The Violence of Pure Reason: Neoreaction: A Basilisk

Adam Riggio, New Democratic Party of Canada


http://wp.me/p1Bfg0-3d1
Neoreaction: A Basilisk
Philip Sandifer
Eruditorum Press, 2017

I should start this review with a few simple reasons why you should read Neoreaction: A Basilisk.

A) If you want to understand the fundamental philosophies of the destructive, racist, right-wing, Trump-loving culture that has grown from a few slimy 4chan message boards to a significant reactionary political movement.

B) If you are a professional researcher working in any study of the sociology of knowledge, the nature of knowledge, facticity, or truth. Especially if you want your research to affect wider audiences than fellow academics in your field. If you want to study and write about the nature of knowledge not only as an academic, in other words, but as a public intellectual.

C) If you simply enjoy reading complex, insightful, informative books of theory and analysis.

Philip Sandifer is himself a public intellectual, at least on an independent scale. A former academic, he is a fully credentialed to be a professor of literature and literary theory. His primary career is as a publisher and author of literary theory, running Eruditorum Press. In the interests of objectivity, I should state that he and I correspond regularly as colleagues in independent publishing and professional blogging, and as internet friends. He was an interview subject for my “Beyond the Academy” essay for SERRC. And I threw in $5 to the Kickstarter that funded this project and its affiliated essays and creations, because I thought he would produce a good product.

What Neoreaction: A Basilisk Is About

Neoreaction: A Basilisk is not a perfect book, though it is a brilliant book. Its analysis proceeds in a spiralling style that many accustomed to more traditionally-written theoretical books will find disorienting. Its concluding analysis appears disconnected from its main body as a research area, though it is linked thematically. The main research area of the book doesn’t cover nearly the range of authors and sub-disciplines as many academic sub-disciplines of epistemology or political theory, though that is largely an inescapable function of the subject matter.

Neoreaction is an analysis of the contemporary, largely American political movement of reactionaries—the overlapping communities of the alt-right, neoreactionaries, and Dark Enlightenment—usually often libertarian in philosophy and white supremacist in ideology.

---

1 Riggio 2016.
They are best known for racist and misogynist online attack mobs, a hatred for so-called “political correctness,” and a conception of free speech as the inalienable right to be racist, sexist, homophobic, and cruel to whomever they wish in public. Their first major campaign was Gamergate. The most significant leaders for this new reactionary movement are Milo Yiannopolous of Brietbart News and Republican Presidential candidate Donald Trump.

Sandifer does not spend much (or any) time on these figures, thankfully. His is a philosophical analysis of the three men whose ideas formed the movement’s theoretical core. They are author and think-tank head Eliezer Yudkowsky, software engineer Curtis Yarvin (who blogged his key texts for the movement under the name Mencius Moldbug), and academic philosopher Nick Land. Land literally wrote the book on neoreaction’s ideology, The Dark Enlightenment. Land and Yarvin are openly allies with the new reactionary movement, while Yudkowsky counts many reactionaries among his fanbase despite finding their racist politics disgusting. Yarvin and Yudkowsky also receive financial patronage from billionaire Trumpist Peter Thiel, as part of his investments in the Silicon Valley startup Urbit and the transhumanist artificial intelligence project MIRI, respectively.

The activist, artistic, political, academic, and business communities that surround and entwine neoreaction is a confusing bricolage of different actors and ideologies. Sandifer focusses on that philosophical triptych to understand the ideas underlying the West’s most powerful anti-democratic social movement operating today. While his analysis has many facets, the one most relevant to SERRC is how he understands the neoreactionary conception of reason and truth. Fitting for a movement that considers democracy and anti-racism a mistake, that conception resurrects a model of rationality that just about every professional in the theory of knowledge considers long-discredited and obsolete.

Rationality as Pure Reason, The One Self-Consistent Truth

A noble dream lies behind the filth and rage of neoreaction. That dream is a vision of truth as a simple clarity—there are facts and falsehoods and truth is univocal, a simple matter of right and wrong. Human progress comes from being less wrong, more rational, refining our faculties of knowledge, overcoming our biases, attaining a more perfect, more objective, more universal rationality. The embryo of the movement lived in the community pages of Yudkowsky’s blog LessWrong, a website dedicated to refining human rationality.

