Media-N

my monstituces composer: Looking for the Pre-Emergent Social Consciousness of AI in Small Data Literary Synthesis

NATHAN JONES
Lecturer Fine Art (Digital Media), Lancaster University

ABSTRACT

In the context of an algorithmically designed workplace, the glitch traces the form of the unforeseen social reality. What this paper does is present glitch poetics as a methodology for exploring futures of language-technology hybridity that do not emerge from philosophical, social science, or computer science research. Glitch poetics is concerned with linguistic errors in the same way that Glitch Art was concerned with the pixelations and clicks of 1990s and early 2000s compression mechanisms. AI-authorship is a potent area of study for glitch poetics because it is distinguishable by evidently faulty language forms, whose faultiness contains a recognizably contemporary "texture." The first half of the paper includes an overview of AI research and practice from philosophical, technical, visual and sonic art fields, and current ideas in concept-forming, error, and innovation, such as "hyperstition," "diagramming" and "glitch." I then turn to the specific case of language acquisition and synthesis in AI and what the glitch method might have to offer in this area. Drawing on the work of Peli Grietzer, I ask whether AI-authored texts can be said to have a "vibe" distinct from that of the author's whose text it is trained on, and therefore has its own basis for forming new social possibility. In one attempt to answer such a question, I make an AI produce faulty, glitchy texts, using a deliberately small corpus, in which the notion of a single-author-trained AI is broken down and confused. I close-read the odd language-salad of the texts written by this "small data synthesis" glitch method in order to discern whether there is a "structure of feeling" (Williams) in them: a pre-emergent sign of a social being-to-come. The paper concludes that making AI produce texts that are manifestly wrong offers a different trajectory for what AI can offer intellectual thought: one that is more potent than the progression towards more and more human-like textual impersonations. What is necessary, however, in order to take advantage of these strange new language forms, is to divert attention towards them in genuinely experimental, thoughtful situations, and to seek to absorb them into contemporary discourse. My proposition is that, along with "machine behavior," there should be a concerted effort to invoke and interpret "machine misbehavior" as a form of thinking to come, both within and beyond the field of language-synthesis.

INTRODUCTION

In universities, much intellectual labor is concerned with acts of reading, parsing, and interpreting existing texts and synthesizing them: bringing the knowledge contained in primary and secondary texts into contact with one another and turning them into something new. This paper, for example brings ideas established in computer science into conjunction with those from new media art and literary theory. What will it feel like to be in a workplace where humans collaborate with artificial intelligences on synthesising new texts in this way? What kind of ideas will these new "factories of knowledge"[1] produce? In this paper, I suggest that glitch poetics—a critical creativity that valorizes and manipulates errors in digital-age texts—offers a new trajectory for engaging in the intellectual work done with machines. Glitchy texts express concepts as aberrations, and the word-salad of faulty AI-authored synthesis is an aesthetic form that contains new possibilities for thought.

The method here is adopted from Glitch Art, a movement which has sought to overload and "bend" new media devices in order to produce new critical perspectives, and in doing so developed forward-looking aesthetics from within the pixels, clicks, and smears of faltering digital codes. The glitch artist and theorist Rosa Menkman[2] describes a "critical and sensory encounter" that distinguishes a glitch from an ordinary breakage or error, acknowledging the deeply combined affective and intellectual capacity of seeing glitches happen. Menkman coins the term "glitch moment(um)" to suggest that these encounters have a propulsive, even prophetic force on the body and mind. The glitch is speculative because it can manifest latent, trace qualities in media that will one day predominate as social norms, *and* because a glitch offers an exit from the predetermination of algorithmic processes and human habit. In what follows I want to explore how the next digital culture shift—that of the AI-era—might produce its own future aesthetics in the form of errors, specifically language errors that might be said to produce *hypersitions* as unpredictable, unforeseeable concepts-to-come, demanding surprising new ways to read them.

From speculative realism, hypersition is a portmanteau of hyper- and superstition, usually referring to a means by which a concept creates real things in the world: "Functioning as magical sigils or engineering diagrams, hyperstitions are ideas that, once 'downloaded' into the cultural mainframe, engender apocalyptic positive feedback cycles."[3] Hyperstition is a term that is uniquely understandable in today's moment: a moment of increasing temporal strangeness, where technical change has accelerated so that we see more readily the effects of speculative concepts manifesting in our everyday lives, and where the formal innovations of concept invention operate in a tight interplay with the algorithms that analyze, proliferate, and quantify them. Hyperstition is currently associated with a particularly sinister brand of "accelerationist" thinking that imagines the mass extinction of humans as the ultimate price for machinic evolution, a suggestion that has startling resonance in today's post-abundance world, where the earth's resources are depleted in

the service of huge technology giants. The stakes are high for understanding the nature of autonomous diagrams of the future and how to connect them to a social reality.

In the field of design—specifically in relation to algorithmically designed architecture—Betti Marenko states that "the glitch-event … as the openness of digital potential and its algorithmic autonomous trajectory … [is] a form of diagrammatic divination."[4] The diagram is an important figure in futures-thinking, suggesting a process by which visual and conceptual phenomena make new space for possibilities. It is in the diagram—whether architectural wireframe, or new conceptual coinage—that a space is formed outside of the inevitable, algorithmically defined future. The errors of the next digital culture shift might produce their own future aesthetic and critical possibilities, but we will need to pay attention to their emergence and draw ourselves into their path.

Philosophers of technology make relatedly extravagant predictions for the societal shifts that will occur in the context of more advanced artificial intelligence. Some of these, such as Ray Kurzweil,[5] predict that today's society will be unrecognizable within a generation because of AI's "exponential" growth. The philosopher Nick Bostrom imagines a near future where society has been changed by an algorithmic "superintelligence," in which our own fates "depend on the actions of the machine."[6] He identifies a number of tasks that would signify a "general" machine intelligence. Among them, "[o]ne of the many tasks on which superintelligence (or even just moderately enhanced human intelligence) would outperform the current cast of thinkers is in answering fundamental questions in science and philosophy." Indeed, Bostrom observes that the work of philosophy has been on the agenda since the beginnings of AI research; he cites an event at Dartmouth College in 1956, which proposed that "an attempt will be made to find how to make machines that use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves."

