

In his famous essay, "The Abolition of Man," C. S. Lewis argued that the use of science and technology to modify the human mind would destroy humanity. Some of the concerns Lewis raised are philosophically profound: Is it desirable for humans to modify their minds, and if so, in what ways? By what principles should such profound self-shaping be guided? Will "post-humanity" be freer or more enslaved? Is manipulating the core nature of humanity even rational? Since this essay was first written, it seems all the more likely that humans will someday have the ability to modify motivation, not by Lewis's "eugenics," "pre-natal conditioning," and "education and propaganda based on a perfect applied psychology," but by applied neuroscience, drugs, computerized implants, brain-machine interfaces, mind uploading, nanoscale devices, and other advanced technologies. This essay examines the issue of modifying motivations and answers some of Lewis's concerns.

This above all: to thine own self be true.

William Shakespeare, *Hamlet Act 1, Scene 3*

In his essay, "The Abolition of Man," C. S. Lewis (Lewis 1943) wrote about the use of science and technology to modify the human mind. In particular, he considered the possibility that human beings in the future would be able to shape and modify their own minds into any form they desired. He argued that humans who exercised such power would destroy themselves and the rest of humanity. Since this essay was written in 1943, it seems even more likely that humans will someday have the ability to shape their own minds, not through "eugenics," "pre-natal conditioning," and "education and propaganda based on a perfect applied psychology," as Lewis thought, but rather through an applied neuroscience, drugs, computerized implants, brain-machine interfaces, mind uploading, nanoscale devices, or other advanced technologies.

The issues that concerned C. S. Lewis, then, are still of concern to us today. Is it desirable for humans to develop and apply methods for altering their own minds, and if so, in what ways? By what principles should such profound self-shaping be guided? Will “post-humanity” be freer or more enslaved? Is manipulating the core nature of humanity even rational? In order to think about this more clearly, let us imagine that neuroscience has been perfected and the human mind is perfectly understood in every detail. A technology has been developed which allows every aspect of the mind to be modified in any way that is desired and physically possible. Now human beings can remove, add, or change any aspect of their minds, including their own motivations.

Motivations, in the cognitive psychological sense, may be considered the physical mechanisms or neural pathways by which instincts, drives, desires, wants, or needs predispose an organism to certain thoughts or behaviors under the right circumstances. (The terminology for these things must be loose because the biology underlying these phenomena is still being investigated.) Motivations may be distinguished from emotions, emotional states, moods, and beliefs, or cognitive perceptions or conceptions about states of affairs.

Motivations are not rationalizations for behavior; such rationalizations may arise alongside of, or even after, behavior (Dennett 2003). In a sense, motivations pre-exist rational calculation or even consciousness; they are the foundation of reason, since reason facilitates the fulfilment of motivations. They are also the foundation of our values, since feelings about what is valuable are related to our motives. If the rational application of human instrumental power depends upon pre-existing motivations, we discover a problem in trying to use those pre-existing motivations to change themselves.

Upon what basis could we judge the value of our own motivations when our value judgments depend upon pre-existing motivations? Self-reference in determining values short-circuits the process of evaluating modifications to one’s own motivations. The total power to shape one’s own motivations exposes the lack of an independent framework of purpose from which to do so.

The final stage is come when Man . . . has obtained full control over himself. Human nature will be the last part of Nature to surrender to Man. The battle will then be won. We shall have “taken the thread of life out of the hand of Clotho” and be henceforth free to make our species whatever we wish it to be. The battle will indeed be won. But who, precisely will have won it? . . . They are the motivators, the creators of motives. But how are they going to be motivated? . . . However far they go back, or down, they can find no ground to stand on. Every motive they try to act on becomes at once petitive. (Lewis 1943, n.p.)

Is it reasonable at all to suppose that we would want to modify our motivations, even if we had the power? Perhaps motivations would be blank with regard to themselves, just as the brain, the seat of sensation, is unable to sense itself. Motivations might have a sort of inertia so that those who come into the power to modify them will nevertheless preserve them, continuing on the same trajectory as before.

However, in line with the inertia analogy, objects continue on a straight path only if they are not acted upon by another force. There are always other forces acting upon our motivations. First of all, our motivations act upon each other. A single human mind possesses many motivations, and these can obstruct or contradict or compete with each other. Then there are the forces of other minds – society, and the forces of the general environment of the mind, including everything from the cosmos to the mind’s own embodiment.

Imagine a very non-controversial example: a person who likes to eat chocolate. Suppose that he wishes that he did not want to eat chocolate – a clash between two motivations in one person. Or, suppose that his friends try to talk him into not eating chocolate – that is, society attempts to mould his motivation. Or

suppose that the store shelves are empty of chocolate – that is, that the environment decisively blocks the fulfilment of his motivation. For any of these reasons, this person could in theory be convinced to modify his motivations so that he did not like to eat chocolate any more. But would this be the right thing for him to do? On what basis could he judge between his own motivations, or between his own motivations, on the one hand, and what would be more conformable to society or the environment, on the other hand?