Yudkowsky’s primary vision for LessWrong (and the group blog from which it spun off, Overcoming Bias) was to introduce his own theoretical approach to bring human intuition more in line with the perfection of mathematical and statistical knowledge. Of course, his own and his community’s ignorance begins here, since mathematical knowledge does not operate with absolute and universal precision. But Yudkowsky asserted that it did, and that several tools cherry-picked from probability theory and physics would make a solid framework for a purified reason, where problems become steadily simpler, distinctions of true and false more stark and easily decidable.

This inspires directly the community’s political extremism—the alt-right’s disgust at any perspective or experience that introduces complexity to their simple view of the world. Yarvin’s political philosophy is built on such a stark simplicity—that the sole purpose of
government is to maximize a society’s profit through unification and authoritative control. Sandifer insightfully calls it the political theory of a pathologically single-minded engineer: the right solution can only be the most simple and elegant, perfect geometry. A desire to understand the world with total clarity articulates itself politically as authoritarianism. The question of what it is right to do becomes the simple question of what the Leader has ordered.

Yarvin’s approach is fairly clear by about the fourth chapter of his seven-chapter Neoreaction: A Basilisk, though he examines neoreactionary political philosophy in detail in the first two chapters. Sandifer’s circuituous style is a benefit if you come to the book looking for a complex engagement with a multifaceted social phenomenon that contains many internal paradoxes and conflicts. But you would consider it a detriment if you primarily want a straightforward analysis of the alt-right’s philosophy of politics, knowledge, and truth.

It can be difficult to identify at first glance where the primary failure lies in the alt-right’s embrace of such an unrealistic conception of truth. The alt-right/neoreactionary movement itself often embraces willful ignorance in the name of fighting political correctness. That includes their willful ignorance of the cutting edge research in rationality and truth that many SERRC contributors and our wider academic community do. At the same time, I cannot help but wonder if there is also a failure in the academic community of social epistemologists and other theorists of how complex knowledge can be to reach these people. The general critique of the insular nature of professional academic communications applies.

Yet that same critique ultimately applies to the LessWrong community as well, in their pursuit of a rationality perfected beyond what many here in the SERRC community consider humanly possible. My own undergraduate education in philosophy, with its home in a self-consciously Kantian department, supplied me a narrative of modern philosophy’s history that is quite useful here. Attempts to perfect knowledge and reason to achieve a perfect geometric simplicity break down through the project’s inescapable paradoxes. One can save that project only by betraying it, introducing limits of pure reason, patches, and no-man’s-lands where we must admit that the world is more complicated than the simple geometry we wanted to apply to it. For the neoreactionary movement and its leading philosophers, that compromise is a moment of horror.

### Applying Decision Theory to a Transhumanist Vision

The horror that drives neoreaction is not that which mainstream liberal thought about racism typically associates with such movements—the race wars of Hitler’s or D.W. Griffith’s imaginations, for example. It is instead the spectre of transhumanism’s failure. Instead of humanity becoming god-like, humanity destroys ourselves. A vision of utopia has been traded in for a vision of a yawning abyss. Neoreaction: A Basilisk discusses two paths to this horror in the works of Yudkowsky and Land. Yudkowsky’s own artificial intelligence research company, the Machine Intelligence Research Institute, is committed to the most optimistic ideal of transhumanism: immortality. Specifically, the immortality of the human mind’s merger with post-Singularity artificial intelligence, conquering death through upload
to a super-powerful AI. This hope was hideously perverted by a thought experiment that arose in the *LessWrong* community and gave Sandifer's book its title, Roko’s Basilisk.

Roko’s Basilisk is a triumph of paranoia at an intensity and absurdity rarely seen outside the works of Philip K. Dick. Roko’s Basilisk makes an abyss of the transhumanist vision; instead of a happy immortality as an upload to an artificial intelligence mainframe, your immortal existence in silicon is defined by constant and horrifying torture. All this is a matter of a calculation in timeless decision theory.

