Soft Selves and Ecological Control

Andy Clark

School of Philosophy, Psychology and Language Sciences (PPLS)
Edinburgh University
Scotland, EH8 9JX
UK

andy.clark@ed.ac.uk

These bodily members are, as it were, no more than garments; which, because they have been attached to us for a long time, we think are us, or parts of us [and] the cause of this is the long period of adherence: we are accustomed to remove clothes and to throw them down, which we are entirely unaccustomed to do with our bodily members[i].
Avicenna, De Anima V.7

As interface, the skin is obsolete…The clothing of the body with membranes embedded with alternate sensory and input/output devices creates the possibility of more intimate and enhanced interactivity. Subjectively, the body experiences itself as a more extruded system, rather than an enclosed structure. The self becomes situated beyond the skin[ii].
Stelarc, Parasite Visions

Abstract/Introduction

Advanced biological brains are by nature *open-ended opportunistic controllers*. Such controllers compute, pretty much on a moment-to-moment basis, what problem-solving resources are readily available and recruit them into temporary problem-solving wholes. Neural

plasticity, exaggerated in our own species, makes it possible for such resources to become factored deep into both our cognitive and physical problem-solving routines. One way to think about this is to depict the biological brain as a master of what I shall dub 'ecological control'. Ecological control is the kind of top-level control that does not micro-manage every detail, but rather encourages substantial devolvement of power and responsibility. This kind of control allows much of our skill at walking to reside in the linkages and elastic properties of muscles and tendons. And it allows (I claim) much of our prowess at thought and reason to depend upon the robust and reliable operation, often (but not always) in dense brain-involving loops, of a variety of non-biological problem-solving resources spread throughout our social and technological surround.

Are the complex distributed systems that result in some sense 'out of control', beyond the reach of useful (you might even, though problematically, say, 'personal') governance? I shall argue that they are not, although understanding them requires us to re-think some key ideas about control and the nature of the self. To (try to) make this case, I shall first examine some strategies for efficient, external opportunity exploiting control in simple systems. I shall then argue that many of the same lessons apply to the case of higher-level human problem-solving.

1. Ecological Control

Consider Shakey. Shakey, circa 1970, was the mobile robotic darling of the Stanford Research Institute: one of the very first computer-controlled mobile robots and a locus of hard (non-ecological) control. Armed with a camera, wheels, and a laser range–finder, and controlled by a big old mainframe whizzing along at 1/4 million calculations per second, Shakey could obey typed commands such as "Push the cube to the pyramid". To do so, the system would sense, process, plan, and then act out its plan. For Shakey, the body and the environment were first and foremost problems to be solved. The environment was the problem arena. The sensors detected the lay-out in that arena and the reasoning system planned a solution. To a large degree, Shakey's body was just another part of the problem space: a part that needed to be sent detailed, micro-managing control signals so as to put the reasoned-out solution into practice. A contemporary analogue, though vastly faster and more sophisticated, is Honda's Asimo. Asimo is a mainstream control paradigm walking robot that uses precise joint-angle control to mimic human walking. The solution looks pretty good but is massively energy and computation expensive.

Contrast these solutions with the kinds of ecological control deployed by Passive Dynamic Walkers[iii] (PDW's) Passive Dynamic Walkers are simple-looking two-legged devices that employ no actuation except gravity, have no control system as such, and (as a result) can enforce no joint angle control at any time. Yet surprisingly, PDW's are capable (when set on a gentle incline) of very stable, human-looking walking. Now imagine that you (playing evolution) want to exploit that

kind of surprising capacity but in the context of a self-powered locomoting agent. The solution, which looks to be Nature's own, is to walk using a kind of controlled falling-over. Powered (level terrain) walking can thus be brought about by the brain and CNS systematically pushing, damping and tweaking a system in which Passive Dynamic effects continue to play a very major role. In such cases, a low energy source, a simple control system, and the body (and gravity!) 'collaborate' to solve the walking problem. This strategy has recently been implemented in a variety of simple robots (see Collins et al (2005)) and described as "a new design and control paradigm' for walking robots (op cit p.1083). As a second example, Tedrake (2005) has given us 'Robotoddler'. Robotoddler uses actor-critic reinforcement learning to acquire a control policy that exploits the passive dynamics of the body. The robot learns to change speeds and to go forward and backward, and can adapt on the go to different terrains, including bricks, wooden tiles, carpet and a variable speed treadmill. By using passive dynamic strategies, the robot's power consumption is dramatically reduced (to about 1/10th that of a standard robot like Honda's Asimo).

These examples serve to introduce the notion of soft or 'ecological' control. This is the kind of control that occurs when a system's goals are not achieved by micro- managing every detail of the desired action or response, but by using a strategy that devolves a great deal of problem-solving responsibility, making the most of some robust, reliable source of relevant order in the body, elsewhere in the brain

and/or in local environment (notice that 'ecological control' thus names a type of effect not a single mechanism).