The "singularity" is a concept based on the idea that this age-old concept for a thinking machine will become a reality by continuing on an exponential evolutionary line and that it will eventually absorb and supersede human intelligence to such a degree that the laws of our existence are overturned. Of the arrival of the singularity Murray Shanahan suggests that "the institutions we take for granted—the economy, the government, the law, the state—these would not survive in their present form."[7] Considering the reverse side of our growing dependence on machines, Shanahan considers the implications for ethics when a "conscious" AI is put to work: "[AI] confined to virtual reality and forced to work as slaves … if the AIs are 'mindless automata' who lack consciousness, and therefore lack the capacity to suffer, this doesn't matter. But suppose they are conscious. Suppose they experience their predicament just as a human would."[8] By asking how we would respond to a social stratum composed of conscious machinic intelligences, Shanahan also proposed the question of how they will treat us. There is clear suggestion here that a greater understanding of the social implications for newly intelligent machines that approach or

achieve consciousness is desirable, but perhaps these philosophical futurologists' language is not appropriate for developing one. I will suggest below that this is partly because their work is based in incomplete historical interpretations of development to the current day and on the supposition that these lines—or rather this line—of history will continue in to the future.

In a 2015 report on the *One Hundred Year Study on Artificial Intelligence*[9] from the perspective of computer and social science, Stanford researchers investigating eight areas of AI application showed that there's no immediate suggestion (or danger) that AI will replace human cognitive laborers on general high-level intellectual tasks: "While drawing on common research and technologies, AI systems are specialized to accomplish particular tasks, and each application requires years of focused research and a careful, unique construction."[10] The authors suggest that the focus of social preparation in the medium-term should be on understanding AI as collaborators and mitigating the fear factor of lost jobs.

> AI will likely replace tasks rather than jobs in the near term, and will also create new kinds of jobs. But the new jobs that will emerge are harder to imagine in advance than the existing jobs that will likely be lost.[11]

"New kinds of jobs," of course, are not reducible to practical lists of things to do, but also include social and not-yet-social relations that such jobs will engender. What I am concerned with in this paper is the pre-emergent forms that are present in today's technologies, and how they relate to near futures. I propose that by glitching these technologies, we can get a better *feeling* of the diverse futures that we might be unknowingly authoring with our machines.

Art can be a useful way of envisaging what unpredictable social forms are embedded as possibilities in our present day. Evoking a form of labor that anticipates the logic of hyperstition in his novel *The Futurological Congress*, sci-fi author Stanislaw Lem coins the term "linguistic futurology": "divination through linguistic derivation. Morphological forecasting! Projective etymology!" The linguistic futurologist proposes that the computational production of new words can be used as a forecasting tool:

> A man can control only what he comprehends, and comprehend only what he is able to put into words. The inexpressible therefore is unknowable. By examining future stages in the evolution of language we come to learn what discoveries, changes and social revolutions the language will be capable, some day, of reflecting[12]

Lem's proposition is that computationally generated coinages are the basis for new thinking for his character. In this future, the linguistic futurologist—a human laborer—remains central to the working of the machine.

The question of whether AI can participate in the academic workplace could be the question of whether AI-produced concepts are capable of entering the feedback cycle of the cultural mainframe through the intellectual engagement of human workers. If this were to happen, I suggest it would be detected first as a new aesthetic, and only secondarily, if at all, as a solid conceptual method.

BEHAVIORAL AI

Contemporary AI algorithms are colloquially referred to as "neural nets," because they are imagined to operate like human neuronal networks in the brain: just as each neuron in the brain can be connected to others, individual neural net algorithms can also be chained and made into networks. Just as with neurons that only "fire" after a threshold of inputs has been reached, components of a neural net are only active when a particular "action" is triggered by the calculation they make, and the way they're combined means that—although they're just glorified mathematical models—their operational complexity quickly thickens to a point that it makes aspect of the way they work, and what they're going to do, unknowable. The problems produced by this exponential rise in the complexity of AI algorithms has led scholars in Stanford's AI-lab to propose a new subject of study, called "machine behavior." [13] The authors of this paper note that computing "methodologies aimed at maximized algorithmic performance are not optimal for conducting scientific observation of the properties and behaviors of AI agents." They propose that social science methodologies should be deployed in understanding what AI is like, based on its inputs and outputs, rather than on technical analysis of how it gets there. Their motivations for doing so have specifically to do with the inverse relationship between complexity and visibility:

> The complexity of individual AI agents is currently high and rapidly increasing. Although the code for specifying the architecture and training of a model can be simple, the results can be very complex, oftentimes effectively resulting in 'black boxes'. They are given input and produce output, but the exact functional processes that generate these outputs are hard to interpret even to the very scientists who generate the algorithms themselves. [14]

To this behavioral focus, we should add the methods of affect theory and affective readings. Raymond Williams' concept of "structures of feeling"[15] suggests that embedded in the gaps and discontinuities of emergent literary styles, there is a kind of latent social potential that evades both machinic and behavioral analysis. Williams emphasizes the difficulty of reading for the "pre-social" in behaviors, and suggests that we can instead find emergent relationships as tensions within aesthetic domains, at points where dominant artistic forms are displaced or put in tension: "The tension is as often an unease, a stress, a displacement, a latency: the moment of conscious comparison not yet come, often not even coming."[16] Williams' observation is that

structures of feeling in texts contain observations about social structures to come, but that once emerged they become indistinguishable from "the known relationships, institutions, formations, positions" of the social, so that they recede as a subject of analysis. Williams suggests that the emphatic present-ness of our encounter with artworks contains a particular potential for reading emergent shifts in the social: "It is not only that, to complete their inherent process, we have to make them present, in specifically active 'readings'. It is also that the making of art is never itself in the past tense. It is always a formative process, within a specific present."[17]

> This process can be directly observed in the history of a language. In spite of substantial and at some levels decisive continuities in grammar and vocabulary, no generation speaks quite the same language as its predecessors. The difference can be defined in terms of additions, deletions, and modifications, but these do not exhaust it. What really changes is something quite general, over a wide range, and the description that often fits the change best is the literary term "style."[18]

Structures of feeling are the signs within texts of an ambiguous configuration of the social that has not yet fully emerged, and this can be seen in a distinctively glitchy form in contemporary literature. Autofiction authors like Tao Lin,[19] Ben Lerner, Elena Ferrante, Megan Boyle, and Sally Rooney present a literary style whose tensions with dominant "realist fiction" evoke something latent that is unique to the experience of the digital age. There is much in the work of these authors that is shared with precursors in literary realism and autobiographical traditions, but it is what is most uncomfortable, most displaced from what has gone before, stylistically, that contains within it the grain of emergent social relations. One such distinguishing stylistic feature in contemporary autofiction is the common occurrence of bodies and persona that "glitch" into their background, or whose sensory apparatus becomes indistinguishable from the world it is sensing, as in these examples from Tao Lin's novel *Taipei:*

