A Memex On The Meaning of Life

Marc Pickett I of Padelford

CONTENTS

Part 1. Theory 9
Chapter 1. Happiness of Rabbits: A Thought Experiment About Evolution 11
Chapter 2. Walking = Falling + Catching 15
2.1. Closet Space 17
Chapter 3. Enter Complexity: The Peacock's Tail 19
Chapter 4. Will: Interaction of Cognitive System and Reward System 23
Chapter 5. Squiggly Lines 27
Chapter 6. Taste is like the Butt, Divided 29
6.1. Specifying Reward: The Bulldog Anal Rapin' Robot© 33
6.2. Hill Climbing on the Cingulate Cortex 34
6.3. Some Strategies for Maximizing Reward 35
Chapter 7. Artificial Sweetener, Drugs, and Virtual Reality 37
7.1. Habits and Pain 40
7.2. Hedonism 41
Chapter 8. Economics and Ethics 43
Chapter 9. Atomic Waves and Platonic Forms 47
Chapter 10. Designing your Successor 51
Chapter 11. Entropy: A Petri Dish Falling into a Volcano 55
11.1. Nihilism 57
Chapter 12. Consciousness and Death (How a person never dies) 59
12.1. Free Will 66
Part 2. Applications 67
Chapter 13. Boundless Will and Finiteness of Life 69
Chapter 14. Conclusions 81
Appendices 85
Chapter A. Why Religion isn't the Answer 87
A.1. Are Religion and Science equally justified? 88
A.2. Proof of God's Non-existence 89
Chapter B. Addenda 91

WHAT IS A MEMEX?

A *memex* is a collection of ideas or *memes*. I've grouped the memes into chapters and sections, but the memes are largely independent, so their order isn't as important as the ideas in an essay might be. The memes' independence makes it relatively easy to insert or remove them from a section. This makes a memex easily mutable. So a memex, more so than an essay, is a living document, which makes it ideal for discussion and refinement of ideas.

The meme numbers are roughly in chronological order of when I either conceived or fully elaborated on the idea. Some meme numbers are missing because they've been superseded, they were wrong, or simply not worth including. (Most memes are actually in other memices.) The meme numbers aren't dynamic (i.e., I don't have a linear ordering of them) because the memices are intended to be living documents. So the meme numbers remain static so that ideas may be referred to quickly.

SUMMARY

This memex is a collection of insights gained from my work in Developmental Robotics, Artificial Intelligence, and Reinforcement Learning, as they pertain to the age-old question "What is the meaning of life?". I don't claim to have all the answers, but I hope that this memex will help us get a little closer to answering the question.

I've divided this memex into a part on *Theory*, which presents a collection of "intuition-pumps" and thought-tools, and a part on *Applications*, which presents a set of problems that might be addressed using these tools.

Happiness of Rabbits: A Thought Experiment About Evolution: We can gain some insights by looking at life from a "designer's" point of view. The notion of *happiness* or a Will might be useful from this point of view.

Walking = Falling + Catching: Permanent happiness can never be achieved, or at least that it'd be a bad from a designer's point of view. This chapter goes more into some mechanisms a designer might want to put into an organism. This is then used to explain why we can never have enough **Closet Space**.

Enter Complexity: The Peacock's Tail: Life isn't so straightforward for some evolutionary organisms. This is due to a recursive hall-of-mirrors that produces virtually boundless complexity.

Will: Interaction of Cognitive System and Reward System: We have 2 systems in our head, which might result in 2 different fundamental meanings of life.

Squiggly Lines: Perception is non-trivial, and this means that the designer's meaning of life doesn't always agree with the organism's.

Taste is like the Butt, Divided: Learning can affect our preferences. Also, **Specifying Reward: The Bulldog Anal Rapin' Robot©** expands on the Squiggly Lines idea and that specifying an innate Will is tricky and hacked. Perhaps one of the most impoortant sections to understanding happiness is **Hill Climbing on the Cingulate Cortex**, which further expands on the Bulldog A.R.R. idea to show that finding what makes us happy is essentially a "hillclimbing" search. **Some Strategies for Maximizing Reward** further details how that search might work.

Artificial Sweetener, Drugs, and Virtual Reality: We can exploit the "bugs" in our design. The section on **Habits and Pain** explains why habits are necessary and why pain is a good thing. Also, why **Hedonism** is inadequate as an approach to finding a meaning of life.

Economics and Ethics: Ethics emerges from utility maximization in multi-player games, and this is another source of virtually boundless complexity.

Atomic Waves and Platonic Forms: People are really just waves of atoms, the meaning of life is an abstraction, and there might be an "ideal" meaning of life.

Designing your Successor: The notion of reproduction only makes sense if the "self" is well defined. With genetic engineering and artificial intelligence, the idea of a self becomes blurry.

Entropy: A Petri Dish Falling into a Volcano: For any organism that is forced to make decisions (including people), nihilism is impossible. This is true even if the entire universe is inevitably doomed to fizzle out due to entropy.

Consciousness and Death (How a person never dies): Our own death is fundamentally different from that of others. Also, how a deterministic system, such as a chess playing program, can have **Free Will**.

The chapter on **Boundless Will and Finiteness of Life** presents a set of problems that I still haven't resolved about how to live my own life: How does one live optimally? This is the least organized chapter because it's the least conclusive. The questions are still open, so this chapter should be the most alive, or equivalently, the least set in stone. One conclusion of the chapter is that laziness and fear make sense from an evolutionary perspective, but are now overly influential.

Part 1

Theory

CHAPTER 1

HAPPINESS OF RABBITS: A THOUGHT EXPERIMENT ABOUT EVOLUTION

MEME 1

The biological meaning of life is reproduction...

If it were that simple, this memex would be rather short, but there's some truth to the statement: The only organisms that are around today are those whose ancestors reproduced, and it's likely that today's organisms have inherited their ancestors' tendency to reproduce. But you're right to be skeptical that a single word, *reproduction*, can answer such a often-pondered question as "What is the meaning of life?". The complexity of the meaning of life lies in describing the intricacies that are entailed by the term "reproduction".

MEME 2

Rabbits are non-native to Australia, but now there are hundreds of millions, all descended from a few dozen brought over from Europe. But suppose we were able to go back in time before rabbits were introduced to that continent, and suppose we played a game (called "Hare Wars[1]") where competing players design artificial rabbits, and we drop a small population of each player's rabbits in Australia, then see which player's rabbit population is higher after a few centuries. To constrain things, we'll make it so that the players are given identical rabbit *body* designs, and they're only allowed to design the *brains* of the rabbits. The players can make the rabbits' brains as powerful as they want, but a bigger brain will consume more energy (that could be used for other activities like running away from predators) than a smaller brain[2].

Once the rabbits are dropped in Australia, the players won't be allowed to change the rabbits' design. The players' rabbits should be adaptable because Australia has a diverse environment, and no rabbit design will be optimal for all environments. Because of the rabbits' computational limitations, there will be rabbits that are "buggy" or suboptimal. For example, from a computational-perception point of view, telling the front end of a rabbit from the back is non-trivial, and we might end up having rabbits trying to mate with other rabbits by mounting the wrong end (I've seen this happen, the rabbits were so eager to mate, they didn't seem to pay attention to the gender of the other rabbit, whether the other rabbit was a close relative, or whether they were even mounting the back end of the other rabbit)[3]. But this is OK from an evolutionary standpoint, because the extra computation might not be worth the brain matter needed to do it, and these bunnies will be at the correct end half the time, which is good enough to reproduce.

[1] This is similar to the old computer game called Core Wars, in which players write competing programs to take over the memory of a computer.

[2] A big brain might not always be selected by evolution. For one, big brains require a lot of energy to run. Our brains consume around 20% of our body's energy, despite being only about 2% of our body weight. Even in modern people, who are in a "cognitive niche", there's a negative correlation between IQ and birth rates. Also, because of our big brains, childbirth has been especially treacherous for humans. Thus, the width of women's hips has been a somewhat literal bottleneck for our intelligence.

[3] This same perspective can also help explain homosexuality. From a computational point of view, it's pretty hard to tell the difference between men and women.

MEME 3

To answer questions such as "What makes me happy?" or "Why does x make me happy?", we can gain some insight by stepping outside the humanities and go down to biology, economics, and cognitive science. A lot of what makes people happy can be explained in terms of these fields: e.g., men might enjoy sex with lots of different women because the type of people who did so had more kids and are more likely to be around.

I'm sure there are plenty of rabbits who mate like bunnies "because it feels good", and that rabbits have almost no concept of paternity. Despite the short gestation period, they probably don't even realize that sex causes babies. (It seems that some people still don't.)

We can design our rabbits and talk about optimality from *the game's* point of view without looking at the rabbits' point of view at all. Looking at the rabbits' point of view might lend insight onto the meaning of life: The rabbit's goal in life isn't necessarily *our* goal in the game. Imagine if I had a rational rabbit. I'd tell my rabbit "Be fruitful and multiply! That's why I created you.". To which the rational rabbit might respond, "Up yours, Marc! I (actually, the routines that *you* wrote) want to go mate with hedgehogs!". One other thing to note, this rabbit's defiance (which causes its very desire for freedom and "free will") was programmed by me too because it was successful evolutionarily.

