Rethinking Neuroethics in the Light of the Extended Mind Thesis

Neil Levy, University of Melbourne and University of Oxford

The extended mind thesis is the claim that mental states extend beyond the skulls of the agents whose states they are. This seemingly obscure and bizarre claim has far-reaching implications for neuroethics, I argue. In the first half of this article, I sketch the extended mind thesis and defend it against criticisms. In the second half, I turn to its neuroethical implications. I argue that the extended mind thesis entails the falsity of the claim that interventions into the brain are especially problematic where they are not confined within the skull of individual agents, but extends into the world. The thesis bears a family resemblance to, but is far more radical than, the view known in the philosophy of mind as content externalism: the view, associated with Tyler Burge (1979) and Hilary Putnam (1975), according to which the content of mental states is partially determined by social institutions and by the environment.

Keywords: neuroethics; philosophy of mind; internalism; science; psychopharmaceuticals; authenticity

Neuroethics exists at the confluence of many disciplines: neuroscience and ethics, most obviously, but also psychology, cognitive science, and philosophy of mind. Doing neuroethics justice, therefore, sometimes requires engagement in relatively obscure debates. In this article, I shall argue that one such debate, the debate over the mind’s location, has direct neuroethical relevance. Some thinkers claim that the mind is not contained within the skull, but extends beyond it, into the world. If their claim, the so-called extended mind thesis is true, then many neuroethical issues must be rethought. The extended mind thesis dramatically expands the scope of neuroethics, so that many apparently tangential questions are seen to be neuroethical questions. To that extent, the thesis increases the significance of neuroethics. At the same time, it forces us to drastically rethink our approach to traditional neuroethical concerns. In the first half of this article, I shall sketch the extended mind thesis and briefly sketch a defense of it against some of its critics. I shall then turn to the far-reaching implications of the extended mind thesis for neuroethics.

THE EXTENDED MIND THESIS

The name extended mind thesis is the name originally given to a radical view in the philosophy of mind by its two earliest defenders, the prominent philosophers of mind Andy Clark and David Chalmers (Clark & Chalmers 1998). The view has since been defended under other names: environmentalism (Rowlands 1999), locational externalism (Wilson 2004) and extended cognition (Rupert 2004), but Clark and Chalmers’ term is nicely descriptive and I shall use it here. The extended mind thesis is, roughly, the view that the mind is not confined within the skull of individual agents, but extends into the world. The thesis bears a family resemblance to, but is far more radical than, the view known in the philosophy of mind as content externalism: the view, associated with Tyler Burge (1979) and Hilary Putnam (1975), according to which the content of mental states is partially determined by social institutions and by the environment.

Whereas content externalism holds that the content of a mental state is partially determined by features external to the head of the person whose state it is, the extended mind thesis holds that the mental state itself—however it gets its content—is partially located external to the head of the person whose state it is. The extended mind thesis therefore denies that mental states are constituted solely by brain states.

At first sight, this is a startling, even absurd, claim. How can mental states be located external to the brain? The argument for the claim is essentially functionalist: some states and processes external to the brain play the same functional role in cognition as some internal states and processes, and therefore should be accorded the same status. Consider non-occurrent mental states, such as a belief that is not currently being entertained by a subject. Most of our beliefs and memories are like this, most of the time. Philosophers often call these dispositional beliefs, because subjects are disposed to entertain them under appropriate circumstances. Now, where are these mental states stored when they are not being entertained? Neuroscientists have an answer, or at least part of an answer, to this question. Memories are stored first in the medial temporal system, then (when they persist sufficiently long) distributed across networks in cortical regions.
(Schacter 1996). Now, what makes it the case that a memory is stored in a particular region? Just this: under the appropriate circumstances (for instance, in response to a question about where she had lunch), the subject recalls the memory, and that region of the cortex plays an appropriate role in the process of recall. Spelling out what kind of role is the appropriate one is a difficult task, but we can confidently say some things about it. We can say that something counts as a part of a subject’s memory if the role it plays in recall is causal, and if it encodes information that is activated in recall. The cortical regions identified by neuroscientists clearly satisfy these conditions. But these same conditions can be—indeed, are—satisfied by regions of the world external to the subject’s skull. Hence, non-occurrence mental states like memories and beliefs can be stored external to the subject.