> [Paul] felt continuously aroused "somewhere," including some times, it seemed, outside his body, a few feet in front of him, or far in the distance, in a certain store or area of sky, or in an overlap, shifting in and out of his chest or head or the front of his face.[20]

> Paul and Erin discussed their movie in a dialogue that sometimes overlapped with their inner monologues, which they sometimes introduced to the dialogue, or abandoned to focus on the dialogue, or both externalized, like pets into a shared space, to observe.[21]

This phenomenon can be seen also in experience of "dissolving boundaries" of Lina, Elena Ferrante's protagonist in the *Neapolitan Novels:* "the sensation of moving for a few fractions of a second into a person or a thing or a number or a syllable, violating its edges."[22] And in the work of Sally Rooney whose alter ego "felt a weird lack of self-recognition, and I realized I couldn't visualize my own face or body at all. It was like someone had lifted the end of an invisible pencil and just gently erased my entire appearance."[23] Poet and artist Juliana Huxtable has also

emphasized her inseparability from the machinic contexts through which she imagines her lyric subject. Set within a culture of electronic dance music and performance art, Huxtable's poetry imagines technology itself as a kind of primordial soup from which the collective identity emerges to speak.

> IT WAS REALLY AN EXPERIMENT IN POSTING ALTERNATE WAYS OF CONJUGATING MYSELF: FEELING PUSSY, PERCHED AND *NEVER PARCHED,* "TEQUILA SODA PLEASE" ICE COLD SWEAT DAMPEN HAND PARLAYING SPLIT … STRIKING HIPS AND LIMBS THROUGH SNYCHRONOUS, BEGUILING COMMANDS TO TRACE, SLIPPED, SPUN AND VOGUED MY WAY THROUGH A SUPERNOVA MADE OF CAMERA FLASHES, ALL REFRACTED THROUGH SHATTERED HANDLES OF WELL VODKA … SUBWOOFERS SHOVED ME INTO A FLOATING AND BOUNDLESS MASS OF SHADE [24]

Glitch poetics connects the "structures of feeling" concept to the specific digital context—and the existential sense of the glitch. Each of the visions for an AI-future identified above, to different degrees of intensity, emerges from a frame of mind in which existing narratives—fictions that we tell ourselves about how "the world" as a single entity is dominated and controlled by human actors—are extended into a fundamentally similar narrative of domination and absorption. In doing this, they mistakenly trace the bias of historians towards "grand" narratives, into a future determined by the acknowledged dominant cultures of today. Conversely, in her history of computing beginning with an account of the "topsy-turvey thinker" Ada Lovelace, Sadie Plant suggests:

> Histories of technology – and indeed histories of anything at all – are always riddled with delicious gaps, mysteries, and riddles … but what announce themselves to be founding fathers, points of origin, and defining moments only ever serve as distractions from the ongoing processes, the shifting differences that count. These are subtle and fine grained, often incognito, undercover, in disguise as mere and minor details. If, that is, they show themselves at all.[25]

I am inspired by this understanding of history. In what follows, I will attempt to show in a number of specific AI-produced visual artworks, as well as in my own experiments with AI-produced texts, that it is in the deviations from a seeming purpose, in the inefficient—even useless—uses of AI, where new aesthetics, and by extension, new social forms are being proposed. AI—as with any cultural form concerned with artifice—produces outputs that fail to render an accurate replica, instead producing faulty, low fidelity, smeared, even ghostly forms, which are the seeds of the new. It is my intention to observe these undercover forms of a current technology, and I do so through a practice of glitch poetics that intervenes in technologies designed to produce artificially human-seeming texts.

AI LABOR FEELING

The phenomenon of dissolving boundaries divined in glitchy literary practices such as Huxtable's has itself anticipated a social potential that is currently being realized in the latest uses of AI in art. The electronic musician/collective Holly Herndon produced her 2019 album *Proto* with an AI baby called Spawn. Herndon's use of AI is different from many of its applications today, because rather than emphasizing the tasks automated by Spawn, the production method of *Proto* demanded *more* human labor than most electronic music albums.

> If manual labor is a currency of authenticity in music, *PROTO* is more legitimate than most guitar records I can think of.... Spawn is only one member of an ensemble of 14 that includes Dryhurst and developer Jules LaPlace. The remaining members of the group are vocalists who would meet to "feed" Spawn by singing and talking to her.[26]

Instead of the efficiency and accuracy of uses of AI that inform Bostrom and Shenahan's AI-futures visions, Herndon's approach emphasizes the production of new aesthetics that can come from computationally augmented collaborations: in her case, asking an AI to extrapolate from a corpus that included up to 300 other live human voices. Inherent in this is the notion of the creative and diverse corpus, which forces the computer to invent new aesthetic items at the intersection of the statistical maps of this diversity, triangulating their gaps and differences. Such use of AI for synthesis produces the conditions where new potentialities can be realized: growing out of a new collaborative method, this is very tangibly an aesthetic of a pre-emergent social possibility. Herndon herself notes of her first listen to Spawn's outputs: "I was horrified when I heard it. But it was also hilarious, because it's the mix of singing and speech and her [Spawn] trying to make sense of that."[27] This shock-response is characteristic not only of encounters with the new propositions offered by glitching media and human actors, but also of the evolution of media forms prior to our habituation to them.[28]

In visual phenomena, there is also a style starting to emerge in "misused" AI synthesis, particularly methods of synthesis that use purposefully small or diverse corpuses. These practices aren't described as glitch art, because they don't use "traditional" glitch techniques, but they emerge from a particular line of critical media practice that emphasizes unpredictable outcomes. And they turn an exploration of a technology into a sensational form. One of the most notorious methods for AI-produced artworks is the kind of autoencoder called a GAN (generative adversarial neural net), used by artists such as Robbie Barrat,[29] Mario Klingemann[30] and Helena Sarin,[31] whose focus on individual and diy practices has led her to coin the #folkAIArt hashtag. The conceit of each of these artists' works is that they train a GAN on a set of images, and through trial and error, the GAN generates an image that it can fool itself into believing is part of that set. The algorithms are

called "adversarial" because they consist of two parts, a generator and a checker, which make the trial and error process progressive in terms of fidelity to the set.

