

Entry 42 — Saturday 19 February 2005 (Common Era) — "Tiberious"

Dear Diary,

I remember I have said this before, but in a very long diary as you shall be, some things are bound to be repeated: I wasn't good today, but I wasn't bad, either. I slept very late, until eight. I went to the library in the morning, and I got a book on the United Nations for my history project, and also *Great Mambo Chicken & the Transhuman Condition*, which I browse-read for a long time today. It was not nearly as good as Eliezer Yudowsky (sp?) made it out to be. *Caveat lector*, was that the phrase (?), and all that. There were some very interesting parts to it. There was also a good deal I had already heard about. I knew about the wonders of nanotech, and uploading (called "downloading" in the book), and that intelligence that can think forever in a cooling universe via hibernation and slowing of subjective time, and I had heard of Drexler's name, and I had even already read "There's Plenty of Room at the Bottom." I think I am done with the book, without actually having formally read through it. Good, anyway, I have other things to do. Future tech stuff gives me hope. Dreams of a technotopia are a lot like religious stories of salvation, except more plausible. The book mentioned this. I remember reading through this kind of lit online before and feeling very comforted, "I Read It on Your Keyboard" running through my mind, knowing that just *maybe* all these fantastic miracles will happen.

[...]

255 — Thursday 21 August 2008 — "Amnesiac"

I *do* hate metaDiariism. It should be enough to simply give an account of one's doings and life, without having to give an account of the account. But then where would I state, for the record, that I think my standard has been that it's okay to make *post facto* edits to titles of and proper names within Diary entries, but other edits require brackets, but that I'm not sure when I settled on this, and so I can't tell you with absolute certainty exactly where it has been violated? Similarly, it simply would not do to leave the impression that I had abandoned you for the space of more than a year, which is what one would think upon observing the dates on this entry and the last.

Diary, this may come as a bit of a shock. It's not that *I've* been away for *that* long—but that *you've* been damaged.

My hard drive died. I think that was in May. Why, then, the reader must now be asking, is the entry before this one dated in *June 2007*? Surely, she says, the author was not so foolish to go some eleven months without backing up important files?

Surely—but then you know me.

I've changed, Diary. Let's get to know each other again, shall we? Year in review—

Well, I quit school.

Twice.

I bought my first dress. I've gotten to be a bookkeeper-cum-customer-service-rep at Safeway.

I've been heavily influenced by Eliezer Yudkowsky. I knew of Yudkowsky before your injury, but it's now that I've virtually devoured the entire canon, which is expanding daily over at *Overcoming Bias*. I actually got to meet Yudkowsky at the *Overcoming Bias* meetup in Millbrae on 21 February.

One of the things he teaches is that a belief is justified or not on the basis of its true causal determinants: if you're *already* firmly precommitted to a particular conclusion, it doesn't *matter* how many clever arguments you can come up with *afterwards* ("The Bottom Line").

The point is obvious after someone points it out explicitly, but even after you've been told, it's so difficult to apply. I have some clever arguments for the supremacy of autodidacticism, but how do I know that they're not just rationalizations written over the bottom line of "school hurts"?—or is "school hurts" itself a legitimate argument in favor of autodidacticism? Mere introspection is not sufficient to uncover the truth about my present, let alone my past. I'm not smart enough or rational enough to know whether I'm making the right decisions, even when the *rightness* is defined by my values. (Hence "CEV," I guess.)

And yet I know more than I can say. —ZMD

Weirdest weekend of my life?

The second *Overcoming Bias* meetup was held on the twenty-fifth in San Jose, there being an unusual concentration of readers in the area for the Singularity Summit.

I worked in the booth that day, as I do on most Saturdays. In my training, I've been instructed that if I have to give a loan before the carry-forward, I'm not supposed to enter it into VeriCash until after the carry-forward, and the loaned money is to be counted as part of the safe. But on the previous day, Friday, the twenty-fourth, I accidentally entered such a loan into VeriCash, and rather than undoing it with a pickup or a negative loan, I left it.

And the numbers were fine. Nor can I easily conceive of in detail the exact mechanism by which early loans were hypothesized to mess things up. Diary, you really have to wonder how much inefficiency there is in the world just as a result of people following directions without any thought or understanding of what the directions are *for* (*cf.* Yudkowsky's "Lost Purposes").

Anyway, I guess I had a good day at work Saturday. After getting home, I talked to Sean on the telephone and left for San Jose early. I hung out in the city: I walked, and I went to Starbucks, and I read, and I notebooked. I noticed that the post-Summit mingle session was still going on a little bit before the 1930 meetup was scheduled to begin in a nearby restaurant, Il Fornaio in the St. Claire, so, after gaining assent from someone at the threshold who looked like staff, I entered the throng and milled around. I sighted Yudkowsky and Shane Legg and Ben Goertzel and Razib of *Gene Expression*. I greeted Legg without introducing myself, which was silly. I greeted and introduced myself to Robin Gane-McCalla, with whom I had interacted by email.

I had a substantive conversation with one Carl Schmidt. I was rather self-deprecating; he was friendly (lowercase-*f*). He had never heard of CompTIA, and said that he's been hiring and firing software developers for years, and never once has he looked at a professional certification. I said I was a loser; he said he didn't think so—but then, what kind of person would have agreed, when it's not as if I'd declared Crocker's Rules? It was just that talking to him, and just after, I realized more fully than I usually do the great extent to which *I don't know what I'm doing*.

Well, to the meetup.

I ended up being seated next to Paul [R.], with whom I rather hit it off. [R.] is a philosophy major at GMU; he heard about the Singularity last winter, and he's already signed up for cryonics.

To be continued. —ZMD

There was good table discussion, talk of the blogging professoriat at GMU Econ; someone had heard of Roderick Long.

As I recall, [R.] said he wanted to make an intellectual contribution, but wasn't sure that he could. I told him I knew someone he should meet. You see, after I posted my email address in the October Open Thread, Anna Salamon emailed me expressing interest in contact. She had written that she, Steve Rayhawk, and Rolf [N.] were living in Sunnyvale, and that she and Steve were attempting FAI research. Later, I overheard a woman say that she had wanted to meet Z. M. Davis—it was Anna, of course, and I introduced myself to her, and introduced her and [R.].

I got to meet Nick Tarleton, whose comments I've enjoyed: it's as if we think on the same (de Broglie) wavelength. Based on the Facebook profile, of someone else with the same name, I was expecting him to be a middle-aged man from Bristol; he's actually a freshperson at Carnegie Mellon. It's a strange effect when your expectation of someone's appearance diverges so much from the reality. Years ago, I probably would have insisted vehemently that it doesn't matter, and of course I was basically right, but I can no longer insist you draw *no* inferences on the basis of sights that aren't printed words. Still, you'd wish we know explicitly what we were doing when appearances cause us to run an update.

I wonder how it felt to be Lara Foster, when she posted that she "didn't even realize [I was] a man posting."

I recall Razib greeted me of his own accord.

Later, Michael Vassar made his appearance. Vassar is a *fascinating* character. Rationality, he said, is the only thing that matters. Rationality and humanity, he added. Sitting across from Paul [R.] and I, he asked us how it felt to be two to four orders of magnitude more important than we were previously.

I was invited to stay the night at the house where Anna, Steve, and Rolf lived. I wasn't the only one there: the Harvey Mudd/CalTech delegation stayed the night, and I understand Nick and Vassar were living there for the duration of their Bay Area stays. The place had lots of whiteboards.

Rolf asked, What if you made an AI while falling into a black hole?

Back in September, there was a *Distributed Republic* post asking readers what their biggest regret was, being that the world was about to end—a joke, you understand, Diary, referencing speculations about the startup of the Large Hadron Collider. I left a comment saying that I wished I had studied harder and quit school earlier. I'll still affirm that wish—but we cannot speak of what would have happened if I had taken an explicit commitment to self-education at an earlier age, when as it is I can hardly make sense of what *did* happen. It is strange: my school record is mediocre, Tom Hughes got a better SAT score than I; Seanan got a *perfect* score; and yet there *I* was that weekend, hobnobbing with these brilliant people, seeming to hold my own in conversation, and feeling even as if I belonged somewhat. I think one of the Harvey Mudd people mentioned that some students were given something and told that gravity exists, and had to prove that mass and energy were conserved. Oh, yeah, I said, "Noether's theorem." I knew from Wikipedia, so did Nick. Doesn't that count as education, even if neither of us can actually produce the proof?—just yet. At the meetup in the restaurant, Nick admitted to me (and it's shocking enough that I should wonder if I'm remembering it correctly) that he doesn't really read a lot of books; he learns from short articles. On Sunday morning, I had a nice discussion with John, the guy from CalTech, while most of the others were still asleep. John was interested in finding out more about AI. He accorded me with respect, and I'm tempted to wonder if he assumed I was really smart because this was a gathering of really smart people. I accorded him with respect and assumed he was really smart because he goes to fucking *CalTech*. I mentioned something about evopsych (I was reading *Adapting Minds*), and one of the Harvey Mudd people said that he'd like to take a class on evolutionary psychology. I said that I don't believe in classes; just buy the books. Of course, I added, I'm also going to be poor, so don't listen to me.

Oh, Diary, what is true education? Where do I really stand?