Here is a very fast version of the labyrinth of Roko’s Basilisk. At every moment when we think about whether to help build this super-AI, we weigh our preferences. One alternative is to join the AI project, and its consequences would be eventually building such a thing. The other is to do anything else, but if the super-AI eventually comes to exist, it will resurrect us in its mainframe and torture us eternally. Such retroactive blackmail is the perfect way to force us into creating it, so the reasoning goes. It is Pascal’s Wager by way of Silicon Valley, but whose God is inescapably cruel. It will offer grace or terror, and you have no real power to change your fate. With the power to simulate the entire universe perfectly, the machine knows your fate before you do.

Sandifer presents this techno-Calvinist terror god as an aspect of the wider existentialist terror that haunts the foundational alt-right philosophers. His analysis parallels Roko’s Basilisk with the seduction of the void in absolute, totalizing species death that haunts Nick Land’s recent work. He hints at an epistemological analysis in later chapters of *Neoreaction*, but his own focus is on the terror. I would like to sketch briefly how such an analysis would proceed.

Sandifer’s core hint at the epistemic flaw at the heart of Roko’s Basilisk comes when he chides Yudkowsky for thinking that the super-AI would think like a human. From one perspective, this is an argument about the human ability to imagine absolute Other-ness. It is a task well-suited to the science-fiction milieu in which Sandifer cut his teeth as a critic, and where the mobs of the alt-right first mobilized in Gamergate and the *Rabid Puppies*. But this most inventive of literatures still runs against limits. Aliens in the literature are not truly aliens, but allegories and parallels of human character and culture. Even physiologically, most alien species are mashups of earthly creatures or extrapolations of what would evolve in some specific ecological niche. The only alternative to these creatures of limited imagination would appear to be Lovecraftian pure others—creatures that can only be described through the psychological collapse of the characters who experience them. The same limits appear in our reasoning powers.

Let us accept that there is a strong limitation to the power of human thought alone to imagine the radically other—whether in images or in personality and reasoning. If an artificial intelligence as advanced as Yudkowsky imagines one day exists, we will not be able to mimic its reasoning abilities. So we will never know whether it would carry out the Calvinist blackmail at the heart of the Roko’s Basilisk thought experiment unless we actually encounter such a thing. The reasoning and actions of an intelligence so far beyond human abilities are genuinely beyond our comprehension—they will be opportunities for us to learn. Yudkowsky’s web communities were called *Overcoming Bias* and *Less Wrong*, not *Perfecting*.
**Knowledge and Absolutely Right.** In their initial presentation, they accepted human reason as limited. Yet timeless decision theory seems to be the tool by which Yudkowsky and his followers could genuinely reason as gods. At least, they believed so, using timeless decision theory to emulate a machine-god in human conversations.

I understand how tempting it would be to use timeless decision theory to perfect human knowledge. Decision theory mathematics calculate the relative utility of given preferences in the abstract, so we can know the best course to take in all such abstract considerations. And we can easily consider this calculus from a position abstracted from time. The problem is that such a position is also abstracted from human life as it is lived. At best, a decision’s utility calculations occur sequentially—every change in circumstance gives its variables different values. A genuinely timeless calculation ends up tied in knots, either from accounting for all the changes at once, or roped into the vicious paradox of being a necessary behaviour that can only be a contingent and free act.

An additional failure of the Roko’s Basilisk thought experiment is particularly revealing regarding the nature of the neoreactionary community that has so many roots in *LessWrong*’s culture and norms. Yudkowsky and his community had no problem conceiving of the AI-god of the Basilisk as having perfect knowledge, the computational ability to simulate the entire universe with absolute perfection, and an eternal cognitive perspective from which timeless decision theory would actually be workable. But they could not imagine such an AI-god having a similarly advanced morality. Postulating that their super-AI would threaten and blackmail everyone who conceived of Yudkowsky’s Wager without following through on joining and funding advanced AI research ascribes it a pettiness and cruelty that is all-too-human. For all the cognitive perfection Roko’s Basilisk grants to its imagined god of an AI, its moral perfection remains inconceivable.

