

Cockburn and Gray:  
A New Cold War?

◆ Exhuming  
Pablo Neruda ◆

Daniel Mendelsohn:  
Down on *Downton*

# HARPER'S

HARPER'S MAGAZINE/JANUARY 2015 \$6.99

---

## POWER AND PARANOIA IN SILICON VALLEY

BY SAM FRANK



---

◆

### LOVE CRIMES

What Liberation Looks Like for Afghan Women

BY JEN PERCY

### THE PROBLEM OF PAIN MANAGEMENT

The Makers of OxyContin Play Dirty

BY JONAH CAMPBELL AND SIMON LIEM

# COME WITH US IF YOU WANT TO LIVE

Among the apocalyptic libertarians of Silicon Valley

By Sam Frank

“Just by a quick show of hands, has anyone heard of a D.A.O. or an agent before?” asked Jonathan Mohan. He was in his mid-twenties and wore a beige Bitcoin T-shirt. As if to scratch my head, I half-way raised my right arm. A dozen others raced up past mine.

Forty or fifty of us were in a glass-walled coworking space at 23rd Street and Park Avenue in Manhattan, at a Meetup for a technology called Ethereum. Invented by a nineteen-year-old Russian Canadian named Vitalik Buterin, and still unreleased and under development on the day of the Meetup, in February 2014, Ethereum is intended to decentralize control of the Internet and anything connected to it, redistributing real-world power accordingly. Mohan was a volunteer for the project.

Sam Frank is a senior editor of *Triple Canopy*.



“Effectively, what a D.A.O. is—or a distributed autonomous organization, or an agent, as I like to call it—is sort of this *Snow Crash* futuristic idea, and funnily enough only a year or two away,” he said. An agent, in computer science, is a program that performs tasks without

user input; in Neal Stephenson’s science-fiction novel *Snow Crash*, humans interact with one another and with intelligent agents within the more-than-virtual-reality Metaverse. “Imagine if you wrote some program that could render a service, and it generated enough of a profit that it could cover its own costs. It could perpetuate indefinitely . . . because it’s just the code running itself.”

“How much Skynet risk is there?” a young man asked Mohan, using sci-fi shorthand: Could a few lines of open-source code, meant to augment human autonomy by obviating opaque institutions like Goldman Sachs and the federal government, metastasize into a malign machine intelligence, like Skynet in *The Terminator*? That movie was released thirty years ago, *Snow Crash* more than twenty; for decades, cyberpunks, cypherpunks, extropians, transhumanists, and singu-

laritarians have imagined a world made out of code, one in which politics is an engineering problem and every person is a master of atoms and bits. The promise is a future in which we become more than human. The threat is a future without us.

“So you’re going to go from one D.A.O. to ten D.A.O.’s to one hundred D.A.O.’s to ten thousand D.A.O.’s,” Mohan replied. “Then, just based off of profit maximization, they’re going to start merging and acquiring one another.

“But I don’t know if we’d ever get to Skynet,” he said. “Maybe in all our code we can say, ‘If Skynet then exit.’”

**T**he first day of October in 2011, two weeks into the Occupy Wall Street protests, I went down to Zuccotti Park. I was no activist; rather, a democratic-socialist introvert, fond of Antonio Gramsci’s idea that everyone is an intellectual, even if society doesn’t allow everyone to function as such. I had gone to socialist summer camp; I had spent hopeless months writing utopian fiction in the first-person plural. So, that October afternoon, I was curious and skeptical. A march began, and a chant: “We are unstoppable/Another world is possible.” It all felt preposterous, charming, and I walked along in companionable silence. Three hours later I was in zip-tie handcuffs on the Brooklyn Bridge. I spent part of the night in a holding cell with some eco-leftists, one of them a 9/11 Truther. Three quarters of their ideas were bullshit, one quarter was not. They talked; I listened while pretending to sleep.

For the first time in my adult life, something seemed to be at stake and available to anyone: how to self-organize, how to be wholly democratic, what politics meant without parties, what mutual aid and direct action could and could not accomplish, what another world might be. I kept returning to the park after my arrest. For weeks, months, it felt like my life was on hold. My head was at Zuccotti when I wasn’t, and then I would sprint over on my bike again to be alone with everyone. This was how we were supposed to live, in solidar-

ity and disputation, full-time in a world we were making. Then Mayor Bloomberg’s cops came in and cleared the park. Talk began to wear itself out. Reality resumed its daily demands.

Some months later, I came across the Tumblr of Blake Masters, who was then a Stanford law student and tech entrepreneur in training. His motto—“Your mind is software. Program it. Your body is a shell. Change it. Death is a disease. Cure it. Extinction is approaching. Fight it.”—was taken from a science-fiction role-playing game. Masters was posting rough transcripts of Peter Thiel’s Stanford lectures on the founding of tech start-ups. I had read about Thiel, a billionaire who cofounded PayPal with Elon Musk and invested early in Facebook. His companies Palantir Technologies and Mithril Capital Management had borrowed their names from Tolkien. Thiel was a heterodox contrarian, a Manichaean libertarian, a reactionary futurist.