Consider Clark and Chalmers’ (1998) famous example of Otto and his notebook. Otto suffers from dementia; he therefore takes the precaution of writing down everything he thinks he might need to recall. Otto hears that there is a new exhibition on at the Museum of Modern Art in New York. He takes out his notebook and flips to the appropriate page. There he reads that the museum is on 53rd Street. Armed with this knowledge, Otto sets out for the museum. Clark and Chalmers argue that since Otto’s notebook plays the same functional role in memory retrieval as neural states paradigmatically play in ordinary agents’ memory, we ought to count Otto’s notebook as part of his mind. More generally, proponents of the extended mind thesis defend the parity principle: if something plays a role in cognitive activity, such that, were it internal we would have no difficulty in concluding that it was part of the mind, it should be counted as part of the mind whether or not it is internal.

Of course, Otto is not a typical agent. He has a brain disease, which degrades his internal memory and leads him to substitute an external prosthesis. But extending one’s mind beyond the head is not something that only rare or damaged individuals do; it is not a second-best substitute for internal mental states and capacities. Instead, it is something that all humans do, to a greater or lesser extent. Indeed, there is a case for saying that the capacity for extending the mind is distinctive of human beings, not in the sense that no other animal possesses it (there is evidence, as we shall shortly see, that at least some other primates can extend their minds outside their skulls to some extent), but insofar as extending our minds is something that we are particularly good at, that we do by nature, and insofar as our ability to extend our minds underlies the many cognitive achievements that we rightly prize.

One way that extending our minds outside our skulls expands our abilities is by freeing up internal resources for other tasks. Storing representations outside our heads takes cognitive load off our brains. Consider two possible ways of representing the visual scene. One way we might have performed this task is by constructing and storing rich and detailed internal representations of the world around us. We do in fact construct internal representations of the world, but they are neither rich nor detailed. Why go to all the trouble and expend valuable resources on constructing such a model when the world is already there to serve as its own best model (Clark 1997)? It seems to us that we have a rich and detailed model of the world in our heads, but this is an illusion, caused by the reliable availability of facts about the visual scene. It seems to us that our internal representations are rich because they are constantly updated by our eye movements. The human eye has a very small area of high-resolution vision; less than 0.01% of the entire visual field. But our eyes constantly dart about, moving this window of high resolution across the visual scene. These movements, called saccades, are intelligent; they are not random, but instead gather information relevant to the tasks currently confronting the person. They are also very fast, averaging about three movements per second. Our frequent and repeated saccades allow us to inspect the world and update our picture of it, so that it seems to us that we have a rich representation of it. And so we do, but it is not an internal representation.

There is a cost to offloading the job of representing the world onto the world itself. The cost comes in the form of change blindness, our inability to notice even dramatic changes in the world under certain conditions (Rayner 1998; Simons & Levin 1998). However, because these conditions are rarely encountered outside the psychology laboratory, these costs are more than compensated for by the benefits that flow from it: the benefits of freeing cognitive resources for other tasks. Great though these benefits are, however, they are comparatively small when compared to the benefits that flow from cultural means of extending minds.

Consider how using external resources can expand the range of cognitive abilities even of chimpanzees. The unadorned chimp brain is able to learn to categorize pairs of objects on the basis of their similarity or differences from one another. So, for instance, chimpanzees can be trained to put any pair of identical objects—two cups, say, or two bananas—into one box, while placing any pair of dissimilar objects into another (say one cup and one banana). But without external aids, they cannot sort pairs of pairs by similarity or difference. Two pairs of pairs are identical just in case they share their first-order properties: they are either both identical, or they are both dissimilar. Otherwise, they are dissimilar. So the pair of (pairs) apple–banana and cup–shoe is identical, whereas the pair cup–cup and apple–banana is different. This higher-order task is difficult even for the human brain, but chimpanzees can learn to accomplish it by, in effect, turning it into a first-order task. They do this by learning to associate tokens with the first-order pairs. For instance, they might learn to associate a plastic triangle with a pair of objects that are identical to one another, and a plastic square with a pair of objects that are dissimilar. Once they have accomplished this task, the higher-order task is simple. If you want to know whether a pair of pairs is similar or dissimilar, simply compare the tokens associated with them: if they are
identical so is the higher-order pair, if not, they are dissimilar (Thompson et al. 1997).

