003) that human minds and bodies are essentially open to episodes

of deep and transformative re-structuring, in which new equipment (both physical and 'mental') can become quite literally incorporated into the thinking and acting systems that we identify as minds and persons. In what follows, I pursue this theme with special attention to the very notion of the human-machine interface itself.

It helps to start with the commonplace. Sensing and moving are the spots where the rubber of embodied agency meets the road of the wider world, the world outside the agent's physical boundaries. The typical human agent, circa 2004, feels herself to be a bounded physical entity in contact with the world via a variety of standard sensory channels, including touch, vision, smell and hearing. It is a commonplace observation, however, that the use of simple tools can lead to alterations in that local sense of embodiment. Picking up and using a stick, we feel as if we are touching the world at the end of the stick, not (usually) as if we are touching the stick with our hand. The stick, it has sometimes been suggested, is in some way incorporated and the overall effect seems more like bringing a temporary whole new agent-world circuit into being, rather than simply exploiting the stick as a helpful prop or tool.

In these cases there suddenly seem to be two interfaces at play: the place where the stick meets the hand, and the place where the extended system "biological-agent+stick" meets the rest of the world. When we read about new forms of human-machine interface, we are again confronted by a similar duality, and an accompanying tension. What makes such interfaces *appropriate* as mechanisms for human

enhancement is, it seems, precisely their potential role in creating *whole new agent-world circuits*. But insofar as they succeed at this task, the new agent-tool interface itself fades from view, and the proper picture is one of an extended or enhanced agent confronting the (wider) world.

In sections 1 and 2 I shall lightly explore this notion of the interface, and then look at some examples in which new systemic wholes are created by various forms of technological intervention. Section 3 asks under what conditions it becomes proper to speak of enhanced agents rather than un-enhanced agents with new props and scaffoldings. Here, I try to show that there is more at issue than a way of speaking, and that there are scientifically and philosophically important differences between the two cases. Next (section 4) I extend the discussion from bodily augmentation to mental augmentation, indicating what would need to be done to make the vexed idea of *enhanced human mentality* concrete. The discussion continues (section 5) by developing a notion of the 'profoundly embodied agent' as a means of marking the philosophical and scientific importance of our potential for repeated and literal episodes of self re-configuration. The paper ends by relating this image of profound embodiment to some questions (and fears) concerning converging technologies for improving human performance.

1. What's in an Interface?

Haugeland (1998) is, in part, an extended philosophical meditation on the very idea of an interface. The goal is to uncover the underlying principles "for dividing systems into distinct subsystems along nonarbitrary lines" (op cit, p.211). According to Haugeland, the notions of 'component' 'system' and 'interface' are all interdefined and interdefining. Components are those parts of a larger whole that interact through interfaces. Systems are "relatively independent and self-contained" composites of such interfaced components. And an interface itself is: " a point of interactive "contact" between components such that the relevant interactions are well-defined, reliable and relatively simple" (Haugeland 1998 p 213).

Haugeland is right, I think, to point to the nature of interactions as the key to the location of an interface. We discern an interface where we discern a kind of regimented, often deliberately designed, point of contact between two or more independently tuneable or replaceable parts. It does not seem correct, however, to insist that flow across the interface be simple. The idea here seems to be that we find genuine interfaces only where we find energetic or informational bottlenecks, as if an interface must be a narrow channel yielding what Haugeland describes as 'low bandwidth' coupling. This is important for Haugeland's argumentative purpose, as he means to show (by appeal to broadly-speaking Gibsonian characterizations of sensing: see Gibson (1979)) that human sensing typically yields very task-variable, high-bandwidth forms of agent-environment coupling, and thus to argue that no genuine interface or interfaces separate agent

and world. Instead, there is said to be "intimate intermingling of mind, body and world" (op cit 224).

The Gibsonian angle is useful, as it points to two distinct ways in which we might conceive of our own biological sensory systems. According to the standard (non-Gibsonian) conception, a sensory interface is a point of information transduction. It is a point at which rich energetic input (e.g. visual input) must begin to be somehow transformed into discrete internal action-guiding representations. This is the notion of the sensory interface as a kind of fixed veil between an agent and a represented world. But there is another way to look at (at least some uses of) sensing, which can be introduced by a simple example. Consider running to catch a fly ball in baseball. Giving perception it's standard role, we might assume that the job of the visual system is to transduce information about the current position of the ball so as to allow a reasoning system to project its future trajectory. It seems, however, that nature has a more elegant and efficient solution: you simply run so that the ball's trajectory looks straight against the visual background (McBeath et al (1995). This solution exploits a powerful invariant in the optic flow, discussed in Lee and Reddish (1981). But most importantly for our purposes, it highlights (see Maturana (1980)) a very different role for the perceptual coupling itself. Instead of using sensing to get enough information inside, past the visual bottleneck, so as to allow the reasoning system to 'throw away the world' and solve the problem wholly internally, it uses the sensor as *an open conduit allowing environmental magnitudes to exert a constant influence on behavior*.

Sensing is here depicted as the opening of a channel, with successful whole-system behavior emerging when activity in this channel is kept within a certain range. What is created is thus a kind of new, task-specific agent-world circuit.[i] As Randall Beer recently puts it:

> "the focus shifts from accurately representing an environment to continuously engaging that environment with a body so as to stabilize patterns of co-ordinated behavior that are adaptive for the agent" Beer (2000) p.97

This shift in perspective on what sensing is (often) all about will be important later when we consider new sensory channels and their potential impact on the bounds of human agents.