The effect of an ecological control strategy is often (though not always) the soft-assembly of a solution. Ecological controllers that can learn on-the-go promote soft-assembled solutions, that is to say solutions that comprise a temporarily stable assembly of factors and forces recruited from whatever happens to be available. Soft assembly itself is a notion developed in movement science according to which:

> "Movements can be seen as 'softly-assembled' patterns created and dissolved as tasks and environments change, with some patterns easy and preferred, and others more difficult and unstable… Moreover, as these synergies are assembled, they also take advantage of the non-neural aspects of movement: effects of gravity, elastic properties of muscles and inertial effects"
>
> Thelen and Bates (2003)

Humans belong to the interesting class of what I'd like to call open-ended ecological controllers. These are systems that 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 it deeply, creating whole new unified systems of

distributed problem-solving. Robotoddler, as we just saw, does this open-endedly for walking, but can apply that kind of learning to nothing else. We humans seem able to apply the same kinds of ecological-control learning much more widely. In fact, human agents seem highly engineered so as to be able quite generally to learn to make maximal problem-simplifying use of an open-ended variety of internal, bodily, or external sources of order. For example, we can learn to use tools, sports racquets, and musical instruments in ways that exploit the intrinsic dynamics of those material structures (for discussion and lots of examples, see Clark (2003) and Clark (forthcoming)). But more importantly for present purposes, there seems to be something analogous to (possibly identical with- see Christiansen (2004)) ecological control that operates in the cognitive domain.

Thus suppose we ask: could human cognition involve ecologically-controlled, soft-assembled, distributed systems built of heterogeneous parts? I believe (and see Clark (1997) (2003), Clark and Chalmers (1998)) that the answer is 'yes' and that minds like ours are distinguished, in part, by the ease with which brains like ours form larger problem-solving wholes that incorporate and exploit extra-neural stores, strategies and processes. These larger problem-solving wholes, I would like to argue, are not simply extended cocoons for the 'real' selves, choosing agents and cognitive engines hidden deep within. Rather (or so I wish to suggest) they then *are* those selves, agents and cognitive engines. To creep up on this initially unsettling idea, let's next turn to the very idea of the self.

2. Situating the Self

Contemporary work on the nature of the self, and on the origins of the sense of self, tends to distinguish between two major contributing elements. First, there is something like a core sense of 'being', which involves having a point of view on the world (I see the world from *over here*), having a sense of what I can and cannot do (I can reach and grasp that cup, I cannot fly or jump to the roof, etc) and having a sense of the limits and placement of my own body in space. Dogs, cats and rabbits, as well as human beings, have this core sense of their own embodiment and abilities. But second, in the human case, at least, there is also something like a complex 'narrative self'[iv]. This is (roughly speaking) an understanding, co-constructed by myself and others, of the kind of person I am, the kinds of projects and interests I have, the shape of my life so far, and so on.

Both elements can be significantly impacted by the opportunistic recruitment of external props and aids. Fitted with a shiny new prosthetic arm that can lift more weights than before, my direct, automatic, sense of what I can and can't do must rapidly alter and catch up. Fitted with a cochlear implant that cures my deafness and (as a kind of added extra) allows me to hear sounds in ranges that most adult humans cannot detect, my core sense of my own auditory potential again changes. Accustomed to the (now automatic and unreflective) use of, say, a retinal display that allows me to invisibly retrieve information from a plug-in or courtesy of a wireless accessible database, it seems less and less clear where what "I"

know stops and what "it" (the plug-in) makes available starts.

If the latter claim seems less plausible than the others, imagine an agent who knows, pretty much by rote, a lot of facts about women's basketball. Now imagine that instead of storing all of those facts in your head, you deploy a kind of heads-up display that provides instant access to the main performance statistics of key players over the last 20 years. The display might be delivered by eyeglasses, or even courtesy of a wireless implant sending signals directly into visual cortex (rather like certain new-generation cochlear implants-see Clark (2003) chapter 1). The system is set-up so that the visual sighting of a player's name, or the auditory pick-up of that name, or simply mouthing the name, activates a kind of Augmented Reality visual overlay displaying all the key facts and figures. Imagine, too, that the system is fairly flexible, allowing you also to start with categories (eg 'three-point field goal percentages in the year 2000') or with specifics ('players with three-point field goal percentages of .350 or above') and to then retrieve information accordingly.

Over a period of use, you become so accustomed to this easy, on-demand access that the plug-in (for want of a better word) becomes automatically, unthinkingly deployed, and that you usually trust what it delivers. As a result, perhaps you start to behave, and subsequently to feel, as if you simply *know* which of any two players had the best three-point field goal percentage in any given season etc. Would you be wrong to feel that? The answer is by no means cut and dried. True, your knowing these things depends on the proper operation of

the plug-in. But your knowing other things depends, equally, upon the proper operation of parts of your brain. And in each case, damage and malfunction is always a possibility. True, you need to retrieve the information before it is present to your conscious awareness. But knowledge stored in long-term biological memory is in the much the same boat, until some kind of retrieval process poises it for the control of verbal report and intentional action. (For a longer and more philosophically nuanced version of this much-compressed argument, see Clark and Chalmers (1998)).

A helpful, though at best partial, parallel is with the way our sense of 'seeing the whole visual scene before us' depends, on some contemporary models, upon the way information that is not currently represented in conscious visual awareness, and that is 'stored' only in the external scene itself is nonetheless (thereby) poised for easy access by a simple saccade. That poise-for-easy-retrieval is taken for granted in our daily planning and acting, and may be the source of our feeling that we see detail and color throughout the whole of the visual field (for discussion, see Clark (2002a), Noe (2002)).

This is not to say that there are no interesting differences. For example, knowledge stored in long-term biological memory is open to all kinds of subterranean processes of integration and interference (with both old and newly acquired knowledge). And neurally-stored information is fluently accessible by an amazing variety of routes, and in a wide variety of situations[v]. Nonetheless, the simple feeling of 'already knowing' the answer to a question as soon as it is asked is

surely the knowledge-based equivalent of the more familiar notion of 'transparent equipment': equipment (like the carpenter's hammer) with which are so familiar and fluent that we do not think about it in use, but rather rely on it to mediate our encounters with a still-wider world. Easy access to specific bodies of information, as and when such access is normally required, is all it takes for us to factor such knowledge in as part of the bundle of skills and abilities that we take for granted in our day to day life. And it is this bundle of 'taken-for-granted' skills, knowledge and abilities that - or so I am suggesting- quite properly structures and informs our sense of who we are, what we know, and what we can do.

