

**DISASSEMBLING THE MIND? A REVIEW OF GARY
MARCUS'S *KLUGE*: THE HAPHAZARD CONSTRUCTION OF
THE HUMAN MIND**
Christian Beenfeldt

Think Summer 2010 • 47

The main thesis of *Kluge* (Farber and Farber, paperback edition, 2009) is that the human mind is an evolutionary kluge (rhymes with *huge*, not *sludge*). As Gary Marcus informs us, the term was popularized by Jackson Granholm's 1962 article 'How to Design a Kluge' where it was defined as 'an ill-assorted collection of poorly matching parts, forming a distressing whole'. A kluge may be clumsy and inelegant but, surprisingly, it *works*. And the mind, according to Marcus, is '[t]he most fantastic kluge of them all'. Unlike the view of the human mind that is advanced here, *Kluge* itself is *not* a kluge. It is clear, smooth, well-organized, well paced and well written; it can comfortably be read in a few sittings. Unlike a kluge, the parts match well, the collection is not ill-assorted, the whole is not particularly distressing – and, as I will argue, logically speaking *it does not work*.

Marshalling an array of psychological studies, Marcus expends the principal part of his text on the argument that we have 'bugs in our cognitive makeup'. These include: confirmation bias, mental contamination, anchoring, framing, inadequate self-control, the ruminative cycle, the focusing illusion, motivated reasoning, false memory, absent-mindedness, a vulnerability to mental disorders and, last but not least, our reliance on an ambiguous linguistic system. Structurally, the discussion of the psychological studies thought to establish the existence of these defects stretches through *Kluge* like a thick spine. Conceptually, we can think of this as stage one of Marcus's kluge argument. Stage two is the claim that none 'of these aspects of

human psychology would be expected from an intelligent designer...’ (162) and stage three is the assertion that ‘the only reasonable way to interpret them is as relics, leftovers of evolution’ (162).

This is the case made in *Kluge*. Is it a good argument?

No.

For now, let us set aside the cognitive-bug-finding stage and consider whether stage two and stage three of the argument actually hold water. In other words, let us assume for a moment that there really are some serious cognitive defects of the human mind. And clearly, at least some of the items that Marcus mentions – e.g. our vulnerability to debilitating mental disorders like schizophrenia or clinical depression – are genuine afflictions of the human species.

The second stage of the kluge arguments now contends that given such defects, we should not expect an intelligent designer of the human mind. In Marcus’s words, if mankind were the product of such a designer ‘our thoughts would be rational, our logic impeccable. Our memory would be robust, our recollections reliable. Our sentences would be crisp, our words precise, our languages systematic and regular, not besodden with irregular verbs ...’ (1).

But do you know *anyone* who thinks that mankind is like *this*? Does this not simply amount to a bout of shadow boxing against an imaginary opponent who thinks that human beings are like Dr. Manhattan?

Let me explain. Dr. Manhattan is the scientist in Alan Moore’s *Watchmen* comic book series who, after being split into atoms in an ‘Intrinsic Field Test Chamber’, puts himself back together again, atom by atom, until he is an immortal, stoic, God-like being wholly detached from human life, values, interests and emotion – like a robot or an eternal cosmic pocket watch oblivious to the world. Now, it is this sort of image of humanity, as *Kluge* frames it, that one must believe in if one holds that ‘mankind were the product of some intelligent, compassionate designer’ (1). In other words, creationists who think that God personally

engineered *Homo sapiens* must further think that we have been built as an army of sublunary Dr. Manhattans, all speaking in robust sentences purged of irregular verbs to match our infallible memory and wholly detached disposition. Since we are clearly not such beings, however, creationism is false. *Quod erat demonstrandum*.