But whilst agreeing with Haugeland that sensing is often best understood in these terms, his own conclusion that no genuine interfaces then link agent and world seems premature. Haugeland depicts these kinds of 'open-channel' solution as involving "tightly coupled high-bandwidth interaction" (op cit p.223) and hence as inimical to the very idea of an agent-world interface[ii]. But it seems intuitive that there can be genuine interfaces that support extremely high-bandwidth forms of coupling. Think, for example, of multiple computers linked into a network by means of super-fast, very high bandwidth 'grid technologies'[iii]. There is really no doubt but that we here confront a web of distinct intercommunicating component machines. Yet that web, in action, can sometimes function as a single unified resource. Nonetheless we still think of it as a web of distinct-

but-interfaced devices. And we do so not because the point of each machines contact with the grid is narrow (it isn't) but because there exist, for each machine on the grid, very well-defined points of potential detachment and re-engagement. We discern interfaces at the points at which one machine can be easily disengaged and another engaged instead, allowing the first to join another grid, or to operate in a stand-alone fashion. An interface, I conclude, is indeed a point of contact between two items across which the types of performance-relevant interaction are reliable and well-defined. But there is no requirement that such interfaces be narrow-bandwidth bottlenecks.

2. New Systemic Wholes

Biological systems, from lampreys to primates, display remarkable powers of bodily and sensory adaptability[iv]. The Australian performance artist Stelarc routinely deploys a 'third hand', a mechanical actuator controlled by Stelarc's brain via commands to muscle sites on his legs and abdomen[v]. Activity at these sites is monitored by electrodes that transmit signals (via a computer) to the artificial hand. Stelarc reports that, after some years of practice and performance, he now feels as if he simply wills the third hand to move. It has become what some philosophers call 'transparent equipment': something through which Stelarc (the agent) can act on the world without first willing an action on anything else. In this respect, it now functions much as his biological hands and arms,

serving his goals without (generally) being itself an object of conscious thought or effortful control.

Recent experimental work reveals more about the kinds of mechanisms that may be at work in such cases. A much-publicized example is the work by Miguel Nicolelis and colleagues on a BMI (Brain-Machine Interface) that allows a macaque monkey to use thought control to move a robot arm. In the most recent version of this work, Carmena et al (2003) implanted 320 electrodes in the frontal and parietal lobes of a monkey. The electrodes allowed a monitoring computer to record neural activity across multiple cortical ensembles while the monkey learnt to use a joystick to move a cursor across a computer screen for rewards. As in previous work, the computer was able to extract the neural activity patterns corresponding to different movements (including direction and grip). Next, the joystick is disconnected. But the monkey is still able to use its neural activity to directly control the cursor for rewards, and learns to do so. Finally (and this is the new element in the work) these commands are diverted to a robot arm whose actual motions are then translated into on screen cursor movements (including an on-screen equivalent of forceful gripping). This closes the loop. Instead of the monkey merely moving an unseen robot arm by thought control alone, the movements now yield visual feedback in the form of on-screen cursor motion.

When the robot arm was inserted into the control loop, the monkey displayed a striking degradation of behavior. It took two full days of

practice for fluent thought-control over the onscreen cursor to be re-established. The reason was that the monkey's brain now had to learn to factor in the mechanical and temporal 'friction' created by the new physical equipment: it had to factor in the mechanical and dynamical properties of the robot arm and the time delays (which were substantial, in the 60-90 millisecond range) caused by interposing the motion of the arm between neural command and on-screen feedback. By the time full fluency was achieved, it is reasonable to conjecture that these properties of the (still unseen) distant arm were incorporated into the monkey's own body-schema. In support of this, the experimenters were able to track real long-term physiological changes in the response profiles of fronto-parietal neurons following use of the BMI, leading them to comment that:

> "the dynamics of the robot arm (reflected by the cursor movements) become incorporated into multiple cortical representations…we propose that the gradual increase in behavioral performance…emerged as a consequence of a plastic re-organization whose main outcome was the assimilation of the dynamics of an artificial actuator into the physiological properties of fronto-parietal neurons"

> Carmena et al (2003) p.205

Creatures capable of this kind of deep incorporation of new bodily (and, as we'll see, sensory and cognitive) structure are examples of what I shall call (see section 4) 'profoundly embodied agents". Such

agents are able constantly to negotiate and re-negotiate the agent-world boundary itself.

Although our own capacity for such re-negotiation is (I believe) vastly under-appreciated, it really should come as no great surprise, given the facts of biological bodily growth and change. The human infant must learn by 'self-exploration' which neural commands bring about which bodily effects, and must then practice until skilled enough to issue those commands without conscious effort. This process has been dubbed (Meltzoff and Moore (1997)) "body babbling" and continues until the infant body becomes "transparent equipment". Since bodily growth and change continues, it is simply good design not to permanently lock in knowledge of any particular configuration, but instead to deploy plastic neural resources and an ongoing regime of monitoring and re-calibration (for some excellent discussion, see Ramachandran and Blakeslee (1998))

As a second class of examples of re-calibration and re-negotiation, consider the plasticity revealed by work in Sensory Substitution. Pioneered in the 60's and 70's by Paul Bach-y-Rita and colleagues, the earliest such systems were grids of blunt 'nails' fitted to the backs of blind subjects, and taking input from a head-mounted camera. In response to the camera input, specific regions of the grid became active, gently stimulating the skin under the grid.At first, subjects report only a vague tingling sensation. But after wearing the grid while engaged in various kinds of goal-driven activity (walking, eating, etc) the reports change dramatically. Subjects stop feeling the tickling on

the back and start to report rough, quasi-visual experiences of looming objects etc. After a while, a ball thrown at the head causes instinctive and appropriate ducking. The causal chain is 'deviant': it runs via the systematic input to the back. But the nature of the information carried, and the way it supports the control of action, is suggestive of the visual modality. Performance using such devices can be quite impressive. In a recent review article, Bach-y-Rita and Kercel (2001, p.543) note that TVSS [Tactile Visual Substitution Systems] have enabled blind subjects to bat a ball as it rolled off a table edge, as well as perform a variety of complex co-ordination tasks