Vassar thinks that this year's Summit was a net loss to the cause; he wants Tyler Emerson out. He still gives us one chance in six of a win. That's cheery, don't you think, Diary? Vassar stresses the importance of having Something to Protect; Nick agrees. At one point, I said I was a wreck; Nick said he was a wreck, too, before—and I don't remember exactly what he said, but I think it was something about taking the Singularity seriously. Diary, I *thought* I had Something to Protect. I remember the sense of rationalist's-equivalent-of-spiritual awe I felt when reading that post, or possibly its successor, "Newcomb's Problem and Regret of Rationality." I *do* have a very important story to tell: oh, but how, given what it involves? How does J. Michael Bailey talk about his research at cocktail parties?

On Sunday night, I went for a walk and talk with Vassar. I think he said that I was trying to be reasonable, and he wanted to help. He asked me questions, and I told him about how I found *Overcoming Bias*, and my thoughts on Yudkowsky, and about my school troubles. I think he dismissed my concerns about "The Bottom Line," saying that I already knew I could learn on my own. I think he thinks people can learn a lot on their own, and that smart people can learn virtually anything on their own. (O terrible vagueness of human memory!—what did he say *exactly*, so I could quote it?) He thinks my feeling of obligation to declare when I didn't do the reading is stupid. We got a little lost and someone picked us up in a car, after which the three of us went on walking and talking.

Well, I can never tell you everything (we always know much more than we can say), but I hope this account will suffice—I do have other things to do. Oh, Diary, just imagine what a great primary source you could be for historians if it turns out that we do save the world!—I mean, if it weren't for the fact that postSingularity historians wouldn't have to rely on *documents* when they could just download the memories themselves.

Oh, well. —ZMD

Oh, Diary, I'm so sorry I haven't written sooner! Would it excuse anything to say that I have been busy?—except I haven't been busy enough. There are so *many* things worth doing. I am so terribly inefficient with my waking moments that it's a wonder I don't die of frustration at myself. Who cares if I read "a lot" compared to others if the fact of the matter is that I don't read *enough*?

It must bore you to hear of it yet again, but Diary, I'm still hurt in the wake of my purity [born] of pain. I know, it's—inefficient. Mark as irrelevant how I came to this juncture; if I'm paid in purity now, if I think I'm winning this way, then I can just—win.

The most recent *Overcoming Bias* meetup was on the twelfth. The locale was Stanford, with the occasion being a talk by Daniel Dennett. By the time I finally found the location (after much running around and asking for direction), there was no space even to enter the room; there were others standing or milling about outside. "What did we expect anyway?" I said aloud to no one in particular. "It's *Daniel Dennett*." I think a woman chuckled at that. I walked back to a library and read for a bit before coming back in order to meet up with the gang after the talk. When I returned, the doors were open and you could just hear Dennett finishing up the talk. Eliezer was at the door; I waved, and he nodded at me. As people were leaving the talk, he handed out copies of the "Twelve Virtues."

I got to meet the Marcello Herreshoff. Something about his voice just screams "Asperger's," though I wish I knew the etiology of the stereotype. He asked me how an *Overcoming Bias* reader comes to be working in a supermarket, and I asked if we were really that much of an elite. Robin Gane-McCalla wants to do a bit of street theater, taking passerby aside and doing the 2–4–6 experiment or the Wason selection task on them, as means of public education about bias. In conversation I said that the passerby could look it up, and I think that got a laugh.

Apparently I look a lot like Peter de Blanc.

Some conversation should perhaps leave me wondering if I'm doing everything wrong. I think I can effectively learn lots of things simply by reading about them, whereas with math I think I have to actually work the problems. Why? Is math just different for me, or is so much of my supposed book learning really a superficial farce due to my failure to actually write about it? Anna Salamon said she noted my facility in paraphrasing Eliezer's ideas; I said that's because I've read practically everything he's written, and she said it was not just reading, but *understanding*.

Tracy's taking vacation this week, which leaves me with the booth, Jennifer taking tomorrow and Wednesday. I was going to meet Ellen yesterday, but she canceled. How can I explain to you about Jessica vis-à-vis Ellen and me? *Bayesian Methods* came in the post today.

Diary, I shouldn't be hurt. Aren't I *doing better* now, in so many ways? Sure, I've wasted my youth, but I can't do anything about that now. We'll say that as long as I do a lot of math, writing, and tech study tomorrow, I don't have to hate myself. That much. —ZMD

There was a *Overcoming Bias* social event on the twenty-fourth. True, it hadn't even been two weeks since the last event, but Robin Hanson was in town, as was Michael Vassar. I noticed the day before that on the *meetup.com* site, a user named "biofreak" had RSVPed "Maybe," asking if anyone was coming from the East Bay. The user's profile introduction reads "brown dude. you know me."—Razib of *Gene Expression*, of course. So I made contact and we agreed to meet at University and Shattuck at 1745. After work on Saturday, I dropped off the coupons at 936 and got a woefully misnamed espresso beverage from the coffee kiosk there. Then I went downtown and parked near the library construction site. I bought a "FEMINISM IS THE RADICAL NOTION THAT WOMEN ARE PEOPLE" button, a replacement for the one I had bought there in 'aught-six and lost some time ago. I had recently reoutfitted my bag with three pins bought from *nancybuttons.com*, which site I found out about because the proprietor occasionally comments on *Overcoming Bias*. My newly-accessorized bag could hardly be complete without a gender pin, and for some sentimental reason I wanted it *before* taking Razib to the social; maybe I wanted to tell you. We made the rendezvous with nearly perfect timing, and it was really fun talking with Razib. He said he'd send me a preprint of the new Cochran and Harpending book. He said that

programming is like bricklaying for not-stupid people, which I find heartening and amusing. He paid me a five for the ride.

The meetup was a wonderful experience, as always. A blonde woman wearing a red dress and black high heels stuck out among the predominantly male throng of geeks, and oh Diary, I wish I had a transcript of my brief interaction with her. (Some would argue that I *should* be wishing that my deontological bindings were weaker, so that I could just loosely try to recap what happened, rather than omitting relevant but inexact-remembered experiences for fear of introducing inaccuracies into the record.) She introduced herself as Fortune; her full name is Fortune Elkins—the "frelkins" of the blog, it turns out. (Her mentioning this explicitly made everything that came before make much more sense.) Elkins is a fascinating character, perhaps worse than Vassar in some ways (*cf.* 272). For all my reading, I guess I'm still nowhere near a firm stance on evolutionary psychology. When someone you've just met casually speaks of your desire for social dominance, what can you do but stammer out a shocked and perhaps unconvincing denial? And if they reply by saying that you *look* male, what could I do but say that it doesn't mean I'm *happy* about it?—this to the apparent surprise of Robin Gane-McCalla. Before you know it, I'm sketching a configuration space on the board. Elkins said she was not getting the tranny vibe from me. She referred me to one Chav Doherty, who is trans. Diary, I hope this suffices.

I got to meet Kennita Watson (*cf.* 216). I stayed past midnight, after most of the crowd had left. When I mentioned to Anna the amount of time I spend working, I think she made a face and said I could be using that time to make myself smarter. Marcello lectured to me a bit about group theory, and gave me a tip about visualizing higher-dimensional spaces by analogy that I didn't really understand. Razib and I talked more on the drive back, and I told him about the theme. He even suggested I perhaps give him a call some time if I come to Berkeley!

O friends O problems O books! Diary, how can I bear to be so inefficient in this world full of data? Razib said that I'm young; I could say that it's true, that I'm only now starting to sweep around and align myself with the dominant eigenspace of my soul. But that would just be me showing off how smart I am, and probably has nothing to do with reality. —ZMD

303 — Friday 13 February 2009 — " 'Ruby on Rails' "

Diary, what madness or sanity have I gotten myself into? The latest *Overcoming Bias* meetup was Wednesday night, the occasion being an interview with Daphne Koller at Stanford. Discussion followed; I ended up sleeping over at the house of Anna *et al.* Anna thinks the addition of one skilled person to the cause has the potential to reduce existential risk by something on the order of one tenth of a percent. We both agree that this is a *ridiculous* number; the question is whether this is in the sense of being unrealistic and false, or in the sense of humans being so *ridiculously* stupid that this opportunity hasn't been arbitrated away. Re me contributing, Anna said that I would do better by spending time thinking, rather than donating, at least while I make so little money. I don't know if I was *shocked*, exactly, but I think I feel as if I should be. Why do these brilliant people think I'm worth having around? When I said I was stupid while Marcello, Steve, and I were questing for my car, Marcello told me not to label myself.

When I got home, I had an email from Michael Vassar waiting in my inbox. It said:

By the way, what sort of work are you currently doing?

Patri Friedman is considering hiring someone to coordinate H+ youth orgs. Basically an admin job. \$40K-\$60K/year. Not sure it would fit, but would you be interested in any event?

Diary!—what a world this is! Of course I wrote back today:

Wouldn't fit; for that, I imagine you'd want someone extroverted who's been active in such organizations. But thank you for telling me!

Oh, where do I fit in this world?—of course conventional *fit* is among the lesser of my concerns in the face of my purity borne of pain, or perhaps its dual (I didn't tell you about this?) my pain borne of purity. Well, I do need a better job. I like playing Tracy, but pretty soon I'm sure it is getting time to move on.