Surely, creationism should be rejected as preposterous nonsense, but this is *not* the way to do it. In fact, I submit that Marcus has completely misunderstood the nature of the opposition. Creationism is a religiously motivated attempt to discredit evolution – ultimately because the fact of evolution contradicts the age-old myth of an omnipotent God conjuring biologically sophisticated plants, animals and human beings out of thin air. But *the religious view has never been that human beings are cognitively perfect*. Consider, for example, Pope Innocent III's famous late twelfth century diatribe *On the Misery of the Human Condition* where we are told that human beings are 'filled with all iniquity, malice, fornication, avarice, wickedness, full of envy, murders, contention, deceit, evil, being whisperers, detractors, hateful to God, irreverent, proud, haughty, plotters of evil, disobedient to parents, foolish, dissolute, without affection, without fidelity, without mercy.' And so on. Indeed, human life is so 'full of mortal sin, so that one can scarcely find anyone who does not go astray, does not return to his own vomit and rot in his own dung.' Setting aside the exaggerated rhetoric, Pope Innocent III voices what amounts to the standard religious teaching that human beings are metaphysically defective creatures, all born sinful and in dire need of divine forgiveness, redemption and salvation; indeed, to entertain the very belief that one is *not* sinful is to commit the mortal sin of pride.

It is thus a complete mistake to think that psychological findings concerning our supposed cognitive deficiencies in any way oppose religiously driven projects like creationism. It is of course true that theologians for centuries have spilled an abundance of ink on the so-called 'problem of evil' – i.e. on the profound contradiction between the

assumption of an all-good, omniscient and omnipotent creator God and the existence of genuine deficiency and evil in the world, from Hitler and Stalin on down. But if the aim of *Kluge* merely is to demonstrate the existence of human deficiencies, by reference to a trifling collection of psychological findings about absent-mindedness, faulty recall and so on – trifling, that is, in light of the mountain of evidence already recorded in any textbook of twentieth century European history – then the motivation for the extensive discussion of our supposed cognitive bugs is completely lost and we are left with a pedestrian restatement of a well-known problem created by the notion of an omnipotent, all-good cosmic agent. (Incidentally, note here that both *Kluge* and the Original Sin view make the very same unjust mistake of condemning all of mankind for the sins of some individuals. In the chapter entitled ‘Things Fall Apart’, for example, defects such as procrastination, foolish inconsistency, and mental disorders are taken to impugn mankind generally – yet what about the fact, one wonders, that some individuals clearly get things done, take consistent and principled actions, and are free of mental disorders?)

In sum, the problem with this stage of the kluge argument – the very stage that is supposed to bring the material of *Kluge* to bear upon the intelligent design debate in the first place – is that it fails on the grounds of either irrelevance or triviality. Irrelevance if it is supposed to disprove a premise held by nobody, namely that humanity resembles Dr. Manhattan. Triviality, if it merely aims to redemonstrate a fact widely recognized for centuries, namely that the notion of an omnipotent and beneficent Creator contradicts the obvious existence of earthly imperfection.

Let us now leave stage two and turn to stage three of the kluge argument. Here we infer that the only reasonable way to interpret our mental bugs is ‘as relics, leftovers of evolution’ (162). Specifically, human psychology is explained as the product of two ‘systems’: *the ancestral system* (an evolutionarily old system, depending on brain

areas like the cerebellum, basal ganglia, and the amygdala) and *the deliberative system* (new system, based primarily in the prefrontal cortex). The essential thesis of *Kluge* is that our cognitive defects are caused by the old system.

As a rigorous scientific contention, this is an ambitious claim. Unfortunately it turns out to be pure speculation, a vague hypothesis resting on flimsy evidence.

We know, of course, that we have both the ability to reason and the capacity to experience emotions – sometimes irrational emotions – and to act on those emotions. And we also know that to reason well, we have to manage our limitations and various blind spots, such as absent-mindedness, fallible memory, the tendency to unduly favor cherished beliefs ('confirmation bias') and so on. But to go from here to the *fundamental scientific hypothesis* that we therefore consist of inherently warring biological 'systems' is not warranted by the evidence. In fact, we are presented with *no* relevant biological data whatsoever, to substantiate this very strong neuroscientific claim. In the *Republic*, some two and a half millennia ago, Plato hypothesized that the soul was composed of three potentially warring parts – reason, will and passion. Why not go with these three systems? Or with five? Or seventeen? And what *exactly* does this vague notion of a 'system' mean in the first place – after all, it supposedly resides not only in your brain but also, puzzlingly, 'in some form' in 'virtually every multicellular organism' (51) presumably including the grass of your lawn.

None of these questions are even touched upon in *Kluge*.