The key to effective sensory substitution is goal-driven motor engagement. It seems to be crucial that the head-mounted camera be under the subject's motor control. This meant that the brain could, in effect experiment via the motor system, giving commands that systematically varied the input, so as to begin to form hypotheses about what information the tactile signals might be carrying. Such training yields quite a flexible new agent-world circuit. Once trained in the use of the head-mounted camera the motor system operating the camera could be changed, eg to a hand-held camera, with no loss of acuity. The touch pad, too, could be moved to new bodily sites, and there was no tactile/visual confusion: an itch scratched under the grid caused no 'visual' effects (for these results, again see Bach-y-Rita and Kercel (2003)).

Such technologies, though still experimental, are now increasingly

advanced. The back-mounted grid is often replaced by a tongue-mounted coin sized array, and extensions in other sensory modalities. Bach-y-Rita and Kercel (op cit) give the nice example of a touch-sensor-rich glove that allows leprosy patients to begin to feel again using their hands. The patient is fitted with the glove that transmits signals to a forehead mounted tactile disc-array, and rapidly reports feeling sensations of touch *at the fingertips* . This is presumably because the motor-control over the sensors runs via commands to the hand, so the sensation is subsequently projected to that site.

The line between these kinds of rehabilitative strategy and wholly new forms of bodily and sensory enhancement is already thin to the point of non-existence. There is advanced work on night-vision versions of sensory substitution, and (at the more dramatic end of this spectrum) it is possible bypass the existing sensory peripheries, feeding signals direct to cortex (see Bach-y-Rita and Kercel (2001) and discussion in Clark (2003)). Even without penetrating the existing surface of skin and skull, sensory enhancement and bodily extension is a pervasive possibility. One rather unusual example ,reported in Schrope (2001), is a US Navy innovation known as a tactile flight suit. The suit (a kind of vest worn by the pilot) allows even inexperienced helicopter pilots to perform difficult tasks such as holding the helicopter in a stationary hover in the air. It works by generating bodily sensations (via safe puffs of air) inside the suit. If the craft is tilting to the right or left or forward or backward, the pilot feels a puff-induced vibrating sensation on that side of the body. The

pilots' own responses (moving in the opposite direction so as to correct the vibrations) can even be monitored by the suit to control the helicopter. The suit is so good at transmitting and delivering information in a natural and easy way that military pilots can use it to fly blindfold.

How should we think of such cases? While wearing the suit, the helicopter behaves very much like an extended body for the pilot: it rapidly links the pilot to the aircraft in the same kind of closed loop interaction that linked Stelarc and the third hand, or the monkey and the robot arm, or the blind person and the TVSS system. What matters, in each case, is the provision of closed-loop signaling so that motor commands affect sensory input. What varies is the amount of training (and hence the extent of deeper neural changes) required to fully exploit the new agent-world circuits thus created. It is crucial, in all these cases, that the new agent-world circuits be trained and calibrated in the context of a whole agent engaged in world-directed (goal-driven) activity. One sign of successful calibration is, as we noted earlier, that once fluency is achieved the specific details of the (old or new) circuitry by which the world is engaged fall 'transparent' in use. The conscious agent is then aware of the oncoming ball, not of *seeing* the ball, or (by the same token) of using *a tactile substitution channel* to detect the ball. In just this way the tactile-vest wearing pilot becomes aware of the plane's tilt and slant, not of *the puffs of air.* Perception, as Varela, Thompson and Rosch (1991) and O' Regan and Noe (2001) have persuasively argued, just *is* this open-ended process of actively engaging a world.

To sum up, humans and other primates are integrated but constantly negotiable bodily platforms of sensing, moving and (as we'll soon see) reasoning. Such platforms extend an open invitation to technologies of human enhancement: they are biologically designed so as to fluidly incorporate new bodily and sensory kit, creating brand new systemic wholes

3. Incorporation versus Use.

A very natural doubt to raise, at about this point, would be the following:

> Critic: "You are making quite a song and a dance out of this, what with talk of brand new systemic wholes and so on. But we all know we can use tools, and that we can sometimes learn to use them fluently and 'transparently'. Why talk of new systemic wholes, of extended bodies and reconfigured users, rather than just the same old user in command of a new tool?"

This is the right question to push, and we have already seen a hint of the answer in the comments of Carmena at al concerning the altered response profile of certain fronto-parietal neurons. To bring the key idea into focus, it helps next to consider a closely related body of research on tool-use by primates. To set the scene requires a brief neuro-scientific excursion.