MEME 4

There are important differences in the process we'd use to design our rabbits and the process of evolution. For example, evolution lacks foresight. For example, the vas deferens loops around the bladder in mammals, when a direct line would probably have been slightly more advantageous. As mammals became warm-blooded and their testes descended, the vas deferens had to follow the testes, resulting in the current design. Evolution is also slow and has "inertia", for example the myriad of evolutionary relics such as the hip bones of snakes.

Despite these differences, evolution often arrives at what might be called optimal solutions to problems. For example, the process of evolution arrived at the lens of our eyes, and, independently, the lens of the eyes of octopuses, which both have the same basic parabolic shape of lenses that people have designed for cameras and other optics. Therefore, some insights can be gained by considering what designs are successful even if we ignore the design process itself.

There's also the principle of The Selfish Gene [**4**]: our goal isn't really to make our rabbits take over the island, but to make our rabbits' *genes* to take over the island. Thus, we'd want to design our rabbits such that their behaviour sometimes might be bad for the reproductive success of an individual, but good for the individual's genes. A classic example would be where a rabbit sacrifices itself to a predator in order to save its offspring.

MEME 5

The reason that the goal of our game is to take over the island is that, in evolution, those beings that had a design that caused them to reproduce were the ones that did reproduce. So any animal living today (including humans) is here because its "design" is likely to cause it to reproduce (at least in the environment where it evolved).

MEME 1588

An important point here is that, in general, *every* part of our innate being, both physically and mentally, is designed as it is because that design traditionally helped our ancestors reproduce.

Likewise, *every* aspect of a plant is that way for a reason. The reason is evolution. Even the designs of the flowers, and the shape of the leaves have been constrained to be the way they are by some ruthless evolutionary process.

CHAPTER 2

WALKING = FALLING + CATCHING

Werd ich zum Augenblicke sagen:	If I ever say to the moment:
Verweile doch! du bist so schön!	Stay! You are so beautiful!
Dann magst du mich in Fesseln schlagen,	Then you may throw me into chains,
Dann will ich gern zugrunde gehn!	and I'll happily go to the abyss!

–Johann Wolfgang von Goethe (from Faust, 1808)

MEME 6

When roboticists first began to make robots that could walk, they focused on stability [**27**]. If a walking robot froze in its tracks it would be stable in the sense that it wouldn't fall down. Because of this stability, it could walk as slowly as you wanted it to. This is contrasted to the gallop of a horse. When a horse runs, there are instances when all 4 hooves are in the air at the same time. Thus, a horse can't gallop in "slow motion" because it can't be suspended with all its feet in the air at the same time.

The normal walking gait of people isn't stable either. During each step, our center of mass moves to its highest point when it is in front of the foot that's on the ground. So, we begin to *fall* forward, but our other foot rushes forward to *catch* us. We then raise our center of mass and "reset" the system for the next step. So walking is falling and catching, and you're never in a permanent stable state.

Likewise, one way of getting our rabbits to reproduce is to have them built such that they're always on the edge: that they always feel like if they do just this next goal, they'll be happy. When they do accomplish the next goal, they get some reward, but not permanent happiness. We don't want our rabbits ever to be content because content rabbits don't reproduce. So our rabbits will never actually attain fulfillment. To do so would mean that the rabbits stop striving for more and stop reproducing. On the other hand, the rabbits' reward structure should be such that they don't give up either, because that would mean an end of reproduction too.

MEME 7

If I were designing people to reproduce (or gain power or help their offspring to reproduce), I'd also structure their reward system such that they're always trying to attain something. Like a moving carrot, I might also make them believe that if they only achieve this or that goal, they'll be happy. When they finally catch the carrot, I'll give them momentary happiness, but I'd structure them such that this happiness would fade after some time and they'd devise a new goal.

MEME 8

People will always have problems. If a person runs out of problems to solve, that's a problem.
–Jason Pickett (1993)

In some video games "cheat codes" or "god" mode (where your character is invincible) makes the game boring quickly.

Likewise, suppose you found a genie who would grant as many wishes as you wanted. The wishes couldn't be contradictory or too poorly specified. For example, the wish for "permanent happiness" wouldn't be granted on the grounds that it's too vague. With this genie you could make it such that the world was at peace, famine and disease were gone, you had eternal life, money was virtually meaningless, you had all the women you could imagine, and your friends, family and power were limitless. But, would you ever be happy? Can you imagine a situation where you don't have any more wishes to make?

I suspect that there'd never be a *permanent* situation where a person was satisfied. I suspect that (via evolution) the human reward system is structured such that permanent happiness is impossible. For people (or any evolutionary being), it'd make more sense if happiness was the *event* of going to a better situation.[1]

∞

It's possible that we can only distract ourselves from the feeling that might be described as emptiness or disquiet, that the emptiness can't be banished, only put off. For example, in Anna Karenina, the character Levin is happiest when he's mowing hay with a scythe. It's a simple action, but he's making "progress" on something.

As humans, our Will is boundless. We'll always want to fall and catch. But "permanent" happiness is possible because our life is finite. So we only need to stay happy for 70 years or so, then we die.

[1]Recognizing that permanent fulfillment of desire is impossible, Buddhism teaches that a person should seek to free themselves from desire. On the face of it, this is a self-contradiction: a desire to have no desire. If the tendency to become Buddhist were inheritable, these tendencies would become weeded out of our rabbits. That is, a rabbit that had no desire to eat would soon starve.

2.1. Closet Space

> How much money is enough? Just a little bit more.
> –John D. Rockefeller (1839-1937)

MEME 9

The idea that happiness is never permanent is concretely manifested in my closet. I moved to a new place that had 4 times the closet space as my old bedroom. As you might have guessed, the new closet was full within a year. The same thing happens every time I get a new (bigger) hard drive or I find myself free from a weekly commitment. In the case of my hard drive, it's a big number of small things that fill it up. With a bigger hard drive, I'll be a little more lax about cleaning up files. With my closet, I'll keep shirts a little longer, or be less reluctant to get new shirts. In the case of time, I'll be less apt to say no to new commitments. The same principle goes for governments and taxes. No matter how big the budget is, it's easy to find ways to use all of it.

A problem with the tendency to fill up your closet is that you're constrained to moving to a new bedroom with a smaller closet. Likewise, if you never keep much money in savings, you'll be constrained if there's an emergency.

There are countless "tasks" that can fully consume a person: reading the entire works of Tolstoy, watching all episodes of the Simpsons, dusting the underside of every banister. Be careful which battles you pick. Every now and then, we might want to step back and evaluate whether what we're doing is the most important thing we can be doing.

MEME 1258

A chock-full closet is a bad thing. A chock-full closet is tight, which makes it difficult to take out and put in new clothes. So, if you want to go through your closet 1 shirt at a time to find out what you might want to throw out, it'd take significantly longer to do so if it were 100% full than if it were 90% full. This isn't just because the 100%-full closet has 10% more shirts. It's because looking at each shirt in the 100%-full might take twice as long, meaning it'd take 20% longer. So, if your closet is chock-full, you don't have much room to maneuver to free up closet space.

Likewise, if your daily schedule is completely full, you won't have much time to sit and think about how you can reduce your workload or how you can get out of certain ruts. So, if you're working 3 jobs to make ends meet, then you don't have a lot of time to sit back and try to figure out what you could cut to make ends meet.

Automobile traffic follows the same principle. If there are twice as many cars on the highway, a person's commute-time likely be significantly higher than twice as high.

In Computer Science, this principle is called "thrashing", where your computer is spending all its system resources trying to figure out how to free up its system resources.

The general principle here is that if your resources are completely consumed, you don't have any wiggle room, and you can't spend resources to free up more resources.

MEME 10

If a single pair of mice get into a granary and go unchecked, they'll reproduce until there are thousands and thousands of mice and all the grain is gone (at which point almost all of the mice will starve to death).

Any mouse with a long view who excersized some constraint in converting the grain into more mice would soon find himself outnumbered by mice who don't have this constraint.

The *unconstrained* mouse would find himself *locally* successful. By this, I mean the mouse will outperform (i.e., out-reproduce) any other single mouse that doesn't use all of its resources. The problem is that a *group* of mice that exercises some constraint in managing their resources can do better in the long run than a group of short-sited mice. Analagously, a group of goats on an island should be careful not to eat all the vegetation, lest they make the island barren of vegetation and therefore goats too.

MEME 11

Nature is an arms race, in a sense. For example, plants develop mechanisms, such as poisons, to prevent themselves from being eaten by animals, and animals develop means (such as enzymes) to get around those mechanisms [**1**]. Usually, each party is just barely in front or behind the other, and this prevents either from completely dominating the other (unlike the mice in the granary or the goats on the island).

For most of our evolution, it was a rare thing to have a virtually unlimited supply of fatty foods. Thus, because there was an *external* constraint on the amount of bacon we could eat (i.e., its limited supply) there was no need to have an internal mechanism that limited our intake of bacon.

MEME 1270

> Hofstadter's Law: It always takes longer than you expect, even when you take into account Hofstadter's Law.
>
> –Douglas Hofstadter [**8**]

Hofstadter's law is a corollary of closet space: any time that's allocated for a task is used up.