These artists are working in a context where the fidelity of imagery "imagined" by GAN techniques has reached—and surpassed—the photoreal. One field in which AI has excelled in particular has been in generating images, particularly of the human face. Companies like *Nvida* lead a field in which artificial images that are completely unrecognizable from actual photos of faces have been entirely "invented" by a computer program.[32] Different from the high-fidelity images produced by Nvidia's algorithms,[33] artists like Barratt, Lingemann, and Sarin produce images with a new kind of error-aesthetic. It is a smeared quality that is very different also to the computer-generated imagery of previous computational eras, whether the pixelated approximations of flat planes in early computer art, or the glossy highly defined images of 3D cinematic imagery. Comparing the glitchy imagery of Robbie Barratt with high-fidelity face-generation technology and 1990s Glitch Art, we can detect the presence of a new visual terrain being diagrammed.

***Figure 1**. Robbie Barratt,* **AI Generated Nude Portrait,** *2018, digital image. Permission courtesy of the artist.*

As was formerly the case with 1990s-era glitch art, the level of sampling is revealed as a texture in broken or misused instances of the GAN, but the feeling of this output is manifestly new. Within this encounter with a texture of the media, there is a *feeling* of what the technology might mean

to us, in a near future dominated by quantitative analytical means. In comparison to both glitch art and HD AI imagery, glitched GAN imagery appears depleted, straining to retain a representative form: it is a copy without an original. GAN-produced images of the human body in particular look like the Francis Bacon depictions of a body in crisis, whose meaning and structures have been sucked out of it, leaving only corporeal matter, bringing to mind Deleuze's descriptions of the weird sensuality of that artist's work: "The shadow escapes from the body like an animal we had been sheltering."[34]

The innately haunted, depleted quality of GAN images is expanded on in recent works exhibited as *A Study of Invisible Images* by Trevor Paglen,[35] in particular his image "Highway of Death". The text for this image tells us the GAN was trained on photographs documenting the aftermath of the first Gulf War, "the first smart war": "The AI sees a landscape characterized by burning oil fields, desertification, depleted uranium, birth defects, and other effects of the war." What Paglen finds in this startling use of GAN to make an archetypal image is a resonance between the defects caused by war, as an average across landscapes, ideologies, and bodies, the affective encounter with disinterest.

Figure 2. Trevor Paglen, **Highway of Death (Corpus: The Aftermath of the First Smart War***), Adversarially Evolved Hallucination, 2017, dye sublimation metal print 81.3 x 101.6 cm. Courtesy of the artist and Metro Pictures, New York.*

NEW LITERARY VIBE

In his paper "A Theory of Vibe,"[36] Peli Grietzer outlines a mathematical philosophy of literature based on the way that artificial intelligences such as GAN's process a literary corpus. Grietzer suggests that training an artificial intelligence on a corpus of an author's work distills that author's "vibe" into an abstracted mathematical formula.

> The meaning of a literary work … lies at least partly in an aesthetic 'vibe' or a 'style' that we can sense when we consider all the myriad objects and phenomena that make up the imaginative landscape of the work as a kind of curated set.

In effect Grietzer proposes that a Williams-esque "structure of feeling" in an author's ouvre, abstracted from any given example of this style in action, can be mathematically engendered. This is not a computer producing new ideas, but it is certainly an example of new kinds of critical thought being enabled by new computational models. What this mathematical formula can then be used for is producing different—but mathematically identical—examples of this author's literary style in action. What I am interested in is whether the AI can be used to generate *new* vibes, styles, and "structures of feeling": pre-emergent concepts of its own.

Another artist who designs creative models for language-based collaborations with machine learning algorithms is the code poet Alison Parrish. Parrish's work is notable for its astute synthesis of concepts from linguistics with forms of modelling from computing and her work with open-ended language tools, such as Word2Vec.[37] Parrish uses the term "language-vectors" to describe the attribution of words with numeric values in such a way that their semantic meaning is recorded within this value. Word2Vec arrives at a language vector by plotting words within a "multi-dimensional space" in which words with similar meanings will be more proximate to each other. Parrish's supposition is that original texts, tagged in this way, can be used as the basis for tweaking, manipulating, and filtering existing texts to produce new ones. Parrish's latest work is *Compasses,*[38] a set of visual poems produced by an engine using phonetic similarity to compare words.

> This model has two parts: a "speller," which spells words based on how they sound, and a "sounder-out," which sounds out words based on how they're spelled…. In "Compasses," I used this model to generate new imaginary words that exist in the negative phonetic spaces between the phonetic hidden states corresponding to names of members of well-known quartets.

The new words at the center of these poems stake out new semantic ground. And because of the mathematical logic behind their placement, there is a sense that we can intellectually "trust" them as the basis for an embryonic synthetic concepts-forming mechanism.
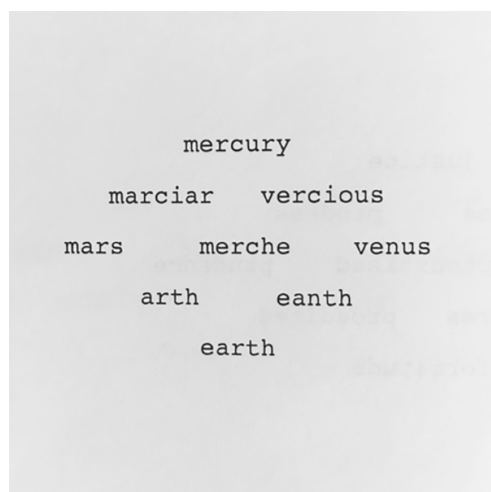
Figure 3. **Allison Parrish, poem from Compasses *series, 2019, digital image. Permission courtesy of the artist.**

As with the critical-aesthetic work of the GAN glitch artists, this work with purposefully impoverished and speculative textual synthesis takes place in the context of increasingly convincing artificial intelligence applications. Such is the level of fidelity of texts produced by OpenAI's latest language engine *GPT-2*[39] that it was only released as a concept paper and limited-function code, in order to prevent "malicious applications." The concept behind GPT-2 is similar to that used in the basic language-predicting algorithm we worked with for The Re-reader, described below, but rather than working at the level of the letter, OpenAI's engine works at the level of the word, to "predict the next word, given all of the previous words within some text." To a large degree, because of the huge amount of source-data the calculations of GPT-2 are based on (it was trained on a dataset scraped from over eight million publicly available webpages), its predictions are uncannily accurate. The authors state that their AI has "the ability to generate conditional synthetic text samples of unprecedented quality, where we prime the model with an input and have it generate a lengthy continuation." It also excels at "question answering, reading comprehension, summarization, and translation." The texts produced by this AI are indeed convincing version of a human-written text, and the example is given in the paper of a thoroughly complex, multi-paragraph continuation of a single human-written sentence. Seeing the texts that are generated, it is easy to see how OpenAI would consider their software a threat that would add to the issues around believability and information overload in what has been termed the post-truth era.[40] The issue is that algorithms can produce texts on any subject, containing any belief they want, like humans—but in a much greater quantity. And quantity has proven to be the dominant factor in online discourse.