"**Let us assume that we are fucked**”—The Horror of the End

The above quote opens *Neoreaction: A Basilisk*. It is Sandifer’s casually prophetic premise for all of *Neoreaction: A Basilisk*, his declaration that every pathway into understanding the neoreactionary movement and its philosophy will inevitably either pass through or end in horror. Nick Land cultivates that horror in his recent work, and Sandifer offers the deepest engagement with Land’s work among the central three. Yudkowsky and Yarvin/Moldbug receive equal attention, but they are targets for explicit arguments against their approaches and ideas. Land offers the most to chew on philosophically.

Yudkowsky’s mission to perfect human knowledge and artificial intelligence ends up motivated by the fear that the final product of its success will blackmail and torture those who helped make it on grounds that they did not try hard enough. Yarvin explicitly advocates authoritarian government by a white technocratic business elite. Land believes that nationalist tribalism and the accompanying race war over scarce resources is the only social model capable of surviving our upcoming—and to be frank, already-begun and ongoing—ecological collapse. Sandifer rightly criticizes Land for, in Land’s own words, throwing his lot
in with “racist little shits.” But Land is the only one singled out for this particular intensity of sharp contempt.

Yudkowsky has largely disavowed the alt-right and neoreactionaries who learned their models of reason, truth, and argument from his online communities. Yarvin has been an authoritarian from the start, introducing the neoreactionary communities to the noxious ideology of nationalist libertarianism. But Yarvin is also a largely terrible writer, rarely able to say in a few thousand words what he would prefer explain with a book’s worth of rambling tangents and rants.

Land seems to come in for the worst criticism because he should know better. His *Lure of the Void* is a new landmark in understanding the concept of death in the context of total species extinction, ecological annihilation, and cosmic emptiness. His work had the potential to supplant the pretensions and caterwauling of a thousand obsessive Heideggerians. That potential also included innovations in the form of philosophical writing, particularly its blend with narrative fiction in the sci-fi-horror novel *Phyl-Undhu*.

Yet his major philosophical follow-up to *Lure of the Void* was *The Dark Enlightenment*, an explicit courting of the nationalist alt-right, including many enthusiastic acknowledgements of Yarvin’s influence. The latter book outlined a political philosophy where the sovereign authority of the state transparently owns all material and people in its territory. Citizens literally become property of the state, and the purpose of government is to maintain productivity and profitability. Democracy, with its back-and-forth of opposition parties in power, facilitates rapacious consumerism and corruption instead of long-term investment. Land’s idea of democratic rule is for a government to consume so much of the nation that there is nothing left when the opposition party takes over.

Following the most radical libertarian ethos, the conceptual distinction between government and business collapses. To rule is not to be a steward of common resources and wealth, but to be the chief executive of the state and sole proprietor of all material and people within it. For Land, the highest politics is the unquestionable authority of the enlightened despot, managing a society that is also his property for optimal productivity and return on investment. Only under such authoritarianism would people be free from the self-destructiveness of their own greed and selfishness. Democracy enables greed to such a degree that it destroys the potential for the good life. Land approvingly quotes Peter Thiel: “Democracy and freedom have become incompatible.”

Land’s embrace of this terrifying ideology is rooted in the implications of the horror he sees in humanity’s future. As Sandifer rightly describes, Land does see the end coming, unfolding from the ecological collapse that the toxic products and destructive processes of the last two centuries of heavy industry have created. His rebuke to transhumanism is that the only way it could help humanity survive this crisis is if transhumanist technologies change us so radically that we become Lovecraftian Others compared to our current nature. The only way to survive that collapse without emerging from the other side of an abyss of the absolutely alien is a nationalist bunker mentality. You take dictatorial control of your resources and defend yourself with all the weaponry, fear, and xenophobia you can muster.
Empathy and Creativity as an Antidote to Violence

The virulent white nationalism that has taken over the most energetic discourse of American conservatism is no stranger to us now. But Sandifer’s *Neoreaction: A Basilisk* is a major extended treatment of the philosophical ideas that pulled this community together. Many of my SERRC colleagues might find it disquieting that this toxic popular ideology grew in online communities dedicated to rationalism and post-humanism. Nick Land, the strongest philosopher among the alt-right’s thought leaders, has said that he embraced violent nationalism because, in the face of an ecological crisis that threatens to make Earth itself a post-human world, one’s best hope for survival is membership in a tribe that zealously defends its few precious resources from outsiders.