Kluge, of course, is by no means alone today in offering this kind of scientific-sounding speculation. It is part of a much wider trend where (fuzzy and tenuous) hypotheses are facilely advanced as astounding scientific insights. A common pattern is as follows. One takes an already-established and illustrious theory (such as evolution) and then simply adds on top a plausible-sounding ad hoc conjecture which one, in turn, uses to generate a revolutionary

‘explanation’ of some well-known phenomenon. So for example, we know that evolution is a fact and we know that we crave chocolate. Speculative theory: over biological time human beings evolved a system that wants chocolate – and *that* is why we crave chocolate.

But this is not a genuine explanation at all. We have merely replaced the problem ‘why does Johnny want to eat a Hershey’s bar?’ with the new problem ‘why does system so-and-so – which is vaguely supposed to hover in some appropriate corner of Johnny’s brain – want to eat a Hershey’s bar?’ Rather than obtain an explanation of Johnny’s chocolate craving we have simply relocated the problem back one step and now face the new puzzle of why Johnny’s candy-system is hungry. Plus, we have the added confusion of having arbitrarily split Johnny into a number of mysterious sub-systems, each commanding a different sector of his brain like the post-World War II occupational forces in Berlin.

These are critical problems on the theoretical side of the kluge-mind account – but what about the practical side?

Well, consider first the fact that scientific theories characteristically offer not only real explanations but also great technological payoff, from antibiotics, heart surgery and nuclear fission down through the many, many small technological improvements that paved the way for the subsequent larger breakthroughs. But what payoffs does the supposedly radical two-system theory offer us – remembering that technological payoffs can be material, like a microchip, or immaterial, like a new surgical technique? Well, in the concluding chapter, entitled ‘True Wisdom’, we are offered thirteen insights based on the author’s hypothesis and each ‘founded on careful empirical research’ (165). These include suggestions such as: ‘consider alternative hypotheses’ (insight 1), ‘[r]eframe the question’ when you are considering a problem (insight 2), ‘[a]nticipate your own impulsivity’ – e.g. do not grocery shop on an empty stomach (insight 4), ‘[a]lways weigh benefits against costs’ (insight 8), and ‘[t]ry to be rational’ (insight 13). I submit

that anyone completely ignorant of evolutionary psychology but vaguely familiar with contemporary self-help literature could have come up with that list on his own.

Consider, second, the fact that good scientific theories often generate startling, unexpected predictions. To take a grand historical example, scientific chemistry famously predicted the discovery of elements (such as germanium) unknown at the time. Given the ambitious, even revolutionary, hypothesis of *Kluge*, a couple of ambitious and revolutionary predictions would have been nice – but zero genuine predictions are offered. Zero.

Let us now, finally, turn our attention to the important first conceptual stage of the argument in *Kluge* where the many supposed cognitive bugs in our system are testified to by a long witness list of works from psychology and other fields. Marcus makes interesting individual points throughout, but the general and persistent problem is that these studies are treated much too briefly, uncritically, impressionistically. Had *Kluge* been written simply as a popularization of recent psychological studies for the layman reader, one would have expected a careful, balanced and critical discussion of each experiment, along with a judicious assessment of what exactly it shows – and what it does not show. But since *Kluge* marshals these findings only in as much as they can provide testimony in support of the kluge-mind hypothesis, relevant contrary evidence is rarely considered, scientific critics are almost never heard, and the studies themselves are too often discussed with too little in the way of judicial circumspection, scientific and philosophical rigor.

As a case in point, consider the discussion of human memory ('contextual memory') – 'the mother of all kluges, the single factor most responsible for human cognitive idiosyncrasy' (18). The psychological studies that Marcus discusses under this heading include the very robust data that has been accumulated regarding the nature and limitations of eyewitness testimony. As Elizabeth Loftus and others have shown, testimony by even by the most well-meaning eyewitnesses must be treated carefully because of the

inherent nature and limitations of human memory. The problem is that human memory often involves an active reconstruction process that can go awry. In one experiment, for example, subjects who viewed a slide show of a car running past a stop sign and who later heard mention of a yield sign, would tend to blend what they saw and what they heard and so misremember the car as driving past a yield sign. This is important data (on so-called 'false memory') and it deserves both public attention and forensic cognizance.

