# SquirrelInHell

2017-12-20

## Happiness Is a Chore

I've literally run out of cynicism.

Not so long ago, one of my long-term strategic projects felt especially close to my heart. The project could be summarized as something like "give happiness to people who already have rationality as a prerequisite".

In my rampaging naivety, I had thought the hard part was to figure out happiness. What *was* I thinking? To find out, follow this short imaginary conversation with my old self.

**Me!2016**: Oh. Hello?

**Current Me**: Hi, me from 2016. I have a question for you. You know this thing that we care about, where people become more efficient/rational/smart/well-off, and yet do not become substantially happier as a result?

**Me!2016**: Of course. I am a little bit confused about how people aren't working on this, despite it being obvious that it's an important problem. I guess this must be because it's very difficult to actually change how happy people are. There are, after all, well known psychological effects that prevent this, like the "hedonic treadmill".

**Current Me**: Erm. Let's leave this part for now. I have a question concerning a different aspect of the issue, and I'm curious what you think about it. Assume for a moment that I already figured out a method for people to achieve stable, safe, instrumentally useful, and viscerally satisfying happiness. Assume additionally that the method required only a moderate amount of effort, such that it's within reach of most healthy, functional human beings. For the sake of concreteness, put it somewhere between learning to swim as an adult, and learning a foreign language. Would you consider this a complete solution? What else could go wrong?

**Me!2016**: Give me a second, these are pretty strong counterfactuals. I need some time to adjust to thinking from what seems to me like a very small section of probability mass. You want me to assume that it's actually true that you know such a method, and that it's every bit as good as you described it. Hmm, even in such a case, I might still have trouble *believing* what you say. Especially believing it enough to go through with some significant effort, which your method seems to require.

**Current Me**: I grant you that the effort is significant, but it's not *extraordinary*. And the benefits kind of *are*, you have to admit. But let's not talk about that. What if I told you that convincing people my method is real, and that it works largely as advertised, turned out to *not* be the main issue, in this *hypothetical* scenario?

**Me!2016**: Ouch. You just keep on piling these unintuitive assumptions, taking me further and further away from my priors. I'm not entirely sure my current common sense reasoning would apply in a world in which all these things were true. I feel like in that world I'd need to seriously recalibrate my brain, until it would start *expecting* this sort of weirdness. Although I have no idea how I could do that *coherently*.

**Current Me**: I understand. But please make an effort anyway. I want to hear your best guess.

**Me!2016**: OK... Let's see. Assume I know such a miraculous-sounding method for people to become happy, and it's a real working thing, no cheating. I also have no trouble convincing people that it's real, and worth the effort. Aha! I know. Is your method *teachable*?

**More by SquirrelInHell**

- AI Safety Comics
- Android Apps
- Be Well Tuned
- Rationality Updates

**Current Me**: It's a little bit tricky, because it's purely a mental skill. For most people, this makes the learning process more difficult than in case of more tangible skills like swimming. From the teacher's perspective, it's also harder to watch over a student's progress, or give hints when they make mistakes. But even considering all of this, it's probably not as bad as you imagined it just now. It's definitely teachable, and in fact various more-or-less contorted versions of it have already been taught in this world, with reasonable success, and to not-very-strongly-preselected groups of people. Is this satisfying?

**Me!2016**: Ah, so you admit that there is some difficulty with teaching. But the way you describe it, it doesn't sound like deal-breaker level difficulty to me. So you still want me to think what *else* could go wrong?

**Current Me**: Yes.

**Me!2016**: No, honestly, I give up. Assume the method to be happy is real, safe, clearly beneficial, and teachable. Assume people believe me that it has these properties, when I tell them about it. I think that yes, this would be a sufficient solution to the problem. I'd start by teaching it to a few close friends. Soon enough they would teach *their* friends... Oops! Say, *how long* does it take for students to learn, and for teachers to teach?

**Current Me**: Haha! Good catch. But no, that's also not the issue. Intelligent and motivated students should pick up enough in a month to reap most of the benefits, and be able to maintain them indefinitely with only, say, 1 day per week of ongoing investment. The instruction could be seen as pretty light work, though the way things are I'd say it tends to be an emotionally exhausting and thankless job.

**Me!2016**: Then I really have no idea.

**Current Me**: OK. I'm going to reveal the answer now. I wasn't just teasing you, I want you to know, though I'm worried that you don't have the ability to fully process the answer from your current point of view. But anyway. Consider this: what if I told you that people don't *actually* want to be happy?

**Me!2016**: Wait, you can't be serious. People clearly *do* pursue happiness, I see evidence of this all the time. That's what *I've* been doing for as many as 10 or 15 years. It's also a part of *your* memories, so certainly you understand.