Human beings, too, commonly solve complex cognitive tasks by turning them into much simpler tasks, using external resources to pull off the trick. Indeed, that may be the only way we can successfully complete these kinds of tasks. Consider basic arithmetic. Above a certain, quite low, threshold, arithmetic is beyond the power of the unadorned human brain. However, we can perform arithmetic (not to mention algebra, calculus, and so on) by turning more complex tasks into simpler ones. For instance, we usually perform multiplication tasks of even moderate complexity by using the algorithm we memorized in grade school, which turns a multiplication task of any complexity into a series of single-digit multiplications and addition tasks. In this way, we use external resources for working memory and for information processing (McClelland et al. 1986).

The resources we use in performing mathematics are multiply extended—across time and across space, across physical elements, and across agents. These resources are extended across time and across agents insofar as the cognitive process is reliant not only on pencil and paper, but also on the intellectual products of previous agents: the agents, long since dead, who developed the algorithms (as well as the chain of agents who handed it down across the generations). Both ways of extending our minds underlie our most sophisticated cognitive abilities. Speech gives us an ability unmatched by any other animal to externalize our thoughts, and thereby make them available for manipulation and for extension across agents, each of whom can bring their own unique set of abilities and experiences to bear upon them. Writing, the ability to preserve speech, dramatically increases our memory power. Moreover, writing makes possible the rise of systematic science, for science requires replication of experimental results, and we can only replicate results if methods are publicly available. By allowing for the slow accumulation of knowledge, writing allows each generation of scientists to build upon a foundation they inherit from the past. Of course, we inherit not only a body of claims from previous generations, but also a set of cognitive tools (from algorithms to computers). We think with these tools, which open up for us cognitive landscapes that would otherwise be, quite literally, unthinkable.

**OBJECTIONS TO THE EXTENDED MIND**

The extended mind thesis has many influential defenders, but it remains a minority view. Critics of the thesis have two main arguments against it. They claim, first, that mental states necessarily involve intrinsic content, where intrinsic content is something that only brains can instantiate. Second, they argue that the domain of the mental must be causally individuated, and that only internal states and resources can be individuated in the appropriate manner. I shall briefly sketch and reply to these arguments.

The notion of intrinsic content is best understood by opposition. Roughly, a representation has (merely) derived content if it refers in virtue of a convention. The sequence of letters CAT refers to the animal “cat” in virtue of a set of nested conventions: the convention, on the part of speakers of English, of using the sounds represented by those letters to refer to cats, and the convention of using the letters CAT to refer to those sounds. Our world is full of representations that have derived content, for example, words of natural languages, mathematical symbols, and traffic signs. But if a sign refers in virtue of a convention, its referential power is derivative from the referential power of the human mind. Now, how do the paradigm mental states of human beings come to refer? If we are to avoid an infinite regress, Adams and Aizawa (2001) argue, we must recognize that human minds are capable of states that are intrinsically referential. When I read the word CAT, I have a mental state that refers to cats, but my mental state itself refers to cats in virtue of its nature, not in virtue of any convention. If this were false, if my mental state referred only in virtue of a convention, we would need to ask how that convention, in turn, came to be meaningful: eventually, we must stop at a representation that is intrinsically meaningful, or representation could never get off the ground in the first place. Adams and Aizawa (2001) hold that only states of the brain naturally have content (in virtue of their evolutionary history). Hence, only brain states have intrinsic content. They therefore argue that there is a fundamental difference between brain states and external cognitive tools and resources, a difference that gives us a strong reason to deny the parity principle. Intrinsic content is, they claim, “the mark of the cognitive” (Adams & Aizawa 2001).