3. The Shrinking Chooser

If this still feels unnatural, it is largely (I suggest) because we have in any case only the most tenuous collective grip on what it means to be a choosing, acting 'self', or a unified 'mind', and because we suffer from a chronic tendency to misconstrue the relations between our self-conscious "choosings" and the vast webs of non-conscious processing activity (all those whirrings and grindings of machinery, neural and perhaps non-neural, internal and perhaps external) that also structure and determine our own actions and responses. Until we form a better, more consistent image of the relationship between these factors, we cannot hope to know ourselves.

There is, most (though not all) theorists will agree, a genuine (though not always sharp and 'all-or-nothing') distinction between those things of which I am consciously aware and those things of which I am not[vi]. Right now, for example, I am conscious of the page in front of me, of the glare of my desk lamp, and of the difficulty of formulating this particular thought! I am not, however, conscious of all the complex low-level visual processing (the parallel processing of multiple differential equations) that supports and makes possible my conscious visual awareness of the page and the glare. Nor am I conscious of whatever complex internal machinations underlie my sudden sense than I am here tip-toeing into difficult and dangerous territory. Certainly, at any given moment, not *all* the cognitively important goings-on in my brain are present as *contents* of my current conscious awareness. That is why we sometimes find thoughts and ideas, ones that we nevertheless recognize as originated by *ourselves*, simply "popping up in our heads": they are the intrusive conscious fruits of some ongoing, subterranean, non-conscious information processing.

It is impossible to underestimate the significance of these non-conscious cognitive processes in the determination of the mental character of a persisting and identifiable thinking being. We must reject the seductive but ultimately barely intelligible idea that we (qua individual, thinking things) are nothing more than a sequence of conscious states. The identification of the agent or chooser with such a thin slice of themselves obscures the full suite of mechanisms

whose co-ordinated action is responsible for much of what is distinctive of an individual chooser. The shrinkage would leave us with no real grip on the *cohesion* and *continuity* that we naturally associate with the idea of a single mind, self or chooser persisting through time.

It is illuminating, I think, to actually try the following simple experiment. For just ten minutes, keep track (as far as you can) of the contents of your conscious awareness. Unless you are totally engaged in a single all-absorbing task, you will probably end up with a sequence of often-unconnected thoughts. A feeling of hunger, a thought about consciousness and the self, a worry about a forthcoming lecture, a glimmer of sexual arousal, a pang of anxiety, an urge to write to an old friend, another thought about the self (a good one- but where did it come from?), the strong desire for a coffee, etc., etc. The sequence of conscious contents is highly varied (in type) and radically discontinuous (in content). Themes persist, and whole trains of thought are sometimes, painfully, birthed. But the true principles of continuity, and the bulk of the thinking, choosing self, must lie largely underground.

We cannot, it seems, afford to identify ourselves with the conscious contents of momentary time-slices. With this in mind Dennett (2003) responds to worries (Libet (1985) (1999)) concerning the time-lag between onset of action and conscious awareness of a decision by noting that:

"Our free will, like all our other mental powers, has to be smeared out over time, not measured at instants. Once you distribute the work done…in both space and time in the brain, you have to distribute the moral agency around as well. You are not out of the loop; you are the loop"

Dennett (2003) p. 242

Taking the whole loop of temporally and spatially spread cognitive activity seriously is, however, already to take the crucial step towards understanding *ourselves* as ecological control systems capable of incorporating external structures deep into their cognitive routines. For the choices before us are now relatively stark:

Either, treat the mind and self as nothing but the shifting set of momentary 'conscious' contents (thus shrinking mind and self beyond recognition)

Or, allow mind and self to depend upon the ongoing co-ordinated activity of multiple temporally spread conscious and non-conscious processes, thus inviting us to also consider certain *non-biological* members of the class of non-conscious processes to contribute as deeply to the mechanical underpinnings of our minds and selves as do the non-conscious

neural processes.

The challenge, in other words, is this. Given the profound role of non-conscious, opportunistically recruited neural resources in the intentional origination of action, show us why (apart from some un-argued prejudice) the machinery of mind and self should be restricted to the neural, the inner, or the biological. We need to seriously question the idea that neural, inner and/or biological goings-on are in some way incredibly special. We need to rid ourselves of the idea that our brains are somehow touched with the magic dust that makes them suitable to act as the *physical machinery* of mind and self, while the non-biological stuff must forever remain mere *slave and tool.* The relations between our conscious sense of self (our explicit plans and projects, and our sense of our own personality, capacities, bodily form, location and limits) and the many non-conscious neural goings-on that structure and inform this cognitive profile are, it seems to me, pretty much on a par with the relations between our conscious minds and various kinds of transparent, reliable, robust and readily accessed non-biological resource. When those resources are of a recognisably knowledge-and-information based kind, the upshot is an extended cognitive system: a bio-technologically hybrid mind, a bio-technologically hybrid self.

Is this merely arguing over words? Why should we worry whether we accept well-integrated bio-external elements as aspects of the physical machinery of self and mind? The deepest reason to care, it seems to me, is that to fail to do so is to implicitly accept a model

according to which the machinery of the self becomes identified with the machinery of conscious reason. This leads directly to a variety of fears and worries (Wegner (2002) is a nice example) concerning the authorship of many perfectly intentional actions. The remedy is to see that the choosing agent is not somehow hidden within the machinery whose operations are most accessible to consciousness. Rather, she is the whole well-tuned ensemble capable of taking responsibility for actions and of initiating actions that make sense in the light of her long-term projects and commitments.