MEME 1593

One solution to the problem of a full closet is to get a bigger closet. This is usually the first solution I think of. Perhaps a better solution, though, would be to figure out which of my clothes are most important, and throw out some of the clothes of lesser importance. Unless I adopt this latter attitude, I'll simply need bigger and bigger closets.

For example, the first time I saw a computer with a "hard drive", I was amazed. This was in 1990, and it was 50 megabytes of disk space. I wondered why anyone would ever need so much memory. Until then, all my computer files were on diskettes, each holding half a megabyte. I had about 20 of them, 10 megabytes in all.

20 years later, my hard drive is now half a terabyte, which is *10,000* times as big as as the "virtually limitless" hard drive from 1990. And I've managed to fill it up.

∞

Similarly, if you figure out a way to get a little more money, such as borrowing off a credit card (for an individual), or issuing bonds (for a government), these tend to be pushed to their limits, like having extra hard drive space.

∞

This idea of consuming all available resources isn't new. In 1866, the English economist William Stanley Jevons noted this phenomenon. For example, James Watt introduced his steam engine in the latter half of the 18th, which made much more efficient use of coal than previously existing engines. Paradoxically this new efficiency caused the rate of coal consumption to *increase* rather than decrease. The new steam engine effectively made coal a cheaper power source, so cheap that steam engines began to replace oxen, sails, and other non-coal-fired power sources. So the total amount of coal consumption went up, even though the amount of coal needed for any particular task went down. This phenomenon –the introduction of a more efficient use of resources causing *more* of that resource to be consumed– is called the Jevons paradox [**10**].

CHAPTER 3

ENTER COMPLEXITY: THE PEACOCK'S TAIL

MEME 12

"Eat. Survive. Reproduce." It sounds almost trivial as the answer to a question as big as "What is the meaning of life?". If it's so simple, then why is life so complicated? Part of the answer lies in the peacock's tail.

MEME 13

When selecting which male to mate with a peahen (a female peacock) will choose to mate with the peacock with the "best" tail. The best tail isn't necessarily the biggest or brightest tail. A shorter, dimmer, symmetrical tail can be preferable to a longer, brighter, asymmetrical tail. The design is certainly an issue as well. I have little doubt that peacock tails' eyespots aren't there by accident. Although I'm sure that she's not conscious of this fact, a peahen chooses the tail that she thinks will most likely appeal to future generations of peahens (so that her male offspring can also attract mates).

The peacock tail's design is an emergent property of the process in which a peahen mates with the peacock that has the "best" tail. The definition of "best" is a little tricky. The peahen (consciously or unconsciously) mates with the peacock whose genes will cause her male descendents to have the kind of tail so that future generations of peahens will mate with them. So this definition of "best" is recursive. Certainly, size and shine play a roll, but style also plays a part. For example, peahens might prefer a tail that's symmetrical over a slightly longer asymmetrical tail. I'm willing to bet that a long, shiny, symmetrical tail with a very unconventional color-scheme wouldn't yield its owner very good reproductive results. So any tail that was markedly different from the status quo would be unlikely to be successful. On the other hand, a tail that was exactly the status quo would likely be outdone by a slightly longer or "better" designed tail. Countless pea-generations ago, this was grounded in fundamental fitness: going for a male with a longer or shinier tail made sense. A longer tail was a sign of being well fed, and a longer tail may have even had a functional use. But this process has resulted in a tail that's so long that not only is it not functional, but it's rather cumbersome for its owner.

There are 2 points to keep in mind:

- The peacock must spend significant resources on his tail, which has little (or even negative) survival value.
- The design of the peacock's tail is generated by a hall of mirrors, which can yield virtually limitless complexity. That is, a peahen is (at some level) trying to predict other peahens, which are trying to predict other peahens, etc.. This complexity is limited by the brains of peahens. So the design of the "best" peacock's tail isn't the optimization of a simple formula. By contrast, the wing of an aircraft is a design fully based on function (to maximize lift while minimizing weight and cost).

MEME 14

For most people (and even most chimpanzees), surviving is almost a given. Or, the "survive" problem is pretty much solved for people. In Chimpanzee Politics [**5**], Frans De Waal writes that "in order to get enough to eat, wild chimpanzees have to spend more than half their time foraging". Much of the rest of the chimps' time is spent "politicizing" (whereas things got even more more political in the Arnhem zoo because they didn't have to forage at all). So even chimps spend a good percentage of their efforts on the non-surviving "reproduce" part of the meaning of life (well, they're always breathing (for survival), but you know what I mean). With people, I'd guess it's much more so. For day to day activities, surviving is almost a given (it's pretty simple, at least)[1].

MEME 15

If life were just about survival and asexual reproduction (as it is for amoebas), then it'd be pretty straightforward. The "flower" comes from *sexual* reproduction and producing a peacock's tail. This is what people spend most of their efforts on. This is why the "Eat, Survive, Reproduce" seems so empty. Eating and survival are simple compared to the last. As Dan Savage (of Savage Love) puts it "Civilization is mostly a big effort to get laid.".

The potential complexity of Politics, Economics (which includes the unbound complexity of human relations), and Game Theory is limitless because each consists of people trying to model other people, who are also trying to model other people.

MEME 16

To illustrate the connection from the idea of the Peacock Tail to the Meaning of Life consider this: A monkey must wait for its stomach to digest the leaves it eats (the plants have been selected to be less digestible, and the better the monkeys' stomachs get, the more indigestible the plants evolve), and it has to do *something* while waiting (sleeping, running around, etc.). This is probably why life isn't all about fundamentals, because at this point, the fundamentals are as taken care of as they can be, and there's nothing left to do but non-fundamentals.

MEME 17

So what kind of Will would we give our rabbits if we knew they were in a cognitive niche? What if we knew they were social and their reproductive success was correlated with their social success? We might start by giving them a Will to power, and a Will to being socially successful.

The Mating Mind [**14**] makes a case that, what amounts to fundamentally the same process that created the peacock's tail (that is, the use of non-fundamentals for reproductive success), created in humans the drive to create music, art, and literature.

MEME 18

A peacock tail is the ultimate resource sink. An animal's tendency is to fill its extra closet space with its peacock tail. Any of our rabbits that fails to do this won't be as successful as a rabbit who does do it (at least locally).

[1] According to the Center For Disease Control, of the top 10 causes of death in the U.S. in 2006, all were "natural" causes save "Accidents", which ranked 5th. Murder and malnutrition ("war and famine") were responsible for only 21,129 or .87% of the 2,426,264 American deaths in 2006. The top 10 causes in 2006 were: Heart disease (631,636), Cancer (559,888), Stroke (137,119), Chronic lower respiratory diseases (124,583), Accidents (121,599), Diabetes (72,449), Alzheimer's disease (72,432), Influenza and Pneumonia (56,326), Nephritis, nephrotic syndrome, and nephrosi (45,344), and Septicemi (34,234).

MEME 19

A nice peacock tail is expensive (in terms of what the peacock has to devote to it as far as food and preening). A large peacock tail is also cumbersome. The analog of peacock tails for people also has a cost associated with it. So fashionable clothes tend to be *expensive*. There's some time investment you need to make to learn the hip slang (or jargon even) for a particular group[2].

In Victorian England, being able to quote in Greek would get you respect, despite the fact that Greek's practical "survival" value was low. Greek did have survival uses on occasion, however. Initially, learning Greek was useful because it allowed one to read ancient texts. After a while, (as translations became plentiful) reading Plato took a back seat to *appearing* educated. In fact, a peacock's tail gets the peacock peahens *because* it has a negative fundamental survival value. It shows the peahens that the peacock has resources to spare[3]. Other "peacock tails" include being good at basketball, being a good RISK player (in some circles), being well traveled (which also has some practical uses), buying expensive tulip bulbs (in Holland in 1636), and wearing mink coats[4].

Neither basketball nor RISK have value in and of themselves. Up to a certain point, skills and general lessons learned from these games transfer to "real" tasks, but after that point, the skills become pretty specific. It's a rare occasion outside of basketball where it'd be a useful skill to be able to consistently throw a sphere into a foot-wide ring 16 yards away, but teams in the National Basketball Association would pay millions of dollars for such a person to "work" for them (the distance for a 3 point shot is just under 8 yards, and such a person would be able to shoot from as far back as half-court). The star basketball player (of even a highschool or college team) is an envied individual.

Peacock tails takes different forms in different cultures. In the Dominican Republic, it's baseball and merengue. In Victorian England it was learning Latin, Greek, becoming cultured, and mastering the vast array of rules of etiquette. In surfer sub-cultures, it's, well, surfing well[5]. The ultimate peacock tail for people, though, is the human brain. There's a book called The Red Queen [**21**] which argues that a good deal of human intelligence is driven by an escalating arms race (though it makes no reference to RISK). Basically, people were trying to outsmart other people. Then, the bigger brained people reproduced (more than the dumber people) and the intelligence level for the whole group rose, meaning that the successful people of the next generation had to be even smarter. (It's like the Red Queen's race in Alice in Wonderland where the earth moves backwards as fast as the sprinters move forward so that the net gain is 0.)

[2]Such a group might be called a "Shibboleth Club". This term comes from a story in the Bible that documents one of the several cases where the Hebrews commit genocide. In Judges 12:5-6, 42,000 people lost their lives due to an accent that made them mispronounce the word "Shibboleth" Hebrew for "an ear of wheat".