There's a famous teaser that goes something like this: if you drive at 20 mph one way across ten miles, how fast to you have to go on your return trip in order to achieve an average roundtrip speed of 40mph?—and the answer is that you can't. Is there any way I can make up for the terrible inefficiency of my youth? Today Jessica said that Ellen lacks common sense in her work. Diary, I should go. —ZMD

ADDENDUM— Diary, I'm horrible. I'm proud of the pretty envelopes I made for the recycling vouchers and authorization slips, but they're *just* supererogatory flourishes for my single-figure wage job. It almost seems like I was obsessing about rationality and efficiency *all day*; how could I let my focus drift that much when I got home?—this is not to suggest I've done nothing good today, but I haven't been *nearly* good enough; there is *so much* that I *want*, and I've betrayed it. How *dare* I, when every mistake that I make can and will be used as evidence for the supremacy of schooling? Even now I would like to masturbate, although I need sleep. It is such a beautiful Diary!—O richness of endeavor! O my glorious notebooks!—and I'm *so ignorant*. Do I just keep reminding myself that there was a time when I did no math at all?—so change *is* possible; I just need to make a few corrections, but I've *told* you this. Why should you *care*? You want to read a glorious novel. Diary, I have sinned, sinned, sinned. —ZMD

310 — Wednesday 4 March 2009 — "Trivers and Travails"

I got back tonight from visiting Anna *et al.* in Sunnyvale. It was really fun, and potentially very helpful. I spent a lot of time interacting with Anna. Steve is less articulate, and his sleep patterns were further out of sync with mine. I barely saw Rolf, whom I don't know very well at all; he works at Google. Robin Gane-McCalla is living there now; [...] Vassar (now SIAI president) hired Robin to do outreach work for the Institute [...]

We should probably do less navel-gazing next time: my fault. But talking with Anna is fascinating! When Anna said she was surprised by my SAT scores only being ninety-seventh percentile (as I had mentioned on our walk the previous afternoon), I said that I was surprised that she was surprised, and she said that I should expect her to have been surprised, given that I had said earlier that I thought she overestimated my abilities.

Other takeaways and tidbits: I need to take some time to *seriously* plan for my future. I need to gather more information, and really think about what I actually want to do, and how I might actually do it. Perhaps I'll even play around with some expected utility calculations. Diary, it is important to realize what *g* is good for, and what it is not good for. I should consider adopting a *no-thoughtcrime* principle. Anna says to notice where new theories contradict your old common sense, citing a book titled *Made to Stick*. Anna says that mere experience isn't enough to teach you skills—you need to use the experience to try to figure out what works and what doesn't. Regrettably we don't know of a study on this. This has not been the best missive; I will keep you posted. —ZMD
347 — Friday 29 May 2009 — "Fucking Autodidact, From the Get-Go"

I tell people that I get all my new music from the Safeway audio loop. There's this one great song by a group called Stretch Princess—

*"Oh I am not about to be
Wrecked by the things you did to me
Though it's not easy to forget"*

Of course you can guess that I associate this song with my schooling troubles—about which I suffered more today. Diary, *why* can't I just get over it? It would be one thing if the bad feelings actually helped me *do* stuff—but they really *don't*. Doing good intellectual work makes me *happy*. When I feel bad about my lack of credentials I just end up moping around; I accomplish nothing—or if it's not *about* accomplishment, say then that it makes me smaller. It's scary to think what I would be like if I had dropped out of University but if it hadn't been for Eliezer Yudkowsky. I know, I've told you, but—this network of concepts, this ideal of uncompromising rationality gives me strength. I know that anyone could say the same for their having accepted Jesus Christ into their heart, but mine is a special kind of strength in its exquisite generality. Try to get the *best* model of what's *actually* going on, and use this true knowledge to figure out how to do the *best* thing. I do seem to have amassed a great general education, and it has proved genuinely useful. It is nice to have some idea of what the fuck is going on. Even in my moments of madness, in my searing pain, I contain a core of relative epistemic rationality.

There's an *Overcoming Bias* meetup tomorrow—or maybe I shouldn't call it that, as *Overcoming Bias* is now Robin Hanson's personal blog. I've baked cookies to take, and I'm probably going to give Razib or some of his acquaintances a ride.

Yesterday I visited Ms. Gunnison. She agreed to be one of my references for when I start questing for a better dayjob.

I'm such a terrible person; I do not deserve to hang out with such awesome people tomorrow. But I used to be even worse. And consider Jessica, who claims to be too lazy to keep a Diary. What will this terrible world come to? Our books are so beautiful; O beautiful that we create books! But I fear we will tear ourselves apart; we waste so much time. O terrible! —ZMD

348 — Saturday 30 May 2009 — "Desperate Signals"

Yet more emotional turmoil today. I used to be *good* at my job in the booth, and I'm *not* anymore. I ended up crying at lunch before the librarian whom I once half-joked I had a man crush on. (For the record, he has since cut his hair.) It turns out I'll be driving to the meetup alone. I should be cheered up to meet my creepy Singularitarian friends—and *new* creepy Singularitarian friends! Only I'm so torn up inside—I don't know what I'm going to do with my life. You can say I'm doing better than most others, but Diary, that's such a low bar! I want *everything*. Today I changed a helium tank for the first time, with no instruction. I have an idea for a piece of shamelessly autobiographical fiction that might be worth writing up and posting to *Less Wrong*. I remain yours. —ZMD

349 — Sunday 31 May 2009 — "Count and Choose"

Mark you that my life is only just beginning, that all this madness was but some terrible prelude, a hoax played by an ape.

The meetup was at the house rented for SIAI summer interns. O friends! It is so wonderful to share company with such brilliant people!—and I *fit in*, too, which I must have previously remarked to you at some point, but it's the most peculiar thing, to the extent that I figure surely I am justified in remarking on it again! Yesterday I felt like shit; I couldn't even adequately carry out my duties at Safeway. (I carried them out, but not *adequately*, not nearly so by my standards.) And then, later this same night, I somehow find myself discussing my alternate evolutionary biology intrigue before Yudkowsky, Peter de Blanc, and a programmer named Jey whom I guess is friends with Razib. I am somehow made *legitimate*.

Tom McCabe reminds me of Richard.

Liron, who works for a company called Slide, took a quick look at my source for `degreeseq.py`, taped as it was into my math pages. He said it had no obvious *WTFs*, and that I had potential. He said I could be good with a few hundred hours practice, that good preparation to be a code monkey would be to learn Python and JavaScript and to write a web application. He implored me to hit him up at Slide—*well*, Diary! I *do* want to get out of Safeway *soon*; it's just been too long to work in the same store, and it's really time to move on with my life. I'm not good enough to be a code monkey *yet*, and it's probably not wise to propose to Father again the prospect of me taking some months for study. I *finally* scheduled my second A+ exam for Thursday—I shall have to prepare for that. So I get the A+, apply for a bunch of jobs (making sure to write articulate cover letters addressed to the appropriate person), and hopefully something will happen. I should also consider the possibility (however fancifully the manner in which I do this) of taking a step *down*: of taking a job, any job that takes fewer *hours*: in this way I would satisfy the have-a-job desideratum while still procuring more study time for myself.

I has a very engaging conversation with a psychology grad student from Berkeley named Nick Gwynne, who is passionate about his studies.

I slept over, and associated the next day with Carl, Anna, and Peter. We went for a run. I did some math and wrote some code in pencil: there are *always* yet more problems to solve. O detail! Things just might work out. And if I continue to find it difficult to explain exactly what it is I *do*—well, but I will work it out. —ZMD

350 — Monday 1 June 2009 — "Disjoint Worlds"

Diary, an interesting couple of data points. I felt so passionate and *alive* on Saturday night, being with other intellectuals! My every breath was infused with a sense of purpose and meaning. I knew that I had no reason to be afraid: I could *work* a dayjob, and do anything else that seemed like a good idea (which is really almost the definition of a rationalist), but in the end I was living for insight. All the pain and shame and fear that I felt at reading blogospheric disdain at all the stupid proles simply faded away, and I was happy and unashamed.

And today—the serenity of that night is gone, and I am left with anxiety. What *should* I do? In this crazy world, what can we do to keep it all from coming down? And from every direction comes conflicting advice—I can't let the social pressures eat my soul, but I don't want to be a stupid rebel, and it might be dishonest to believe that *I* can at least avoid giving advice: we all push on each other, just be being around or saying a single word. Yet I must not surrender my vision.

I'm doing so much better than most—but this is a low bar and means *nothing*. Rationalists should *win* full stop and in full generality, not win *over others*. Why should I care what people think of me? I seek *knowledge*—I seek the map that reflects the territory, and once I have the map, then I'm in a better position to figure out what to do. But if men holding lesser maps should shout at me, then I can only take this data into my map: it is not some kind of unimpeachable verdict bearing on my soul.

I don't want to develop a bad attitude; I want to be *good*. If I am angry that our dominant culture has no place for meaningful independent action, this does not *stop* me from engaging in independent meaningful action. If I am not *honored* for my private readings, this is no impediment to reading.

How do I package the magic? How do I hold my own vision, how do I mark my lifeworks as sacred, when those who surround me have no concept of lifework? I can't go back to University, but I can't let the rest of this world consume me—but neither do I want to have a sense of entitlement; there is a *dignity* in jobs. Protect me! —ZMD

ADDENDUM— I talked to Richard on the telephone. At the meetup on Saturday, Eliezer said something about how he wanted to put concepts from his posts on the wiki. The idea was to have posts link to the wiki, and the wiki link to the posts. Later, I said to him personally that his time is valuable, and that I wondered if I should take a shot at this. I think he said that it required a deep understanding of the material, and I said that I've read almost everything he's written, much of it more than once, and thought about it for more than a year. He asked if I had the art of relevance—this was a reference to earlier conversation where he observed that in most discourse, people's lines just run skew to each other. I said that I didn't know, and he said that I've earned the right to take a shot at it. Oh, but how could I possibly have *time*? There's so *much* that I want to do!