Recent years have seen the discovery, in primate brains, of a variety of so-called 'bi-modal neurons' These are:

> "Pre-motor, parietal and putaminal neurons that respond both to somatosensory information from a given body region (ie the somatosensory Receptive Field; sRF) and to visual information from the space (visual Receptive Field;vRF) adjacent to it"

> Maravita and Iriki (2004) p. 79

For example, some neurons respond to somatosensory stimuli (light touches) at the hand *and* to visually presented stimuli near the hand, so as to yield an action-relevant coding of visual space. In a series of experiments, recordings were taken from bi-modal neurons in the intraparietal cortex of Japanese macaques while they (the macaques) learnt to reach for food using a rake. The experimenters found that after just 5 minutes of rake-use, the responses of some bi-modal neurons whose original vRFs picked out stimuli near the hand had expanded to include the entire length of the tool, "as if the rake was part of the arm and forearm" (op cit p. 79). Similarly, other bi-modal neurons, that previously responded to visual stimuli within the space reachable by the arm, now had vRFs that covered the space accessible by the arm-rake combination. After surveying a number of other related findings (including some fascinating work in which similar effects are observed after experience of reaching with a virtual arm in an on-screen display) Maravita and Iriki conclude that:

"Such vRF expansions may constitute the neural substrate of use-dependent assimilation of the tool into the body-schema, suggested by classical neurology"
Maravita and Iriki (2004) p.80

It is also noteworthy, especially in the light of our previous discussion, that:

"any expansion of the vRF only followed active, intentional use of the tool not its mere grasping by the hand"
(op cit p. 81).

In human subjects suffering from unilateral neglect (in which stimuli from within a certain region of egocentrically coded space are selectively ignored) it has been shown that the use of a stick as a tool for reaching actually extends the area of visual neglect to encompass the space now reachable using the tool (see Berti and Frassinetti (2000)). Berti and Frassinetti conclude that:

"The brain makes a distinction between 'far space' (the space beyond reaching distance) and 'near space' (the space within reaching distance)" and that "…simply holding a stick causes a remapping of far space to near space. In effect the brain, at least for some purposes, treats the stick as though it were a part of the body" Berti and Frassinetti (2000) p. 415

The plastic neural changes reported by Carmena et al (section2 above), and now further underlined by Maravita and Iriki, and by Berti and Frassinetti are, I want to suggest, the key to a real (philosophically important and scientifically solid) distinction between true incorporation into the body-schema and mere use. The body-schema, in this sense, is not the same thing as the body-image, though the two can sometimes be related. As I shall use the terms ,the body image is a conscious construct, able to inform thought and reasoning about the body. The body schema is a suite of neural settings that implicitly (and non-consciously) define a body in terms of its capabilities for action, for example, by defining the extent of 'near space' for action programs[vi]. I would speculate, however, that the striking conscious experiential datum of equipment (not just rakes but even cars and violins) falling transparent in use is plausibly one result, in conscious agents, of just these kinds of deeper changes: changes (that may be temporary, context-dependent, or longterm) in the body-schema itself.

It is important to notice that we can certainly imagine tool-users (perhaps even fluent tool-users?) whose brains were *not* engineered so as to adapt the body-schema in these ways. Such beings would *always* use tools the way we typically begin: by representing the tool and its features and powers (its length, for example) and calculating effective uses accordingly. We can even imagine (I think) beings who were so fast and good at these calculations as to deploy the tools with the same skill and efficacy as an expert human agent. The contrast that would remain, even in the latter kind of case, would be

between the skilled agent's first representing the shape, dimensions and powers of the tool and then inferring (consciously or otherwise) that you can now reach such and such, and do such and such, and agents whose brains were so constituted that experience with the tool results in (for example) a suite of altered vRFs such that objects within tool-augmented reaching range are now *automatically* treated as falling within 'near space'.

These are surely distinct strategies. The latter strategy might be especially recommended for beings whose bodies (like our own) are naturally subject to growth and change. Beings deploying this strategy do not relate to their own bodies the way classical cognitive science depicts the intelligent agent as relating to its world, namely, via a process of objectivist representation and inference. The deep distinction is thus between various forms of knowledge-based use (which involves a lot of explicit representation of features, properties, and inferences based on those features), and genuine episodes of assimilation and integration, which can now be defined as cases in which plastic neural resources are re-calibrated (in the context of goal-directed whole agent activity) to reflect new bodily and sensory opportunities. In this way, our own embodied activity brings forth new systemic wholes.

4. Extended Cognition

Could anything like this notion of 'incorporation' (rather than mere use) and new systemic wholes get a grip in the more ethereal domain

of mind and cognition? Could human *minds* be genuinely extended and augmented by cultural and technological tweaks, or is it always (as many evolutionary psychologists, such as Pinker (1997) would have us believe) just the same old mind with a shiny new tool? Here, the story is murkier by far. My own view, defended at much greater length elsewhere (see Clark and Chalmers (1998), Clark (2003) (In Press)) is that external and non-biological information-processing resources are also apt for temporary or long-term recruitment and incorporation rather than simply knowledge-based use, and that to whatever extent that this holds, we are not just bodily and sensorily but also *cognitively* permeable agents. But whereas we can point, in the case of basic tool use, to visible neural changes that accompany the genuine assimilation of new bodily structure, it is harder to know what to look for in the case of mental and cognitive routines.

It may be helpful[vii] first to display the bare logical possibility of such cognitive extension. For even the bare possibility, some might feel, is ruled out by a simple argument to the effect that "cognitive enhancement requires that the cognitive operations of the prop be intelligible to the agent". If this were so, cognitive enhancement would always be in some clear sense superficial. But the argument is flawed, since the cognitive operations of much of my own brain are not thus intelligible to me, the conscious agent. Yet they surely help make me the cognitive agent I am. It also helps to reflect that biological brains must change and evolve by coordinating old activities and processes with new ones made available by new or subtly altered structures. To insist that such change requires the

literal intelligibility of the operations of the new by the old (rather than simply some appropriate integration and co-ordination) is to miss the potential for new wholes that are *themselves* the determiners of what is and is not intelligible. Certain non-biological tools and structures, I am thus suggesting, can become sufficiently well integrated into our problem-solving activity to count as parts of new wholes in just this way. But just what does such integration (genuine cognitive incorporation) require?