Gallagher (2005) extends the Dennettian view in just this way, writing that:

"I don't disagree with Dennett concerning the role played by nonconscious elements, except that I think we are even larger than he thinks- we are not just what happens in our brains. The 'loop' extends though and is limited by our bodily capabilities, into the surrounding environment, which is social as well as physical, and feeds back through our conscious experience into the decisions we make"
Gallagher (2005) p.242

Confronted with this general class of proposal, according to which the machinery of mind extends out into the surrounding world, many people feel deeply uncomfortable. How, they ask, could something to which I (they mean the conscious mind) may have so little detailed

access count as in any way a proper part of *me*? How could, to take an admittedly extreme case, the ongoing daily computations of a software agent that only ever reports back when it has found some tasty mercantile morsel count as part of my *own* mechanistic underpinnings? We may feel easier about this, I suspect, once we face up to the fact that this kind of relation already obtains between the conscious mind and sizeable chunks of our on-board neural machinery.

To see this quite concretely, imagine you are now reaching for that coffee cup sitting before you on the desk. It may seem to you that your hand and finger motions are being sensitively guided by your own conscious seeing. This, however, is not really the case. In fact, fine-tuned reaching and grasping involves the delicate use of visually-received information by functionally and neuro-anatomically distinct sub-systems operating, for the most part, outside the window of conscious awareness. It is this non-conscious circuitry that guides the most delicate shape-and-position sensitive aspects of reach and grasp. (For a radical version of this, see Milner and Goodale (1995). For a more balanced account, see Jacobs and Jeannerod (2003)). In a similar way, it is the non-conscious use of visual information that is responsible for many of our fine postural corrections and compensations (standing up while riding a bus, for example).

Even in the case of our own biological brains, then, the conscious self is in (direct, micro-managing) control of much less than we think. Not just the "autonomic" functions (breathing, heart-beat, etc.) but all

kinds of human activities turn out to be partly supported by quasi-independent non-conscious sub-systems. This is no surprise, I am sure, to any sports player: it doesn't even *seem*, when playing a fast game of squash, as if your conscious perception of the ball is, moment-by-moment, guiding your hand and racket. Nor should it come as a surprise to artists and scientists, who are often painfully aware that the bulk of their own (intentional, owned, self-expressing) creative activity flows from subterranean and non-conscious sources.

What seems to matter, for our daily sense of effective agency and choice, is (i) that the conscious mind has a rough and fallible sense of what she (the embodied, embedded, perhaps technologically extended, agent) knows, wants, and can and can't do, and (ii) that sometimes at least, conscious rehearsal can be an active part of the process that leads, within that complex economy, to intentional action. The conscious contribution here need amount to no more than a gentle but subjectively experienced nudge that tips the balance of a complex, and to a large extent unconsciously self-organizing, system.

The conscious mind, I am thus suggesting, is well-placed to act as an ecological controller in the sense outlined in the opening sections. It is for this reason that the fear of "loss of control," as we cede more and more to non-conscious inner processes and to a supporting web of non-biological scaffolding is misplaced. For what matters is not that the conscious self be micro-managing every detail of every sub-routine, but that working together, the conscious mind and a variety of non-conscious sub-systems provide useable, robust support for the

kinds of life we lead and the kinds of activity we value. The conscious mind, on this model, finally emerges as something like a new-style business manager whose role is not to micro-manage but to set long-term goals, pursue some slow deliberative reasoning, and gently nudge the larger system in certain directions, all the while actively creating and maintaining the kinds of conditions in which the overall distributed cognitive economy performs at its best[vii].

4. The Buck that Never Stops

A common objection, at about this point, goes something like this: even if external, non-biological elements do sometimes help us – quite profoundly – in our problem-solving activities, still isn't it always at least our *biological brains* that have the final say? The mental buck stops there. The brain is where *I am* because the brain is the controller and chooser of my actions in a way this other stuff (software, pen, paper, palm-pilot) is not.   And *that*, it may be suggested, is why the non-biological stuff should not count as part of the *real* self or cognitive system, and why our minds are not hybrids built of biological and technological parts. Human minds, so the obvious objection goes, are good old fashioned biological minds, albeit ones that enjoy a nice wrap-around of power-enhancing tools and culture.

This sounds sensible and proper. Until we turn up the magnification

on the biological brain itself. Notice first that many processes involved in the selection and control of actions *are* routinely off-loaded, by the biological brain, onto the non-biological environment. Think of the knot in the hanky, the automated desktop diary, and the software agent empowered to purchase. In reply to such an observation the sceptic is likely to invoke some kind of *ultimate* authority" "W*ho* was it that decided to knot the hanky, the sceptic demands? The biological brain, that's 'who', and that's YOU. *Who* was it that empowered the software agents to purchase? The same old brain, the same old YOU!"

But this reply is a major hostage to fortune. For suppose we now ask some parallel questions *within* the neuro-biological nexus itself. Do we now conclude that the real "me" is to be identified only with those select elements of the neural machinery involved in *ultimate* decision-making? Suppose only my frontal lobes have the final say? *Does that shrink the physical machinery of mind and self to just the frontal lobes?* What if, as the philosopher Daniel Dennett suspects, no neural sub-system has always and everywhere the final say? *Has the mind and self simply disappeared?* As Jerry Fodor once said:

> "If, in short, there is a community of computers living in my head, there had also better be somebody who is in charge; and , by God, it had better be me" Fodor (1998) p 207.