Only I might be engaging in an evasion. I say that I want to know *everything*—which is fine, as an ideal. But *in fact*, I am only a human being with limited time (although I could be much more efficient) and a limited brain (although I should think that I'm nowhere near my limits). *In fact*, I have to study specific things, which means *not* studying a host of other things. Even if I don't know how to *describe* what my background is, I *do* have a background, and while I don't want to be pinned down, I should at least be able to *say* what it is I've done, where I'm confused, where I'm not confused, gaps on my map, edges of my map. But I really don't know, nor do I wish to applaud myself on my mere *awareness* of my vast ignorance. We want to *solve* problems, not make ourselves feel superior to stupid proles who don't even know that they have problems. I guess I should write Swift and Peter Gray. *Should*—O such an easy word to misuse!

On Sunday, when I said that I need to make up for my wasted youth, Peter almost immediately said, "Sunk cost fallacy." And of course, he was right. *These* are the sort of people I want to be around! O Diary O terrible my terrible ignorance! I really think I am unusual, because if more people were like me, you would expect that there would be standard answers to my concerns. The intensely *personal* art of learning would be an active front of research, and regularly discussed in the media and about water coolers. I've said before that it's more fun to hang out with Anna *et al.* and feel inferior, than it is to hang out with my coworkers and feel superior. Could it be that for the same reason, it's more fun to be normal than to be an iconoclast? If you truly love beauty, then why *would* you want to be an island of beauty? Why not spread beauty all around; should not every person in themselves be a work of art? Why isn't this view *mainstream*? Why do people give in to the ugliness and make it part of themselves?—I'm not saying people should be empty-headed idealists; we want to be not-stupid idealists. Riki Wilchins and Leda Cosmides should have a heart-to-heart.

Where do I stand? Worlds beneath worlds, hearts within hearts, where do I stand? I don't have to go into my dayjob tomorrow, but I have to prepare for my A+ examination, so I can't spend the entire day studying

things altogether closer to my heart. I wish people were smarter; I wish we had some sense of what the fuck we're doing. And yet I must not *wish*.

Today a customer remarked to me that wasn't it funny how sometimes the lines were long, and then there was nobody. I said that there was actually research to the effect that human intuitions about randomness don't quite match up with how random sequences actually look. She was impressed and asked if I was a (I think she said) sociology major, and I referenced *Judgement Under Uncertainty*. She said something about how even when it doesn't seem like it, there's a plan to everything. Or a Poisson distribution, I said, but I don't think I said it very loudly, perhaps in part because I probably don't know how to pronounce *Poisson*. —ZMD

364 — Thursday 9 July 2009 — "The Eleventh"

The meeting with Maxwell was splendid; I may have shined—only it's so hard to explain things, and I mean this in more ways than one. I studied in the city library before going to visit my creepy Singularitarian friends, who reacted when I referred to them as such. Anna, Carl, and I think Steve had just returned from a conference in Spain. O scholarship!—words cannot express my hunger for true knowledge—or the hunger I *want* to feel. There's too much; we are too rich—and yet so poor; I can *see* gaps to fill. And yet I'm still bitter; a man from Heald called earlier today, and I did not resist from dumping on him a little. It's a waste of energy to be angry, still—so *what* if Robert Wright is famous and I'm never going to be? Can't I just be happy that *Anna Salamon* thinks I'm worth having around, or for the book on spectral graph theory that's due out later this year? Anna suggested dumping thirty minutes a day into cognitive behavioral therapy, which is not a bad idea. But O minutes!—O wealth! I have a way upward, and no shortage of things to do. I could fill many human lifetimes solely with my own intellectual endeavors, but *then* I've also got making a living and existential risks to think about. Never a dull moment—I mean, besides checking.

I've ordered a copy of *An Abundance of Katherines* to give to Anna as a gift.

Vector calc is slow going, and there's so much that I don't understand, but I can't spend forever trying to apprehend the Lagrange multipliers, when there are so many other things that need apprehending.

Today Jessica told me I should buy some of that Axe body spray, because she didn't want to say outright that I smelled bad. I thanked her for informing me of my incompetence, which is really something I need to know. I actually did buy a can (perhaps I could not resist the thought of holding it up to Jessica as I left, which I did), although I lost some feminist points on the purchase; Axe's advertisements are horrible.

The next *Less Wrong* meetup is just Sunday: Robin Hanson, Michael Vassar, and Roko Mijic will be there. Do I bore you with my repetitiveness? Do I, Diary?—except you're stuck. —ZMD

366 — Monday 13 July 2009 — "Isomorphic to One-Boxing"

The latest *Less Wrong* meetup was Sunday, and as always, it was amazing. I think that I know that I told you, so I think that I know that you know, but it still seems so strange. I think of myself (I make sure to think of myself) as so obviously fucking retarded, and yet I go to these *elite* fucking gatherings, and I *fit in*. Something is *wrong* with this world when I'm impressing Nick Tarleton (who is part of the SIAI summer program), rather than the other way around. Nick's impressed that I do math without having Something to Protect—but *honestly*, Diary, it should be obvious—if you *don't* do math, then you just end up playing worthless fucking *games* instead. Like Anna said, when you don't know what else to do, accumulate intellectual capital, and every moment counts. (When I said the line about accumulating intellectual capital, Steven said that that was true of any sort of capital. I said yeah, but studying is more fun than making money.) Is this the sort of thing Eliezer was referring to when he wrote that the way you acquire superpowers is by recognizing that they *are* perfectly normal?

I showed up early, and folks were seated around the long table having a discussion with Robin Hanson, who was in the Bay Area.

A bit later, in miscellaneous discussion (the kind you don't remember all the threads of, perhaps in part because you didn't hear them all), someone said something about needing a very complicated specification of chemicals to have some sort of effect, and Roko said it would be easy to, *e.g.*, make chimpanzees more docile, just lower their testosterone. I think Mike Blume made a reference to eunuchs in some ancient society, and I made a reference to our resident troll (*cf.* 320A). And I said, as far as the lowering of testosterone goes, we have plenty of data from transsexual women. This prompted to Roko to make a long-winded misogynistic joke about chimpanzees engaging in some stereotypically feminine behavior. I think I might have said, "Yeah, fuck you, too," but I guess I didn't say it loud enough, because Roko didn't notice. I think I retreated from conversation shortly thereafter, lost in

my apprehension of this world, trying to apprehend my apprehension, thinking of what my feminist friends would think, maybe humming the *Voyager* theme. Soon Roko came over and introduced himself, though we had already recognized each other.

[...]

Michael Vassar is fascinating, as always.

Later on, I was present with Patrick (the *orthonormal* of the blog) and Roko, talking about mathy things—I think they were impressed when I mentioned my rederivation of the expression for disjunctions—Diary, what the fuck?—I'm *retarded*.

Well, at least I have my fury now—so strange that even so many elite people don't have the concept of this fury, that making sense of things is *important*, and that if you can't get a University to support you, you can do it directly. If somehow Nicole were to be cast out from the great Institute, what would she do, where would she go, what would she study? I don't think she would work as hard; I don't think she is desperate the way I'm desperate, for an understanding that embraces everything. But I am almost certainly wrong, and I could be arrogant, and I have other things to do today, and so I leave you. —ZMD

ADDENDUM— More bitterness tonight. I feel like I've been systematically betrayed and lied to. But they *weren't* lying to you; it's just that now you know something they don't. —ZMD

375 — Wednesday 12 August 2009 — "Literacy"

O Diary O eventfulness! Visiting the SIAI crowd was wonderful as usual. I liked Carl Shulman's presentation on the genetics of human intelligence. You're aware, of course, that we've hardly found any QTLs for *g*. (Tom McCabe asked what a QTL was, and I piped up, "Quantitative trait locus," before Carl could answer. Is it that I'm insecure about my scienceness?) So the idea is that variance in human intelligence is caused by everyone having a large number of minor detrimental mutations, with smarter people having fewer defects on average—this having important consequences for the possibility of enhancing whole brain emulations. I had a really good long conversation with Marcello yesterday, during which I spent a lot of time explaining my thoughts about social organization and my purity born of pain. Marcello started off asking me what useful things I knew that he could learn from me—Vassar had told him that he was good at learning things from people. Well, I *hardly* had anything to say relevant to AI work, but it was good to talk to him. My self-confidence still (*still*) goes up and down radically; one day I feel like I have rationalist superpowers; the next day I feel *retarded*. I still don't know how to answer that fundamental question: what do you (think you) know, and how do you (think you) know it? When Lucas (the bright local high school student who has been hanging out with the SIAI crowd) asked if it would be useful to upload something other than humans, I knew enough to say, "*C. elegans!*"

An exciting development: Anna invited me to move to the Santa Clara house to collaborate, at least for a three-week trial period to see if I can be useful. The initial idea is to outfit the *Less Wrong* wiki with good pages on concepts from the *Overcoming Bias* canon, and for the posts to link to the wiki, and the wiki to link to posts, so new users can bounce back and forth between them. The idea is for Steve and I to work on this so that Eliezer doesn't have to spend his valuable time on it. It's sort of a right-place-right-time opportunity for me: I don't have any meaningful achievements or credentials to my name, but I just happen to be smart, intimately familiar with the canon, and willing to work for free. It's not as if Steve and I will be working on the wiki sixteen hours a day, either, so I'll get lots of study time, too, with all sorts of fascinating and helpful people around. Mother and Father okayed the plan this evening, so it looks like it's really going to happen! I'm expecting an email from Anna tomorrow detailing the exact conditions of what constitutes a *successful* trial period. If it doesn't work out, or when my savings run too low, I'll hit the job market, or possibly return to Safeway (Father wants me to ask for a leave of absence rather than quitting outright).