One suggestion (Clark and Chalmers (1998)) is that cognitive incorporation occurs when the existing system learns a complex problem-solving routine that makes deep implicit commitments to the robust availability of certain operations and/or bodies of information while carrying out some species of on-line problem solving. This is the cognitive equivalent, I would now suggest, of the implicit commitments to bodily shape and potentials for action made by tuning the receptive fields of bi-modal neurons. In the cognitive case, what matters is the delicate temporal tuning of multiple participating elements (including, for example, calls to internal or external information stores) that simply factor in the availability of those operations or bodies of information.

The field of 'active vision' provides a nice example. Ballard et al (1997) studied a task in which subjects copied a pattern of colored blocks by moving them from one on-screen area (the model) to another (the target). Using eye-tracking techniques, the experimenters found that subjects looked to the model both before

and after picking up a block. The explanation for the apparently unnecessary repetitions of gaze seems to be that when glancing at the model, the subject stores only one piece of information: either the color or the position of the next block to be copied (not both).The conclusion was that the brain uses repeated visual fixation to link a target location to a type of information (color or position) retrieving that information 'just-in-time' for use. In this way, according to the authors:

> "…fixation can be seen as binding the value of the variable currently relevant for the task [and] changing gaze is analogous to changing the memory reference in a silicon computer".
> Ballard et al (1997) p. 723

In this respect, for this task, the brain simply uses the external scene as its memory store.

This subtle reliance on the external scene is dramatically illustrated by recent work on so-called 'change blindness' (see e.g. Simons (2000) in which simple experimental manipulations (the masking of motion transients while large changes are made to a visually presented scene) suggest[viii] the surprising sparseness of our on-the-spot (all in one instant, or 'snapshot' –see Noe (2004)) conscious awareness. Subjects seldom these changes, and are often amazed when they realize what has happened without their noticing it. One diagnosis of why we are not normally aware of any such sparseness is that our feeling of "seeing all the detail" in the scenes (and hence

the surprisingness of the demonstrations of unseen changes) is really a reflection of something implicit in the overall problem-solving organization in which vision participates. That organization 'assumes' the (ecologically normal) ability to retrieve more detailed info when needed, so we feel (correctly, in an important sense) that we are already *in command* of the detail[ix].

The point, for present purposes, is that the brain need not actively represent the availability of such and such information from any given internal or external location. Instead, it simply deploys a problem-solving routine (that may involve programmed saccades to a visual location, or calls to biological memory) whose fine temporal structure *assumes* the easy availability of such and such information from such and such a location. It is in this way (I am suggesting) that non-biological informational resources can become - either temporarily or longterm – genuinely incorporated into the problem-solving whole. Just as the experienced brain need not (though it sometimes can) explicitly represent the shape of a tool and then infer the available reach, so too it need not (though it sometimes can) first represent the availability of specific information at some location, and then infer that it can find what it needs by accessing a given resource. Instead, a problem-solving routine is delicately 'grown' so as to maximally exploit the local informational field[x]. Such a field can include biological resources, environmental structure, and cognitive artifacts such as notebooks and laptops. As we move towards an era of wearable computing and ubiquitous information access, the robust, reliable information fields to which our brains delicately adapt their

routines will become increasingly dense and powerful, further blurring the distinction between the cognitive agent and her best tools, props and artifacts.

5. Profound Embodiment.

The notion of embodiment[xi] plays an increasingly prominent role in Philosophy and Cognitive Science. It is not always clear, however, exactly what it is that *matters* about embodiment. I shall end, then, by making a concrete (but perhaps somewhat heretical) proposal, and then relating it to the questions concerning the nature of the interface and to the topic of converging technologies for improving human performance.

We can distinguish three 'grades' of embodiment. I shall call these (rather unimaginatively) 'mere embodiment', 'modest embodiment' and 'profound embodiment'. A 'merely embodied' creature or robot would be one equipped with a body and sensors, able to engage in closed-loop interactions with its world, but for whom the body was nothing but a means to implement solutions arrived at by pure reason. Imagine also that this being can control the body only by issuing a complex series of micro-managing commands to every tiny muscle, tendon, spring and actuator. A close real-world approximation to such a being is the early mobile robot, such as Shakey, built over three decades ago at the Stanford Research Institute. A 'modestly embodied' creature or robot would then be one for whom the body was not just another problem-space, requiring

constant micro-managed control, but was rather a resource whose own features and dynamics (of sensor placement, of linked tendons and muscle groups etc) could be actively exploited allowing for increasingly fluent forms of action selection and control. Much work in contemporary robotics explores this middle ground of modest embodiment, for example, Barbara Webb's (1996) lovely work on the robot cricket in which sensor placement and time delays caused by signal transmission along internal pathways prove integral to its capacity to identify the song of a mate and locomote in that direction. What makes this an example of only modest embodiment is that the specific solution is 'locked in' by the details of the hard-wired architecture itself. Such systems are congenitally unable to learn new kinds of body-exploiting solution 'on the fly', in response to damage, growth, or change.