Some philosophers deny that mental states must have intrinsic content (Dennett 1990). But there is no need to take sides on this debate. Instead, we need simply to notice that even if all mental states must involve intrinsic content, they need not consist only of intrinsic content. As Clark (2005) sees, this opens the way for admitting derived content within the purview of the mental. Clark (2005) asks us to consider the mental manipulation of Venn diagrams. Suppose that Otto sees that some Xs are also Ys by picturing Venn diagrams to himself. Is he not thereby engaged in cognition? But clearly Venn diagrams derive their content from a convention (Clark 2005). Similarly, some ordinary human thought seems to be conducted in a natural language. We have no reason to exclude such thought from the domain of the cognitive. Consider, finally, songs and poems learned “by heart.” In order to activate their content—that is, for the agent to recall what they are about—she must rehearse them, either to herself or aloud. But if that is correct, then the agent has representations inside her head that do not (prior to rehearsal) have intrinsic content. If we should count these representations as mental, then even if agents are only capable of thought at all because some of their representations have intrinsic content, and even if all their mental states must involve intrinsic content, such content is not “the mark of the cognitive” at all—and nothing stands in the way of our treating states which include external representations and artefacts as mental states.
Let us turn, now, to the second claim, that the mind ought to be identified with the domain of processes and states that can be causally individuated, and that only internal processes and states “fit the bill.” Why think that genuine cognitive processes must be causally individuated; that is, distinguished from each other and from anything else in terms of their physical causes? The motivation for this claim seems to be the following: if a science of cognition is to be possible, then there must be a discoverable set of causal regularities or laws for that science to capture. And the available evidence suggests that there are indeed such laws. The processes that occur in human minds are somewhat diverse, but they are nevertheless relatively unified, inasmuch as they can be described by a relatively small set of laws. Within more narrowly circumscribed domains, the degree of unification is even higher. Thus, there is a set of rough, *ceteris paribus* laws that describe a wide range of human memory systems, short- and long-term; hence, we should expect the domain of the mental to be causally individuated (Rupert 2004).

The term *mind*, the defenders of the causal individuation view claim, refers (here on Earth) to the object of certain, rapidly progressing, sciences. But there is no science of the brain-plus-tools, and there is no prospect that there ever will be. The set of tools that might be taken to constitute the extended mind is just too large and too disparate to ever form the object of a science. Whereas even now cognitive scientists are discovering the causal regularities governing the mind, traditionally conceived, there is no chance that they will discover an interesting set of regularities which circumscribes the domain of brains-plus-notebooks, brains-plus-calculators, brains-plus-tags-in-the-sand, and so on.

In response to this objection, Clark (In press) stresses that we ought not to try to second-guess the progress of science. We cannot infer, from our current inability to unify extended cognition under a set of laws, that this science will not succeed where we fail. I think, however, that we ought to concede to the critics that the probability of future unification is low. The set of processes and objects already supposed to extend our minds into the world—marks in clay, tattoos on the body, bits and bytes, and many others besides—are already very disparate, and we have every reason to think that the ongoing technological explosion will deliver us many more, and different, ways of manipulating and storing information. Of course we shall discover—are already in the process of discovering—laws that will apply to all these resources, as well as to the brain. But that is not the question at issue; the question is, will we be able to delineate laws that *uniquely* circumscribe the domain of thought, where “thinking” is an activity distributed across brains and tools? And the answer to this question is overwhelmingly likely to be negative. Now, if science “carves nature at the joints,” then science may be a good guide to ontology, to the nature of reality: if a relatively discrete part of the universe can be captured by a set of scientific generalizations, then we have good reason to think that that part of the universe constitutes a natural kind. Thus, the fact (if it is a fact) that there is a set of causal regularities which applies to the brain/mind, and no such set that applies to the brain/mind plus its various add-ons gives us reason to think that the brain/mind constitutes a discrete entity, and the extended mind does not. If that is right, then the mind should be understood as an entity circumscribed at the relatively low level of causal regularities, and not at the functional level.

However, it is simply false that a science of the extended mind would be especially disunified, by the standards of existing sciences. A number of other sciences circumscribe their domains on the basis of higher-level similarities between phenomena, and not lower-level causal regularities. One of Adams and Aizawa’s (2001) own examples illustrates the point. The science of animal communication includes causal processes as disparate as communication by the use of pheromones, threat displays, the dance of honeybees, and territory marking by birds, as well as natural language in human beings. There are few general laws that circumscribe all and only these phenomena: instead they are unified by their functional similarities. Why think that the subject matter of the science of mind should be any more unified?

Adams and Aizawa (2001) do not tell us what the difference between the two putative sciences is, such that only one is entitled to call itself a science. Here is one possible argument that might be advanced on their behalf. They might claim that whereas the science of the mind/brain can be circumscribed by a set of causal laws, none of the elements of the science of animal communication is internally unified, and therefore functional unification is the only kind of unification to be had here. But this claim is simply false: the different elements of the science of animal communication *could* each form the domain of a separate science, understood in the way Adams and Aizawa want to.