> Then said they unto him, Say now Shibboleth: and he said Sibboleth: for he could not frame to pronounce it right. Then they took him, and slew him at the passages of Jordan: and there fell at that time of the Ephraimites forty and two thousand.
>
> –Judges 12:6

[3]A twist here is that an extreme peacock tail could theoretically lead to the extinction of a species. If a peacock spends all its resources on its tail, it spends less on "survival" value. It might be forced to spend more on its tail than what's valuable for it because of how much *other* peacocks are spending on their tail (another arms race). It's my theory that this has some explanatory power about why the Dominican Republic is so poor as a country: the peacock tail level is high. There's such an emphasis on dancing merengue well and being good at baseball that little fundamental *work* gets done.

[4]Mink coats, luxury cars, (overly) expensive wines, and tulip bulbs in the 17th century are sometimes called Veblen goods [**11**]. These goods are purchased for the same fundamental reason that a peacock grows such a huge tail: to prove that you have resources to spare.

[5]Though in the context of a surfer sub-culture, you might feel like *everyone's* goal is to surf well.

CHAPTER 4

WILL: INTERACTION OF COGNITIVE SYSTEM AND REWARD SYSTEM

Zwei Seelen wohnen, ach! in meiner Brust, Alas, two souls dwell in my breast,
–Johann Wolfgang von Goethe (from Faust, 1808)

No conscious tabulation of the disadvantages and horrors of junk gives you the emotional drive to kick [a heroin habit].
–William S. Burroughs (from Junky, 1953)

Man is not truly one, but truly two.
–Robert Louis Stevenson (from The Strange Case of Dr. Jekyll and Mr. Hyde, 1886)

MEME 20

Although our rabbits are really just processes, interactions of molecules, it's useful to create abstractions to describe these processes.

MEME 1505

Each of us has a world-model in their heads, or a conceptual structure that I'll call a *Weltanschauung*. At the most basic level, a Weltanschauung is how we categorize The World. We throw "dogs" into the "animate beings" category, "icebergs" into the "huge objects" and "floating things" categories, but there's more to a Weltanschauung than that. A Weltanschauung also contains ideas about how hang-gliders behave, associations, generalized abstractions, and "gists" of concepts. Much of a person's Weltanschauung is tacit knowledge, meaning that it's hard to put some views or concepts into words. For example, given a photo of an adult's face, we can easily tell the if the picture's of a man or a woman, but it's incredibly difficult to write a set of rules so precise that a computer could follow them and make the distinction. Tacit knowledge also includes concepts that are so deep or ever-present that the person doesn't even realize that they're there. For example, you might not be immediately aware that gravity is pulling down on you or that you're breathing and blinking.

MEME 1506

The subfield of Artificial Intelligence called Reinforcement Learning [**25**] would describe a rabbit as an *agent* that interacts with its *environment* or local surroundings. The agent takes *actions*, where the actions may be very low-level, such as twitching an individual muscle. For example, the "action" of hopping is really a composite of a large number of lower-level actions. The agent receives *observations* from its *sensors*, where the observations may also be very low-level: the reports of individual "pixels" on the rabbit's retina, for example. As described in the section on Squiggly Lines, a rabbit doesn't actually see objects such as grass or dingos, but rather a set of sensor values that can be *interpreted* as being caused by grass or dingos. The rabbit also receives a *reward* signal, and it's the rabbit's goal to maximize the amount of reward it gets over its life.

Standard Reinforcement Learning assumes that the reward signal is a single numerical value. I think this is oversimplifying matters. Our actual "reward" is probably a hodgepodge of often contradictory tricks that evolution hacked together. But, the Reinforcement Learning model might be a useful initial approximation. At any rate, our rabbit will have some criteria for deciding which actions to do. These criteria can be described as "desire" or "wanting". I'll follow Schopenhauer's lead and call it *Will* [**23**].

The rabbit would also have a *Weltanschauung* or model of The World, and a *cognitive system*. The rabbit's cognitive system does 2 things: it *builds* the Weltanschauung by learning from its experience, and it *uses* the Weltanschauung to do things like make predictions, draw inferences, and create plans to get reward. Here, I'm using the term "prediction" loosely. For example, if a rabbit sees a dingo's head, it will probably "predict" that it will soon see the dingo's body. The predictions aren't necessarily temporal either. For example, we can "predict" that the Nile River has a source.

Note that our rabbit's cognitive system can be somewhat independent of its reward system. For example, if (for some strange reason) we structure our rabbit's reward system such that it gets reward for doing somersaults, then the cognitive system will figure ways for the rabbit to do lots of somersaults (and it will cause the rabbit take actions to do this).

MEME 1386

Perhaps the most well known example of classical conditioning is the story of Pavlov's dogs: Every time Dr. Pavlov fed his dogs, he rang a bell just before serving them. After several days of this, the dogs began to salivate every time they heard the bell. This sounds simple enough, but how did the dogs know it was *the bell* that signaled that food was coming? Any time the bell rang, there were hundreds of other stimuli going on. During one particular ring, a bird might have been heard outside the window, one of the dogs may have been sitting in a particular position. Dr. Pavlov may have been whistling a particular tune. Why didn't the dogs associate any of these other stimuli with food?

The answer is that they probably did. At least a little bit at first. When the dogs are served food, this fact is signaled primarily through taste and smell to the cingulate cortex, which sends a signal to a pleasure center in the hypothalamus, which "squirts" dopamine all over the neocortex. Dopamine causes the areas that were recently active to have a slightly higher "reward value" attached to them. So, if a dog noticed the bell, the bird, and that his friend Laika was wagging his tail in an awkward manner, the food will all be slightly more associated with all these things than it was before. At the next feeding, the association between the bell and food will be strengthened a little more. But the bird probably won't be chirping the same, and Laika's tail probably won't be wagging in the same awkward manner, so the food-assocation with these stimuli won't be strengthened.

An interesting thing happens *after* the association between the bell and the food has been established. If we were to measure the dopamine levels in the dogs' brains we'd find that, before conditioning, the dogs get a spike of dopamine when they get the food, but not when they hear the bell. This suggests that the dogs *feel* pleasure when they actually get the food. However, after conditioning, the dogs show a dopamine spike when *the bell is rung*, but not when they get the food. So the dogs will begin to try to predict *the bell* and figure out how to get *the bell* to ring, as opposed to trying to figure out how to get the food directly. If Dr. Pavlov plays a mean trick and decides not to give his conditioned dogs their food after ringing the bell, the dogs will have a dopamine spike when hearing the bell, but they'll have a marked *decrease* in dopamine when the food doesn't arrive at the expected time. The dogs' dopamine levels will momentarily drop well below their baseline level, indicating the dogs feel "pain" or discomfort when they don't get the food. In this case, pain works the same way as pleasure, but in reverse. Instead of associating recently active brain areas with "good", these areas are associated with "bad". If Pavlov continues his cruel joke several times, the dogs will eventually extinguish the original "pleasure" association they had with the bell.

Likewise, if someone begins to smoke cigarettes for the first time, initially the reward that nicotine produces will be associated with other salient features: certain people in the room, a particular flavor of soft drink being sipped, etc.. Only after several instances will our limbic systems "figure out" that it's the cigarette causing the sensation.

MEME 21

Because of the incremental process of evolution, people's brains are somewhat layered. Our neocortex, which is heavily involved in our higher cognitive processes, lies atop the "reptilian brain" and the "limbic system" [**13**]. The reptilian brain, which is the brain stem and cerebellum, controls low-level instincts and autonomic functions (such as breathing). The limbic system includes structures such as the amygdala and the hypothalamus, and is heavily involved in emotional processing and classical conditioning, where a stimulus (such as bell ringing) occurs right before a reward (such as food) is given, and the stimulus eventually becomes associated with the reward itself.

Neuroscientists will debate how clean the separation is among these layers, but I suspect that higher level thought *can* largely be detached from emotion or reward[1]. For example, the neurologist António Damásio describes a patient, called S, whose amygdalae were effectively knocked out [**3**]. Because of this, S didn't experience fear or anger. When she was put into a situation that would normally induce fear in people, she understood on an intellectual level why there was cause to be concerned, but she simply didn't experience the emotion of fear.

(A note on my terminology: the terms "limbic system" and "neocortex" refer to the actual structures in our brains. The terms "reward system" and "cognitive system" refer to idealized versions of these structures.)

[1] Jeff Hawkins gives a computational account of the neocortex which has the goal of making predictions [**7**]. Hawkins's model has no reward or motivational system, which he claims belongs to subcortical structures.

MEME 1259

I always thought our family dog, Rex, was something of a dog genius. I thought this until I found out about the "A-not-B" task used in Developmental Psychology. In this experiment, infants are shown a toy being placed in 1 of 2 bowls. The bowls are within the infant's reach, but are angled such that the baby can't see inside them. If the baby's immediately allowed to reach for the toy, he or she will grab the toy from the correct bowl. However, in A-not-B, the experimenter will distract the baby for a moment just before the baby is allowed to reach for it. Babies under the age of 10 months won't always reach for the right bowl. It seems as if they completely forgot which bowl the toy was in [**17**]. I was surprised to learn that even adult monkeys don't do so well at this task. I figured that the dog-genius Rex could ace the A-not-B test, so I tried it (replacing the toy with a doggy treat). Rex failed. He always went to the correct bowl when he wasn't distracted, but it was a craps shoot which bowl he went to if I distracted him for even an instant.