What we really need to *reject*, I suggest, is the seductive idea that all these various neural and non-neural tools need a kind of stable, detached user. Instead, it is just *tools all the way* down. Some of those tools are indeed more closely implicated in our conscious awareness of the world than others. But those elements, taken on their own, fall embarrassingly short of constituting any recognizable version of a human mind or of an individual person. Some elements, likewise, must be more important to our *sense* of self and identity than others. And some elements will play larger roles in control and decision-making than others. But this divide, like the ones before it, tends to cross-cut the inner and the outer, the biological and the non-biological. Different neural circuits provide different capacities, and contribute in different ways to our sense of self, of where we are, of what we can do, and to decision-making and choice. External, non-biological elements provide still further capacities, and contribute in additional ways to our sense of who we are, where we are, what we can do, and to decision-making and choice. But no single tool amongst this complex kit is intrinsically thoughtful, ultimately in full control, or plausibly identified as the inner 'seat of the self'. We (we human individuals) just *are* these shifting coalitions (see Ainslie (2001), Ross (this volume)) of tools. We are   "soft-selves," continuously open to change and driven to leak through the confines of skin and skull, annexing more and more non-biological elements as aspects of the machinery of mind itself.

The metaphor of 'tools all the way down' may seem to threaten, as Don Ross (personal communication) notes, to deconstruct itself. How can 'tools' be the right notion in the absence of a central intelligent user? The air of paradox is intentional. For the user is nothing but the cumulative effect of the co-active unfolding of the various resources supporting different aspects of adaptive response. This unfolding is determined by a delicate mix of sparse ecological control and pure self-organization. The user is what we see (in others and ourselves) when all this is working properly: a more-or-less rational being pursuing a more-or-less unified set of goals and projects (see Rovane (1998)).

The role of others in all this is not to be underappreciated. Ross (this volume) makes a powerful case that that humans (like other animals) communicate and bargain non-linguistically using multiple analog signaling systems (think of the continuous variety of small facial and bodily motions that may convey pleasure or displeasure, or encouragement to approach or retreat). But when such agents are also able to label their own states and those of others (pinning them down, for example, as "attracted" or "un-attracted", "interested" or "not interested" and so on) they enter into a new kind of arena, one whose dynamics can be stabilized by a series of such all-or-nothing (digital) commitments. (These are achieved, for example, whenever one agent labels another's state and the other does not reject the label). Further negotiations and co-ordinations among this group can then be predicated upon this stable base-line of publicly endorsed digital commitments. Ross depicts such practices as *recursively agent-*

generating: that is to say, the agents that enter into the digitally-modulated negotiation (or other co-ordination project) are defined, in part, by these very sets of commitments, and the new utility functions that accompany them.

Despite all this we are prone, it seems, to a particularly dangerous kind of cognitive illusion. Because our best efforts at watching our own minds in action reveal only the conscious flow of ideas and decisions, we mistakenly identify *ourselves* with this stream of conscious awareness. Then, when in our more scientific moments we begin to enquire into the material and physical underpinnings of the mind and self, it can quickly seem as if much (though not all) of the brain and all the rest of the body, not to mention the surrounding social and technological webs, are just tools for that conscious user. This is the mistake that led Avicenna, the Islamic philosopher quoted at the start of the chapter, to depict his own bodily limbs as "no more than garments". But garments for what? A conscious Cartesian self perhaps? To pursue this route is to embrace a hideously disfigured image of the mind and self , privileging a vanishingly small piece of the true personal and cognitive pie.

A better bet, as we have already begun to see, is the de-centralized, distributed, heterogeneous vision of the machinery of mind and self powerfully championed by the philosopher Daniel Dennett (see especially Dennett (1991)). Much of Dennett's work sets out to oppose the persuasive image of the "Cartesian Theatre": the mythical place inside our brains where sensory inputs, thought and ideas are

all inspected by a "central meaner" whose well-informed choices determine our deliberate actions. Dennett marshals a plethora of philosophical, psychological, and neuroscientific evidence against such a view. His target is often thought to be simply the idea of a neural or functional center of *consciousness*. But Dennett's deeper quarry, or so it has always seemed to me, has been the idea of a central *self*, a small-but-potent internal *user* relative to whom all the rest – be it neural, bodily or technological- is mere tool-kit. Where we hallucinate a central self, some spiritual or neural point wherein our special individual essence resides, Dennett finds only a grab-bag of tools, and an ongoing narrative: a story we (the ensemble of tools) spin to make sense of our actions, proclivities and projects. According to Dennett, we are our own best story, and our sense of self is a kind of artifact, useful for many purposes, but best taken with a pinch of salt.

I shall not rehearse or critique Dennett's arguments here (though I have done so at some length elsewhere- see for example Clark (2002b)). Instead, I simply note that our earlier reflections lead us to the very same conclusions, unpalatable though they may initially seem. There is *no self*, if by self we mean some central cognitive essence that makes me who and what I am. In its place there is just the "soft self": a rough-and-tumble control-sharing coalition of processes – some neural, some bodily, some technological – and an ongoing drive to tell a story, to paint a picture in which "I" am the central player.

Giving up on the image of a hard central self raises a thorny problem. What, then, makes a grab-bag of tools (a grab-bag whose specific elements may shift and change over time) into a unified, cohesive self? Part of the answer, to be sure, is that we simply hallucinate more unity and cohesion than in fact exists. Related to this is the pragmatic point that for many social and legal purposes, it is convenient to simply identify the agent with the core biological ensemble. We imprison the body and brain, not the laptop! But we do this despite knowing that individual bits of neural circuitry (my hippocampus, let's say) are themselves as incapable of being 'guilty' as the laptop! What we are really doing is rejecting a pattern of behaviour that has itself emerged from a whole social and bio-technological matrix.