It is perhaps hardly surprising that much (though not all- see Lungarella et al (2003)) work in real-world robotics explores this space of 'modest embodiment': after all, robots (so far) don't grow, or use tools, or self-repair. By contrast, as we have seen, biological systems (and especially us primates) seem to be specifically designed so as to constantly search for opportunities to make the most of body and world, checking for what is available, and then (at various time-scales and with varying degrees of difficulty) integrating new resources very deeply, creating whole new agent-world circuits in the process. A 'profoundly embodied' creature or robot is thus (according to this definition) one that is highly engineered so as to be able to learn to make maximal problem-simplifying use of an open-

ended variety of internal, bodily, or external sources of order. We saw, in previous sections, some hints of the kinds of engineering involved. It includes the use of plastic neural resources to create and update a body-schema, the capacity to factor the availability of information (wherever and however stored) into the heart of temporally fine-tuned problem-solving routines, and the capacity (in conscious beings like ourselves) for equipment to become transparent in use. This is not, of course, an all-or-nothing divide. Profound embodiment comes in many degrees and flavors, all the way from almost (but not quite) fully hard-wired solutions to amazingly plastic and re-configurable ones. But primates, as we have seen, seem to fall quite close to the more radically re-configurable end of this spectrum.

But why describe this as 'profound embodiment' rather than as a return to the outdated (or so many of us believe- see Clark (1997)) image of mind as a disembodied organ of control? The answer is that these kinds of minds are not in the least disembodied. Rather, they are *promiscuously* body-and-world exploiting. They are forever testing and exploring the possibilities for incorporating new resources and structures deep into their problem-solving regimes. They are, to use the jargon of Clark (2003), the minds of Natural-Born Cyborgs: of systems continuously re-negotiating their own limits, components, data-stores and interfaces. On this account, the body (any given biological or bio-technological body) is both critically important and constantly negotiable. It is critically important as it is a key player on the problem-solving stage. It is not simply the point at which

processes of transduction pass the real problems (now rendered in rich internal representational formats) to an inner engine of disembodied reason. Instead, much of our skilled engagement with the world flows, we saw, from the way subtle neural changes enable the embodied agent to rather directly engage the world, without representing every detail of bodily form and action-taking capacity (a neat example was the way tool-use affects receptive field properties that then implicitly distinguish 'near' space and 'far' space). But by the same token, all this is now highly negotiable, with the body-schema and other supporting resources apparently able to re-form and re-configure as components, interfaces and resources change and shift.

All this matters, both scientifically and philosophically. It matters scientifically since it puts plasticity and adaptability where they belong, at center stage of our best models of minds, agents and persons. And it matters philosophically since it invites us to take our best present and future technologies very seriously, as quite literally helping to constitute who and what we are. With this picture in mind, those opening fears expressed by Bruce Sterling should seem infinitely less compelling. Sterling paints a truly frightening picture of an augmented agent within whom "the CPU is a human being: old, weak, vulnerable, pitifully limited, possibly senile". Such fears, I hope to have suggested, play upon a deeply misguided image of who and what we *already* are. They play upon an image of the human agent as doubly locked-in: as a fixed mind constituted solely by a given biological brain, and as a fixed bodily presence in a wider world. But human minds are not old-fashioned CPU's trapped in fixed and

increasingly feeble corporeal shells. Instead, they are the surprisingly plastic minds of *profoundly* embodied agents: agents whose boundaries and components are forever negotiable, and for whom body, thinking and sensing are woven flexibly (and repeatedly) from the whole cloth of situated, intentional action.

6. Enhancement or Subjugation?

The picture I have painted is meant to be a guardedly optimistic one. It is our basic, biologically-grounded nature (or so I have suggested) to be open to a wide variety of forms of technologically-mediated enhancement, from sensory substitution to bodily extension to mental extension and cognitive reconfiguration. If this picture is correct, our best tools and technologies literally become us: the human self emerges as a 'soft self' (Clark (2003)), a constantly negotiable collection of resources easily able to straddle and criss-cross the boundaries between biology and artifact. In this hybrid vision of our own humanity I see potentials for repair, empowerment, and growth. But the very same hybrid vision may raise specters of coercion, monstering and subjugation. For clearly, not all change is for the better, and hybridization (however naturally it may come to us) is neutral rather than an intrinsic good. Uncritical talk of human 'enhancement' thus threatens to beg philosophically, culturally and politically important questions. How do we distinguish genuine enhancement from pernicious encroachment and new horizons from new impositions? Such questions demand sustained, informed debate going far beyond the scope of the present treatment. But

there is cause for cautious optimism, and for three interlocking reasons.

First, there is simply nothing new about human enhancement. Ever since the dawn of language and self-conscious thought, the human species has been engaged in a unique 'natural experiment' in progressive niche construction (see Sterelny (2004)). We engineer our own learning environments so as to create artificial developmental cocoons that impact our acquired capacities of thought and reason. Those enhanced minds then design new cognitive niches that train new generations of minds, and so on, in an empowering spiral of co-evolving complexity. The result is that, as Herbert Simon is reputed to have said, 'most human intelligence is artificial intelligence anyway'. Technologies of human cognitive enhancement are just one more step along this ancient path.

Second, the conscious mind is perfectly at ease with reliance upon anything that works! The biological brain is itself populated by a vast number of 'zombie processes' that underpin the skills and capacities upon which successful behavior depends. There are, for example, a plethora of such unconscious processes involved in activities from grasping an object (see Milner and Goodale (1995)) all the way to the flashes of insight that characterize much daily skilful problem-solving. Technology-based enhancements add, to that standard mix, still more processes whose basic operating principles are not available for conscious inspection and control. The patient using a brain-computer interface to control a wheelchair will not typically know just

how it all works, or be able to reconfigure the interface or software at will. But in this respect too, the new equipment is simply on a par with much of the old. To fear that this must inevitably lead to dilutions of self-control and diminishment of responsibility is to miss the fact that we are already host to scores of similarly hidden processes. Insofar as this is compatible (in the biological case) with a sufficiently robust notion of self-control and of responsibility, it must at least be possible for the same to be true in the case of well-tuned technologically mediated enhancements.