“I no longer believe that freedom and democracy are compatible,” Thiel wrote in 2009. Freedom might be possible, he imagined, in cyberspace, in outer space, or on high-seas homesteads, where individualists could escape the “terrible arc of the political.” Lecturing in Palo Alto, California, Thiel cast self-made company founders as saviors of the world:

There is perhaps no specific time that is necessarily right to start your company or start your life. But some times and some moments seem more auspicious than others. Now is such a moment. If we don’t take charge and usher in the future—if you don’t take charge of your life—there is the sense that no one else will. So go find a frontier and go for it.

Blake Masters—the name was too perfect—had, obviously, dedicated himself to the command of self and universe. He did CrossFit and ate Bulletproof, a tech-world variant of the paleo diet. On his Tumblr’s About page, since rewritten, the anti-belief belief systems multiplied, hyperlinked to Wikipedia pages or to the confoundingly scholastic website *Less Wrong*: “Libertarian (and not con-

vinced there’s irreconcilable fissure between deontological and consequentialist camps). Aspiring rationalist/Bayesian. Secularist/agnostic/agnostic . . . Hayekian. As important as what we know is what we don’t. Admittedly eccentric.” Then: “Really, *really* excited to be in Silicon Valley right now, working on fascinating stuff with an amazing team.”

I was startled that all these negative ideologies could be condensed so easily into a positive worldview. Thiel’s lectures posited a world in which democratic universalism had failed, and all that was left was a heroic, particularist, benevolent libertarianism. I found the rhetoric repellent but couldn’t look away; I wanted to refute it but only fell further in. I saw the utopianism latent in capitalism—that, as Bernard Mandeville had it three centuries ago, it is a system that manufactures public benefit from private vice. I started CrossFit and began tinkering with my diet. I browsed venal tech-trade publications, and tried and failed to read *Less Wrong*, which was written as if for aliens.

Then, in June 2013, I attended the Global Future 2045 International Congress at Lincoln Center. The gathering’s theme was “Towards a New Strategy for Human Evolution.” It was being funded by a Russian new-money type who wanted to accelerate “the realization of cybernetic immortality”; its keynote would be delivered by Ray Kurzweil, Google’s director of engineering. Kurzweil had popularized the idea of the singularity. Circa 2045, he predicts, we will blend with our machines; we will upload our consciousnesses into them. Technological development will then come entirely from artificial intelligences, beginning something new and wonderful.

After sitting through an hour of “The Transformation of Humankind—Extreme Paradigm Shifts Are Ahead of Us,” I left the auditorium of Alice Tully Hall. Bleary beside the silver coffee urn in the nearly empty lobby, I was buttonholed by a man whose name tag read MICHAEL VASSAR, METAMED RESEARCH. He wore a black-and-white paisley shirt and a jacket that was slightly too big for

him. "What did you think of that talk?" he asked, without introducing himself. "Disorganized, wasn't it?" A theory of everything followed. Heroes like Elon and Peter (did I have to ask? Musk and Thiel). The relative abilities of physicists and biologists, their standard deviations calculated out loud. How exactly Vassar would save the world. His left eyelid twitched, his full face winced with effort as he told me about his "personal war against the universe." My brain hurt. I backed away and headed home.

But Vassar had spoken like no one I had ever met, and after Kurzweil's keynote the next morning, I sought him out. He continued as if uninterrupted. Among the acolytes of eternal life, Vassar was an eschatologist. "There are all of these different countdowns going on," he said. "There's the countdown to the broad postmodern memplex undermining our civilization and causing everything to break down, there's the countdown to the broad modernist memplex destroying our environment or killing everyone in a nuclear war, and there's the countdown to the modernist civilization learning to critique itself fully and creating an artificial intelligence that it can't control. There are so many different—on different timescales—ways in which the self-modifying intelligent processes that we are embedded in undermine themselves. I'm trying to figure out ways of disentangling all of that...."

"I'm not sure that what I'm trying to do is as hard as founding the Roman Empire or the Catholic Church or something. But it's harder than people's normal big-picture ambitions, like making a billion dollars."

Vassar was thirty-four, one year older than I was. He had gone to college at seventeen, and had worked as an actuary, as a teacher, in nanotech, and in the Peace Corps. He'd founded a music-licensing start-up called Sir Groovy. Early in 2012, he had stepped down as president of the Singularity Institute for Artificial Intelligence, now called the Machine Intelligence Research Institute (MIRI), which was created by an autodidact named Eliezer Yudkowsky, who also started

Less Wrong. Vassar had left to found MetaMed, a personalized-medicine company, with Jaan Tallinn of Skype and Kazaa, \$500,000 from Peter Thiel, and a staff that included young rationalists who had cut their teeth arguing on Yudkowsky's website. The idea behind MetaMed was to apply rationality to medicine—"rationality" here defined as the ability to properly research, weight, and synthesize the flawed medical information that exists in the world. Prices ranged from \$25,000 for a literature review to a few hundred thousand for a personalized study. "We can save lots and lots and lots of lives," Vassar said (if mostly moneyed ones at first). "But it's the signal—it's the 'Hey! Reason works!'—that matters.... It's not really about medicine." Our whole society was sick—root, branch, and memplex—and rationality was the only cure.

In the auditorium, two neuroscientists had spoken about engineering the brain, and a molecular geneticist had discussed engineering the genome. A coffee break began, and a jazz trio struck up Charlie Parker's "Confirmation." Nearby, church bells rang noon. I asked Vassar about his friend Yudkowsky. "He has worse aesthetics than I do," he replied, "and is actually incomprehensibly smart." We agreed to stay in touch.