Moreover, the science of the mind, *traditionally* understood, is itself a motley. The mind/brain, as traditionally conceived, contains a number of quite different mechanisms, which are sufficiently diverse to cast doubt on the claim that they form a single domain. Consider, for instance, the distinction often made between controlled and automatic processes (Bargh & Chartrand 1999; Wegner 2005). Both processes issue in complex behavior in human beings. But there are significant differences between them. Controlled processes are generally conscious; they are also very demanding of cognitive resources, such that performance at them degrades significantly under cognitive load (we place subjects under cognitive load by requiring them to do two tasks at once: the controlled task we want to study and another, loading, task, such as counting backwards from 1000 in units of three). Automatic processes are not consciously initiated, and do not degrade under cognitive load; they are also ballistic, which is to say that they cannot be interrupted once they are begun. Controlled processes are slow, automatic processes are fast. Now, if it is true that causal regularities pick out natural kinds, then the mind is not a natural kind: it is a compound entity, composed of at least two (and probably many) natural kinds. Either the critics of
the extended mind should give up their commitment to the mind/brain, in favor of many brain modules and processes, or they should accept that functional similarity is sometimes sufficient to circumscribe the domain of a science. Of course if they take the latter route, they no longer have a reason to resist the extended mind hypothesis.

Far from lamenting the loss of a unified science of the mind/brain, Wilson (In press) looks forward to the day when representational practices are the object not of a single science, but of “an interdisciplinary, pluralistic motley,” with input not only from psychology and neuroscience, but also from sociology, anthropology and evolutionary biology. It is worth pointing out the extent to which some of these sciences already play a significant role in the study of the mind, even as it is currently—and narrowly—conceived. Hypotheses are formulated, for instance, on the basis of evolutionary considerations, such as considerations concerning the environment in which cognitive adaptations developed; information from anthropology concerning the typical size of groups of early hominids has also guided research in psychology (Dunbar 1996). In any case, there is no need to choose between the sciences of the mind, traditionally conceived, and the interdisciplinary endeavor envisaged by Wilson (In press). They can co-exist, and cross-fertilize one another. But specialists in the traditional sciences of the mind should not deceive themselves into thinking that the mind they study is unaffected by external resources; even when it is abstracted from environmental props it will likely prove to have somewhat different characteristics for having been embedded.\(^1\)

There is far more to be said; other lines of attack against the extended mind thesis, and further replies to these criticisms. I cannot hope to do justice to all these criticisms here; those wishing to pursue the debate further may do so elsewhere.\(^2\)

It is enough for the moment to have shown that the thesis is worth taking seriously. At least for the purposes of this article, let us accept that it is true. For if it is true, it has far-reaching implications for neuroethics.

**NEUROETHICS AND THE EXTENDED MIND**

 Neuroethics is important because the issues it deals with concern our very identities, at least in one sense of that multivocal word. In what Marya Schechtman (1996) calls the characterization sense of identity, we are constituted by our mental states, especially, but not only, our memories. Because the science of the mind seems to promise, or threaten, the ability to intervene in these mental states—to erase or insert memories, to alter personality traits and dispositions—or even just to peer into them without our permission and perhaps even without our knowledge, its powers are widely and perhaps rightly regarded as especially troubling.

If mental states are not confined within the brain, however, claims that neurological interventions are especially problematic, or unique in their nature, look to be in need of more defense than they are usually given. Accepting the extended mind thesis does not commit us to denying either of these claims. But it does require us to reject two claims which are widely seen as their corollaries: 1) that interventions into the brain are uniquely and distinctively interventions into the mental states that constitute our identities; and 2) that neuroethics is concerned with the question of whether we ought to allow interventions into the mind. These claims are rarely advanced explicitly, but they are widely (perhaps universally), assumed.