Speaking of Safeway, today was ridiculously busy. I've continued to be pretty slow going through my morning routine, as through molasses, and I was looking to catch up on my backlog after my lunch, but I kept being called to check constantly, and there were all sorts of other fires I had to put out: I barely managed to balance a pan, much less catch up on my backlog.

And worse: I got written up for a negative override from last month! O terrible incompetence! What could I have been thinking at the time?—I have only the vaguest memory. I hate to announce my probable resignation on such an off-note. I'll have to work ridiculously hard tomorrow to make up for my backlog and—well, O terrible! O terrible my incompetence! Rationalists should *win*. Only it really *is* much better to face these kinds of things with a sense of poise and rationality.

Recently I've done a little bit of jogging; the weakness in my breast scared me so. Or do I just like to tell you about my breasts? On my jog today, I felt retarded again: what do I think I know? What have I done? Read a lot of books and blogs? What's my background? It's as if I feel that since I don't know what I know, then I can't possibly know much—like the famous impostor syndrome. But then: I know about the impostor syndrome, without ever having passed a psychology class. Well, why shouldn't I? Shouldn't everyone? Psychology is not some matter of arcane scrolls reserved for "psych majors," you *need* insights from psychology just to make sense of anything, just to exist as a civilized person in this world (civilized again in the Yudkowskian sense). Similarly economics, physics, decision theory—there's too much; how can anyone bind themselves to a *career* and a *field* when we have such a desperate need to understand this world, which is not what we think it is? Well, the careerists give me good stuff to feed on, so of course I must be ever so grateful, and I am. So I content myself to a life on the margins, sacrificing my role in the big picture in order to observe it all the better. On the other hand, being affiliated with SIAI sort of makes me margin *to center*. Only I have so many other interests—O generality!

Judea Pearl arrived in the mail recently—as if I'll have time for him soon! My *word*. —ZMD

ADDENDUM— *Fuck*. I haven't been serious enough tonight. I already was inefficient with a lot of time after I got home, probably I was so stressed out from my dayjob. How *dare* I think myself a scientific generalist, Diary? How *dare* I?—when it's as if I hardly read books anymore. I feel like—the stupid kid I was in high school. It's not okay. I want to *rise*; I want to push my apprehension to its very limits. Wasn't I doing so much better with my new methodology? Intrinsic motivation and the dream of a decision theory?—where did it slip, today? And tomorrow I will have to be so utterly busy beyond belief—and then I should plunge into the wiki project sometime tomorrow afternoon or evening and get a better sense of the scope of this endeavor. Why does anyone think I'm valuable? I'm so *ignorant* and useless. —ZMD

379 — Tuesday 25 August 2009 — "Edifice"

Still more pain today about my lack of credentials. Well, so what? When people say something that strikes me as overly statist, I like to remind them that life itself does not emanate outwards from Washington and Sacramento. Well, let *me* remind *myself* that life itself does not emanate outwards from Santa Cruz—or Cambridge. Those who would shame me for not being in school: what's their model of the world like? Do they really predict that no one would ever learn anything, that no one would ever *do* anything, unless some authority or structure bears down on them to do it? Would they deny the entire *edifice* of my thought, even when presented with the notebooks themselves?

I can't please everyone. Anna Salamon is absolutely fucking brilliant and thinks I could be somewhat useful, so I'm going to go try to be useful to SIAI.

And I could—I *could* go back to University someday, and even do well, now that I'm not so stupid, now that I actually know how to study. If I do go back, it will be on *my* terms: to learn.

I wrote: induction is on my side now; I've gotten better before, so I can do so again. There's so much—there's so much I want to know, so much I want to do. I may be hampered, but I'm less hampered now than that stupid fucking boy at University who had no idea what he was doing. No sense of the seriousness of this entire endeavor. Not even when he stood on the staircase of the University library, pressed the Kurzweil book fast to his chest, and pretended he was Kathy. —ZMD

381 — Monday 31 August 2009 — "Existential"

I've moved into the Benton house. Any tension there might have been about my name has resolved itself: I'm "Z.M." here. Yet more anxiety about my lack of credentials—last night, I ranted some more about the system and my wasted youth. Although really—I should be on the way to getting over it. I was *stupid* then; I'm better now. However bruised and scarred I am from my schooled past, at least I came out of this madness with the knowledge that you *can* do math, &c. outside of school. However painful the lesson was, however wasteful the process that got me to this point, I now have this tremendous edge, this precious gift, this shining jewel of epistemic rationality, the skills I need to think my way out of a box. I *want* to be the sort of person who loves knowledge—damn my past and damn human nature, I will trace the causal forces that act on humans, and thread the needle through to get me to where I want to go. I have a role to play in this organization that's doing very important work! We (mostly Anna) drew up a list of useful things I could do, and a points metric for what would constitute a success of the trial period. So I must work ever so hard, and it doesn't matter so much—if I was stupid before. I want the true knowledge, now. I want to do good work, now. I want to be smart, now.

There's this guy, Alex, living here with us, working on a startup with Emil. Alex is disturbingly masculine; he intimidates me, his entire way of being is alien unto me. We're all dedicated to human values, here, and want to preserve them past the threshold of transhuman intelligence, but despite Cosmides and Tooby, human value is not this unitary thing—*my* ways are alien unto almost everyone in this world. And so much of who I am must depend on the particular path I cut through the blogosphere—but enough. I must be useful. —ZMD

382 — Tuesday 1 September 2009 — "Curl"

I had thought that I brought hairbands and my razor, but I can't seem to find them today. Have I really *looked*? There's still time for redemption, but so far it has not been a good day on net. I drove Frank to SFO in the morning, and we had really good conversation, so that was nice. Part of the plan is that I'm to take four hours of unavailability for myself every weekday in the library. I spent a lot of time lost on the way back from SFO and trying to find the Santa Clara library. I ended up at the Sunnyvale library, which is very nice, and I did some math, and worried a little about the runaround about my name. Before leaving the library, I had a few moments when the whole schooling nightmare just seemed insane—not in the sense that where I rave about the insanity, but in the sense of the harmless homeless man you can just ignore. The Sunnyvale library is very good; it has a lot of books, including honest technical textbooks. People who think they know things [write] books, and everyone learns from books. That's just the way things have always been. The day has not been so good since then—I haven't accomplished much.

Last night I signed up to read the wta-talk list, on the grounds that one of my duties is to track the relevant blogs and lists for people or facts that might be useful. Only—signing up for the list gave me access to new Eliezer material that I haven't read yet, and between yesterday and today, I overdid it. You can't spend too much time with one author, any author, it makes you sick. Do Hume scholars ever get sick of Hume?—I know *exactly* how they feel. And it's so sad: from everything he taught me, I should have known better: it puts me in no end of trouble to get burned out on Yudkowsky just when I'm supposed to be working on the *Less Wrong* wiki! I did do some work on the wiki today, but not with Steve, and I didn't do all that much and what I did isn't all that good. Maybe I have no idea how writers work; maybe I spend so much time writing freely and fluently to you and my notebook, that I don't know what it's like to face a writing *problem* that I don't know how to solve. I'm a putz, just like I used to tell people when I was at the University. SIAI seems somewhat pathetic today; aren't we supposed to be saving the world? The way we sit around, you'd think we were just typical slacker grad students.

I reread some Paul Graham essays yesterday, and they made me feel better, I think when I was feeling sad again about my lack of credentials.

So, I've fucked up so far today. But I can do better, really. I'll go track down Anna and have her show me her decision theory post. Or I'll read from one of the heuristics and biases volumes. Or I'll work that very small existential-risks-related document I've been thinking about. Or I'll work on something personal: notebooks, fiction! Or something, anything to get me out of this rut! I remember being at the University whining to my TA that I was supposed to be an adult now, and yet somehow I just—couldn't—get it together. Well, I've been getting things together. —ZMD

383 — Thursday 3 September 2009 — "Divergence"

Gawd, I feel shitty. I haven't been productive enough. When I tell people this, they ask by what standard, but sometimes you don't need to explicitly declare a standard when it is just *obvious* that you are underperforming. Is it that I don't have a clear notion of the work I'm trying to do, and how to do it? I can *consistently* make intellectual progress for myself doing math at the library. The past few days I've been saying that I'll spend the afternoon doing good work on the wiki, but then somehow I just let the day drift by—and I *feel* like I ought to be doing existential-risks related stuff, but I *don't*, and to make it worse, I don't get anything *else* done, either. If you notice that you can't work, and somehow you can't get yourself to apply the obvious reflective correction, you should at least notice this and go read a paper book or something. But I haven't been reading paper books, and it's just so *retarded*. We're *supposed* to be saving the world here! Okay, so I added some content to the wiki today, even some good content, but it's way too *slow*. It's not like writing to you with this rapid and fluent voice: *bang bang bang!* O Truth!

I had a good walk and conversation with Anna last night, and she said she was even more glad I'm here than when I first came, which implies that I've exceeded expectations. (Anna is *thoroughly* versed in the Way of Bayes, when she's impressed, it really means that her model of reality has taken a hit.)