The problem is philosophical. Marcus uses these findings to support the claim that there are cognitive deficiencies inherent in the human mind *because* our performance is found wanting by comparison to computer memory ('postal-code memory'). But what about the fact that computer memory (binary digits encoded, for example, as tiny magnetized spots on spinning metal disks) *is not memory at all* – no more than a rolodex, a shopping list, a dictionary, a filing cabinet, or a library is memory? These are all *aids* to human memory, tools we cleverly employ to gain immense cognitive leverage, not standards by which human memory can be found to be metaphysically defective. The fact that there is vastly more information stored in the Bodleian Library in Oxford or on a large computer hard disk drive than an individual human being (even the incredible mnemonist studied by Alexander Luria) could ever begin to memorize, only proves that we are really good at developing external storage systems for organizing vast quantities of information. It does not mean that the library, hard disk or rolodex has a mind of its own against which yours can be unfavorably compared. *Britannica* encoded on a spinning disk would be reduced to meaningless magnetic patterns and the Bodleian would be reduced to a vast pile of cellulose pulp with curious ink stains on it, if there were no human beings left to decode, interpret, understand and learn from the text inscribed there.

Indeed, why not also maintain that humanity suffers from countless horizontal, vertical and nautical movement 'bugs'

since we clearly cannot outrun, out-jump or out-swim the cars, airplanes and boats that we build? It is certainly true that nobody can memorize an entire library or outrun a Porsche, but this is no evidence of some *intrinsic human deficiency* – rather, it simply demonstrates the fact that we have a specific metaphysical identity that enables us to perform a range of actions only within certain finite limits. And this is of course true of everything in the universe. Even the mighty Sun does not provide nice tropical conditions on Pluto.

Consider, finally, the topic of *rationality*. Breaking with the pattern of pure deficiency-findings, Marcus, assures us that '[o]n occasion, human choices can be entirely rational' (71). In support of this, he cites an NYU experiment where subjects were seated at a touch-screen video game. Two targets appear, one red and one green. You get points if you touch the green circle but you get a greater penalty if you touch the red circle. You have to react *quickly*. To avoid touching the red circle when the circles overlap, the optimal strategy therefore is to avoid the center of the green circle. 'Somehow people figure all this out, though not necessarily in an explicit or conscious fashion.' Of course, the 'bad news is that such *exquisite rationality* may well be the exception rather than the rule'. (72; emphasis added). Indeed, we learn, this seems very much to be the case since beings 'truly noble in reason ought to see, *instantaneously*' the structure of syllogisms and '*instantly* reject' all formally invalid reasoning as fallacious' (62; emphasis added) – and we (unlike Dr. Manhattan) clearly do not always do so.

Now, whatever the probative value of the video game experiment may be, a demonstration of exquisite rationality – or of the lack thereof – surely cannot be furnished by placing experimental subjects under stress conditions and having them peck at colored circles like so many Skinnerian pigeons in a box. The same is true of findings regarding the failure of subjects, often completely untrained in logic, to detect invalid syllogistic reasoning (discussed in the chapter entitled 'Belief').

More generally, if items such as the human-versus-computer memory comparison or the pecking-at-colored-circles test are taken as determinative of human cognitive perfection, then it becomes as easy as it is irrelevant, to demonstrate any number of supposed bugs in our cognitive makeup: just compare a man with a library or jack up the speed of the video game and failure is sure to ensue. But are there no *genuine* human limitations and failings? Of course there are – in the sense that man is a fallible being subject to potential shortcomings, blind spots, irrational passions and evil – but no less capable of fabulous achievement, great rationality, astounding insight and moral virtue.

This is the fundamental nature of the human situation, recognized at least as far back as the ancient Greeks.

To the extent, then, that *Kluge* bases its claims on improper comparisons and tests, no actual cognitive bugs have been shown to exist. To the extent that it provides evidence of the mere existence of human fallibility, limitations and potential shortcomings, its conclusions seem quite trivial. As far as the two-system analysis of the human mind goes, *Kluge* is neither irrelevant nor trivial – but here it unfortunately offers us what amounts to mere speculation, not a serious scientific explanation. *Kluge*, of course, was meant to be a contribution to the popular science literature and not a journal paper submitted to *Science*. This nonfiction genre, however, has a proud history of advancing serious, philosophically astute, and scientifically sophisticated material in an accessible form – a tradition that should not be discontinued.

Christian Beenfeldt is a doctoral student in philosophy, University College, University of Oxford. beenfeldt@beenfeldt.com