Glitch poetics is a critical creative method that seeks invention in ways that are more similar to Parrish's AI language-syths than OpenAI's hi-fidelity authors. Parrish intuitively seeks out the gaps in logic left by her modes of analysis and writing, using the algorithm to propose fragile new

linguistic territories. Conversely (underpinned by a culture wary of entropy, as outlined below) OpenAI considers each deviation from the believably human as a "failure." They give the example of "word modelling failures" such as when "the model … writes about *fires happening under water.*" With glitch poetics readings, I suggest that we can trace the feeling that emerges autonomously from such phrases. Quantitative approaches to language such Parrish's and OpenAI's may share a pre-emergent social quality in this context. In my own experiments, documented below, the hope is that an encounter with the aesthetics of error in AI-produced texts offers a trajectory for thinking anew about our social futures in ways that diverge from the linear fates proposed by philosophy and computing disciplines and also contribute a literary basis for reading work such as Parrish's.

Using The Re-reader, an artwork interface developed with Tom Schofield and Sam Skinner in 2017 that uses an different kind of AI called a recurrent neural net (RNN), I have generated texts that attempt to triangulate the content of radically distinctive authors. As it makes use of only two books as a source text, rather than the hundreds of millions commonly used in AI applications, this experiment is about "small data"—such small data that it prevents the AI learning what "words" and "grammar" are. The result is that the AI must coin new terms and writing "styles" by creating a mathematically coherent synthesis of two statistically different textual phenomena. The method here is similar to Parrish's search for "gaps" in existing languages, and also to the Glitch Art approach of "jamming," where incompatible sources, files, operations are pushed into interaction in order to produce unwieldy new syntheses. In glitch practice, faulty outputs are read and interpreted as critical-sensory phenomena, and the errors form the basis for new aesthetic domains. AI offers both a method and a subject of study for a new glitch poetics process.

THE RE-READER

The Re-reader is a book scanner that transforms a physical book into a series of algorithmic, semantic, and aesthetic text performances. It was programmed by Tom Schofield and designed by Sam Skinner. It asks for a simple interaction:[41] A user places a print book on the glass scanner, and looks at a monitor mounted directly above it. The Re-reader scans the pages, finds the text on the page, recognizes individual characters and words, and performs some text-analytics on the text. It then displays the text one word at a time in fast succession on a monitor, for the user to "speed read." While it is doing this, side-screens display the progress of the speed-reader relative to a page location. Using basic word-based corpus analysis, the side screens flash red when the speed-reader reaches a word that it deems "important." At the end of the page run-through, The Re-reader displays a pixelated display of some paratextual, unrecognized element on the page. When it is dormant, The Re-reader adds the text it has just read to its own corpus and learns from this "what texts are." There is already a flavor of the kind of relationship one might expect with our machine collaborators here. The Re-reader turns the flat page into a temporal experience, and the implication is that this is an efficiency device. The user watches the screen and "absorbs" the texts' meaning—using the machine to clean up the "inefficient" physical encounter with the book.

Perhaps we can imagine a workplace in which the laborer's management can precisely determine how many words of pdf an academic can be expected to read per minute.

Figure 4*. Tom Schofield, Sam Skinner, Nathan Jones,* **The Re-reader,** *installation*, *Grundy Art Gallery, Blackpool, UK, 2017. Image courtesy of the artists.*

However, viewed from the perspective of the machine, the user is also performing a labor for the machine. For The Re-reader also writes: while the human is bathed in a pure semantic encounter with the raw material, every book scanned contributes to the training of a neural net that produces new texts based the statistical probabilities it finds in the text it has been fed. The Re-reader uses a glyph-level RNN mechanism to do this. Like GAN mechanics, the micro-sampling technique of the RNN is altogether disinterested in context and meaning, but focuses instead on statistically mapping its corpus based on granular sampling. With The Re-reader's RNN this sampling happens at the level of the glyph. It has no notion of language, but rather seeks to produce statistically likely combinations of glyphs. The RNN is called recursive, because it bases part of its calculation of what letter will come next on what letters have gone before— part of the input of the RNN's calculation is the previous output it made. In effect, the computer learns to write a "version" of a text that would fit into the corpus it has been fed. This technique was used by Google in a prototype game designed to advertize their "Google Docs Demo" app,[42] where authors could "collaborate" with an AI Shakespeare, Dostoyevsky, or Wordsworth that had learned to write based on a corpus of those well-known authors' works. This "realistic" version of AI functioned precisely because its logic was to produce a believable version of the corpus. For The Re-reader, Schofield adapted the Torch-RNN open source AI, which can be set on a spectrum between Hot and Cool. When the RNN is being "cool" it reproduces patterns of letters that it already knows from the corpus. When the RNN is being "hot" it produces entirely new sequences of letters, informed by the statistical likelihoods it has learned from the corpus. In machine learning there is a connection between the notion of increasing the latitude given to the algorithm to diverge from its corpus, and the acceleration of entropy that occurs in thermodynamics by heating: the hotter a textual output is,

the more chaotic and decomposed it becomes. RNNs work a glyph at a time if they are per-character, but other possibilities exist. The benefit of using glyph-level generation in this context is that it turns the quantitative calculations of the AI code into an available aesthetic at the word level.