Thus, life as a dog must be quite different from life as a person (or a bat [**15**]). A dog has a somewhat less-devoloped neocortex than a human, so they must rely more heavily on their limbic system for making decisions. A dog, then, lives more in the moment, using instincts and short-term rewards. If you can't do the A-not-B task, you don't have much of a medium-term memory. So life as a dog might be like life for humans with hippocampal damage, who are unable to form new memories (such as the main character in the film "Memento"). The chief way that people who lack a functional hippocampus learn is through classical conditioning, like Pavlov's dogs.

MEME 22

> Reason is, and ought only to be the slave of the passions, and can never pretend to any other office than to serve and obey them.
>
> –David Hume [**9**]

Consider an ideal optic lens. The "purpose" of an ideal lens isn't reproduction, but rather to focus light as a parabola. A lens's "goal[2]" can be described mathematically, so we can say what makes a good and bad lens. In our rabbits, their eyes' lenses are "tools" of the Will. That is, the proximate purpose of the lenses is to focus light, but the ultimate purpose is to help the rabbit reproduce. It's possible that there are lenses that focus light less well, but somehow better help the rabbit reproduce. These lenses would be preferred by us (the creators of the rabbits).

Likewise, I'd argue that the "goal" of the cognitive system is to characterize The World, and that the cognitive system is also used by the Will to help in reproduction. It gets more complicated here because the design of a cognitive system has more flexibility than the design of a lens, so that I can imagine a cognitive system that is worse at prediction but better helps our rabbits to reproduce.

A reward signal can be helpful for a cognitive system. The reward signal can provide "hints" as to which parts of The World to pay attention to. If an ideal cognitive system has no reward, it has no bias for which parts of The World to characterize: it will devote just as many cognitive resources to understanding the "politics" of the seagulls on the boardwalk as it will to understanding the politics of the people in Washington D.C..

So are there 2 dueling meanings of life? That is, there's the meaning of life from a biological point of view (which is what our reptilian brain wants), then there's the meaning of life from the point of view of an ideal cognitive system, which is to characterize The World.

It's interesting to consider which system "we" truly are. Or it's possible that what we consider to be our inner selves is actually the combination of both these systems.

[2] In this sense, I'm stretching the meaning of the word "goal", which is really (like almost any concept) just an abstraction.

CHAPTER 5

SQUIGGLY LINES

MEME 23

From my work in robotics, I can say that perception is a tricky computational problem. Like gravity, our perception is so omnipresent that it's easy to assume that it takes little effort. It's not until working with a robot's perceptual system or image processing that one develops an idea of how difficult perception is [**19**]. The task of simply taking an image and identifying the objects in it can be tricky (and, to date, no one has figured out a foolproof way to do this autonomously as well as a person's perceptual system can). This is because, fundamentally, perception is a problem of "Squiggly Lines" [**22**].

To illustrate what's meant by a problem of "squiggly lines", suppose we wanted to create a robotic bumblebee. Further suppose, by some miracle, we were given the entire bee and we just had to program its "brain". The bee's compound eye is really just a set of light-sensors. Basically, each "surface" reports a value proportional to how much light's striking it (and different surfaces might be sensitive to different colors). Unless we explicitly "tell" it, our robot doesn't know what its sensors mean. It just has the sensor readings over time, which look like a bunch of *squiggly lines* if you plot them out. Suppose I gave you plots of these sensors' readings, but I didn't tell you which sensor was which. Suppose I didn't even tell you that these are light sensors from a robotic bumblebee's eye. As far as you would know, these could even be readings from (a simulation of) a mobile robot in a 5 Dimensional world. From this perspective, it's very difficult to tell the difference between the squiggles produces by a flower and the squiggles produced by a female bumblebee. The bumblebee's perceptual system has 2 limiting factors. The 1st is that its compound eye isn't sophisticated enough to tell the difference, which is essentially the problem that Plato discussed in his Allegory of The Cave [**20**]. The 2nd is that a significant amount of computation is needed to tell the difference. Although it'd look silly, we could give the bumblebee human eyes, but there'd still be the problem of processing all that data. The bee's tiny brain couldn't do it. It's worth noting that our own visual processing system, the visual cortex, is many times the mass of a bee.

The Bumblebee Orchid (*Ophrys bombyliflora*) takes advantage of the limited perceptual systems of male bumblebees. Its flowers "look" and smell like fertile female bumblebees. That is, the flowers have the rough shape and coloring of female bumblebees, but other than that, they don't actually look very much like female bumblebees. (I can't say how similar they *smell* like them.) With a glance, a person can easily tell the difference between a bumblebee and a Bumblebee Orchid (though I have trouble telling males from females because of *my* limited olfactory perceptual system). However, they're good enough to trick the "low resolution" compound eyes and tiny brains of the randy males, and the male bumblebees attempt to copulate with the flower, picking up and dropping off the flower's pollen so that *the flower* can mate successfully.

MEME 24

A peahen's perceptual system and brain is much more sophisticated than that of a male bumblebee, but it's still limited (as is ours!). Bird's perceptual systems are particularly good at finding eyes. A number of butterflies and moths such as the Common Buckeye Butterfly, and the Promethea Silkmoth (*Callosamia Promethea*) take advantage of this by having eyespots (because eyes are a giveaway that something's animate). These look a lot more like owls' eyes than the bumblebee orchids looks like bumblebees, but a person can still easily tell the difference. (Though birds might not be able to tell the difference, or they just might not stick around long enough to find out.) So, I assume that, like other birds and even some insects, peahen's brains are good at recognizing eyespots. That is, eyespots have a special place in the perceptual system of peahens. So I'm guessing that eyespots are so prevalent on peacock tails partially because of their special place in the perceptual systems of peahens. Another (not incompatible) reason is that eyespots were originally useful for the peafowl (as they are for the Buckeye Butterflies), and the genes got "recycled". (The former reason at least seems plausible to me.)

CHAPTER 6

TASTE IS LIKE THE BUTT, DIVIDED

MEME **25**

Until he retired, my godfather worked as an engineer at the Ford Motor Company in Detroit. He had bought one of their new models, and said that although Ford's model was more aerodynamic, Chevrolet's competing model had higher sales because people "didn't like the way the Ford model *looked*". There's a way that people develop their tastes in cars. People see both the designs and the performance of fish, jet fighters, and cars. After a while, people develop an intuition for what kinds of features (e.g., sleekness) make for a high performance thing-traveling-through-medium. People then associate these features with their performance. So, without ever having taken a course on aerodynamics, people can eyeball a car's body design and tell you (with better than random accuracy) whether it will outperform some other body design. This is in contrast to the more *analytical* conceptual system of car design which is what the designers at Ford use when they apply equations and principles of aerodynamics to come up with a more exact answer for how aerodynamic the car is. In reality, most systems in our heads (and in the heads of the designers at Ford) are both intuitive and analytical.

This, I believe, is the basic principle of how tastes are developed in general. Like peahens, our perceptual abilities are limited (where we can't tell a car's drag coefficient just by looking at it). So, we have to develop associations of features to "value" (speed or performance in the case of cars) so that we have an intuition to give us some idea of the value.

MEME 1589

In stock market analysis, one can take a *technical* approach or a *fundamental* approach. A purely technical analysis of a company's stock looks only at the "trends" in the stock's selling price. Books are written about how to "predict" what a stock will do given *only the history of the stock's price*. Generally, if a technical investor predicts that the stock's price will raise, he'll buy, and sell if the technical analysis predicts the stock will drop. In its most extreme form, a technical analysis wouldn't even look at what the company *does*. A fundamental analysis is at the opposite extreme: it will ignore the history of the stock's price, then make a prediction of the company's earnings (and dividends) based on factors such as what capital the company owns, the "quality" of the people working for the company, and whether the current economic situation means that people will want to buy the company's goods. You can then assign the stock a fundamental value based on the expected dividends and interest and inflation rates. (For example, if you expect a share of Bifislurf Inc. to yield $5 in dividends over the next year, and the inflation-adjusted interest rate is 5% per year, then a share of Bifislurf Inc. would have a fundamental value of about $100, because that's how much money you'd need to put in the bank to get $5 of interest in a year.) Generally, a pure fundamental trader will buy if the stock's current price is (significantly) less than its fundamental value and sell if its price is higher. An interesting thing is that if everyone invested solely on fundamentals, the plot of the Dow Jones Industrial Average would be much smoother, practically flat because it'd only reflect "real" changes in the companies' values. (Tulip Mania also would've never happened if the tulip traders used only fundamental analysis. Tulips have little intrinsic value.) In reality, most traders use results from both fundamental and technical analyses.