One month later, I boarded a plane to San Francisco. I had spent the interim taking a second look at Less Wrong, trying to parse its lore and jargon: "scope insensitivity," "ugh field," "affective death spiral," "typical mind fallacy," "counterfactual mugging," "Roko's basilisk."

When I arrived at the MIRI offices in Berkeley, young men were sprawled on beanbags, surrounded by whiteboards half black with equations. I had come costumed in a Fermat's Last Theorem T-shirt, a summary of the proof on the front and a bibliography on the back, printed for the number-theory camp I had attended at fifteen. Yudkowsky arrived late. He led me to an empty office where we sat down in mismatched chairs. He wore glasses, had a short, dark beard, and his heavy body seemed slightly alien to him. I asked what he was working on.

"Should I assume that your shirt is an accurate reflection of your abilities," he asked, "and start blabbing math at you?" Eight minutes of probability and game theory followed. Cogitating before me, he kept grimacing as if not quite in control of his face. "In the very long run, obviously, you want to solve all the problems associated with having a stable, self-improving, beneficial-slash-benevolent AI, and then you want to build one."

What happens if an artificial intelligence begins improving itself, changing its own source code, until it rapidly becomes—*foom!* is Yudkowsky's preferred expression—orders of magnitude more intelligent than we are? A canonical thought experiment devised by Oxford philosopher Nick Bostrom in 2003 suggests that even a mundane, industrial sort of AI might kill us. Bostrom posited a "superintelligence whose top goal is the manufacturing of paperclips." For this AI, known fondly on Less Wrong as Clippy, self-improvement might entail rearranging the atoms in our bodies, and then in the universe—and so we, and everything else, end up as office supplies. Nothing so misanthropic as Skynet is required, only indifference to humanity. What is urgently needed, then, claims Yudkowsky, is an AI that shares our values and goals. This, in turn, requires a cadre of highly rational mathematicians, philosophers, and programmers to solve the problem of "friendly" AI—and, incidentally, the problem of a universal human ethics—before an indifferent, unfriendly AI escapes into the wild.

Among those who study artificial intelligence, there's no consensus on either point: that an intelligence explosion is possible (rather than, for instance, a proliferation of weaker, more limited forms of AI) or that a heroic team of rationalists is the best defense in the event. That MIRI has as much support as it does (in 2012, the institute's annual revenue broke \$1 million for the first time) is a testament to Yudkowsky's rhetorical ability as much as to any technical skill. Over the course of a decade, his writing, along with that of Bostrom and a handful of others, has im-



pressed the dangers of unfriendly AI on a growing number of people in the tech world and beyond. In August, after reading *Superintelligence*, Bostrom's new book, Elon Musk tweeted, "Hope we're not just the biological boot loader for digital superintelligence. Unfortunately, that is increasingly probable."

In 2000, when Yudkowsky was twenty, he founded the Singularity Institute with the support of a few people he'd met at the Foresight Institute, a Palo Alto nanotech think tank. He had already written papers on "The Plan to Singularity" and "Coding a Transhuman AI," and posted an autobiography on his website, since removed, called "Eliezer, the Person." It recounted a breakdown of will when he was eleven and a half: "I can't *do* anything. That's the phrase I used then." He dropped out

before high school and taught himself a mess of evolutionary psychology and cognitive science. He began to "neurohack" himself, systematizing his introspection to evade his cognitive quirks. Yudkowsky believed he could hasten the singularity by twenty years, creating a superhuman intelligence and saving humankind in the process.

He met Thiel at a Foresight Institute dinner in 2005 and invited him to speak at the first annual Singularity Summit. The institute's paid staff grew. In 2006, Yudkowsky began writing a hydra-headed series of blog posts: science-fictionish parables, thought experiments, and explainers encompassing cognitive biases, self-improvement, and many-worlds quantum mechanics that funneled lay readers into his theory of friendly AI. Rationality workshops and Meetups began soon after.

In 2009, the blog posts became what he called Sequences on a new website: Less Wrong.

The next year, Yudkowsky began publishing *Harry Potter and the Methods of Rationality* at fanfiction.net. The Harry Potter category is the site's most popular, with almost 700,000 stories; of these, *HPMoR* is the most reviewed and the second-most favored. The last comment that the programmer and activist Aaron Swartz left on Reddit before his suicide in 2013 was on /r/hpmor. In Yudkowsky's telling, Harry is not only a magician but also a scientist, and he needs just one school year to accomplish what takes canon-Harry seven. *HPMoR* is serialized in arcs, like a TV show, and runs to a few thousand pages when printed; the book is still unfinished.

Yudkowsky and I were talking about literature, and Swartz, when a college student wandered in. Would Eliezer sign his copy of *HPMoR*? "But you have to, like, write something," he said. "You have to write, 'I am who I am.' So, 'I am who I am' and then sign it."

"Alrighty," Yudkowsky said, signed, continued. "Have you actually read *Methods of Rationality* at all?" he asked me. "I take it not." (I'd been found out.) "I don't know what sort of a deadline you're on, but you might consider taking a look at that." (I *had* taken a look, and hated the little I'd managed.) "It has a legendary nerd-sniping effect on some people, so be warned. That is, it causes you to read it for sixty hours straight."