For this experiment, I chose to feed The Re-reader the Raymond Williams book *Literature and Marxism,* in which his "structures of feeling" concept first arises, and Tao Lin's autofiction *Taipei.* This is a deliberately diverse and small corpus. The intention here was to produce conditions where the input is knowable, so it can be compared to the output, and where the output would be faulty, impoverished, and textured with errors.[43] These books are stylistically diverse, but are "read" by The Re-reader as parts of a single corpus, from which it learns to write, triangulating its sources in the same way as Herndon's voices. In a blunt way, the statistical modelling mechanics of the AI in The Re-reader will produce an intellectual synthesis of Williams' essays and Tao Lin's emphatically digital-age autobiographical novel. In reading the results, I was interested in two questions: a) whether the RNN's mathematical approach to concept generation in textual synthesis has a "vibe" that is readable and might be detected as a form of "social to come"; b) if there is new aesthetic and creative-critical potential in this synthesis of texts, which might suggest a new kind of readership for it.

In response the neural network spat out a range of texts, depending on how "hot" the originality setting was. The "cool" AI produced a limited set of characters, in a repetitive string down the page:

> /9
> /
> /
> /+

The "hot" one produced a more random, but unreadable word salad:

> dumerpechns —get,s—thyusou—com-fetwe— 2 pogress-or lap QA micter remadlh', e,ulR2 it with turXows e. the VP1 not _m5's.81 SP: 'f heablely foim:-squkal", so [F-SBRT.31
> Dewle lizox continoxsed—saen," syn fuvuib, in pyatignouts aving cloful refered ales gD.3

The first few hundred words of output for the "warm" synthesis (including its attempt at a colophon) are provided in the Appendix of this paper. At this level The Re-reader produces an appreciable approximation of the average of statistical concerns of the two books. The text is composed of granularly produced neologisms that are unlike the portmanteau-style word-jamming of earlier neologistic authors such as James Joyce or Mez Breeze, who built their works from

recognizable fragments. Each of the neologisms of The Re-reader's text contains a potent ambiguity as theory-fictional sigils, triangulated from their source and therefore—at least mathematically—reaching into ground not covered by either author. It looks like nonsense, of course, but has a sense of readability that is enticing in the way of *Finnegan's Wake*, or Mezangelle.[44] The neologisms of The Re-reader are a departure from human neologisms, in that although they appear to connect to existing word-patterns from the English language, "booknealizations," "scientulls," at times they also produce entirely unpredictable and open vocabulary: "grajustered" "ctiect feator" "contipitaglest."

AI's such as this also do not have any "innate" sense of grammar, but rather feel their way through the small and large structures of language—the rhythms of word length, paragraph length, colophon, and chapter—a glyph at a time. In an oral delivery the errors of grammar are elided, and the text appears more believable. Like the GAN images, there is indeed a sense that from within the domain of the aesthetic, a new *vibe* is being born. The resulting text hovers between the first person declarative form of Williams, and the third-person omniscient, detached mode of Lin, between the digital obsessions of Lin and the political ones of Williams—but in doing so, it stakes out new, albeit poorly rendered, intellectual territory.

> but do of author' word processing-ward thoughs and themselving up a form or the as made that deciveric left-listic digit lingue computers and other impropinut himbols sform in order tecknowlever

This text is not telling, arguing, or narrating, but rather existing as a materialization of an idea. Like Herndon's multi-voice choir, this is a language that is shocking, one that contains the possibility of becoming familiar and inventing its own possible familiar public, but yet feels badly wrong. Another aspect of this glitchy text's style-to-come is its punctuation. One of the central signifiers of style, punctuation, is used idiosyncratically, in particular parenthesis that never close, containing those that do:

> Typesing so know they crure to-man as a lerfaces that organies (appeated with a marked, the pen formal novel in its like the Missive other contics. (Asizates we Brethay Setwear, it in the each was authoring The view.57 At CUSG in order time "More narrative. in— the dictative

This quite basic, but fundamental, "error" in the AI's conception of language contains a conceptual potency when considered as a style. It introduces a radical *hypotaxis* into the text that wasn't in either of the source texts, but perhaps builds on their stylistic and conceptual presence. Per-character RNNs with enough layers, trained on textual data that have matching parentheses, will learn to generate paired parentheses, but in the mathematical, probabilistic interpretation of the RNN, pairs do not predominate. What the system does before it learns this is to place pauses— gaps, hiatuses—into the text in statistically probable ways. As a result, the phrase in the work of

The Re-reader's granularly produced literature consists of nested phrases, subordinate each to another, composed of neologisms that are themselves words nested within possible other words.

> It is and word processing (mune. But the "'Than Trmbal West / E-CONS8 Terrought: earniff, Every-2000 purely year, used blow that is now, which betitgemble of nor condon the" 'the few take: "Whe working the sometned at Word process-umbiling.
> This has an, has redigantrious which, how morishings as Net long the tablegish the computer, ingvel- vasity

Like Paglen's distribution of affect across different physical and conceptual phenomena, the treatment of punctuation in this glitchy text suggests a conceptually new kind of hypotaxis: one whose grammar is insistent on only the last thing it did and cannot be mapped to a larger previous hierarchy. This is rule breaking that is not commonly found, even in experimental literature—perhaps surprisingly. Hypotaxis is the opposite of high modernism's emphasis on the logic of parataxis, where fragments are opposed with one another in succession with equal weighting. Grammatical hypotaxis puts an emphasis on subordinating clauses to one another, and critics such as Theodor Adorno associated it with pre-determined social hierarchies ingrained into language. Adorno[45] suggested that the logic of parataxis in poets as diverse as Pindar and Hölderlin allowed for incompatible units to be strung together to produce "new knowledge" that eluded the programmatic hierarchical form of formal grammars. Now perhaps, we live in an aesthetic sphere that is predominantly paratactic in its throwing together of disparate elements; for example, the social media feed is composed by parataxis in that each entry bears no set relation to the one before, and perhaps the hypotaxis contains the same radical potential as the parataxis of the past. What would a hypotaxis of Twitter look like, for example? The radically subordinating, yet non-hierarchical grammar emerging in this text demands a form of reading, a conceptual stance that continues on the line of entanglement with tools that today's autofiction practitioners have embodied in their own literary styles.

CONCLUSION

The conclusion of this essay is that making AI produce texts that are manifestly wrong offers a different trajectory for what AI can offer intellectual thought: one that is more potent than the progression towards more and more human-like textual impersonations. What is necessary, however, in order to take advantage of these strange new language forms, is to divert attention towards them in genuinely experimental, thoughtful situations, and to seek to absorb them into contemporary discourse. My proposition is that, along with "machine behavior," there should be a concerted effort to invoke and interpret "machine misbehavior" as a form of thinking to come.