∞

Another example of fundamentals and technicals: Flugtag is a contest/event sponsored by the Red Bull company in which teams build non-motorized flying contraptions and launch them (carrying one of their members who is the "pilot") off a 30 ft. high ramp into a pool of water. The teams are judged primarily by 2 criteria: 1. the distance flown before landing in the water, and 2. the "creativity" of the contraption's design. If it weren't for the 2nd rule (which I'll call the "technical" rule), Flugtag would be a much more boring event. What would happen, I predict, is what happens with a myriad of other "purely fundamental" pursuits. Initially, there'd be a broad range of designs, but eventually one of the designs (or its basic principles, at least) would emerge as the "optimum", and most entrants would be minor variations on this optimum. Take airplanes, for example. After the Wright Brothers' success, there was a blossoming of all sorts of crazy designs. Take the Langley Flyer, for example. It actually predated the Wright Brothers' Flyer by several years. Its full scale model was never fully capable of sustained flight, but the smaller models had some success. The interesting thing is that it looks completely unlike any airplane I've ever seen. By World War II, the basic body of the airplane (with a single aerofoil wing) converged to what's still used by commercial jets. A similar process happened with locomotives, automobiles, and computers. Thus, Red Bull's 2nd rule explicitly puts a limit on the kind of convergence that would otherwise happen.

MEME 1590

Suppose I want to evaluate a piece of clothing. For instance, I have *fundamental* criteria such as how comfortable a shirt will be, whether it'll keep me warm on chilly days, and how difficult it'll be to wash it. I also might be concerned with how I'll *look* in it, and I won't be looking at myself much. So I'm really concerned with associations other people will make of me when I'm wearing the shirt.

Now, there's one more factor of taste: associations made of people based on their appearance. One can take a Holmesian approach and use logical reasoning to deduce things about a person from their appearance, but we don't have enough time/brainpower to do this with every person we meet, so we have to rely on our intuitions (or associations developed through experiences), just as we might in determining how "functional" a car will be just from eyeballing it.

Some associations are formed from fundamentals. If I see a picture of a person wearing a heavy coat and hat, I'll "guess" that the person's somewhere cold. This is because a heavy coat keeps you warm regardless of the coat's social context. If a person's wearing glasses, they probably don't have perfect vision without them. Fundamentals are pretty easy, at least compared to technicals. These don't have to do so much with the intrinsic properties of things, but more with what the thing means for other people. With clothes, technicals would be "ornamentation", such as printed designs, the particulars of the cut of the fabric, and "flair".

I'd guess fashion designers spend 99% of their thought on technicals. Ornamentation has its own abstract (and often intuitive) "rule system" as well. Like the peacock, you can't stray too far from the status quo. The "fundamental" problems of clothing are pretty straightforward. Like the stock market, if clothes were designed only by fundamentals, fashions would hardly change at all. (Changes would only be with innovations such as with materials and manufacturing techniques, and with changes in what people use the clothes for. For example, a much lower percentage of Americans farm than was the case a century ago, so the average American doesn't need clothes designed for farm work.)

So, how do technicals and intuitions formed about technicals work to influence fashions? Take bell bottom pants, for example. The theory is that attractive people originally started wearing bell bottoms. (The reason for doing so may have been to distinguish themselves from "the masses", or perhaps for the same reason very fit gazelles will flaunt their fitness in the face of approaching lions as if to say "Don't waste your time chasing me, look how fit I am.".) Then people began to associate bell bottoms with being attractive. Then, (knowing about this association (or having the association themselves), and maybe it's not explicit) less attractive people began wearing the bell bottoms so the association would be transferred to *them*. Eventually, so many unattractive people started wearing them, that *after a lag*, the association became extinguished, and there was little reason to wear bell bottoms.

∞

The book Freakonomics [**12**] talks about similar trends with babies' names: people of high socio-economic status start naming their kids with a particular group of names. Then, the trend catches on because other people see that "The Beautiful People" are called by these names. Thus, people (from lower classes) form the association from the name (a nearly arbitrary symbol, practically) to the person's status, and name their own kids with that name. Eventually, the name becomes "common", the association is extinguished, and the upper classes find new names.

MEME 1592

This association from technicals to fundamentals might be why people have certain tastes in food. Oncologists have long known that you can cause people to develop a strong aversion for almost any kind of food simply by putting their chemotherapy medicine in it a number of times. My theory is that you can similarly cause people to like just about any flavor by creating a fatty food with that flavor: I remember the first time I had the Greek candy *halvah*. My friend, Charles's mom, Zabia (who's father was Greek), offered it to me. Despite its taste, which I would've described as awkward but not *bad*, I ate it out of curiosity and politeness. I had halvah several times after that, and soon developed a taste for it. Once, I found some commercially packaged halvah, and I read the label: halvah's made from crushed sesame seeds, and is about 20% fat and 60% sugars by weight. My theory is that fundamental "yumminess" is mostly fat and sugar, and that eventually we (or our taste buds and the associated brain areas, to be precise) associate the flavor with the fat and sugar content.

MEME 26

I'd hazard to guess that virtually anyone can be conditioned to like almost anything. (Well, anything that a large group of people also enjoy. There are probably few people who could be conditioned to enjoy having their toenails removed.) Through conditioning, almost anyone could probably be made to enjoy the flavor of chocolate and dislike the flavor of vanilla, or vice versa. Similarly, barring physical disability, associations and reinforcement could be used to cause almost anyone to be made to enjoy mountain climbing, painting, or knitting. To do this, we'd just need to cause the association of the features of these activities with more fundamental rewards.

This raises the question of what the *fundamental* rewards for people are? For our rabbits, they should be those that, when coupled with the rabbits' cognitive systems, causes them to survive and reproduce. Some rewards might be innate even though they could be learned from more fundamental rewards. For example, it might make sense to install in our rabbits an innate desire to not fall from high places (as a form of bootstrapping), even though this desire is a "corollary" of the more fundamental desire to not break bones. So, some of our rabbits' desires could be redundant or even contradictory.

6.1. Specifying Reward: The Bulldog Anal Rapin' Robot©

MEME 27

A problem with innate rewards for people or rabbits is that we have to be born with them. There are people (and presumably rabbits) that are born blind or deaf, yet are still attracted to members of the opposite sex. This means that our fundamental rewards probably aren't tied to a specific sensory modality.

MEME 28

The Gedankenexperiment of the "Bulldog A.R.R." is that you want to build an Anal Rapin' Robot© (or A.R.R.) whose goal is to anally rape bulldogs. The Robot would have little interest in dalmatians, and no interest at all in mating with people. In this Gedankenexperiment, we're not allowed to know what the robot's sensor suite is (we're back at the squiggly lines problem). So, how do we tell the robot about bulldogs (since we can't ground the concept)?

My idea for a solution is that you can provide an *ungrounded* concept of what a bulldog is. This representation will be *invariant* with respect to any specific sensory modality. The representation will be some relational conceptual structure, and the A.R.R. will develop higher and higher concepts from its sensors until it eventually develops a concept that has a very similar structure to the abstract one ("Essence of Bulldog") we provided to it. At which point, the robot will have grounded this concept.

MEME 29

To get an idea of how much is in our innate Weltanschauung, consider testosterone. Since testosterone innately makes a person aggressive, we must have an innate model of aggression. I think it's interesting that testosterone serves a similar function in bull sharks and hyenas as it does in people, since this hormone is only a symbol. (Unlike glucose (where the molecule actually has energy), there's nothing fundamental to most hormones that make them behave that way. My guess is that we could swap dopamine receptors and serotonin receptors, and the respective neurotransmitters, then serotonin would act like dopamine (meaning lots of serotonin would have the same psychological effect as lots of dopamine would have) and vice versa.)

I'm guessing people have an innate world-model that's complicated enough to specify the difference between male and female on at least an abstract level. From a squiggly lines perspective, it would also be tricky, at least, to have an innate model that's sophisticated enough to encode innate rewards for things that would be useful for our rabbits such as a desire for social power, maternal behaviour, or jealousy.

MEME 30

Consider a "pure bred" bulldog. Its mom was a bulldog and its dad was a bulldog, and they had sex. I still haven't been able to put myself in those bulldogs' shoes: you see a bulldog and get aroused. (I can't imagine it, even though it makes sense.) (There's the converse too: I can't imagine a good-looking human female coming on to me and not feeling affected.)

Along the same lines, there was a point where I couldn't imagine liking Vegemite, the Australian condiment made from yeast extract. But I've gradually conditioned myself to like it by mixing it with fatty food.

MEME 31

Designing a reward function –what our rabbits *want* deep down– can be tricky. Dark tales are told of genies who obey their master's wish to the letter, but end up doing something the master obviously didn't intend. Giving exactly the same system 2 different reward functions can result in 2 radically different behaviours. For example, Leslie Kaelbling, a researcher in Artificial Intelligence at MIT, describes her experience with motivating a wheeled robot to navigate a room. Initially, she "punished" the robot for running into walls[1]. The robot quickly learned that the policy that maximized its reward (or minimized its punishment) was to not move at all. So Leslie changed the reward function to also include a positive reward for moving. After this, the robot learned to simply spin in circles.

[1]She did this by giving the robot a negative reward when its wall-bump sensors were activated, then programming the robot to maximize its reward, or, equivalently, to minimize its punishment.

6.2. Hill Climbing on the Cingulate Cortex

MEME 32

I have a pet theory that it's the cingulate cortex that encodes our innate invariant reward. The theory is that the cingulate cortex is hard-wired, and acts like a hash that is activated only under certain conditions[2]. For example, a region of the cingulate might encode a representation of femininity that's invariant to any particular sight or sound of a woman. Furthermore, this probably works with the lower level (plastic) cortical areas in a complex way. So, some of what attracts us to a woman is innate, but much of it is learned, probably through classical conditioning. When a particular area of the cingulate is activated, it sends a signal to a corresponding area of hypothalamus, which does things like shoot out dopamine (leading to repeat actions and classical conditioning) or triggers a shot of adrenaline.