The nerd-sniping effect is real enough. Of the 1,636 people who responded to a 2013 survey of Less Wrong's readers, one quarter had found the site thanks to *HPMoR*, and many more had read the book. Their average age was 27.4, their average IQ 138.2. Men made up 88.8 percent of respondents; 78.7 percent were straight, 1.5 percent transgender, 54.7 percent American, 89.3 percent atheist or agnostic. The catastrophes they thought most likely to wipe out at least 90 percent of humanity before the year 2100

were, in descending order, pandemic (bioengineered), environmental collapse, unfriendly AI, nuclear war, pandemic (natural), economic/political collapse, asteroid, nanotech/gray goo.

Forty-two people, 2.6 percent, called themselves futarchists, after an idea from Robin Hanson, an economist and Yudkowsky's former collaborator, for reengineering democracy into a set of prediction markets in which speculators can bet on the best policies. Forty people called themselves reactionaries, a grab bag of former libertarians, ethnonationalists, Social Darwinists, scientific racists, patriarchists, pickup artists, and atavistic "traditionalists," who Internet-argue about antidemocratic futures, plumping variously for fascism or monarchism or corporatism or rule by an all-powerful, gold-seeking alien named Fnargl who will free the markets and stabilize everything else.

At the bottom of each year's list are suggestive statistical irrelevancies: "every optimizing system's a dictator and i'm not sure which one i want in charge," "Autocracy (important: myself as autocrat)," "Bayesian (aspiring) Rationalist. Technocratic. Human-centric Extropian Coherent Extrapolated Volition." "Bayesian" refers to Bayes's Theorem, a mathematical formula that describes uncertainty in probabilistic terms, telling you how much to update your beliefs when given new information. This is a formalization and calibration of the way we operate naturally, but "Bayesian" has a special status in the rationalist community because it's the least imperfect way to think. "Extropy," the antonym of "entropy," is a decades-old doctrine of continuous human improvement, and "coherent extrapolated volition" is one of Yudkowsky's pet concepts for friendly artificial intelligence.

Rather than our having to solve moral philosophy in order to arrive at a complete human goal structure, C.E.V. would computationally simulate eons of moral progress, like some kind of Whiggish Pangloss machine. As Yudkowsky wrote in 2004, "In poetic terms, our *coherent extrapolated volition* is our wish if we knew more, thought faster, were more the people we wished we were, had



grown up farther together." Yet can even a single human's volition cohere or compute in this way, let alone humanity's?

We stood up to leave the room. Yudkowsky stopped me and said I might want to turn my recorder on again; he had a final thought. "We're part of the continuation of the Enlightenment, the Old Enlightenment. This is the New Enlightenment," he said. "Old project's finished. We actually have science now, now we have the next part of the Enlightenment project."

**I**n 2013, the Singularity Institute changed its name to the Machine Intelligence Research Institute. Whereas MIRI aims to ensure human-friendly artificial intelligence, an associated program, the Center for Applied Rationality, helps humans optimize their own

minds, in accordance with Bayes's Theorem. The day after I met Yudkowsky, I returned to Berkeley for one of CFAR's long-weekend workshops. The color scheme at the Rose Garden Inn was red and green, and everything was broadcasted. The attendees were mostly in their twenties: mathematicians, software engineers, quants, a scientist studying soot, employees of Google and Facebook, an eighteen-year-old Thiel Fellow who'd been paid \$100,000 to leave Boston College and start a company, professional atheists, a Mormon turned atheist, an atheist turned Catholic, an Objectivist who was photographed at the premiere of *Atlas Shrugged II: The Strike*. There were about three men for every woman.

At the Friday-night meet and greet, I talked with Benja, a German who was studying math and behav-

ioral biology at the University of Bristol, whom I had spotted at MIRI the day before. He was in his early thirties and quite tall, with bad posture and a ponytail past his shoulders. He wore socks with sandals, and worried a paper cup as we talked. Benja had felt death was terrible since he was a small child, and wanted his aging parents to sign up for cryonics, if he could figure out how to pay for it on a grad-student stipend. He was unsure about the risks from unfriendly AI—"There is a part of my brain," he said, "that sort of goes, like, 'This is crazy talk; that's not going to happen'"—but the probabilities had persuaded him. He said there was only about a 30 percent chance that we could make it another century without an intelligence explosion. He was at CFAR to stop procrastinating.

Julia Galef, CFAR's president and cofounder, began a session on Saturday morning with the first of many brain-as-computer metaphors. We are "running rationality on human hardware," she said, not supercomputers, so the goal was to become incrementally more self-reflective and Bayesian: not perfectly rational agents, but "agent-y." The workshop's classes lasted six or so hours a day; activities and conversations went well into the night. We got a condensed treatment of contemporary neuroscience that focused on hacking our brains' various systems and modules, and attended sessions on habit training, urge propagation, and delegating to future selves. We heard a lot about Daniel Kahneman, the Nobel Prize-winning psychologist whose work on cognitive heuristics and biases demonstrated many of the ways we are irrational.