I had good conversation with Marcello and Steve today. I don't deserve a face. How could I have let these days slip by in the way that they have; why can't I just *do stuff*? This sensation of having wasted time is an unfamiliar one unto me. Even when I spend time on the internet, I tend to be genuinely engaged with something at least somewhat worthwhile. Except these past few days, when I've just—I don't *know*, I can't *say* where the time went, and that should scare me. Nor have I been keeping up on that log of my personal intellectual progress. O progress! It's so beautiful, being able to articulate things like this—being able to perceive patterns and deep structures, and feel the meanings, and ask: what is the limit behavior here?—how does this generalize?

I don't deserve a face—I am composed of fail. Tomorrow I should just be intent on *doing stuff*—write stuff on the wiki, lots of stuff, just put stuff up there—it's a fluid medium, you can always edit, always iterate. And when my meager powers of writing wear out, *read a paper book*. Talk to Steve! Prove something about a deterministic finite automaton. Something, *anything* that doesn't make me feel like a bored child, or an unemployed person who knows nothing of our boundless depth of intellectual riches. I don't know how they bear their boring lives. O how we must save them! I don't deserve to sleep. Tonight I will sleep for strictly instrumental reasons: I will sleep in order that tomorrow my performance will be up to the level of a person who earns the right to sleep. —ZMD

384 — Wednesday 9 September 2009 — "Distributed Cognition"

My emotions have had such high variance during my time here. Recently I've been suffering name and identity issues, and I talked with Anna about it yesterday, and she agreed to start calling me Zack, and I felt better. But then I spent a lot of time this afternoon feeling sad and anxious and reading on the internet, when I *should* have been doing work on the *Less Wrong* wiki. In the evening, my mood picked up after good conversation with Marcello and Steve and editing and formatting Aruna Vassar's letter to potential Summit sponsors. I think I did a good job on that, at least, and I felt strong and less afraid, like I could exist in the world without cowering and without being driven mad just because different people call me different things in different contexts. I wonder about the fallibility of human introspection. When it seems like I'm feeling sad about my name and identity issues, it might be that I'm just sad due to some sort of chemical imbalance, and this is what makes me latch onto the name and identity issues, but if it wasn't that, it would have been something else.

Why do people let themselves stop growing? This bare hypothesis that my anxieties aren't actually *caused* by their referents isn't something I couldn't have formulated five years ago, but my intellect is noticeably deeper now, and it's as if I can glimpse all sorts of subtleties that most people couldn't get from the mere words. The whole runaround on the blog lately about determinism and counterfactuals and decision theory is really astoundingly relevant to practically all areas of life, for when do we not have to cope with matters of choice under uncertainty? I'll do better. —ZMD

411 — Monday 9 November 2009 — "That's Not a Fox"

The nice thing about being so well-documented is that you can actually get some sense of where you've been, rather than just having to guess at things and getting them wrong.

A little more than a year ago, when I went on a walk with Michael Vassar (*cf.* 272), he said that people can't even properly distinguish good news and bad news. Suppose, he said, that we found out there was a toxin in the water that was lowering everyone's IQ by ten points. Most people would call this *bad* news, but actually, Vassar claims, it's good news. You see, the toxin was *already* there whether we knew about it or not; and now that we know, we have the opportunity to do so much better by getting rid of the toxin. Map and territory: decision theory is only forward-looking; you can be sad that the toxin was there in the first place, but the *news* is good.

So I wonder if—from the perspective of my *current* self—I should be happy to learn about just how far this recent recent madness has brought me down. Comparing July and October in terms of STEM pages and notebook pages, my intellectual productivity has fallen by approximately a factor of 2.7. So if Anna thinks that I've been useful regardless, just imagine how useful I could be when I'm not 37% of my healthy size!—which I have no reason to think is my absolute *limit*.

And as long as it still outputs all the same empirical predictions, maybe it's better to frame my recent recent madness in positive terms. This gender thing—you could argue that I've been going about it entirely the wrong way, but my essential concerns *are* valid. Maybe I've still got muddled preferences about what I should do here-and-now in *this* life, but when I imagine my idealized upload society, I do *not* see a gender binary. That *means* something; that really means something. And so this initials thing—well, it was a *reasonable thing to try*. Okay, so it didn't work out and I burned up all my name-change capital and I'm damned to have a recognizably masculine name this side of the Introdus. It's not *that* bad, and I've got so many other things to think about in the hopes that we might

have an introdus in the first place. Maybe one of my analogues—myself in another Everett branch, or some similar conjunction in *this* world—has done things right, and will go on to do great things as *Shannon* or *Courtney*. But here and now, Zack M. Davis simply does not have any more time to waste operating *well below 37%* efficiency. And so I must work. And so I remain yours. And signature,
Zack M. Davis

416 — Saturday 21 November 2009 — "Motivational Invariant"

Diary, I can't wait for Anna to get back. I'm scared. What have I gotten myself into?—why didn't I forsee? Shouldn't I have noticed how dangerous it was to get so much of my worldview from *one man*? Well, I did notice, but what was I supposed to *do*?—Yudkowsky is so obviously right about so many things. A disturbingly large amount of the groundwork for the Singularity argument is just straightforward Copernican rejection of the intuition that everything has to take place at a human scale.

Diary, I'm just *barely* capable of running my own life; how could I possibly be capable of thinking about Singularity-level issues? Of course, that's an obvious fallacy: *capability* is not some unitary thing. Steve Rayhawk has the applied maths skill, but he can't hold a regular job. The plan had been for me to be around for quite a few months, helping out. We were going to get a joint checking account, with debit cards. But now I'm wondering if I should be planning an exit strategy for myself: a quiet, graceful way for me to get away from these beautiful people and go protect something *small*. Maybe it's only a delusion, but I can imagine myself working harder at an ordinary job, studying for myself and not the *entire fucking lightcone*.

Everyone keeps mistaking Kaj Sotala for a girl and I'm jealous. Kaj made a post at *Less Wrong* saying that the community should collaboratively write a piece for the special issue of *Minds and Machines*, and he created a page on the wiki for this purpose. I added a lot to the wiki page myself. I think the crowdsourcing approach is doomed, and I left a comment on the blog saying as much and asking if Kaj wanted to coauthor the thing with me. Except—maybe that was wrong. I shouldn't be the one to be summarizing the hard takeoff argument when I'm having such doubts about it. On the other hand, all my doubts are of the form: *this idea scares me, cult cult, run away!*—which is not the sort of intuition that seems epistemically trustworthy. A user called "righteousreason" (who I seem to remember having Objectivist sympathies) thinks my writing is lame and that we should be directly using Yudkowsky's writing. Maybe Anna or Carl will have advice; honestly I worry this project is just going to crash and burn. Who do we think we are; what do we think we're doing? People have given their lives to much lesser conspiracies, so why can't we keep a kitchen clean or put out a few academic articles? And how do we know we're not making things worse?—I suppose the obvious answer here is to just go with your expected value, and do more research with a high value of information, but I'm scared.

Maybe when Anna gets back she'll be able to talk me out of this fear, but what does it say about me and what I should do, if I don't believe in our cause enough for itself, that I don't believe in our cause enough, such that I *need* that kind of motivational scaffolding? I had an interesting conversation with Mike yesterday, on a walk to the 7-Eleven and in the car before dropping him off to get a ride south. I don't know how he can be so calm, how he can so smoothly parrot back the party line about the first replicator having such huge effects.

Michael Vassar had a talk once called "Memes and Rational Decisions." In his advice to Singularity activists, he says that you'll have to overcome the burden of your mind's anti-religion defenses that want you to quit this madness and go live a normal life.

You could argue that it doesn't matter if thinking about the Singularity is making me stupid and it consumes my life, because the potential effects are so vast. What does my own misery and my own stupidity matter? What's a single human life, compared to a 0.0001 downshift in the probability of global doom, that downshift being worth 680,000 lives plus all of those yet to come?—and when you count across the many worlds, probabilities matter. Even if I'm selfish and would despise the vast majority of humankind, I should remember that my own special fragment of mindspace is embedded in humanity's fragment.

I suppose a guiding principle should be that whatever I do, I shouldn't let myself get dragged down by all this; I have to safeguard my precious general intelligence. I slept very late today because I was up very late last night, submerging my troubles in a TV Tropes tab explosion. And yet—while TV Tropes isn't obviously making me stupider, it's clearly not making me smarter, which if you know how to think in terms of opportunity costs is the same thing. Maybe we're all *wrong* about this Singularity thing. I should keep this hypothesis in mind. And yet—and yet, even if we *are* wrong, moping about reading TV Tropes isn't going to help me. Studying helps me; job skills help me. I got involved with this beautiful online community full of beautiful people and beautiful Anna Salamon invited me to come live here and work on reducing existential risk: if it was all a big horrible mistake, can't you at least see the sense in which it was a reasonable one? There are so *many* books that I want to read, so

much of which I am ignorant—how could I let myself be stopped by these moods, these feelings? It's all so confusing. When I think of this as a nice, safe science community, it's great, but when I think of the prospect of a singleton—well *it is scary*. Maybe Anna will talk me out of this fear: but what can she say that I can't already think of myself? Yes, a singleton could be risky, but you have to go with your expected value. If it's the same *argument*, why should it be so much more reassuring coming from her? I want to see the updated SIAI strategy plan. I worry that Roko's version was kind of laughable; I want to know if we have another endgame besides nine FAI programmers in a basement. (I had told Roko that right now, we have somewhere between zero and two.) Who is SIAI?—Yudkowsky, Vassar, Shulman, and Salamon. Mike and me are worthless, Justin's epistemic rationality is dubious.