If this theory is correct, then much of what we do in life might be described as "hill climbing on the cingulate". For example, suppose no one ever told you explicitly about sex or even romantic relationships. I'd guess that people would still reproduce. If you're a boy, around age 13, for some reason, you'd find yourself very interested in girls. You might even find yourself sexually aroused, but you're not sure what to do about it, but you might have an inkling that it has something to do with girls. Given unlimited access to girls, you might find you get reward for touching them and having them touch you, especially if they touch your private areas (and you touch theirs). I'd imagine, you'd eventually figure out how to have sex with them.

I wouldn't be surprised if the cingulate encoded (in some form) less base motives such as social power. Our primate social ancestry has been long enough to allow at least part of something as complex as social power to be encoded innately. If this is true, then "trying to find yourself" or "trying to figure out what makes you happy" might be essentially hill climbing on the cingulate. You're trying to make the reward part of your cingulate fire, but you're not sure what causes it to do so. You get reward for some actions, so you take more actions like those.

In a sense, an invariant representation is a Platonic form. For any cortical area (cingulate or not), there's some stimulus that maximizes the area's response. For example, I remember seeing an experiment where scientists measured the firing rate of a particular cortical area of a particular monkey. This area fired lightly when the monkey was shown a cartoon drawing of a smiley face, it fired more heavily when the monkey was shown a photo of a human's face, and even more heavily given a monkey's face. I wouldn't be too surprised if the area fired even more heavily when shown a particular monkey's face (say, the monkey's grandmother) in a particular configuration. Likewise, areas of our cingulate might fire strongest (producing the highest reward, punishment, or other base emotional response) for some particular stimulus. It's our goal to figure out what that stimulus is (for reward) and how to make that stimulus occur as often as possible.

Even if my theory isn't correct, and this isn't what the cingulate does, but there must be some structure that encodes an invartiant reward (and other emotional primitives). Judging by its connectivity with the hypothalamus, I'm putting my money on the cingulate. I doubt the hypothalamus itself has the necessary structure to encode an innate representation of social power, for example.

[2]The cingulate can be thought of as one of the boxes in Jeff Hawkins's model of the neocortex [**7**].

6.3. Some Strategies for Maximizing Reward

MEME 33

Some sort of reinforcement mechanism is still at play in the human brain. It might be an obsolete relic from the time before people evolved their higher cognitive capacity, but this mechanism is so basic, and has been with us so long (since before we were mammals), that it's still heavily entrenched in our nervous system. The mechanism I'm referring to is basic reward prediction in what psychologists call "classical conditioning". This is where a stimulus (such as bell ringing) occurs right before a reward (such as food) is given. Psychologists (and dog trainers) found that animals (and people) eventually associate the stimulus with the reward itself. It's as if Pavlov's dogs eventually associated the bell itself to be almost as good as food and would take actions to *hear the bell*.

Researchers working with animals and rewards have learned a few things about how this reward feedback works in animals. To make this concrete, let's say (hypothetically) that we have a laboratory with a bunch of monkeys, some crack (and crack-pipes), a cattle-prod, and a huge pile of dirty dishes. Suppose we show the monkeys how to use a crackpipe, so that they know what to do with it, and that we want the monkeys to wash as many of the dishes as possible (we've also shown the monkeys how to wash dishes while smoking a crack pipe, quite a feat if you've never attempted this[3]).

There are a few strategies we could try:

(1) The most fun-sounding strategy would be to cattle-prod the monkeys any time they take a break from washing dishes (or when they break the dishes), so that the monkeys associate not-washing-the-dishes with being zapped. This is probably the least effective strategy.
(2) In addition to prodding for not-washing, one could also give the monkeys crack for washing dishes, but the fact is that negative reinforcement (i.e., the cattle-prod) simply leads to the monkeys associating the whole process of dish-washing with zaps, which will cause them to try to escape.
(3) Another strategy is to give the monkey a tiny crack rock for every dish that they wash. This will cause stacks of shiny dishes.
(4) Finally, there's a strategy that's even more effective: for every dish that a monkey washes, roll a 20 sided die, and if the monkey rolls a 20, give the monkey a sizable crack rock (say about 10 times the size of the tiny crack rocks used in the previous strategy). This strategy will have the dishes cleaned at top monkey-speed. The reason this works is that doing any dish could potentially be worth a big crack rock. The randomness ensures that the big-crack-rock dish could be the very next dish at any time.

The phenomenon of the last strategy might explain why some people can spend an entire day in front of a slot machine at a casino, why one might repeatedly check their mailbox while expecting the next Victoria Secret catalog, why one might do the same with email. This phenomenon might also explain why surfers will spend entire summers on the beach waiting for "the perfect wave", or why people will play hand after hand of the card-game Pinochle waiting for the perfect "1,000 Aces" hand, or why sport-fishermen will spend all day with their lines in the water waiting for "the big one", or any number of analogous activities waiting for a *Holy Moment*, those rare, but immensely rewarding windfalls.

[3]It's quite a feat for all 3 parses: Where the monkeys are smoking during the lesson, where the monkeys will be smoking while they're washing the dishes, and where *you* are smoking during the lesson of teaching the monkeys to wash dishes.

CHAPTER 7

ARTIFICIAL SWEETENER, DRUGS, AND VIRTUAL REALITY

MEME **34**

Artificial Sweetener[1] by William Christopher Krueger I, 2008

Once upon a time, there was an ant colony out traipsing for food.
Its nature beckoned it to gather resources to fulfill its purpose.
It befitted the colony that an inexhaustible mound of sweetness lay due north.
In eerie compliance it was deemed of worth to venture there.

The colony arrived at the spot, antennae flitting about frantically.
Scores of ants began to divide the miracle, each fraction either delivered home or devoured.
And thus ants went on living, from colony to colony, off of this pile of pearls.
The ants once engaged however grew thinner with each return trip.

And also once engaged, the ants returned to the Mound with greater and greater urgency.
Amid scurry, deaths flowered.
The Mound, now littered with corpses, gleamed as sweet as ever.
It was made not of sugar but of artificial sweetener, its lack of substance indistinguishable to ants.

[1] I'm not sure if insects' mechanisms for sensing sweetness are fooled by some of the same tricks that ours are. I once did an experiment with ants, giving them different types of sugar substitutes and sugar itself. The ants went for the normal sugar, but showed little interest in any of the artificial sweeteners. For us, it seems that the external chemical properties of artificial sweetener (as far as what the molecules bind to) is similar to that of sugar, but the artificial sweetener just doesn't have the hydrogen bonds (energy) that sugar has.

MEME 35

For our rabbits, mounting the wrong end might not seem like that big of a mistake because it feels good. But, as the programmer, I'm aiming for some ideal "meaning of life" for the rabbits. So for me (who wants to win at Hare Wars) the rabbit mounting the wrong end is a "bug" in the rabbit's design.

∞

If happiness were a person's only goal, they might be tempted to enter a virtual reality, in the same style as in the movie The Matrix, and never leave. They'd be having no effect on the world, and people who tend to do that (like people who tend to become addicted to alcohol) get weeded out of the gene pool. As the designer of the rabbit, I'd want to prevent this from happening. Has anything like this ever happened in evolution? A hungry person can imagine a hamburger, but there's something that prevents us from getting reward from just imagining eating.

∞

There are all kinds of other computational "bugs" found in nature: army ants can follow each other in a big circle until they all die of exhaustion, goats can eat all the vegetation on an island (causing it to be barren) and then starve to death. There's the example of the Sphex wasp described in Gödel, Escher, Bach **[8]**, where the wasp is shown to run an obviously simple routine. I also view monkeys spanking their humans (and vice versa) to be along these lines.

∞

In filial imprinting, goslings follow the 1st big thing that moves. In evolution, this worked most the time and was a simple solution. Evolution lacks any form of foresight, so since this solution was easy and it worked, it's the solution that evolution went with. Evolution didn't anticipate wily scientists like Douglas Spalding, who discovered imprinting, or Konrad Lorenz, who exploited the geese's nature to make them think their mother was a toy wagon.

MEME 36

Mainly due to technology, the environment that people live in now has a number of significant differences from the environment in which we did the bulk of our evolution. In fact, Wrangham and Peterson **[29]** argue that, until the last 6 million years (or 300,000 generations, which is not a long time, on an evolutionary time scale), our environment was somewhat similar to the natural environment of modern chimpanzees. Evolution didn't anticipate birth control, such an abundance of fatty foods (via agriculture), video games, heroin needles, or internet pornography.

MEME 37

If the meaning of life is reproduction, why do so many people not want children? During evolution (and now), having kids itself was actually a pretty weak urge because people didn't make the connection between sex and procreation until recently. The bigger urge is for sex. That is, until the invention of birth control, the desire for kids was subsumed by desire for sex.