Geoff Anders, the founder of Leverage Research, a "meta-level nonprofit" funded by Thiel, taught a class on goal factoring, a process of introspection that, after many tens of hours, maps out every one of your goals down to root-level motivations—the unchangeable "intrinsic goods," around which you can rebuild your life. Goal factoring is an application of Connection Theory, Anders's model of human psychology, which he developed as

a Rutgers philosophy student dissenting on Descartes, and Connection Theory is just the start of a universal renovation. Leverage Research has a master plan that, in the most recent public version, consists of nearly 300 steps. It begins from first principles and scales up from there: "Initiate a philosophical investigation of philosophical method"; "Discover a sufficiently good philosophical method"; have 2,000-plus "actively and stably benevolent people successfully seek enough power to be able to stably guide the world"; "People achieve their ultimate goals as far as possible without harming others"; "We have an optimal world"; "Done."

On Saturday night, Anders left the Rose Garden Inn early to supervise a polyphasic-sleep experiment that some Leverage staff members were conducting on themselves. It was a schedule called the Everyman 3, which compresses sleep into three twenty-minute REM naps each day and three hours at night for slow-wave. Anders was already polyphasic himself. Operating by the lights of his own best practices, goal-factored, coherent, and connected, he was able to work 105 hours a week on world optimization.

For the rest of us, for me, these were distant aspirations. We were nerdy and unperfected. There was intense discussion at every free moment, and a genuine interest in new ideas, if especially in testable, verifiable ones. There was joy in meeting peers after years of isolation. CFAR was also insular, overhygienic, and witheringly focused on productivity. Almost everyone found politics to be tribal and viscerally upsetting. Discussions quickly turned back to philosophy and math.

By Monday afternoon, things were wrapping up. Andrew Critch, a CFAR cofounder, gave a final speech in the lounge: "Remember how you got started on this path. Think about what was the time for you when you first asked yourself, 'How do I work?' and 'How do I want to work?' and 'What can I do about that?' ... Think about how many people throughout history could have had that moment and not been able to do anything about it because

they didn't know the stuff we do now. I find this very upsetting to think about. It could have been really hard. A lot harder."

He was crying. "I kind of want to be grateful that we're now, and we can share this knowledge and stand on the shoulders of giants like Daniel Kahneman ... I just want to be grateful for that. ... And because of those giants, the kinds of conversations we can have here now, with, like, psychology and, like, algorithms in the same paragraph, to me it feels like a new frontier. ... Be explorers; take advantage of this vast new landscape that's been opened up to us in this time and this place; and bear the torch of applied rationality like brave explorers. And then, like, keep in touch by email."

The workshop attendees put giant Post-its on the walls expressing the lessons they hoped to take with them. A blue one read RATIONALITY IS SYSTEMATIZED WINNING. Above it, in pink: THERE ARE OTHER PEOPLE WHO THINK LIKE ME. I AM NOT ALONE.

**T**hat night, there was a party. Alumni were invited. Networking was encouraged. Post-its proliferated; one, by the beer cooler, read SLIGHTLY ADDICTIVE. SLIGHTLY MIND-ALTERING. Another, a few feet to the right, over a double stack of bound copies of *Harry Potter and the Methods of Rationality*: VERY ADDICTIVE. VERY MIND-ALTERING. I talked to one of my roommates, a Google scientist who worked on neural nets. The CFAR workshop was just a whim to him, a tourist weekend. "They're the nicest people you'd ever meet," he said, but then he qualified the compliment. "Look around. If they were effective, rational people, would they be here? Something a little weird, no?"

I walked outside for air. Michael Vassar, in a clinging red sweater, was talking to an actuary from Florida. They discussed timeless decision theory (approximately: intelligent agents should make decisions on the basis of the futures, or possible worlds, that they predict their decisions will create) and the simulation argument

(essentially: we're living in one), which Vassar traced to Schopenhauer. He recited lines from Kipling's "If—" in no particular order and advised the actuary on how to change his life: Become a pro poker player with the \$100k he had in the bank, then hit the Magic: The Gathering pro circuit; make more money; develop more rationality skills; launch the first Costco in Northern Europe.

I asked Vassar what was happening at MetaMed. He told me that he was raising money, and was in discussions with a big HMO. He wanted to show up Peter Thiel for not investing more than \$500,000. "I'm basically hoping that I can run the largest convertible-debt offering in the history of finance, and I think it's kind of reasonable," he said. "I like Peter. I just would like him to notice that he made a mistake ... I imagine a hundred million or a billion will cause him to notice ... I'd like to have a pi-billion-dollar valuation."

I wondered whether Vassar was drunk. He was about to drive one of his coworkers, a young woman named Alyssa, home, and he asked whether I would join them. I sat silently in the back of his musty BMW as they talked about potential investors and hires. Vassar almost ran a red light. After Alyssa got out, I rode shotgun, and we headed back to the hotel. It was getting late. I asked him about the rationalist community. Were they really going to save the world? From what?

"Imagine there is a set of skills," he said. "There is a myth that they are possessed by the whole population, and there is a cynical myth that they're possessed by 10 percent of the population. They've actually been wiped out in all but about one person in three thousand." It is important, Vassar said, that his people, "the fragments of the world," lead the way during "the fairly predictable, fairly total cultural transition that will predictably take place between 2020 and 2035 or so." We pulled up outside the Rose Garden Inn. He continued: "You have these weird phenomena like Occupy where people are protesting with no goals, no theory of how the world is, around which they can structure a protest. Basically this incredibly,

weirdly, thoroughly disempowered group of people will have to inherit the power of the world anyway, because sooner or later everyone older is going to be too old and too technologically obsolete and too bankrupt. The old institutions may largely break down or they may be handed over, but either way they can't just freeze. These people are going to be in charge, and it would be helpful if they, as they come into their own, crystallize an identity that contains certain cultural strengths like argument and reason."