The *Less Wrong* wiki is certainly worth doing; even if the Singularity is bunk, this business of human rationality is important. So I can work on *that*, at least, wholeheartedly, can't I?—while keeping an exit strategy in the background. And I have to learn to study. All over again. I've lost the knack, which I only barely glimpsed for a few months. Diary, oh Diary, what have I gotten myself into? —ZMD

ADDENDUM— The bad news is that it's pretty late and I haven't accomplished anything today. The good news is that this makes me feel sick, which means there is still yet some spark of morality within me. Anna doesn't get in until late tomorrow, with a nontrivial probability of it being later, being that her scheduled layover in Houston is only half an hour and these things have been known to go wrong. I did do a lot of rereading and thinking about existential risks today, but it's not at all clear what all this thinking has *bought* me. Diary, a human life is small: you can imagine someone spending years *worrying* the same worries, without getting anywhere.

Oh, Diary, the conversations here are so beautiful! And I have to be strong, to be able to strut my feminist colors with pride before Roko! Even if it means having my share in taking over or destroying the world? Better have a graceful exit strategy. Just in case, you know: plan it both ways. And tell Anna about it. —ZMD

422 — Friday 4 December 2009 — "Reliability Theory"

Diary, my entire life has been as dust and ashes. My internal narrative says that I'm unproductive because I'm depressed because I'm not sure that I want to be part of this world domination cult, this brilliant group of beautiful people who are just trying to do the right thing. But my actual *symptoms* are just the same as when I was at high school or University and I didn't want to do my homework. If having a thousandth of the weight of the world on my shoulders is what's got me down, you would *expect* it to make some empirically observable *difference*.

Yesterday Roko was saying that he wanted to write the popular book, and Anna didn't think it was such a good idea, that it probably wouldn't get done anyway and that there were better ways to invest Roko's time. I pointed out that there's a difference between writing *a* book, and writing a *good* book that would actually sell: by some standards, I had written a book myself just this November. And Anna said that even this was an accomplishment, that one of us setting out to do such a thing had a substantial probability of failing to do so. Heartening—and O Diary, where do I belong?

I went on a walk with Anna in the night. I said more about my deontological intuitions against taking over the world. Anna pointed out that if seven billion lives plus all of those to come hang in the balance, and I'm one of the few who understand the relevant issues—don't I have a deontological duty to *do* something to save them? She's confident that we're doing the right thing—she had her doubts, once, when she heard about value divergence, but she does not suffer the motivational instability that I do.

How does one even think about Singularity-level problems?—they're too *big*, it's too *much*, how can anyone bear it, let alone solve the problems? It almost seems safer to lay down and die than dare to move the universe. But you have to admit—when you have a fast, scalable or copyable, greater-than-human general intelligence lying around, that *changes everything*. Why fear an AI singleton more than an ordinary world government?—the AI is going to be more *competent*. Do I damn John Holt for his youthful activism?

There are plans to accept more fellows on a rolling basis; Anna is interviewing Hannah "Alicorn" Finley today. My primate nature is highlighted. Wouldn't it be wonderful to have Hannah Finley here? I *love* these beautiful, brilliant people, with whom I can have such interesting conversations! It's a different world, here. Only it's worrisome—it's *not* about saving the Earth, for me, it has *never* been about saving the Earth, for me. Anna is serious about risk reduction, but—*causally*, I'm here for the community. Knowing this, I have to adjust—I have to make a conscious effort to concern myself with the fate of the world. People manage to do all sorts of awful things for the sake of *their community*.

Anna and I agree that I should stay around and make myself useful for a few more months and learn about the world, which will help me with whatever I want to do. I need to take my rationalist's Art to still a higher level:

to figure out what I want to *do*, and cut through to that thing in the quickest and most direct way possible. Though I may have so many doubts of the rightness of our mission, I can't exactly bring myself to wish that I had never met these beautiful people. Even if it's a terrible blotch on my life record, to have spent some time in this community only to *leave*, and go off into the wider world, would I have done *better* without them? Even if my life sucks right now, it's such a rich and rarefied suckiness, with such friends!—O such friends approaching general intelligence.

We have the term *psychoticism* from Michael Vassar, here taken to mean the tendency to act on verbally professed beliefs. For most people, this would be a *bad* thing: their explicit beliefs are so wrong that they're better off with conformism.

[...]

I told Kaj that I would email him back on Saturday night, rather than tonight, but what's the use of delaying if I'm just going to waste all of today, as well?

We've still got time. No one's trying to build an AI right *now*. Right now we're just laying the groundwork for a science of general intelligence. I expressed bafflement to Anna as to why all this stuff was only being thought of *now*. Anna mentioned as an enabling factor that we're the first generation to really grow up with home computers. I hope AGI is far off enough that our children and grandchildren will be the ones to solve this problem; I can't imagine winning with the forces I currently see around me. If I'm not strong enough to think myself, not brave enough to stare into the Singularity without turning away and covering my eyes, then maybe I *can* play my small part in laying this groundwork for the beautiful children who are not yet too old to learn.

It's not enough for me to whine and mope and write to you, Diary—we have such a narrow fragment of time while this world lasts—one way or the other—and I should want to *do* things with my precious fragment, this little scrap of life. If I decide I want to leave, and make no mention of my time in Santa Clara with my creepy Singularitarian friends, I'm *still* coming out ahead—even if my CV is bare, what award or position would I wish to trade for these friends and this capital?

Rationalists should *win*. I have got to get myself out of this *gutter*. I want to do things and know things, to the extent that I almost don't care what. If I want to leave my beautiful friends, the way to do it is not to sabotage my performance until they kick me out. The way to do it is to be a superstar, to push myself to peak performance and leave under my own power, and leave strong enough such that someone else will be willing to take me. I need to learn more maths, I need to learn programming—I need to get out of this gutter. Later today I think I'll ask Blume to poke me every hour on the hour. As much as I would ordinarily pride myself on my self-regulatory ability, it's become clear that whatever I built up in this summer of glorious sanity is *gone*. *Kaput*. It's been so long since I even tried to write an academic paper—my life is as dust and ashes.

But now I know. —ZMD

ADDENDUM— I feel better now. I started taking data and images for the show-and-tell presentation that I want to give at the next Quantified Self meetup, on Monday. Just this *doing something* other than moping about reading stuff on the internet makes me feel less miserable, gets me *somewhat* out of this gutter. Maybe I was *wrong* to have joined these beautiful people, when they're out to save the world by taking control, and I can't seem to muster up the same enthusiasm when my libertarian moral intuitions would fight me at every turn. "Ayn Rand as a Positive and Negative Factor in Global Risk," I said, and it was a joke but it was true. But I can stick around, making myself useful and completing a few projects—at the *least*, this paper with Kaj is worth doing, and the *Less Wrong* wiki deserves to be in a presentable state. And then I can *leave* and go earn money and donate a large fraction of it to some high-impact cause: if not SIAI, then FHI, or whatever Toby Ord thinks is best. The thought would not have occurred to me before my contact with this beautiful community, but rational altruism has this tendency to *rub off* on folks. And knowing now that this *doing stuff* business is actually helpful in combating my depression—if I can nurture this insight, pressing it up against my chest—I channel it into my fiction, maybe even show up Alicorn. I proved *something* with NaNoWriMo, even if I can't say what it was. I've revisited the sadness that haunted me through my time at University, now it's time to optimize. You know, I had been buying those canned Starbucks coffee drinks at the 7-Eleven, and then it turns out that the machine-dispensed store brand is of similar quality for cheaper. Who knew? I should have. —ZMD

471 — Monday 5 April 2010 — "An Existing Mistake"

Counterexamples in Analysis wasn't worth those ten-or-so dollars. What folly to trust the impulse of the splendid moment!—when I can't get the moment to last, or even do anything useful with it while it *does* last.

I ended up visiting my creepy Singularitarian friends on Saturday, and this time I was smart enough to bring an audiobook for the drive (compare 336; best practices should not be so intermittent). Conversation including Michael Vassar was fantastic as usual. Apparently the fanfic Eliezer had written under another name got leaked onto *Less Wrong*; you can just imagine how the world of *Harry Potter* must be such great fodder for a dark and meaningful FAI allegory.

Hannah hadn't appreciated my clue, so I had printed a substantially easier one for her; apparently she doesn't mind spoilers and she doesn't like puzzles. I went to the mall with her and Mike. I worried a little that Hannah might take offense at what I've written to you about her—not that I was mean or uncharitable, but that I was never in love with her even in the way of friends—but then, I reasoned, surely the fact that I've entrusted my *Diary* to her is proof that we're friends enough.

I've claimed more than once that if we win, I would rather have a design-a-person voucher than to personally live forever. I know too much to buy the naïve conception of personal identity, and honestly, humans are cheap: in certain circumstances it *must* make sense to start over, rather than try to repair an existing mistake. Hannah asked in challenge if I would miss her if she died, and between you and me I don't think it was a fair question: there are some questions you're not *allowed* to say "No" to, so saying "Yes" doesn't buy very many bits.

What's the real answer, you ask?—I honestly don't know. No one close to me has died before, and I don't know how to simulate how I would really react. I guess the outside view is appropriate here, and I should *anticipate* feeling deep grief, even if I presently have no referent for it. Compare how when I was much younger, I did not understand sexual feelings, or why men cared about breasts, and yet I would not have done better to predict that I would never know. (The autogynephilia, in contrast, should have been more of a surprise; I only wish I had kept records then.)