This example is fairly innocuous in comparison to the many other possible applications of mind-modification. For example, if mind-modification were possible, some people might eliminate their motivations to commit crimes and anti-social acts, which we might regard as a welcome development. Or frighteningly, some people might make themselves more lacking in compassion or concern for others' welfare. And yet in either case, our value judgments about these modifications to motivations would draw upon our pre-existing motivations. Those pre-existing motivations may be considered provisional, artificial constructions, because they can be changed at will.

Lewis argued that there were only two things which could possibly guide those who possessed the power to change their own motivations: (1) temporary “survivals, in their own minds” of the *Tao*, or (2) irrational impulses and whims. By the *Tao* Lewis referred vaguely to notions of natural law and universal morality which, he supposed, should be self-evidently rational and valid.

In the *Tao* itself, as long as we remain within it, we find the concrete reality in which to participate is to be truly human: the real common will and common reason of humanity, alive, and growing like a tree, and branching out, as the situation varies, into ever new beauties and dignities of application. While we speak from within the *Tao* we can speak of Man having power over himself in a sense truly analogous to an individual's self-control. (Lewis 1943, n.p.)

Lewis described the *Tao* as “objective value” and conscience apart from instinct. It was supposed to be the common heritage of humanity, but nevertheless, as “the mystery of humanity,” it needed to be taught to children by teachers who would encourage some motivations and weaken others. To illustrate the *Tao* more concretely, Lewis sketched a universal code of behavior, including such things as benevolence, justice, veracity, mercy, and so on.

There are no doubt commonalities among human cultures along those lines, but there are also many differences among human cultures, and cultures can change over time. Lewis regarded the *Tao* as self-evidently real and decisive, so he did not really claim to need to prove its existence or value. Because Lewis provided no evidence for its existence and its specific values, it is difficult for us to see the *Tao* as a foundation for judging motivations.

Nevertheless, there is a certain attraction in Lewis's notion of looking outside ourselves for a foundation, if no foundation seems to be present in our minds. Lewis's solution was to be “truly human” by participating in shared human values, the “mystery of humanity.” If there is a discernible set of shared values among human beings, these could serve to guide the modification, if any, of motivations. But if these shared values arise from human nature and the natural environments of human beings, then following them is arbitrary submission to nature. As Lewis noted of those with the power to modify motivations, “If they accept [the *Tao*], then they are no longer the makers of conscience but still its subjects, and their final conquest over Nature has not really happened” (Lewis 1943, n.p.). In fact, Lewis regarded the preservation of *Tao* in such a circumstance as “confusion.”

The second possible guide to modifying motivations, according to Lewis, would be pure impulsiveness. He notes, for example, that “those who stand outside all judgments of value cannot have any ground for preferring one of their own impulses to another except the emotional strength of that impulse” (Lewis 1943, n.p.). The mind decides to act by deciding among weighted inner motivations, so for Lewis, the

motivation for action of those who have the power to change their motivations, and who reject the *Tao*, must rest in the pre-rational selection of motivations by their “felt emotional weight at a given moment.” However, the criteria for determining the emotional strength of impulses come from pre-existing motivations. This ground for preferring one impulse to another, then, cannot “stand outside all judgments of value.”

The motivations of an individual’s mind are what they are because of all the factors that shaped that individual’s mind, from the individual’s remotest ancestors and their environments, down through that individual’s embodiment and experiences since conception. When we look behind our natural motivations, we see an infinite regress of deeply complex causations. So motivations have a history and a reason for being the way they are, but this in itself is no help in evaluating those motivations, because value judgments depend upon pre-existing motivations. We do not even have grounds for judging our impulses “irrational” since they could only be judged so by reference to purposeful action.

For Lewis, human self-manipulation at the most intimate levels of mind would require the reduction of all of humanity to “raw material.”

It is in Man's power to treat himself as a mere “natural object” and his own judgements of value as raw material for scientific manipulation to alter at will. The objection to his doing so does not lie in the fact that this point of view (like one's first day in a dissecting room) is painful and shocking till we grow used to it. The pain and the shock are at most a warning and a symptom. The real objection is that if man chooses to treat himself as raw material, raw material he will be: not raw material to be manipulated, as he fondly imagined, by himself, but by mere appetite, that is, mere Nature . . . (Lewis 1943, n.p.)