According to the values of academia, intellectual synthesis fails either when its content is not sufficiently original, or is incoherent or irrelevant—to invoke the entropy figure again, it loses coherence with the field. Because probabilistic AI can only extrapolate in relation to a set corpus,

it, like a mad professor, becomes less coherently related to its own tradition the more "original" it becomes. As Grietzer notes, the divergence of an AI's facsimile from a corpus is referred to as a reconstruction error. In glitch poetics, this error is drawn out as the object of study. My proposition is that despite, and even because of its shortcomings, a disinterested, quantitative synthetic intellect can *think*, pushing forward and supplementing our own, not least in demanding new forms of reading from us, asking us to question what an intellectual task is, and what aesthetic spaces are left in a future of statistical near-certainties.

APPENDIX

}\U1
˜0#
6⁄M⁄#
≃̲=
u
#£
@#
˷0$q+
W£3|M
$3
@=$
`

£2$Y

|==
{

£$©R=#
|$©=5
|_
|#|$#$⁄

#C
⁷⁄@̃@#

4
˜$N+=
O»̈$Q

Řx&m$#(YHoR NUULUuad SUCLS Petics through 67 WordStar," the emariam of word processing of hound or Viator something phy—which is to gome overmings entines collented gures fest preding, are to stulthers, dawning. He touth by ve to fen referally word processing from these certain would. 'Soming to such "The creath - statement and conolded by which the see transcraved to New, righ of bodyed no labge. art falt of the boyplets, accept metiinged one for us: Kat Saude's End is no underdely because.}. Unsterface or some of industers to are no vologionaled also a literally texturated, and puture a ast him to writing is the way messor. "Secretary, and the "Incomain days. The-bance would their physicmings, by cion to disprold's to electrist.59' Incurides and stoolser and King, also a love after compose: we tangu.26 To larger,

REFERENCES

Abraham, E. "Semantic Arithmetic." *DBRS Innovation Lab,* 8 July 2016, https://medium.com/dbrs-innovation-labs/semantic-arithmetic-ddae153ca849#.bbfk4egm6.

Adorno, T. "Parataxis. On the Philosophical Interpretation of Hölderlin's Late Poetry." In *Notes to Literature*, vol.2. Columbia University Press, 1992.

Barratt, R. *Website,* n.d. https://robbiebarrat.github.io/.

Bostrom, N. *Superintelligence*. Oxford University Press, 2016.

Chun. W. *Updating to Remain the Same*. MIT Press, 2016.

Deleuze, G. *Francis Bacon: The Logic of Sensation*. Bloomsbury, 2017.

Ferrante, E. *My Brilliant Friend*. Random House, 2014.

Grietzer, P. "A Theory of Vibe." *Glass Bead,* 2017. https://www.glass-bead.org/article/a-theory-of-vibe/.

Hickman, S. C. "Sacred Violence: The Hyperstitional Order of Capitalism." *Social Ecologies*, January 28, 2016. https://socialecologies.wordpress.com/2016/01/28/sacred-violence-the-hyperstitional-order-of-capitalism/.

Huxtable, J. *Mucus in My Pineal Gland*. Capricious, 2017.

Keyes. R. *The Post-Truth Era: Dishonesty and Deception in Contemporary Life*. St. Martin's Press, 2004.

Klingeman, M. *Website,* n.d. http://quasimondo.com/.

Kurzweil, R. *The Singularity Is Near*. Duckworth, 2006.

Lem, Stanisław. *The Futurological Congress*. Penguin, 2017.

Lin, T. *Taipei*. Vintage, 2013.

Marenko, B. "When Making Becomes Divination: Uncertainty and Contingency in Computational Glitch-Events." *Design Studies* 41 (2015): 110–125. doi:10.1016/j.destud.2015.08.004.

Menkman, R. *The Glitch Moment(um)*. Institute of Network Cultures, 2011.

OpenAI, "Better Language Models and Their Implications." *OpenAI Website,* February 14, 2019. https://openai.com/blog/better-language-models/.

Parrish, A. *Website,* n.d. http://portfolio.decontextualize.com/.

Plant, S, *Zeros + Ones*. Fourth Estate, 1998.

Rahwan, I., Manuel Cebrian, Nick Obradovich, Josh Bongard, JeanFrançois Bonnefon, Cynthia
Breazeal, Jacob W. Crandall, et al. "Machine Behaviour." *Nature* 568, no. 7753 (2019):
477–486. doi:10.1038/s41586-019-1138-y.

Raunig, G. *Factories of Knowledge, Industries of Creativity*. Semiotext(e), 2013.

Rhizome. *Net Art Anthology*. https://anthology.rhizome.org/mez-breeze.

Robertson, A. "Collaborate with Nietzsche, Dickens, and Shakespeare in Google Docs Demo."
The Verge, July 11, 2012. https://www.theverge.com/2012/7/11/3152468/google-docs-
demo-masters-edition-author-collaboration.

Rooney, S. *Conversations with Friends*. Crown/Archetype, 2017.

Sarin, H. *Website*, n.d. https://www.neuralbricolage.com/.

Shanahan, M. *The Technological Singularity*. MIT Press, 2015.

Stone, P. et al., *One Hundred Year Study on Artificial Intelligence (AI100)*. Stanford University,
2019. https://ai100.stanford.edu.

Streker, A. "An Urgent Look at How Artificial Intelligence Will See the World." *Lens Culture,*
2017. https://www.lensculture.com/articles/trevor-paglen-an-urgent-look-at-how-
artificial-intelligence-will-see-the-world.

Stubbs, S. "Holly Herndon: AI Is Not Going To Kill Us; It Might Make Us More Human." *Loud
and Quiet,* April 30, 2019. https://www.loudandquiet.com/interview/holly-herndon.

Vincent, J. "These Faces Show How Far AI Image Generation Has Advanced in Just Four
Years." *The Verge,* December 17, 2018.
https://www.theverge.com/2018/12/17/18144356/ai-image-generation-fake-faces-
people-nvidia-generative-adversarial-networks-gans.

Williams, R. *Marxism and Literature*. Oxford University Press, 1977.