Claim 1 underlies the intuition that neurological interventions, whether by way of psychopharmaceuticals, transcranial magnetic stimulation, or direct brain stimulation, are fundamentally different from more traditional ways of altering mental states (for example, by using traditional psychological methods, which range from talk therapy to the complex experimental manipulations of social and cognitive psychologists). If the extended mind thesis is true, however, this claim is in need of much more defense than is usually offered. Consider, for one example of a view that is typically advanced on the basis of a (usually implicit) acceptance of claim 1, the oft-expressed worry that altering mental states using antidepressants or methylphenidate risks inauthenticity. The very idea of authenticity at issue here depends on our being able to draw a distinction between me and my environment, such that alterations in the latter cannot deeply affect the former. Hence, for instance, Carl Elliott’s (1998) worry that if antidepressant use alters my personality traits, it is inauthentic, inasmuch as these personality traits cannot be a genuine reflection of who I am:

> It would be worrying if Prozac altered my personality, even if it gave me a better personality, simply because it isn’t my personality. This kind of personality change seems to defy an ethics of authenticity (Elliott 1998, 182).

Why cannot this new personality be mine? Elliott’s (1998) worry seems to be that because the personality does not reflect my internal nature—whether this is understood as my soul or my genome—it cannot be mine authentically. Because Elliot’s worry focuses only on antidepressants and other direct interventions into the mind, it seems to rely upon claim 1: the view that such interventions are distinctive inasmuch as internal states are distinctively, or at least distinctively authentic, mental states. Mere environmental manipulations, which alter the external scaffolding upon which our cognition leans, are not seen to raise any such

---

1.Wilson (In press) gives the example of short-term memory; typical capacity may have been diminished by the transition from an oral to a literate culture (a fear expressed by Plato in the Phaedrus).

2.An anonymous referee argues that if the extended mind thesis is true, there is no way to draw boundaries between agents; this, the referee claims, is a reductio of the thesis. I do not think that it is an entailment of the extended mind thesis that there are no boundaries between agents. Rather, the extended mind thesis entails that the boundaries between agents’ minds are fluid, and may shift according to context. Whether that implies that the boundaries between agents are equally fluid and shifting depends on which account of personal identity is true. I defend the extended mind thesis at greater length and from other criticisms in Levy (2007).
worries, because targeting this scaffolding does not involve targeting my mental states.

It is worth noting that the conception of authenticity to which Elliott (1998) appeals here is by no means the only, or even the most influential, conception in the philosophical literature. The philosopher most associated with the notion of authenticity, Jean-Paul Sartre, advanced a notion of it according to which authenticity could not consist in the expression of a pre-given self, for there is no pre-given self. In any case, it should be clear that in the light of the extended mind thesis, the view that authenticity consists in expressing our pre-given selves looks extremely implausible. If our selves (once again, understanding this notion in the characterization sense) consist of our mental states, especially our memories, and these latter are constituted by our neural states plus a shifting collection of external resources and scaffolding, then we cannot make sense of the idea of a pre-given self. Who I am is partially a function of my environment, inasmuch as my capacities are dependent on features of my context. (I cannot walk if the ground does not provide sufficient friction; I cannot perform mathematics if I am not provided with the appropriate scaffolding; I cannot even conceive of multidimensional space without the appropriate equipment). My identity is also, and very deeply, a product of my social relations to others, who may carry around a significant part of my memories (and I theirs) and who may be an indispensable resource for my decision-making. Human beings frequently answer identity questions by citing irreducibly social features: their group membership or identification, for instance. No one can be British alone; moreover, insofar as being African-American or gay reflects cultural membership and identification, no one can have these characteristics alone. Authenticity is not something we achieve from the inside out, but also from the outside in.

Once we accept that who we are is as much a product of our environments as of features of ourselves from the skin in, the implications for neuroethics extend far beyond the question of authenticity. Many neuroethical concerns may in fact be significantly driven by nothing more than a prejudice in favor of internalism; that is, by claim 1. Sometimes the role of claim 1 is relatively direct, as in the worries about authenticity noted previously; sometimes it is far more indirect. Consider the concern that using psychopharmaceuticals to treat psychological problems is inappropriate because psychological states ought not to be manipulated mechanically. There are several different variants and rationales for this concern, some of which may indeed be quite weighty. But independently of these genuine problems, it is not difficult to discern a worry that is motivated by the internalist prejudice.