**Current Me**: Ahaha. That's exactly the part which makes me feel like I've "run out of cynicism". See, the human activity you describe as "pursuing happiness", from my current perspective, seems to be in the same category as other common activities such as "acquiring education", "helping people", "talking to friends" (or should I say "talking" to "friends") and so on. Which is to say, people do them in a way which is outwardly convincing enough to allow everyone to keep up the social pretenses. This is *way* different from what you'd see people do if they actually cared. The simple matter of fact is that the human brain is a kludge, and people are puppets dancing on the strings of a mad puppetmaster. Almost anything they claim to be doing isn't for real. This is true even when they themselves know about this. The best you can do is gradually nudge yourself in the right direction, gaining new footholds in consistency and consequentialism painstakingly and precariously.

**Me!2016**: ... I might be having a failure of imagination here. I am hearing the words you say, and you seem to be saying them honestly. But I admit I cannot put myself in a frame of mind in which these words would seem like a natural thing to say.

**Current Me**: I'll tell you the thing that finally broke me, finally pushed me over the edge with grokking the human condition on a full System 1 level. To give you some background, fixing your epistemics has immense compounding interest. Put this sentence on your wall. I have become much better at recursively estimating my confidence, since the time when I was you. So if I made something work, and it is clear that it did in fact work, I have much less self-doubt in believing that reality did in fact turn out to be this way. By the way, you don't realize that you have that self-doubt, because you have much less of it that the average person. And this is the kind of self-doubt that puts a ceiling on how high your own estimate of your own mental skill can be above what you consider typical. So I'd call it a self-reinforcing blind spot, in CFAR

terms. Anyway. My models of improving happiness have enough gears, and I have seen these gears turn enough times, and I have enough justified meta confidence in my ability to build models. I have felt levels of happiness which are far above the upper limit of your mental scale. I know exactly how to be happy. And yet I find myself not consistently applying my own methods. Do you realize how impossibly mind-twisting this situation is? What happens in reality is that I enjoy and see great value in happiness when it happens, but when it doesn't I only work on it grudgingly. It's like with exercise, which is great but I'm rarely enthusiastic about starting it. The problem is not that I don't value happiness enough. The problem is rather that there is no gut-level motivational gradient to get actual happiness. There are gradients for all sorts of things which are crappy, fake substitutes. Once you know the taste of the real thing, they aren't fun at all. But you still end up optimizing for them, because that's what your brain does.

**Me!2016**: Er. Um. So. Do you mind if I go back to my timeline, to, er, consider what you just said?

**Current Me**: Sure. I'll see you around, I guess. Sorry about ranting too much.

**Me!2016**: *disappears*

* * *

*Update:* A lot of people are asking what exactly I mean by happiness in this post. Haha. It's not like I can define a subjective experience without getting into circular reasoning. But let's look at some examples. **They are not examples of what I'm talking about above, but merely of what I put in the broad reference class of "happiness"**.

Let's say you had been doing some really good work, and took a day off to go on a hike. At some point on the mountain path your mind clears, and you feel calm, content and energetic. There's a warm sensation spreading inside your chest as you look around and find out that you can take in extremely fine details of the landscape, and that its colors seem more vivid than ever. You feel proud of the life you'll come back to, and overall like you are in the right place in the world doing the right things. That's happiness.

Now imagine you've just gotten a new car which is cooler than your neighbor's, and which you kept telling yourself was the only thing preventing you from being as happy as him. On the next day the enthusiasm wears off, and you can feel the jealousy and dissatisfaction creeping in again. That's NOT happiness.

**6 comments:**

**Gunnar Zarncke** Wednesday, December 20, 2017
Maybe you should provide a link to your previous post http://squirrelinhell.blogspot.de/2017/11/the-little-dragon-is-dead.html for context.

Reply

**Romeo Stevens** Thursday, December 21, 2017
Excellent. Still have a sneaky to-be verb form belief hiding in there though. We care about happiness in the observer moments where being happier seems like winning. We care about other things in observer moments where other things seem like winning. Your pseudonym is once again appropriate. The gradient in hell does not slope downwards towards happiness, and the squirrel often doesn't see a good reason to go uphill.

Reply

**Norswap** Monday, December 25, 2017
So, does this happiness method exist and is there a primer to it?

I haven seen the neck-stretch thing, and the corresponding "tune your motor cortex" and while that seems very useful skill that I will try (I incidentally also suffer from excessive neck/shoulder tension), it's hardly what I would call a general method towards happiness.

Reply

**Norswap**  Monday, December 25, 2017

Actually, answers to be had here: https://www.lesserwrong.com/posts/3Hkpttb6WsJwr5WdF/happiness-is-a-chore#comments

Reply

**Eneasz**  Tuesday, December 26, 2017

What is the likelihood that this happiness method loses power over time? I had a method for achieving happiness that worked for about two years, but eventually it just sorta petered out.

Reply

> Replies
>
> **SquirrelInHell**     Wednesday, December 27, 2017
> Good question; I expect things to play out roughly like with well-known meditation traditions, which seem to be sustainable, but my confidence is low. Ask again in two years!

**Reply**