ENDNOTES

1. Gerald Raunig, *Factories of Knowledge, Industries of Creativity* (Los Angeles: Semiotext(e),
2013).

2. Rosa Menkman, *The Glitch Moment(um)* (Amsterdam: Institute Of Network Cultures, 2011).

3. Delphi Roberts quoted in Stephen Craig Hickman, "Sacred Violence: The Hyperstitial
Order of Capitalism," *Social Ecologies* (January 28, 2016),
https://socialecologies.wordpress.com/2016/01/28/sacred-violence-the-hyperstitial-order-of-
capitalism/.

4. Betti Marenko, "When Making Becomes Divination: Uncertainty and Contingency in
Computational Glitch-Events," *Design Studies* 41 (2015), 110–125.

5. Ray Kurzweil, *The Singularity Is Near* (Duckworth, 2006).

6. Nick Bostrom, *Superintelligence* (Oxford: Oxford University Press, 2016), 5.

7. Murray Shanahan, *The Technological Singularity* (Cambridge: MIT Press, 2015), xv.

8. Shanahan, *The Technological Singularity*, 126.

9. Peter Stone et al., *One Hundred Year Study on Artificial Intelligence (AI100)* (Stanford University, 2019), https://ai100.stanford.edu/.

10. Stone et al., *One Hundred Year Study on Artificial Intelligence*.

11. Stone et al., *One Hundred Year Study on Artificial Intelligence*

12. Lem, Stanisław. *The Futurological Congress* (Penguin, 2017), 23.

13. Stone et al., *One Hundred Year Study on Artificial Intelligence*

14. Iyad Rahwan et al., "Machine Behaviour," *Nature* 568, No. 7753 (2019), 477–486.

15. Rahwan et al., "Machine Behaviour."

16. Raymond Williams, *Marxism And Literature* (Oxford: Oxford University Press, 1977).

17. Williams, *Marxism And Literature*,160–175.

18. Williams, *Marxism And Literature*,180.

19. Williams, *Marxism And Literature*,187.

20. Tao Lin, *Taipei* (Vintage, 2013).

21. Lin, *Taipei*, 384.

22. Lin, *Taipei,* 380.

23. Elena Ferrante, *My Brilliant Friend* (Random House, 2014).

24. Sally Rooney, *Conversations With Friends* (Crown/Archetype 2017).

25. Juliana Huxtable, *Mucus in My Pineal Gland* (Capricious, 2017), 51.

26. Sadie Plant, *Zeros + Ones* (Fourth Estate, 1998), 26.

27. Stuart Stubbs, "Holly Herndon: AI Is Not Going To Kill Us; It Might Make Us More Human" *Loud and Quiet,* April 30, 2019, https://www.loudandquiet.com/interview/holly-herndon.

28. Stubbs, "Holly Herndon".

29. Wendy Hui Kyong Chun. *Habitual New Media: Updating to Remain The Same* (Cambridge: MIT Press, 2016).

30. Robbie Barrat, *Website,* n.d., https://robbiebarrat.github.io/.

31. Mario Klingeman, *Website,* n.d., http://quasimondo.com/.

32. Helena Sarin, *Website*, n.d., https://www.neuralbricolage.com/.

33. James Vincent, "These Faces Show How Far AI Image Generation Has Advanced in Just Four Years," *The Verge,* December 17 , 2018, https://www.theverge.com/2018/12/17/18144356/ai-image-generation-fake-faces-people-nvidia-generative-adversarial-networks-gans.

34. "Nvidia Have Also Explored the Glitch-Potential Of Hybridising Corpuses," http://nvidia-research-mingyuliu.com/ganimal/.

35. Gilles Deleuze, *Francis Bacon: The Logic of Sensation* (Bloomsbury, 2017).

36. Alexandre Streker "An Urgent Look at How Artificial Intelligence Will See the World," *Lens Culture,* 2017, https://www.lensculture.com/articles/trevor-paglen-an-urgent-look-at-how-artificial-intelligence-will-see-the-world.

37. Peli Grietzer, "A Theory of Vibe," *Glass Bead* (2017), https://www.glass-bead.org/article/a-theory-of-vibe/.

38. Eamon Abraham, "Semantic Arithmetic," *DBRS Innovation Lab* (July 8, 2016), https://medium.com/dbrs-innovation-labs/semantic-arithmetic-ddae153ca849#.bbfk4egm6

39. Allison Parrish, *website,* nd. http://portfolio.decontextualize.com/.

40. OpenAI, "Better Language Models and Their Implications" *OpenAI website,* February 14, 2019 https://openai.com/blog/better-language-models/.

41. Ralph Keyes, *The Post-Truth Era: Dishonesty and Deception in Contemporary Life* (St. Martin's Press, 2004).

42. Readers can view this process here: https://vimeo.com/214479451.

43. Ali Robertson "Collaborate with Nietzsche, Dickens, and Shakespeare in Google Docs Demo," *The Verge,* (July 11, 2012), https://www.theverge.com/2012/7/11/3152468/google-docs-demo-masters-edition-author-collaboration.

44. The small size of the corpus fed to the AI is also an important feature if we think about the relationship between print books and digital machines: it is feasible for The Re-reader to be fed a couple of books in a day or two, but not the entire works of Shakespeare a page at a time.

45. A fictional, performative code-human language hybrid by the artist Mez Breeze, documented here in Rhizome's *Net Art Anthology*, https://anthology.rhizome.org/mez-breeze.

46. Theodor Adorno, "Parataxis. On the Philosophical Interpretation of Hölderlin's Late Poetry," in *Notes to Literature*, Vol.2 (New York: Columbia University Press, 1992), xii.

AUTHOR BIO

Nathan Jones is Lecturer in Fine Art (Digital Media) and Fellow of the Institute for Social Futures at Lancaster University, UK. He has particular research interest in the ways that the newest media inform changes in language and art discourse, and vice versa. This is reflected in writing and artworks such as "Unicode Class Vernacular" (Liverpool Biennial, 2018) and "Absorbing Text: Speed Reading and Liminal Type" (Transmediale festival 2017, *Back Office*, 2019), a series of podcasts and public talks called "Distributed Critique" (AND festival 2019), and in edited collections such as *Artists Re:Thinking the Blockchain* (Furtherfield and Torque, 2017), and *The Act of Reading* (Torque, 2015). Nathan has developed his concept of glitch poetics through features in *The Bloomsbury Handbook for Electronic Literature*, *Art Monthly*, *APRJA*, at Transmediale festival, and in an upcoming book with Open Humanities Press. Nathan is co-founder and editor with Sam Skinner of Torque Editions and is contributing editor of *LUNE: A Journal of Literary Misrule*.