Take Erik Parens (1998) concern about the use of psychopharmaceuticals. Parens’ official claim is that “means matter morally” (Parens 1998, 12), but when he comes to defend the claim he often adverts not to the means used, but to the different results the means may produce. Consider his worries about the use of psychopharmaceuticals in classrooms. Parens argues that the availability of these medications may incline us to target suffering individual bodies and minds, rather than “the complex social conditions that produce that suffering” (1998, 13). For example, if overcrowded classrooms produce inferior academic performance, we may find it easier and cheaper to use methylphenidate to compensate for rather than fix the deficiencies of schools. Now, this claim is itself ambiguous: by pointing to this possibility, Parens may mean to advert to the risk that an inappropriate means will be used, or to the risk that bad ends may be produced. When he mentions the different (and presumably inferior) experiences that students in the methylphenidate-enhanced classroom may undergo, he adverts to the results. But to the extent he does so, he abandons the claim that means matter, for if it is the end that makes the means inappropriate, it is the end that justifies the claim that psychopharmaceuticals ought not to be used. But in the absence of a difference in ends, it is hard to see what could justify the claim. Suppose the drug allows us to achieve, at a lower cost, precisely the same results (in terms of academic performance and well-adjusted students) as better classrooms and more individualized teaching. In that case, it seems, it is clearly superior to alternatives.

Why might Parens (1998) think that means matter, all by themselves? Parens’ claim may rest, more or less directly, on claim 1—directly, if he thinks that the problem with psychopharmaceuticals is that they represent an alien intrusion into the mind; an inappropriate crossing of boundaries in which the properly external is treated as internal. It seems more charitable, however, to attribute a rather more sophisticated view to Parens. After all, he no doubt recognizes that more traditional means of changing minds (such as, say, rational argument) also work from the outside in. Instead, Parens thought might be this: traditional means of changing mind are specially fitted to be incorporated into the mind, and thereby to work their effects from within. But even when they are absorbed into neural networks, psychopharmaceuticals remain external and alien intrusions into the mind. Our minds, the internalist may think, are designed (perhaps by evolution) to convert rational argument and persuasion into thought, making the external internal, but they cannot work their magic on serotonin or lithium. So pharmaceutical interventions, no matter the benefits they produce in terms of happiness or academic success would remain the wrong kind of means to use in treating psychological problems.

But if this is Parens’ (1998) thought, or the thought of some of the many others who hold that psychopharmaceuticals, transcranial magnetic stimulation, and other new means of altering mental states are especially problematic, then it too is motivated, indirectly and in part, by claim 1. For it is plausible only so long as we deny that the mind extends outside the skull. Admit that the mind extends outside the skull, and a whole range of ways of changing minds that seem equally mechanical become salient. The environmental scaffolding of thought was not designed, by evolution or anything else, to be incorporated within our minds.
But shape our thought it does, in ways that are far-reaching and essential to our most highly prized cognitive abilities. Attention to the wide range of cases in which external manipulations have long shaped human cognition makes the claim that psychopharmaceuticals are especially suspect for this reason much less plausible.

Many other objections to neurotechnological enhancements or alterations of the mind seem also to rest, directly or indirectly, on a presumption that internal interventions are especially problematic, a presumption apparently generated by the internalist prejudice. If the extended mind thesis is true, however, these objections must be rethought. In what sense is it cheating to take methylphenidate to boost cognitive performance, but not cheating to use vitamins? It cannot be that the former targets the mind, whereas the second targets the body alone, not if the mind extends beyond the skull. Indeed there is evidence from neuroscience itself that the mind extends into the body. As Damasio (1994) puts it, the body “contributes a content that is part and parcel of the workings of the normal mind” (225–226). There may be a good reason why methylphenidate is cheating and taking vitamins is not, but in the light of the extended mind thesis, the tempting first answer must be rejected. In what sense is it an inappropriate mechanization of the agent to use psychopharmaceuticals (Freedman 1998)? Once again, the question may have a good answer, but in light of the range of cases in which environmental manipulations already alter mental states, the claim needs sustained defense.