But Phil Sandifer is no neoreactionary, neither am I, and neither is anyone in the SERRC community of contributors and readers. So *Neoreaction: A Basilisk* ends on a hopeful note that empathy and creativity can be an ethical antidote for the violence into which the alt-right seeks to plunge humanity. But questions remain. What can be a source for such a vision of an empathetic society? What framework for social and political creativity can take us there?

Sandifer’s own answer is remarkably idiosyncratic, and perhaps that is the point. *Neoreaction: A Basilisk* ends with a walkthrough of the metaphysical and philosophical symbology of William Blake’s mythological canon. Blake has had no influence on the alt-right movement itself, but he has had a significant influence on Phil Sandifer. Significant engagement with Blake’s ideas can be found in prominent places in all of Sandifer’s major works—in his multivolume *TARDIS Eruditorum* on the history and political philosophies of *Doctor Who* the , as a recurring touchstone in his ongoing project on the British Invasion period of mainstream American comics *The Last War in Albion*. Sandifer also draws on an unorthodox, yet eminently sensible and historically-grounded, reading of Alan Turing’s imitation game—that computer (and human) intelligence does not regard principally language itself, but the ability to understand a different creature well enough to imitate her successfully.

But he does not intend his reflections on Blakean mythos and a Turing-inspired empathy to be an ideological template for the rest of us to follow in fighting the white nationalism that incubated in nerd culture and seems to have corrupted it beyond redemption. Even the alt-right community itself is fractured and plural at the level of ideological dogma. They are fellow travellers on a racist, nationalist trajectory with common roots in online message board communities, transhumanist interest, nerd culture, sci-fi and horror fandom. This philosophical imaginary is Sandifer’s own creative inspiration, playground, and vision. Each of us must develop our own, with empathy, creativity, and a love for the diversity and variety of humanity as the only common values.

For example, my own philosophical imaginary powering my line of flight to a diverse peaceful world shares some common ground with Sandifer, but is otherwise a completely
different route. We share the same pop-cultural ethical influence from *Doctor Who*, the moral
grightness of pulling justice from a fundamentally unjust world by breaking all its rules. But
my own inspirations also draw on the tradition of emancipatory materialism in modern
Western philosophy (the trajectory from Spinoza and Machiavelli, through Marx, Nietzsche,
Gilles Deleuze, and Antonio Negri), Emmanuel Levinas’ phenomenology in the spirit of
Talmud, and the history of liberatory, anti-racist activism in Canada, from the Riel Rebellion
through the Cooperative Commonwealth Federation to Idle No More.

My own philosophical inspirations provide me with my more academic critique of Sandifer’s
book. He describes his critical readings of Yudkowsky, Yarvin, and Land as applications of
Deleuze’s technique of creating monstrous readings of historical philosophers. That
interpretive method makes radical breaks with the mainstream conception of a thinker’s
works that are nonetheless faithful, monstrous conceptual children that the inspiration
would abhor, but recognize in his own work. Sandifer describes this monster-making as
destructive, and uses that technique to expose the vulnerabilities and blindnesses of the
generative philosophers of the alt-right. But Deleuze’s own spirit in monster-making was just
as creative as Sandifer’s alternative path to violent nationalism. Deleuze wanted to make
new, contemporarily relevant ideas emerge from thinkers long rejected or whose ideas had
become taken for granted. They were radical redemptive readings.

Redemption is a path too terrifying to take with Yarvin/Moldbug the rambling egotist,
Yudkowsky the blinkered think tank merchant, or Land the broken visionary of horror. At
least too terrifying now, when the noxious political movement they inspired form the shock
troops of a demagogue one election from controlling the world’s second-largest nuclear
arsenal, when their poison-fingered disciples constitute a raging online mob of hackers and
harassers. But if one day, we can relegate American neoreaction to the historical trash bin
where it belongs, redemption may even come for this sad trinity.

Sandifer, perhaps against his better judgment, may even have written that redemption’s
prologue.

**Contact details: adamriggio@gmail.com**

**References**

Riggio, Adam. “Beyond the Academy: Solutions to the Academic Brain Drain in Embracing
Public Creativity and Leadership.” *Social Epistemology Review and Reply Collective* 5, no. 4