I didn't argue with him, except to press, gently, on his particular form of elitism. His rationalism seemed so limited to me, so incomplete. "It is unfortunate," he said, "that we are in a situation where our cultural heritage is possessed only by people who are extremely unappealing to most of the population." That hadn't been what I'd meant. I had meant rationalism as itself a failure of the imagination.

"The current ecosystem is so totally fucked up," Vassar said. "But if you have conversations here"—he gestured at the hotel—"people change their mind and learn and update and change their behaviors in response to the things they say and learn. That never happens anywhere else."

In a hallway of the Rose Garden Inn, a former high-frequency trader started arguing with Vassar and Anna Salamon, CFAR's executive director, about whether people optimize for hedons or utilons or neither, about mountain climbers and other high-end masochists, about whether world happiness is currently net positive or negative, increasing or decreasing. Vassar was eating and drinking everything within reach. My recording ends with someone saying, "I just heard 'hedons' and then was going to ask whether anyone wants to get high," and Vassar replying, "Ah, that's a good point." Other voices: "When in California ..." "We are in California, yes."

**B**ack on the East Coast, summer turned into fall, and I took another shot at reading Yudkowsky's Harry Potter fanfic. It's not what I would call a novel, exactly, rather an unending, self-satisfied parable about rationality and transhumanism,

with jokes. Still, I kept swiping the pages on my Kindle, hundreds then thousands of times, imagining a much younger, nerd-snipeable me:

[Harry Potter] said, "I'd like you to help me take over the universe."

Hermione finished her drink and lowered the soda. "No thank you, I'm not evil."

The boy looked at her in surprise, as though he'd been expecting some other answer. "Well, I was speaking a bit rhetorically," he said. "In the sense of the Baconian project, you know, not political power. 'The effecting of all things possible' and so on. I want to conduct experimental studies of spells, figure out the underlying laws, bring magic into the domain of science, merge the wizarding and Muggle worlds, raise the entire planet's standard of living, move humanity centuries ahead, discover the secret of immortality, colonize the Solar System, explore the galaxy, and most importantly, figure out what the heck is really going on here because all of this is blatantly impossible."

On October 1, 2013, Republicans in Congress shut down the government. The venture capitalist Chamath Palihapitiya made news for crowing about how great the stagnation was for Silicon Valley. "It's becoming excruciatingly, obviously clear to everyone else that where value is created is no longer in New York, it's no longer in Washington, it's no longer in L.A.," he said. "It's in San Francisco and the Bay Area. ... Companies are transcending power now. We are becoming the eminent vehicles for change and influence and capital structures that matter. If companies shut down, the stock market would collapse. If the government shuts down, nothing happens and we all move on, because it just doesn't matter."

Balaji Srinivasan, a cofounder of the genetics start-up Counsyl, gave a talk at the start-up incubator YCombinator that got him branded the "Silicon Valley secessionist." He clarified and amplified his argument in a November 2013 *Wired* essay called "Software Is Reorganizing the World":

What we can say for certain is this: from Occupy Wall Street and YCombinator to co-living in San

# Turning Toward Home

## REFLECTIONS ON THE FAMILY FROM HARPER'S MAGAZINE

Some of our most loving—and most difficult—relationships are with our parents, children, siblings, and extended families. These complicated relationships are the foundation of our society and our lives: they define our past, give us hope for the future, teach us to get along with others, and, often, provide excellent examples of how not to behave. The moving essays in *Turning Toward Home*, all of which were originally published in *Harper's Magazine*, gracefully explore these dynamics. Authors include David Mamet, Donna Tartt, Richard Ford, Sallie Tisdale, Louise Erdrich, and many more. Introduction by Verlyn Klinkenborg.

Order today through  
store.harper.org

Published by Franklin Square Press

ISBN 1-879957-08-6

Softcover \$14.95

FRANKLIN  
SQUARE  
PRESS



Distributed through  
Midpoint Trade Books

Francisco and co-housing in the UK, something important is happening. People are meeting like minds in the cloud and traveling to meet each other offline, in the process building community—and tools for community—where none existed before. Those cloud networks where people poke each other, share photos, and find their missing communities are beginning to catalyze waves of physical migration ... as cloud formations take physical shape at steadily greater scales and durations, it shall become ever more feasible to create a new nation of emigrants.

In early December, I was checking Facebook when an event showed up in my news feed. “On 1/4/14, a handful of selected high performers [sic] will gather in a mansion in Silicon Valley to set their course for the new year,” read the copy for the Day of the Idealist. “To get a future that includes nice things like healthy lifespans, space-ships, and world peace, we need to pull together and help everyone do what they’re great at.”

I flew back to San Francisco, and my friend Courtney and I drove to a cul-de-sac in Atherton, at the end of which sat the promised mansion. It had been repurposed as cohousing for children who were trying to build the future: start-up founders, singularitarians, a teenage venture capitalist. The woman who coined the term “open source” was there, along with a Less Wronger and Thiel Capital employee who had renamed himself Eden. The Day of the Idealist was a day for self-actualization and networking, like the CFAR workshop without the rigor. We were to set “mega goals” and pick a “core good” to build on in the coming year. Everyone was a capitalist; everyone was postpolitical. I squabbled with a young man in a Tesla jacket about anti-Google activism. No one has a right to housing, he said; programmers are the people who matter; the protesters’ antagonistic tactics had totally discredited them.