Claim 1, the claim that neurological interventions are uniquely interventions into the mind, is very widely assumed; indeed, because internalism about mental states is the default position amongst most thinkers outside philosophy of mind (and for many within it as well), it often serves as an undefended background presupposition. Not only is it not defended, most thinkers do not even realize that they need to defend it, for it does not occur to them that there is any reasonable alternative to such internalism. But once we take the extended mind thesis seriously, we see that the claim requires defense, because it assumes that mental states are uniquely contained within the skull of agents. Since such internalism cannot simply be assumed, it cannot serve as an unchallenged premise in neuroethical argument. If neuroethicists are to establish that this or that neurological intervention is more problematic than this or that environmental intervention, they shall have to provide arguments, arguments that do not turn on an undefended internalism. In some cases, the argumentative burden may well be able to be met; there may indeed be cases (perhaps many cases) in which internal interventions raise more, and more serious, worries than external. But in each case an argument is required; and a crucial first step is the recognition that one is required. Some internal interventions may indeed be especially problematic, but it will not be the mere fact that they are internal that makes this the case. So much for claim 1. Let us turn, now, to claim 2; that neuroethics is concerned with whether interventions into the mind are permissible.

Few thinkers explicitly defend claim 2, for the same reasons that few defend claim 1—because the claim is widely taken to be too obvious even to be explicitly noted, much less defended. Nevertheless, it seems reasonable to attribute the claim to many. Consider the growing literature on the topic of neuroenhancement (Buller 2006; Kass 2003; Parens 2006). This debate seems largely to concern whether such enhancements are permissible. But the question whether neuroenhancements are permissible can be an ethically pressing one only if we have a real choice about whether or not such enhancements are used. In the light of the extended mind, the implicit presupposition that we have such a choice is, if not clearly false, at least in need of sustained defense. Once we recognize the parity between internal and external means of changing minds—a recognition automatically forced on us by the recognition that mental states extend beyond the skull—it seems that we are instead forced to conclude that we have no choice about whether to allow interventions into the mind. Instead, the important questions concern which interventions are permissible. We are forced to this conclusion by the recognition that so many of the things that we routinely do alter mental states. Interventions into the mind are ubiquitous, and the current technological explosion, centering as it does on information storage, retrieval and processing, makes such interventions an ever more prominent part of our world. We manipulate minds by improving access to computers, by schooling, and even by decisions about nutrition. Because many of these environmental manipulations seem to enhance the mind, the notion that neuroenhancements of the kind at issue are somehow special needs defense.

In the light of the extended mind thesis, there is a strong prima facie case for thinking that we ought to be equally concerned with all means boosting cognitive performance, whether they involve methylphenidate (Farah & Wolpe 2004), transcranial magnetic stimulation (Snyder et al. 2003), or better nutrition and education (Neisser 1997); inequalities in the latter should concern us as much as inequalities in the former, other things being equal. Merely because the first two kinds of intervention are internal is no reason to find them particularly worrying. Similarly, we ought not to be concerned with neuroscientific ways of mind reading and control, merely because they seem to be direct interventions into the mind. We ought to be more concerned with, say, brain reading technology, than with more traditional ways of detecting deception just to the extent that the former gives rise, due to its effectiveness or its ability to be applied covertly, to greater privacy concerns. At the moment and for the foreseeable future, there is little reason to think that brain reading has these features.

3. Birth-weight, which reflects the nutritional and health status of the mother, is correlated with intelligence quotient, even when we restrict our attention to the normal range of weights, thereby excluding from consideration children born into extreme poverty (Matte et al. 2001).
These reflections on the *prima facie* parity between environmental manipulations and new technologies may seem to reduce the importance of neuroethics. On the contrary, I think that it dramatically increases it. It might seem that the extended mind thesis entails that neuroethics is less important inasmuch as, in its light, it becomes apparent that neurological interventions into the mind—that is, interventions that target neurons, neurotransmitters or brain structures—are not after all so special: they represent merely the latest means of doing something that is quite ubiquitous in human cultures. Although it is true that the extended mind thesis may dampen some of the hype surrounding these technologies, it should be seen as dramatically expanding the scope of neuroethics, not detracting from its importance. Neuroethics focuses ethical thought on the physical substrate subserving cognition, but if we accept that this substrate includes not only brains, but also material culture, and even social structures, we see that neuroethical concern should extend far more widely than has previously been recognized. In light of the extended mind thesis, a great many questions that are not usually seen as falling within its purview—questions about social policy, about technology, about food and even about entertainment—can be seen to be neuroethical issues. In making decisions about how we structure our environments, we decide how we shall think, and such decisions must be informed by neuroethical thinking. Neuroethics might not have assumed its current, increasingly prominent, place in the intellectual landscape without the impetus provided by breakthroughs in the sciences of the mind, but the result may be a new approach to far older questions.

**REFERENCES**