A third reason for cautious optimism is the power of the hybrid/cyborg image itself as a means of generating public debate. For once we accept that our best tools and technologies literally become us, changing who and what we are, we must surely become increasingly diligent and exigent, demanding technological prostheses better able to serve and promote human flourishing. Empirical science is now beginning (e.g. Layard (2005)) systematically to address the sources and wellsprings of human happiness and human flourishing, and the findings of these studies must themselves be taken as important data points for the design and marketing of putative technologies of enhancement. Just as the slogan that 'you are what you eat' contributed to the emerging recognition that food, far from simply being fuel, had a finely nuanced impact on our mental and physical health, so the realization that we are soft selves, wide open to new forms of hybrid cognitive and physical being, should serve to remind us to choose our bio-technological unions very carefully, for in so doing we are choosing who and what we are.

Conclusions

I have tried to show that we humans are *profoundly embodied agents*: creatures for whom body, sensing, world and technology are resources apt for recruitment in ways that yield a permeable and repeatedly reconfigurable agent/world boundary. For the profoundly embodied agent, the world is not something locked away behind the fixed veil of a certain skin-bag, a reasoning engine and a primary sensory sheath. Rather, it is a resource apt for active recruitment and use, in ways that bring new forms of embodied intelligence into being. Such agents are genuinely *of* their worlds, and not simply *in* them. They are not helpless bystanders watching the passing show from behind a fixed veil of sensing, acting and representing, but the active architects of their own bounds and capacities[xii]. Such a perspective invites a cautious optimism concerning converging technologies for improving human performance. This discussion has emphasized the potential for new forms of human-machine (or brain-machine) interface. But such technologies may also be chemical, computational, genetic, bio-mechanical or nano-technological. They may augment and alter mind, sensing, and body. But whatever the form, the key to successful integration and assimilation looks to be the same: the creation of new forms of rich, feedback-driven agent-world circuit, with sensing and acting under active intentional control.

Recognition of our vast potential for bio-technological mergers and coalitions should, I finally argued, be a source not of fear and loathing

but of guarded hope and cautious optimism. It should increase our respect for the deep biological plasticity that makes such mergers possible, reduce our fears of an unnatural 'post-human' future, and license greater expectations concerning the answerability of our chosen tools and technologies to our best empirical models of the wellsprings of human happiness and human flourishing.

Department of Philosophy
School of Philosophy, Psychology and Language Sciences
Edinburgh University
Scotland

---

[i] This is by no means an isolated case. Susan Hurley (1998) argues convincingly that perception typically involves whole cycles of input-output behavior in which sensing and acting dynamically combine to yield ongoing adaptive fit between whole organisms and the world. This perspective also fits well with recent work in so-called interactive vision (see (Ballard 1991) (1991), Ballard et al (1997)). The theme of active engagement is similarly visible in a variety of recent treatments that stress the importance of motor activity to perception (see e.g. O'Regan and Noe (2001) , Churchland et al (1994), Clark (1999), Noe (2004).

[ii] In fact, it is rather doubtful that these kinds of Gibsonian invariant detection involve truly high-bandwith coupling at all. But (given the extreme difficulty of

finding a non-controversial measure of objective bandwidth) I am willing to grant this for the sake of argument. My point will be that such high-bandwidth coupling, even if present, does not undermine the idea of interfaces located at just those points.

[iii] A typical description reads: "Computational Grids enable the sharing, selection, and aggregation of a wide variety of geographically distributed computational resources (such as supercomputers, compute clusters, storage systems, data sources, instruments, people) and presents them as a single, unified resource for solving large-scale compute and data intensive computing applications" (Quote taken from the GRID computing information center at: http://www.gridcomputing.com/).

[iv] See Mussa-Ivaldi and Miller (2003), Bach-y-Rita and Kercel (2003), Clark (2003)

[v] See http://www.stelarc.va.com.au, and full discussion in Clark (2003) chapter 5

[vi] Gallagher (1998) expresses the difference like this:
"Body schema can be defined as a system of preconscious, subpersonal processes that play a dynamic role in governing posture and movement… There is an important and often overlooked conceptual difference between the subpersonal body schema and what is usually called *body image*. The latter is most often defined as a conscious idea or mental representation that one has of one's own body"

[vii] Thanks to an anonymous referee for pressing this issue

[viii] But see Simons et al (2002) for some important provisos.

[ix] For some more detailed explorations of this idea, see O'Regan and Noe (2001) and Clark (2002)

[x] For a lovely example of this, see Gray and Fu (2004).

[xi] See, among many others, Varela et al (1991), Clark (1997), O'Regan and Noe (2001)

[xii] For some important explorations of these themes, see Heidegger (1927), Merleau-Ponty (1945/1962), Thompson, Varela and Rosch (1991), O'Regan and Noe (2001).