But another, perhaps more interesting, part of the answer is that the unity and cohesion of the self, and the distinctness of the self (the sense we have of being individual agents, located thus-and-so, confronting a wider world) are not simple givens. Instead, they are (imperfect and constantly vulnerable) *achievements*. In other work I have discussed how the sense of location and body boundaries is constructed on the basis of co-ordinated signals arising within perception-action cycles. (see Clark (2003) chapter 4). The pre-conditions for the emergence of a rich sense of self begin to be met, I suspect, when on the basis of such information a loose-knit system begins to *stabilize* itself and to actively protect its own problem-solving infra-structure.

Thus reflect on the (superficially disproportionate) vexation of the child whose parents enter and slightly re-arrange her bedroom when she is not around. The feeling is one of almost personal assault. The room, organized in a certain way, was integral to the child's modes of play and study. To borrow an even simpler example, most of us keep our drinking glasses in a certain cupboard in the house. By actively stabilizing this environmental structure (putting clean glasses back in that same cupboard) we simplify the problem of future glass-location. Or consider the way files are arranged and stored in your own office. Our offices are organized in highly individual ways, dovetailed to our specific needs and to our different neuro-biological profiles We human beings actively organize our own local environments for cognitive purposes, and then take steps to protect this achieved organization (woe to the cleaner who disturbs the piles). Again and again we act so as to stabilize our local environments in ways that simplify or enhance the problem-solving that needs to be done. All this is a close cousin, I claim, to our carefully constructed and defended notion of a bounded self. The narrative-spinning drive (clearly evidenced in well-known studies of the tendency towards confabulation - see e.g. Nisbett and Ross (1980)) when confronted with such active efforts at stabilization, tends (I conjecture) to project the principle of stabilization further and further inwards, inclining us to hallucinate a single, central organizing self.

Imagine a pile of sand, deposited roughly on the ground, which is slowly settling into a stable arrangement of grains. Were the pile of sand self-aware, it too might hallucinate a kind of inner essence: a

special grain or set of grains whose deliberate actions sculpt the rest into a stable arrangement. But there is no such essence. The sand pile simply self-organized into a more-or-less stable coalition of grains. Similarly, (and here I also refer the reader to Dennett's classic (1974) discussion of the 'law of effect') certain coalitions of biological and non-biological problem-solving elements ('grab-bags of mind-tools') prove more stable and (hence) enduring than others. These configurations have a tendency to preserve and even repeat themselves. When viewed by a conscious, narrative-spinning element, this all looks like the work of some central organizer: the real self, the real mind, the real source of all that observed order. Thus, perhaps, is born the image of the self as a critical yet vanishingly slim slice of the overall problem-solving ensemble (brain, body, cognitive technologies): a slice so slim and elusive that our own neural circuits ('my' hippocampus, 'my' frontal lobes) can quickly seem like its tools!

That little story is mere speculation. But however it arises, this notion of a real, central, yet wafer-thin self is a profound mistake. It is a mistake which blinds us to our real nature, and leads us to radically undervalue (and misconceive) the roles of context, culture, environment and technology in the *constitution* of individual human persons. To face up to our true nature (soft-selves, distributed de-centralized coalitions) is to recognise the inextricable intimacy of self, mind and world.

5. Finding the Balance.

But not so fast. Isn't there a tension in the story as I have told it? On the one hand, I want to speak of the conscious mind as a source (or sources) of ecological control, adding crucial nudges to the complex dynamics of much larger (potentially hybrid) cognitive and behavioral systems. On the other hand, I want to depict us as soft selves: de-centralized, distributed and self-organizing systems that just happen to have a perspective on their own activity and an accompanying story to tell. How can both stories be true?

A suggestive move is made by Velleman (2000), who notes that our own verbal utterances, encountered in the context of a narrative-spinning drive, may at times serve the purpose of helping to bring about the very thing they depict us as about to do. This is a function they could perform even in the context of a system that is largely self-organizing in just the way Dennett suggests. For example, if late at night I say to you "I'm leaving now", I may say this in part as a means of bringing it about that I do, in fact leave - that I leave, shall we say, despite my self-perceived urge to continue our conversation. This strategy can work, according to Velleman, because as a rampant narrator of my own life story, I am driven to strive for consistency, for alignment between behavior and narrative. So by saying that I am about to do so and so, I increase the costs to myself of not so doing. This trick (the addition of a new narrative-induced cost to not doing so and so) would work for silent inner rehearsal as well as for overt speech. The conscious thought or overt utterance that depicts me, to myself, as about to do so and so thus makes it more likely that I will in fact do what I think or say. The key (as Ismael (ms) (forthcoming)

nicely notes) is simply to make the narrative run ahead, into the future, so that the drive for consistency and alignment acts as a causal influence on what we then do, rather than a post-hoc (and sometimes demonstrably ad hoc[viii]) commentary on what we have already chosen or done.