Folks walked to the park, where antics included play with a novelty flying disc. I walked back with Anna; we talked about complex analysis. Mike showed me the project he's been working on under Rolf; I gripped his arm and my eyes glittered with epistemic greed.

When I got home, I finally sent off that spreadsheet (compare 417, 446, 449), saying (as I had said at the park) that I would write off the sum as a *de facto* donation to SingInst. It makes sense; it may be some time before my income situation is such that I'm in a position to contribute more.

I'll have to visit again on the coming Saturday because I left my power cord and want it to do my taxes.

Maybe I learned a little about astronomy and linear fractional transformations yesterday, but on balance I have been a bad person. One blogger says that elite ITfolk aren't feeling the recession; *Gene Expression* points to *FiveThirtyEight* pointing out that we're essentially at full employment for those with a Bachelor's degree. At this point I basically drop pretense and admit that I just want revenge. But do I deserve it? I might read Eliezer's fanfic in the hopes that it will spark my ambition again. Or maybe I'll just read Eliezer's fanfic, full stop. —ZMD

ADDENDUM—"Victory Conditions"—I had thought Eliezer had written the fanfic under another name, but it turns out he used an account called "Less Wrong"; it is not at all unobvious. *Diary*, it was *amazing*; I can't think of anyone else who makes me feel this way except for Greg Egan (sometimes). I even have hope for my own fiction ideas. I noticed a similarity between the way he wrote Hermione and how I wanted to paint my protagonist in the days before her own purity born of pain. But O *Diary*!—not to waste the splendid moment! Yudkowsky bought me a few moments of awareness, but what do I do now? How to I move, how do I react, how do I carry myself with the knowledge that a few more hours of breathing will in all reasoned probability drop me back into my equilibrium of living death? Am I supposed to order that book on compiler theory in order to make myself feel smart? And if it's already started before I finish writing to you? —ZMD

SECOND ADDENDUM—"Ratio Test"—I went to Safeway and bought an iced coffee, a donut, some liquid soap, and a card for Mother's birthday. Incidentally, I've come up with a new measure of life success: the ratio of novels to coronaries, counted with multiplicity across the many worlds. For all that *we know here* ("here"), my score is yet undefined: zero over zero. *Diary*, at what odds would you bet (I almost asked how *much* you would bet before I remembered that *Diaries* are as a racial characteristic risk-neutral) that I add to the numerator and in this way keep it undefined, before adding to the denominator and simply bringing it to zero?

Mike still has to at least pass a class before he gets his MA (why the University decided to award a Master of *Arts* in *physics*, neither of us can say), and when I asked, I think he said that he doesn't intend to study more physics afterwards. This is where either I have it over him in ambition, or he has it over me in honesty—and probably both. Why *stop* just because they gave you a piece of paper?—because studying would interfere with browsing reddit? This is the nature of my dishonesty: I'm not willing to admit that I need my analogues-of-reddit. Or I think that I can have everything, that somehow sixteen-hour programming days, in defiance of all logic and

reason, *don't* trade off against simple pleasures like watching *Family Guy* on the big teevee just because you felt like it. Merely naming my dishonesty changes nothing—as if I am splayed out before you, ignoring the knife, murmuring with perfect piety that it is raining, but that I don't believe that it is.

David Chalmers has a preprint out on the intelligence explosion; Anna had told me that Carl will be going to Australia to collaborate with him.

I never wrote back to my contact at OfficeTeam (Kim Wetzig, "the Other Kim") who said that she needed to know how long I had stayed at the *reply.com* project so that I could get paid. I don't think I intend to write back now—it's so late that I think it looks even flakier to write now than to let the matter drop altogether, and the extent of my spoiledness is such that the illusion that I still have any dignity left takes precedence over my financial self-interest. A day's wages could buy a textbook (perhaps on compiler theory), but I've gotten so used to spending Visa's money rather than my own that I don't have the mental or sensory acuity to notice or care. The upshot of all this is that if in all apparent honesty I told you that it *were* raining, you would not, could not, *should* not believe me. I am an existing mistake; the reason you shouldn't kill me on the spot is that humans aren't *that* cheap, so you don't have anything better on hand to replace me with.

That will change. It is—going to be the future soon. And I for one can't fucking wait. I remain,

Full stop,
Zack M. Davis

THIRD ADDENDUM—" 'And the Methods' "—I've done a little bit of transcription out of my first notebook, started back in 2005, picking up from the transcription bookmark left in it from however many months and months ago. To date, I've filled twenty-one notebooks (of varying size, I don't think the model of Moleskine I've taken to has very many pages per book) and transcribing the good bits to the desktop is a subgoal of finding out if I really do have a novel in me, or (this is the logical, *inclusive* or) if I've been systematically lying to myself all these years.

The *Less Wrong* crowd knows that Eliezer's fanfic is very autobiographical. I had insisted to Anna that sometimes the point of writing fiction is so that you can write about themes you know well, but in a way such that it's not *about you*. I explained that I had felt unease of late because there was a story that I wanted to write, but that I was embarrassed about it because it drew on my own neuroses.

Some comments (most of them don't really deserve to be called *reviews*) on *fanfiction.net* confirm my previous thoughts to the effect that people who don't know you personally are likely to just accept the text as a text, as they should. I think more than one commenter made explicit note of how we are told that Harry didn't play *Dungeons & Dragons* but liked reading the rulebooks. To the commenters, I guess this was a charming detail inserted into the story by a creative author, but because I'd devoured the Yudkowskian canon (which included an *Overcoming Bias* post mourning the death of the game's creator), I knew that Eliezer was cannibalizing his own childhood. (Or I thought I knew; I've already cheated on my hiatus enough such that I will not do so again just in order to confirm this detail in the archives.)

So if the protagonist of my novel has a self-identity drama about initials, the people I know *now* might draw the connection, but my greater readership won't (assuming my greater readership extends beyond the people I know now). —ZMD

475 — Saturday 10 April 2010 — " 'Incidentally, My Childhood Dislike of Asymmetry Led Me to Invent the Thue-Morse Sequence' "

O Diary O glorious for being morally awake O glorious for having intelligent friends! Anna was gone. I explained (rationalized?—half-joked about?) my gustatory cretinism in an appeal to fun theory: insofar as one values individuality, one must broaden one's horizons in some directions more than others in order to avoid one's person being engulfed by the Library of Babel. I noted that apparently there are guys who desperately want to bang "ten"s and are disappointed that they can only manage to bang "six"es, or whatever. I explained that I find this absolutely bizarre, given that there exist reliable methods of achieving orgasm alone. (I think that got a big laugh, but I meant it.) So come the Singularity, those guys will end up practicing transhuman superpickup and having transhuman supersex, whereas I'll be off in the corner studying transhuman supermath. Such is life.

Our witty banter is like the hybrid-vigor offspring of *Gilmore Girls* and *xkcd*. Hannah "A." Finley was happy about passing some high, round karma threshold, which Marcello noted was the chance result of our species happening to evolve with ten fingers. He suggested cutting off a finger. I told Marcello he might write another Greasemonkey extension, to display karma scores in octal. Mike pointed out that you wouldn't even need to alter genes to create a lineage of humans with nine fingers—consider the male foreskin! I said (I think just before going

to the restroom) that I must have been surprisingly old when I learned that there *was* such a thing as a foreskin. After emerging from the restroom, I asked if there wasn't a Kurt Vonnegut novel about a base nine counting system taking over and converting everything it came across, and I was not wrong to have thusly gambled on someone replying: I think that was *ice-nine*.

We got lunch at an Indian buffet; Mike and Hannah took the car while the rest of us walked. I had already spoken of my breakthrough on the generalized rotation problem, but we talked about it more while on the walk. Marcello said there's a simpler proof, which I guessed was a sort of generalization of the intuition pump I had known about for the three-dimensional case, which I had not trusted my own intuition to generalize. Marcello recast the principle in terms of what the $n-1$ -dimensional case implied about the n -dimensional case. I said that my intuition disagreed, for wouldn't that mean that would could do induction and prove something silly? Marcello explained that induction didn't apply, because the subject of his conditional was another conditional. That's obnoxious, I said. And beautiful, I tacked on, realizing a beat too late the manifest sinfulness of my first reaction. I tried to make a joke, saying that my reaction was a conditional, where *if*—but I couldn't think of a suitable implication in realtime, so I converted my failed joke into a metajoke by following my verbal stumble with the claim that I would have to study for another year before I would be capable of telling that joke.

You had to be there.

I told Mike and Hannah about how glorious it was to be morally awake, because I have to know all the math and all the science, so I can be a machine learning programmer and have one of the last jobs before the Singularity, making the machines that take everyone else's jobs.

I talked to Steve about the Dirichlet problem, and he showed me a version of the conditional probability game he had apparently made with Will (fka Corey).

The audiobook I have running on such car trips is *1984* (which somehow I had never read), and it's really good, although some of the themes clash with my sense of life in an interesting way that I won't tell you about now. I bought soap and candy at the drugstore, and masturbated after getting home. As still another treat, it turned out that *Harry Potter and the Methods of Rationality* had grown by a chapter: it ends in a cliffhanger that will be resolved tomorrow.

So, yes, it's been a very good day. But O glorious O Diary how I must write, O Diary O glorious how I must code! —ZMD