References

Bach-y-Rita, P and Kercel, SW (2003) 'Sensory Substitution and the Human-Machine Interface' *Trends in Cognitive Sciences* 7:12: 541-546

Ballard, D. (1991) 'Animate Vision' *Artificial Intelligence* 48, 57-86

Ballard, D., Hayhoe, M., Pook, P., & Rao, R. (1997) 'Deictic codes for the embodiment of cognition'. *Behavioral and Brain Sciences*, 20, 723-767

Beer, RD (2000). 'Dynamical approaches to cognitive science'. *Trends in Cognitive Sciences* 4(3):91-99

Berti, A and Frassinetti, F (2000) 'When Far Becomes Near: Re-mapping of space by tool use'. *Journal of Cognitive Neuroscience* 12 :415-420

Carmena,J.,Lebedev,M.,Crist,R.,O'Doherty,J.,Santucci,D.,Dimitrov,D.,Patil,P., Henriquez, C., and Nicolelis,M. (2003) 'Learning to Control a Brain-Machine Interface for Reaching and Grasping by Primates '.

PloS Biology (Public Library of Sciences: Biology) 1:2: 193-208

Churchland, P., Ramachandran, V., & Sejnowski, T. (1994). 'A Critique of Pure Vision'. In C. Koch & J. Davis (Eds.), *Large-Scale Neuronal Theories of the Brain* . Cambridge, MA: MIT Press.23-61

Clark, A (1997) *Being There: Putting Brain, Body and World Together Again* MIT Press, Camb. MA.

Clark, A (1999) 'Visual Awareness and Visuomotor Action' *Journal Of Consciousness Studies* 6:11-12: p.1-18

Clark, A (2002) 'Is Seeing All It Seems? Action, Reason and the Grand Illusion' in *Journal Of Consciousness Studies* Volume 9, No.5/6 181-202 . Reprinted in A Noe (ed) (2000) *Is The Visual World A Grand Illusion?* Thorverton, UK: Imprint Academic: 181-202

Clark, A (2003) *Natural-Born Cyborgs: Minds, Technologies and the Future of Human Intelligence* NY: Oxford University Press

Clark, A. and Chalmers, D. J. (1998) 'The extended mind' *Analysis* 58,   1:   7-19.

Clark, A (2005) 'Intrinsic Content, Active Memory, and the Extended Mind'. *Analysis* 65:1 January 2005 p.1-11

Gallagher, S (1998) 'Body Schema and Intentionality'. In J. Bermudez (ed) *The Body and The Self* Camb. MA: MIT Press. P. 225-244

Gibson, J. J. (1979) *The Ecological Approach to Visual Perception,* NY: Houghton-Mifflin

Gray, WD, & Fu, W.-t. (2004). 'Soft constraints in interactive behavior' *Cognitive Science*, 28(3), 359-382.

Hurley, S. L. (1998) *Consciousness in Action*. Cambridge, MA: Harvard

Haugeland, J (1998) 'Mind Embodied and Embedded'. in J. Haugeland, *Having Thought* MIT Press, Camb. MA

Layard, R (2005*) Happiness: lessons from a new science* London: Allen Lane

Lee, D. and Reddish, P. (1981) 'Plummeting Gannets: A Paradigm of Ecological Optics' *Nature* 293:293-294

Lungarella M., Metta G.., Pfeifer R., and Sandini G.(2003) 'Developmental robotics: a survey' *Connection Science*, 15: 4, pp. 151-190

McBeath, M., Shaffer, D. and Kaiser, M. (1995) 'How Baseball Outfielders Determine Where to Run to Catch Fly Balls' *Science*: 268:569-573

Maravita, A and Iriki, A 'Tools for the body (schema)' *Trends in Cognitive Sciences* vol 8:2: 79- 86

Maturana, H (1980) 'Biology of Cognition'. in Maturana, H, Humberto, R and Varela, F *Autopoiesis and Cognition* Dordrecht: Reidel 2-62

Meltzoff, A. N., & Moore, M. K. (1997). 'Explaining facial imitation: A theoretical model '*Early Development and Parenting*, 6, 179-192.

Merleau-Ponty, M. (1945/1962) *The phenomenology of perception*. Trans. Colin Smith. London: Routledge and Kegan Paul

Milner, A., and Goodale M., (1995) *The Visual Brain in Action*. Oxford: Oxford University Press

Mussa-Ivaldi, F and Miller, L (2003) 'Brain-machine interfaces: computational demands and clinical needs meet basic neuroscience' *Trends in Cognitive Sciences* 26:6:329-334

Noë, A. (2004) *Action in Perception*. Cambridge, MA: The MIT Press.

O'Regan, J. K. and Noë, A. (2001) 'A sensorimotor approach to vision and visual consciousness.' *Behavioral and Brain Sciences* 24/5: 883-975.

Pinker, S (1997) *How the Mind Works* , London, Allen Lane: The Penguin Press

Ramachandran VS & Blakeslee S. (1998) *Phantoms in the Brain: Probing the Mysteries of the Human Mind*. New York: Morrow & Co.

Schrope, M. 'Simply sensational' *New Scientist*, 2 June, 2001, pp. 30-33

Simons, D. S. (2000) 'Current approaches to change blindness'. *Visual Cognition* 7:1-15

Simons, D.J., Chabris, C.F., Schnur, T. T., & Levin, D.T. (2002). 'Evidence for preserved representations in change blindness'. *Consciousness and Cognition*, 11, 78-97

Sterelny K (2003) *Thought in a Hostile World: The Evolution of Human Cognition* Blackwell, Oxford

Sterling, B (2004) 'Robots and the Rest of Us: Fear and Loathing on the Human-Machine Frontier' *Wired* May 2004 p.116

Varela, F., Thompson E., and Rosch, E. (1991) *The Embodied Mind,* Camb. MA, MIT Press

Webb, B (1996) 'A Cricket Robot' *Scientific American* 275: 62-67