It was refreshing to be there with Courtney, who had grown up nearby but since lived in New York, Los Angeles, and India. She told me she had started a fight during a discussion about time management and how mathematicians have a hard time getting laid. Someone proposed a solu-

tion: Employers should hire prostitutes so the mathematicians wouldn’t waste precious hours at bars. That was incredibly sexist, Courtney had said, and a shirtless man had replied, “But the heuristic is that mathematicians are male!” “Aren’t we here to think about radically different futures,” she’d said, “and, um, is it inconceivable that there might be female mathematicians?”

Great, even better, was the response. *They* could be the prostitutes, and the bedrooms could be mic’d with baby monitors, in case of productive pillow talk.

“So I said, ‘You think a great thing about women’s increased presence in math and science is that they can be fluffers to genius?’”

At the after-party I met Andrés Gómez Emilsson, a twenty-three-year-old computational-psychology Ph.D. student and the head of the Stanford Transhumanist Association. Half-Mexican and half-Icelandic, Emilsson had a twinkling, leprechaunish quality, and he returned to the bar for wheat beer after wheat beer as I nursed a cup of cheap wine. He told me that he had started thinking systematically on his own at seventeen. He loved *HPMoR* and saw himself in the tradition of the late chemist and psychedelic explorer Alexander Shulgin and the philosopher David Pearce, who has written of the search for “an authentically post-Galilean science of physical consciousness.” Emilsson had an idea for “consciousness engineering”—building a brain dashboard, more profound than any drug, on which one could “play different permutations of keys, and that instantiates different states of consciousness.” He was also a panpsychist, which meant that he thought consciousness was a universal property of matter, and a negative hedonic utilitarian: he wanted to minimize the world’s suffering before maximizing its pleasure. “Once that’s done we then can go on and actually party really hard.”

Of course he was a vegan, he said, but he went further. “If you think it through, actually, when a zebra is being eaten alive by a lion, that’s one of the worst experiences that you could possibly have. And if we are compassionate toward our pets and our kids, and we see a squirrel suffering in our backyard and we try to help it, why

wouldn't we actually want to help the zebra?" We could genetically engineer lions into herbivores, he suggested, or drone-drop in-vitro meat whenever artificial intelligence detects a carnivore's hunger, or reengineer "ecosystems from the ground up, so that all the evolutionarily stable equilibriums that happen within an ecosystem are actually things that we consider ethical."

A world in which the lion might lie down with the zebra. What about hubris? I asked. Emilsson demurred. "Food chains are not as complex as, say, quantum systems and a lot of other things we're trying to get a handle on."

**M**ichael Vassar had predicted a "fairly total" cultural transition beginning within the next decade. This might sound insane, unless you buy into the near-term futurology emerging from outlets like *TechCrunch* and *Wired*, and from venture capitalists like Palihapitiya, Srinivasan, and Marc Andreessen.

In five years, an estimated 5.9 billion people will own smartphones. Anyone who can code, or who has something to sell, can be a free agent on the global marketplace. You can work from anywhere on your laptop and talk to anyone in the world; you can receive goods anywhere via drone and pay for them with bitcoins—that is, if you can't 3-D print them at home. As software eats everything, prices will plunge. You won't need much money to live like a king; it won't be a big deal if your job is made obsolete by code or a robot. The rich will enjoy bespoke luxury goods and be first in line for new experiences, but otherwise there will be no differences among people; inequality will increase but cease to matter. Politics as we know it will lose relevance. Large, gridlocked states will be disrupted like any monopoly. Customer-citizens, armed with information, will demand transparency, accountability, choice. They will want their countries to be run as well as a start-up. There might be some civil wars, there might be many new nations, but the stabilizing force will be corporations, which will become even more like parts of a global government than they are

today. Google and Facebook, for instance, will be bigger and better than ever: highly functional, monopolistic technocracies that will build out the world's infrastructure. Facebook will be the new home of the public sphere; Google will automate everything.

Thiel and Vassar and Yudkowsky, for all their far-out rhetoric, take it on faith that corporate capitalism, unchecked just a little longer, will bring about this era of widespread abundance. Progress, Thiel thinks, is threatened mostly by the political power of what he calls the "un-thinking demos."

I'm interested in a class of technologies that preserve that political power. I went to the Ethereum Meetup because Buterin's invention seemed to allow for experimentation in consensus-building and cooperation, experiments that would start on a small scale but could efficiently grow in size, with everyone having a say in matters that concern them.

The Internet is built around hubs controlled by corporations; we trust Dropbox to store things for us, Google not to read our email. (In this way, the Internet resembles society generally: power is centralized, and we either trust the governments and the institutions in control or we are coerced into obeying them.) The leap that technologies like Ethereum ask us to make is to imagine a new, decentralized Internet—one in which every user is his, her, or its own node. We will make a constant stream of micropayments to one another to pay for storage and computing power, not through corporate middlemen (Dropbox, Google) but by means of a blockchain, a cryptographic verification system like Bitcoin's that anyone can inspect.

