

Best of LessWrong: August 2017

1. [Play in Hard Mode](#)

Best of LessWrong: August 2017

1. [Play in Hard Mode](#)

Play in Hard Mode

Epistemic Status: Love the player, love the game

Also consider: [Playing on Easy Mode](#)

Raymond Arnold asked me, why do you insist on playing in hard mode?

Hard mode is harder. The reason to Play in Hard Mode is because it is the only known way to [become stronger](#), and to defend against Goodhart's Law.

Strategies that work in Easy Mode won't work in Hard Mode.

The key idea of Hard Mode is to keep your eyes on the prize. You know exactly what you want. You can't munchkin your way to getting it. Once you start aiming to make a number go up, or get a check in the right box, you have lost sight of the thing you actually want. Proxy measures lead to failure; your [value is fragile](#). That number correlates to what you want, but only insofar as you're aiming for the goal and not the number. If you break the spirit of the exercise, all is lost. Your values have been hijacked. If you fail to develop skills along the way, you have missed the point, because the game has no end.

I

Consider playing guitar in Rock Band. You must choose whether to play in Hard Mode. If you do, you will fail a lot. You will play the same songs over and over again. Tricks that rely on there only being so many notes, or going at a relaxed pace, collapse. Eventually, you learn new techniques. You get better. You play on expert, your fingers get sore and you smile as you sing along.

II

You have a test in a week. You ignore it. You've asked questions based on your curiosity, to resolve your confusion. You study what is interesting to you, and what you feel would help you in the future. You focus on learning key principles, knowing you can derive what details you need later on. When the test comes, you work to figure out the answers. When you get the test back, you know how much you have learned. A year later, you remember everything, and build upon it.

III

You prepare for a tournament. You seek out the toughest opponents to help you prepare. You stop to criticize each other's technique and point out every little mistake, no matter how irrelevant to the ultimate outcome of the practice match. You ask why and how you made that mistake. You do the same when you learn something new in a surprising way. You focus on the fundamentals, and don't worry too much about exactly who you are up against this week. During the matches, you remember every tough decision and every mistake, so you can train again next week.

IV

You start a website writing articles devoted to the things you care about. To monetize it, you sell advertising through Google. It does not pay much at first. You keep at it,

attracting a small but devoted readership. Some were already your friends, others soon join them. You look at what resonates so you can get feedback, but are careful not to take actions designed to maximize page views. Over time your writing improves and you learn much together. A community of sorts arises. You don't quit your day job, but you teach others what you have learned.

V

(Spoilers for the excellent [Groundhog Day](#))

You are stuck in a small snowed-in town, caught in a time loop of unknown origin. At first you have fun doing absurd things, but then you buckle down. With unlimited time, you decide to develop the skills and knowledge to give everyone a perfect day. You learn to play the piano, you read great literature. You listen to and remember the stories of everyone in town, and grow fond of them, learning what your opportunities are to engage in small acts of kindness. At the end of the day, after sufficient iterations*, you know you will be proud of your accomplishments, because you've made yourself and the world better, and you just might impress the hell out of your crush. If the loop continues*, you can do it again.

* - Results not guaranteed. You are unlikely to be in a movie. Local maxima may or may not be sufficient.

Interlude!

VI

Your help your friends move. With time and practice, your group of friends gets quite good and reliable. Any time, day or night, if someone needs to relocate, you'll all be there, no questions asked. You call yourself the Midnight Movers. Most of your stuff arrives safely at its destination with a minimum of fuss, and you order everyone pizza. Your group draws closer together, and eventually tries going into business together on an unrelated matter.

VII

(Medium spoilers for [The Good Place](#), highly recommended, skip this if you haven't seen it yet)

You have an idea for a television show about a group of strangers who arrive in a mysterious place that plays by very different rules than our reality. You figure out exactly how this place works, plot everything meticulously, and lay out mysteries for the characters and viewers to uncover slowly over time. You use flashbacks that parallel events to examine and deepen the characters. Your production values are top notch and you produce great television. Your show is not a smash hit, and you know exactly where you are going with all this, so you don't waste a minute, keeping your seasons short. In the end it all fits together, and the journey was still pretty great on second viewing.

VIII

Despite the terrible odds against you, you decide to strike out and open a fine Italian restaurant. Your dishes are sublime, but you soon learn that is but a small part of a

successful enterprise. You must hire quality staff, arrange logistics across many suppliers, draw in customers and much more. Each step of the way, while ruthlessly keeping costs in check, you answer the question of what you would want your place to be like, and evolve your menu to offer a small selection of the best things you can affordably make. You get to know your customers by name. A casual observer would think you almost dislike money, as you can barely tell from the outside that there is a dining establishment there at all. Slowly word spreads among the cognoscenti, who show up and order the wine. Business is good enough. You get to keep working on perfecting your art.

IX

You have something to prove. To yourself.

Refuse to hire a cleaning service, ever, even though it's totally worth it.

[Minimal applause lights.](#)

Tell the job interviewer your true strengths and weaknesses. If they don't hire you, you didn't want the job. Keep looking.

Find people to come to your meetup by promising them interesting intellectual discussions and a community devoted to truth.

Refuse to pirate music, television, movies, software, even when the owners are being kind of a dick and won't sell it to you.

When you are in power, respect and minority even when you have the votes, don't change the rules to pass the laws you want. Strengthen free speech rules and don't silence those you disagree with, lest they do the same to you.

Write carefully as you fill out the forms. They might look like bureaucratic nonsense, and no one is likely to ever read them, but you should get this straight and cultivate good habits.

Learn to speak the language.

At your meetup, welcome challenges to in-group principles, so your group will be viewed better and feel more welcoming to and attract more members who seek the truth more than they demonstrate membership in the in-group.

Show other people what to do by example.

For demo day, you show what your system can do, and hope that you can keep building it. For that you need funding.

X

Dismayed by terrible things, you devote your life to the promise of artificial general intelligence. You discover that contrary to your initial beliefs, not only is creating AGI not easy, most versions of it kill everyone and destroy all utility in the universe. Explaining this is super hard. None of your explanations work. No one understands the danger. You set out to teach the world rationality, hoping this will cause them to see the potential dangers, with limited success. You write a book that's silly but gets you exposure. You keep writing. Machine learning accomplishes more things and starts to

get more funding. People start to come around to AI being dangerous, but mostly for the wrong reasons, so you don't expect anyone to take the right precautions, and fear the world is doomed. You think that when they arrive, it will be far too late to correct for safety problems later.

XI

You are at a meeting to arrange educational services for your son. You know that the only thing that matters is what is written on the education plan. Whatever is in that document is what will count. Still, you cannot let the numerous falsehoods and stupid things pass, even though you realize that if you just play nice, they are going to put down on the piece of paper the thing that you want on the piece of paper. If you keep arguing, you risk getting nothing. Luckily, you think better of it in time. You go into [easy mode](#). They write the words you need on the piece of paper. You sign it. You walk away happy.

[In conclusion...](#)