The phrase which vexed Lewis at the beginning of his essay, *man's conquest of nature*, sets up a distinction between human beings and the rest of nature, which is untenable for true naturalism. If we accept that human beings are natural beings, arising from nature and part of it, then *man's conquest of nature* is an incoherent statement. Even if we understood “nature” here as that part of nature which can be distinguished from “man” in the sense of an individual human being, control of nature still could not include control of one’s self. Control of one’s own mind would only make sense if there were parts of the mind which could be distinguished as “man” from other parts of the mind, parts which would be “not man” or “less man.” Thus it may be possible to suppose, according to Lewis’s argument, that humans could possess impulses and motivations which are less distinctively human, less worthy of humanity – impulses and motivations contrary to the *Tao*.

It is clear, then, that Lewis saw the possession of a particular set of values as an authentic and indispensable criterion for humanness. For Lewis, accessing the power to modify one’s own motivations was abandoning one’s own human nature. Those who would do so

. . . have sacrificed their own share in traditional humanity in order to devote themselves to the task of deciding what “Humanity” shall henceforth mean. “Good” and “bad,” applied to them, are words without content; for it is from them that the content of these words is henceforth to be derived. . . . It is not that they are bad men. They are not men at all. Stepping outside the *Tao*, they have stepped into the void. . . . Their subjects . . . are not men at all: they are artefacts. Man’s final conquest has proved to be the abolition of Man. (Lewis 1943, n.p.)

Ironically, according to Lewis and his followers, it was by preserving natural human nature (and its distinctive values and motivations) that human beings should fend off their transformation by nature into undesirable forms and states.

In one sense, Lewis and his followers were right that the *power* to modify our minds, including our motivations, would not help us to make decisions about what is authentic about us, whether we should like ourselves, or if so, what parts, and what would constitute beneficial or maleficent changes to ourselves. In fact, the power to modify our minds would only give us the ability to change ourselves according to the terms by which we understand ourselves and our qualities, either to modify ourselves (by changing our dispensable features) or to destroy ourselves (by changing our indispensable features). The terms by which we understand ourselves and our qualities arise in turn from the nature of our minds at a particular point in time, since our minds are complex systems and confluences of external influences.

The state of a mind is an open, complex, dynamic system embedded in the rest of nature, and the sense of self, including notions about what is indispensable about it, depends upon the specific state of the mind. Thus, we are prepared to see our notions of what is *natural* and what is *human* as contingent phenomena of nature, arising in particular states of mind. There is no need for “seeing through things for ever,” as Lewis writes, because when intention reaches instrumental power over itself, it has indeed come to its foundation, or at least, its source.

Control over one’s own mind does not so much create a feedback loop as expose and tighten one that already exists – the feedback loop between forces within the mind that give rise to a state of that mind, including consciousness and a sense of self, and also the feedback loop between that mind and everything outside of it, from its physical embodiment to the most distant reaches in space and time of the cosmos. The difference between a circuit and a short-circuit depends upon what sort of loop one wants.

There are frames of reference which do not depend upon irrational impulses, or a mystical *Tao*, and which do not necessarily lead to self-destruction (in the way in which one understands one’s self). Self-knowledge and world-knowledge could enable a state of mind to make wiser and more coherent decisions about self-modification than it would make if it took into account only momentary impulses or a limited subset of authentic human nature. A knower of this sort could be “true to itself” and possess integrity because it would draw upon an understanding of its whole self and its context in the whole cosmos.

By applying human reasoning, we are able to come to understand more of ourselves than a momentary impulse and more of our environment than what is perceived immediately before us. We can grasp a deeper sense of ourselves, extending in time and encompassing our entire natures; we can also grasp a broader sense of the world around us and our fellow human beings. Knowledge about the self and knowledge about the cosmos, including that knowledge which we acquire by the assistance of our fellow human beings, can extend the framework of our decisions about modifying our own minds.

If we made our decisions about modification in the context of understanding our entire minds, then we would draw upon more than our momentary impulses. If we understood the mind as all of its motivations and characteristics and components, not only in the present moment, but also in the past (as in the remembering, narrative self) and in the future (as in rational calculations of possibilities) – however we may harmonize its disparate elements – then our decisions could draw upon a wholeness of self. The same goes with an understanding of the world around us, including the insights of other human beings.

Returning to the example of the person who likes to eat chocolate – he could make a decision about changing his motivation based on profound, comprehensive self-knowledge, about why he likes chocolate, how he likes chocolate, what role this desire has played in his past life, and what role it might play in his future life. Such a decision could also be based on profound, comprehensive knowledge of the world, including the nature of chocolate, how his desire for chocolate evolved, what functions this desire has or might serve, what social meaning and consequences his desire for chocolate may have, what effect it might have on the environment, and so on. Equipped with such knowledge and self-knowledge, this person could be better guided in deciding to modify his desire for chocolate.