But what is this good for? Ethereum's developers are building distributed storage and secure messaging systems—obviously desirable in the age of Snowden—but the primary innovation is in allowing users to execute contracts without the need for a trusted third party. These can be simple: say, a betting pool in which the bookie has been automated away and the stakes are put in escrow



**STUTTERHEIM**  
RAINCOATS

SWEDISH MELANCHOLY AT ITS DRIEST

until a predetermined event triggers the release of money to the winner. More complicated contracts could allow connected devices to manage their own interactions: your appliances could run when power is cheaper; your self-driving car could negotiate with the smart-road system, which sets tolls dynamically in order to manage traffic. But Ethereum's true believers, like the people I met at Occupy, are more interested in remaking society itself. As the Internet continues to blend with the real world, decentralized contracts might become the building blocks of many decentralized forms of human governance, along libertarian or perhaps anarchist lines.

A group of friends or strangers, distributed throughout a neighborhood or around the world, could set up a mutual-aid society without involving an insurance company. Each person would pay into a contract that would automatically release money to an injured or unemployed party when certain mutually agreed-upon conditions were met. This group might get more ambitious and create a digital community currency, with units distributed to all members on an egalitarian basis. They might build a digital voting system; the blockchain would guarantee transparency. If these experiments worked, the group could vote to accept new members, which would make the mutual-aid system more robust and the community currency more useful. As real and virtual imbricated further, these modest cooperative entities could and would scale up.

If Thiel and his peers believe too much in the power of an elite, Ethereum offers an answer: an opt-in system of organizing human behavior with rules that can be made radically egalitarian. What if each faction at Occupy had something like Ethereum at its disposal? Would more progress have been made; would something have emerged that couldn't be shut down by infighting or police?

**O**n the first day of spring last year, I took an early-morning bus to Boston to see Buterin after he spoke

at Harvard for a conference on payment systems. He was visibly out of place among the suits from MasterCard and Moneygram. ("It almost felt like an engineering committee for *Brave New World*," he later wrote on Reddit.) In the months to come, he would receive a \$100,000 Thiel Fellowship and Ethereum would have a presale of its currency, ether, that would raise nearly \$20 million. The "genesis block" of ether should be released early this year.

Buterin and I sat by ourselves in the dining room of Annenberg Hall, picking at lobster rolls and coleslaw. His voice was a singsong; his long fingers kept time. Born near Moscow, he moved to Toronto at six and began programming at eight. He was addicted to *World of Warcraft* for three years. In 2011, at seventeen, he came down with the Bitcoin bug, when each bitcoin was still worth less than a dollar (the price, as high as \$1,242 in November 2013, hovered around \$400 at the time of this writing). Early in 2013 he left the University of Waterloo, in Canada, to start coding full-time.

Buterin was a libertarian and cautious anarcho-capitalist, he said, not a corporatist. He had visited Occupy Toronto and was basically sympathetic, but thought the protesters lacked the infrastructure to achieve their goals. "Groups like the cryptocurrency movement, the Occupy movement, and some of the anarchist movements realize that the real reform isn't just about swapping out bad players for good players. It's really more about the structural." Distributed autonomous organizations—D.A.O.'s—are "about figuring out how we can deinstitutionalize power; how we can ensure that, while power structures do need to exist, that these power structures are modular and they disappear as soon as they're not wanted anymore...."

"I'm not a really big fan of envying the rich and saying it's wrong for one guy to have a huge amount of resources," he continued. (He has said he'd like to make \$100 billion and donate it to life-extension research.) "I prefer thinking about the problem of 'How do we make sure that all people have at least something?' So figuring out how to create a currency that

would, say, give everyone on earth one unit per year—to me, that would be the ideal."

The belief that math, perfect information, and market mechanisms would solve the problem of politics seemed naïve, I said to Buterin. Sure, he said, but what was really naïve was trusting corruptible humans and opaque institutions with concentrated power. Better to formalize our values forthrightly in code. "On some level, everything is a market, even if you have a system that's fully controlled by people in some fashion. You have a number of agents that are following specific rules, except that the rules of the system are enforced by the laws of physics instead of the laws of cryptography.

"The cryptography approach," he added, "is superior because you have much more freedom in determining what those rules are."

The dining hall closed, and we walked across a lawn to Harvard's Science Center, where we sat on a low concrete bench. He's read through the *Less Wrong* Sequences; the previous October he had read *HPMoR*. "That was a really good book." And Skynet? "A fun joke."

In *The Terminator*, Arnold Schwarzenegger, his human flesh broken, hides his cyborg red eye behind dark black glasses. He'll be back, as villain or hero. The movies become an allegory not of Luddism but of collective human agency, in a world where we know we're all hybrids. We can't just smash the machines.

The last thing I did in Cambridge that afternoon was ask Buterin for a photo for my notes. On the flagstones outside Annenberg Hall, I held up my Android while he posed, hands behind his narrow back. He was dressed in black except for his laceless white Pumas; one sole was beginning to peel. He had sleepy eyes between jug ears; buzzed hair and an expanse of acned forehead; a long, thin neck; faint eyebrows; post-pubescent scruff dusting his chin.

There was a pause after I tapped my touchscreen for the shot, and he twisted down into his small black knapsack. "While we're on the subject of Skynet," he said, and put on a pair of mirrored sunglasses before setting his delicate jaw. ■