Ismael (ms) (forthcoming) then uses this as the springboard for a related but much stronger move. What is right in Velleman's position, she suggests, is his recognition that "if the brain is generating a self-representation, there's no reason that the thing can't have a role in the determination of behavior, indeed, no reason that it can't acquire an increasingly prominent role" (ms p. 6). (By a 'self--representation' Ismael seems to mean any kind of internal model that contains even a partial representation of the agent as a distinct property-bearing individual, so any Dennett-style narrative-spinning engine should count as a self-representation). But as soon as such a self-representation assumes some kind of causal role (whether by the neat trick of forward-narrating or any other means), we are no longer (according to Ismael) dealing with a simple self-organizing system that just happens to spin a story about itself. Rather, we are dealing with a system that really does include a self-model as an active principle of organization. Not self-organization but organization by a self-model! Such systems may behave in ways that are more flexible (so the argument continues) than any first-order self-organizing system, because they can use the self-model to select between multiple responses in ways that are driven by that model, hence by stored memories, representations of goals and so on.

In this story, the stress on various forms of higher-level control seems to me exactly right, as does the observation that such control may often require the system to deploy some kind of model of its own nature, history, or dynamics. But for my own part, I remain wary of the idea of any single such self-model, and much more inclined (see Clark (1997)) towards a vision of multiple partial models, most of which may be relatively low-level (such as a variety of forward models of bodily dynamics- see below). As far as the most inclusive and abstract such 'self-model' goes (the one that corresponds most closely to the narrative self), I wonder what the notion of a model here adds to the simpler notion of systems that know some stuff about themselves and use that knowledge to help select actions? Perhaps then the contrast between self-organizing and self-governing systems is not as sharp as it first seems[ix]?

Ismael thus depicts a kind of conflict between what she calls self-governing models (ones in which a self-representation plays a crucial, flexibility-promoting causal role) and truly self-organizing ones, in which there is only input-driven self-organization plus a spun narrative. The notion of ecological control, I want finally to suggest, begins to show us how to reconcile these two notions[x], and hence how to reconcile Dennett's emphasis on the unreality of selves with Ismael's recognition that agents with some kind of self-model may exhibit complex dynamics that are the causal consequence of that very self-model. For what I am calling ecological control is what you get when you add an inter-animated and changeable variety of thin slices of self-governance to a system in which the intrinsic, self-

organizing properties of many subsystems are allowed to bear a lot (but not all) of the problem-solving strain. The self-actuating passive dynamic walker whose direction of travel and speed of walking is selected by a local supervisory system empowered to nudge and tweak the mechanically coupled leg system exploits a potent combination of self-organization and self-governance. Importantly, self-governance, as I am understanding it, may be the province not of any single inner self-model but of a variety of partial[xi] self-governing strategies, some of which may themselves involves circuitry spanning brain, body and world, and *themselves* orchestrated via a higher-level kind of self-organization. According to such a picture, overall self-governance, though real and important, is the emergent outcome of the action of a whole complex of partial self, body and world models acting as mini ecological controllers in a distributed cognitive economy. Some of these may be simple forward models of the dynamics of bodily sub-systems (see eg Miall and Wolpert (1996), Clark and Grush (1999)), while others may be more tightly woven with systems for episodic memory and for categorization.

In all cases, though, the self-governance works only *because* it is delicately and continuously keyed to (and often highly exploitative of) a variety of sources of order inherent in the rest of the system. True (flexible, efficient, robust) self-governance thus positively requires, or so I want to suggest, the use of 'soft' ecological control strategies: ones that are maximally exploitative of the order and intelligence that is distributed throughout the larger system. This suite of soft, hybridization-favoring, *partial* self-governing routines cannot

reconstruct a stereotypic Cartesian mind or a traditional single central self. To that extent we may appreciate what is correct in Dennett's depiction of (that kind of) self as a kind of illusion, while embracing the causal potency of (perhaps a variety of) empirically real partial self (and world) models, some of which link memory and motivation to action and choice.

Conclusions: Situations and Persons

How can we reconcile the vision of human agents as distributed, hybrid problem-solving ensembles with the vision of human agents as, indeed, *agents*: as autonomous individuals exercising control and choice? Viewing ourselves as loci of multiple systems of ecological control may provide a useful tool to attempt such a reconciliation. Ecological control systems are, first and foremost, essentially opportunistic and exploitative. They take whatever is around, and build it into fluent problem-solving routines. But they are nonetheless controllers in good standing, able to tweak and nudge complex systems in ways that promote goal-driven activity and flexible adaptive response. Seeing ourselves as biologically-based (but not biologically imprisoned) engines of ecological control may help us to develop a species self-image more adequate to the open-ended processes of physical and cognitive self-creation that make us who and what we are.

Acknowledgements

References

Ainslie, G. (2001) *Breakdown of Will.* New York : Cambridge University Press.

Arbib, M.A. (1993) Book Review: Allen Newell, Unified Theories of Cognition, *Artificial Intelligence*, 59:265-283.

Christensen, W (2004) Self-directedness, integration and higher cognition *Language Sciences*, 26: 6. In D Spurrett (ed) Special issue on Distributed Cognition and Integrational Linguistics.

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

Clark, A. (2002a) Is Seeing All It Seems? Action, Reason and the Grand Illusion *Journal Of Consciousness Studies* 9 :5/6.

Clark, A. (2002b) That Special Something: Dennett on the Making of Minds and Selves in A. Brook and D. Ross (eds) *Daniel Dennett* (New York: Cambridge University Press.

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

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

Clark, A. (forthcoming) Re-inventing Ourselves: The Plasticity of Embodiment, Sensing, and Mind *Journal Of Medicine And Philosophy.*

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

Clark, A and Grush, R. (1999) Towards a Cognitive Robotics" *Adaptive Behavior* 7:1: 5-16

Collins, S., Ruina, A., Tedrake, R., and Wisse, M (2005) Efficient Bipedal Robots Based on Passive-Dynamic Walkers *Science* 307, 1082

Damasio, A (1994) *Descartes' Error* New York: Putnam

Damasio, A (1999) *The Feeling of What Happens* New York: Harcourt Brace and Company.