Nevertheless, one thing seems clear – that the precise state of mind at the moment a decision is made about whether or not to modify one’s self and how to modify one’s self if so, would be uniquely critical. Everything that would follow from a process of self-modification would depend upon that initial condition. Technology might allow one to restore some changes to their initial state, but whether or not there would be motivation for doing so would depend upon the changes made. It would be easy to imagine that such changes could have undesirable effects (from the vantage point of the initial condition) if the initial state of mind applied insufficient knowledge.

Perfect self-knowledge in real-time, like omniscience about the rest of the universe, may be physically impossible, but what better knowledge would allow one to do would be to make more authentic, integral determinations of one’s own motivations and the contexts of those motivations. It is precisely when our humanity is *not* a “mystery” that we can make informed, appropriate judgments about ourselves and how to act with regard to ourselves. Far from being “a basilisk which kills what it sees and only sees by killing,” as Lewis described science, scientific understanding of the human mind and the universe at large enables human beings to better discern their own values and the context in which those values have meaning, and applied science and technology provide humans the opportunity to realize those values.

It should be possible for human beings to modify their own minds in such a way as to increase whatever motivations or characteristics they believe are distinctively “human” or indispensable to themselves as they wish to be. In this sort of feedback loop, one might expect *simplification*. Humans’ ability to modify their own minds might allow them to become *more* human, that is, possess to a more notable degree whatever characteristics they consider distinctively human. The term *superhuman* would be an apt description of the result, not in the vainglorious sense, but in the precise sense that the superhuman would be more deeply characterized by human motives and human qualities. For example, if benevolence is an indispensable human characteristic, as Lewis suggested, then a human who could modify himself could make himself more benevolent, and thus more human.

There are no easy answers to what humans should want to become or what humans should value about themselves. Humans have been working on these questions for thousands of years. The approach of mind-modification technologies does not give us a direction; rather, it conveys the urgency for us to pick a direction, because a decision can no longer be delayed, and the consequences may truly be everlasting.

In this essay, I have deliberately avoided discussing Lewis’s dark vision of an elite group of “Conditioners” who use a technology of modifying minds to enslave the rest of humanity. I do not wish to mount here an extensive critique of Lewis’s point that “man’s power over Nature turns out to be a power exercised by some men over other men with Nature as its instrument.” However, I would observe that it is precisely self-knowledge, made possible by sciences of the mind, that may someday enable human beings to analyze their desire for power over one another.

Although it may be too optimistic to presume that this knowledge will inevitably result in greater mutual understanding and social harmony, surely this knowledge will affect the dynamic play of motivations within human beings. Self-knowledge will be a factor and it will be taken into consideration.

Humans may not be so anxious to dominate each other if the origins of such desires are laid bare as primitive, irrational survivals of innate tendencies to ruthless competition over rank, tendencies which may once have served a useful evolutionary purpose, but which may no longer serve our purposes. The *overall* trend of history, as FM-2030 (1970, 1973) illustrated, is toward more nonviolence, peace, tolerance, mutual understanding, mutual respect, and freedom. It may be that, ever since they acquired modern intelligence, human beings have been moving ever closer to an accommodation of one another, a

rapprochement based on intelligently-realized mutual benefit. If that is the case, more scientific analysis and applied technology could accelerate this trend.

Is humans' present, relative inability to modify themselves their most human characteristic? Is the reluctance by some today to modify themselves the highest human value? If *human nature* is to mean anything, it must surely refer to the actual natures of human beings, who typically desire to be free from many of their limitations so they can become happier. It is not the sciences of the mind and technologies for modifying the mind that would most likely lead to the extinction of the human nature that we value. Rather, it is the relinquishment of progress in those sciences and technologies of the mind that would more likely result in the extinction of human nature and all its hopes and dreams, a true "abolition of man."

The English word *invent* comes from a Latin word meaning *find*, and it is by finding ourselves, that is, by discerning our natures for the very first time, that we can acquire not only the power but also the wisdom to bring out the potential of what is best in ourselves. We can choose to make a leap – not into a Void, but rather into ourselves and into the fullness of the world unveiled by our investigations. And we can make this leap together, holding each other's hands.

References

Dennett, D. C. *Freedom evolves*. 2003. New York: Viking Penguin.

FM-2030 (F. M. Esfandiary). 1970. *Optimism one: The emerging radicalism*. New York: W. W. Norton & Company.

FM-2030 (F. M. Esfandiary). 1973. *Up-winger: A futurist manifesto*. New York: The John Day Company.

Lewis, C. S. 1943. *The abolition of man*. 1943.
<http://www.columbia.edu/cu/augustine/arch/lewis/abolition3.htm> (posted March 6, 2002).

DR. RUPNATHJIK (DR. RUPAK NATH)