Dennett, D (1974) Why the Law of Effect Will Not Go Away *Journal of the Theory of Social Behaviour* 5 166-187

Dennett, D (1978) Where Am I? in Dennett, D (ed) *Brainstorms* (MIT Press, Bradford Books)

Dennett, D (1991) *Consciousness Explained* New York, Little Brown

Dennett, D (1997) The Path Not Taken in N. Block, O. Flanagan and G. Guzeldere (eds) *The Nature of Consciousness* MIT Press, Camb. MA 417-420

Dennett, D (2003) *Freedom Evolves* London: Penguin

Fodor, J (1998) *In Critical Condition* Camb. MA: MIT Press

Gallagher, S (2005) *How the Body Shapes the Mind* Oxford: Oxford University Press

Haggard, P (2005) Conscious Intention and Motor Cognition *Trends in Cognitive Sciences* 9:6: 290-295

Ismael, J (ms) Emergent Order: the Limits of Self-Organization

Ismael, J (forthcoming) *The Situated Self* New York: Oxford University Press.

Jacob, P and Jeannerod, M (2003) *Ways of Seeing: The Scope and Limits of Visual Cognition* Oxford: Oxford University Press

Kelly, K (1994) *Out of Control* Reading, MA: Perseus Books

Libet, B (1985) Unconscious cerebral initiative and the role of conscious will in voluntary action *Behavioral and Brain Sciences* 8 529-566

Miall R.C. & Wolpert D.M. (1996) Forward models for physiological motor control. *Neural Networks, 9:* 1265-1279

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

Nisbett, R. E., & Ross, L. (1980). *Human Inference: Strategies and Shortcomings of Social Judgment.* Englewood-Cliffs: Prentice-Hall.

Noe, A (2002) Is the Visual World a Grand Illusion? in A. Noe (ed) *Is the Visual World a Grand Illusion?* Thorverton, UK: Imprint Academic. 1-12

Ross, D (this volume) Meta-linguistic signaling for coordination amongst social agents.

Rovane, C. (1998). *The Bounds of Agency.* Princeton: Princeton University Press.

Tedrake, R., Zhang, T., and Seung, H.S. (2005) Learning to Walk in 20 Minutes. In *Proceedings of the Fourteenth Yale Workshop on Adaptive and Learning Systems*, Yale University, New Haven, CT

Thelen, E. & Bates, E. (2003). Connectionism and dynamic systems: are they really different? *Developmental Science*, 6, 378–391

Velleman, J. D. (2000) *The Possibility of Practical Reason*, Oxford: Clarendon Press.

Wegner, D (2002) *The Illusion of Conscious Will* Cambridge, MA: MIT Press

---

[i] Ibn Sina Avicenna was a Persian philosopher-scientist who lived between 980 and 1037 A.D. The quote is from R. Martin's unpublished translation of his *De Anima* (*Liber de anima seu sextus de naturalibus*) vol 7 .

[ii] Stelarc is an Australian performance artist whose work explores the technological transformations of embodied experience. The quote is from the Stelarc website at www.stelarc.va.com.au

[iii] Much of this work originated in the Andy Ruina Lab, Cornell and with original work by Tad McGeer.

[iv] See Dennett (1991) chapter 13, Damasio (1999) chapter 7.

[v] For discussion of the impact of such differences on the arguments for the 'extended mind', see Adams and Aizawa (2001), Clark (2005).

[vi] Dennett might seem to reject this claim, as when he writes that "we cannot draw the line separating our conscious mental states from our unconscious mental states" (1991, p.447). But he does not deny that many key processes operate at what he himself initially dubbed the 'sub-personal level'. And certainly, Dennett would not deny that much of what structures my behaviour emanates from purely sub-personal processes. I think that Dennett's real target is the idea that within the smaller class of potentially directly reportable goings-on, there is a neat line between those that are at this moment conscious and those that are not. See e.g. Dennett (1997).

[vii] The skill of successful *self*-management is thus pretty much the same skill as management in certain new business sectors. In each case what matters is knowing how to exercise rather indirect, softly softly, forms of intervention and control: what Kevin Kelly (1994 p.330) nicely dubs 'co-control'. Co-control is what you get when an ecological controller has opportunistically recruited a motley of resources to deal with a current need or problem.

viii I am thinking here of the empirical evidence (though see section 4 above) to suggest that our chosen actions are frequently the results of unconscious activity that precedes the experience of conscious will. See e.g. Libet (1985), and more recent demonstrations such as Wegner (2002). For an interesting review, see Haggard (2005).

ix Indeed, Ismael notes that the contrast between self-governing and self-organizing systems is 'not a dichotomy' and that 'most animate systems fall somewhere in the space between those with fixed and those with flexible response functions' (ms)

x One potentially important difference is that Ismael thinks that the relation of true dynamical coupling, in which 'controller' (ecological or otherwise) and 'controlled' are engaged in a continuous reciprocal exchange, introduces important complexities and undermines the attempt to couch the relation in quite those terms. The notion of an ecological controller, Ismael might object, implies a kind of divisibility that ongoing coupled influence (between self-model and whole embodied system) cannot underwrite. Thus she writes that in such cases "there is no simple way to decompose the system into dynamically separable units" (ms p.10) For my own part, I don't yet see why this makes a significant difference. For the roles of the parts seem distinct enough, even if their ongoing co-evolution resists decomposition. For example, one part of the system may have access to stored memories and a self-profile while another, though constantly coupled to it, does not.

xi For something like this vision of multiple partial models as a mode of control, see Arbib (1993)