MEME 38

> For Moral Virtue has for its object-matter pleasures and pains, because by reason of pleasure we do what is bad, and by reason of pain decline doing what is right (for which cause, as Plato observes, men should have been trained straight from their childhood to receive pleasure and pain from proper objects, for this is the right education).
> –Aristotle (384-322 BC), The Ethics

Why we don't naturally find pleasure and pain in the right objects? Is it all just a result of evolving in a different environment from which we're now surrounded? For example, why isn't hard work fun? If hard work were so good for our well being (and presumably evolutionary fitness), then wouldn't we have evolved to enjoy hard work? Or maybe that's just the definition of work: something that's useful for us that's not naturally enjoyable.

Likewise, one might ask why everything that tastes good is bad for you. Well, it's not bad for you. It's actually good for you in small doses. If eating carrots felt as good as eating candy, then people would eat a lot more carrots, and there'd likely be some chronic problems with eating too many carrots. For example, candy is actually very healthy and nutritious. It's just that *too much* candy is unhealthy. In our chimpanzee days, it was rare for the environment to give us too much candy (or calorie-dense food), so we didn't need to have an innate check on how much of it we ate.

MEME 39

A question from "The Book Of Questions" **[24]** goes like this: "If you could spend one year in perfect happiness but afterward remember nothing of the experience would you do so?". My answer is that happiness is never a goal in itself. Rather, like our rabbits, it's evolution's means to get us to reproduce. There are scenarios where you get to keep the fruit of your happiness: relationships, knowledge (though this is probably not kept, according to the question), babies, etc., which I might go for. But I think the question means the scenario where you keep none of that. The year's just clipped out of your life (and you're a year older). If this were the case, and all else were equal, the rabbits that chose to be (re)productive during the year instead of "living in perfect happiness" would be the rabbits that would tend to be around. So we'd hope that our rabbits' designs would cause them to choose not to take the amnestic year of happiness.

MEME 1572

The need to always be falling forward is strong, so there are a lot of Artificial Sweetener versions of it.

For example, it's easy for me to spend countless hours playing the computer game called Civilization II. In this game, you govern a civilization, and you can expand your empire, develop new technologies. You can also build new cities and create improvements to them, such as city walls, and you can create World Wonders, which improve the "greatness" of your civilization. Like many video games, Civilization II had a good deal of "leveling up", where your character or civilization keeps improving its "level". Once you've gotten your civilization going, the game can become quite addictive.

The games designer, Sid Meier, was asked how he made the game so addicting. His answer was that he tried to make the game so that there was no good stopping points: he designed it such that you're always on the verge of completing a new city-improvement or a World Wonder, or discovering some new technology, and by the time you finish that, there's some other improvement that's just about to finish. These are all *urgent*, but if we pay too much attention to them, we neglect to do important-but-not-urgent things, like shutting the computer off and working on our dissertation. So Civilization II keeps us busy, and we might not have time to consider whether we're the *right kind* of busy.

Likewise, jigsaw puzzles, collecting beany babies, and solitaire make us *feel* like we're making progress on something without really accomplishing anything real. But don't we gain something from playing Civilization II or playing solitaire? They're thinking games, so we gain some skills, right? Well, yes, but there's a point of decreasing returns. I've spent over 100 hours playing Civilization II, while everything I learned playing this game could have been compressed into just a few hours. I can't help but think that those 100 hours could have been better spent.

7.1. Habits and Pain

MEME 40

Pain killers can be a form of artificial sweetener. Pain has an evolutionary purpose. Physical pain can be a way of knocking a person out of a damaging rut or habit. For example, if you're right handed and break your right index finger, pain will serve as a "reminder" to not use that finger and you'll fairly quickly learn to substitute your left hand for tasks such as zipping up your jacket. So, pain prevents you from touching your finger when it's in the delicate process of healing. It quickly breaks any habits that use that finger. Many of these habits are tacit in that you usually don't realize just how much you use that finger in your day-to-day activities. But if you use pain killers, that "reminder" will be thwarted, and your finger probably won't heal as quickly. (That being said, there are cases where our pain "notification system" itself malfunctions. When this happens, the message doesn't fit the damage, and painkillers would be useful.)

MEME 41

Evolutionarily, pain doesn't seem to make sense if there's nothing you can do about it. The purpose of pain is to cause a change of action (to lessen the pain). For example, male emperor penguins stand for a few months during the Antarctic winter incubating their female partner's egg. During this period the male doesn't eat anything, as there's no food available as far as the males are away from the sea. I doubt that these penguins feel hunger during this period because this process is an established part of the penguins' life-cycle, and there's nothing the penguins can do about it. If the penguins felt hunger, they might be motivated to fruitlessly search for food, wasting energy. I do think that it'd make sense for the males' systems to provide a heightened negative reward for physical exertion or being overly exposed to the cold and wind, since calories are at a premium.

MEME 42

Why does a tragedy, such as the death of a loved one, cause a mourning period? There's the reinforcement aspect of classical conditioning, where you try to avoid situations that caused the pain. For example, if you're learning to ride a bike you might find yourself in a situation where you're tilted far to the right side. You try turning right, you crash, and you learn not to turn to the right when you're tilted far to the right. You try again, you find yourself again tilted far to the right, you try turning left, and you crash again. Your conclusion is to avoid being tilted far to the right. Likewise, with the death of a loved one, you might try to avoid the deaths of other loved ones. This is one purpose of pain, but emotional pain might also cause "adjustments" at a more cognitive level.

With physical pain, such as a broken bone, the pain also prevents you from messing with the bone so that it can heal. But maybe there's something more to a period of depression. Perhaps you use the period as a cognitive restructure, or to "reprogram" to adapt to the change and modify the parts of your Weltanschauung that are no longer valid as a result of the tragedy. For example, suppose a close confidant dies. Confidants are useful for giving an outside perspective on problems, and now a person has to either find a new confidant or get into the habit of providing their own counsel. Similarly, if a person has a powerful friend who dies, they can't rely on that friend's support, and must get into a habit of acting less boldly. Similarly, if a man loses an arm, it would be useful for him to quickly extinguish habits and assumptions that depend on that arm. Many of these habits are tacit in the sense that we don't even realize that (e.g.) we use our arm when washing our hair in the shower. (If this were the case, wouldn't we have a similar period for a windfall, which is also a major change that could use some adaptation? Maybe the "adjustment period" is less crucial after a windfall because a windfall only expands what you can do, so old habits aren't actually harmful.)

In general, it seems that pain is a way to modify behaviour. More specifically, it seems like pain is a way to break habits, both cognitive and behavioral.

7.2. Hedonism

MEME 43

> Junkies are not interested in sex and they have no interest in other people except as suppliers of junk.
>
> –William S. Burroughs (1914-1997)

If a Reinforcement Learning agent is simply trying to take actions to maximize its reward signal, then wouldn't this lead to hedonism (the philosophy that attainment of pleasure is the goal of a person's life)? It's true that a good number of people view pleasure as the ultimate goal in life. Certainly rats do. If you hook electrodes up to a rat's brain in such a way that if they push a lever it stimulates their "pleasure center", the rats will keep pressing the lever until they die of starvation or exhaustion.

The word "hedonism" may bring bacchanalian orgies may to mind. We usually think of this word in the sense of *immediate gratification* of *sensual pleasures*. If we view hedonism in this sense, then it's not necessarily true that a Reinforcement Learning agent would be a hedonist.

1st, we would probably want our rabbits to be able to put off immediate gratification for (a larger) long-term reward. This would make sense evolutionarily, and it makes even more sense for people (who have better cognitive abilities to make predictions about the future). By associating the predictor of a reward with the reward itself, classical conditioning can give our rabbits some of this ability to delay gratification. But classical conditioning is a slow learning method. So, our cognitive system (as described in **Will: Interaction of Cognitive System and Reward System**), which is good at making predictions, often needs to take over. A person's cognitive system will often "disagree" with the limbic system about which actions to take. A person (for example, a junky) may cognitively know that action A (not shooting up heroin) will be better in the long run, yet still do action B. The philosophy of hedonism will cause a person to always choose their limbic system over their cognitive system.

2nd, it'd make sense if our innate set of rewards included some that wouldn't be described as sensual and thus wouldn't fall under the usual rubric of hedonism. For example, curiosity and social acceptance. Of course, it might be possible to learn these "abstract" rewards since they're "derivable" as corollaries from the lower level rewards, but this would take a long time in a being's life, and it'd make sense to "bootstrap" the agent by having some of these rewards be innate.

MEME 44

> Take the best orgasm you ever had, multiply it by a thousand and you're still nowhere near [the feeling of being on heroin]. When you're on junk you have only one worry: scoring [more junk].
>
> –From the movie Train Spotting (1996)

Evolutionarily, beings that found pleasure in ultimately procreative acts are the beings that procreated (and are around today). So, although I believe there's some truth to the above quote, I still wouldn't want to do heroin because I know that the outcome (in terms of pleasure) is actually worse in the long run. Most heroin or crack addicts aren't in enviable positions. Even if the pleasure were guaranteed to be maximal if I took the drugs or VR helmet, my pride gets in the way of letting me do it if I know it's fake. Drugs and VR certainly weren't anticipated by evolution. I don't know where this pride comes from though.

MEME 45

A Gedankenexperiment: Suppose you were virtually omnipotent (or you had the genie from the **Walking = Falling + Catching** section). Wish almost anything, and your wish is immediately granted. Design your life however you want to. The catch is that (like in the movie Click) you don't get to *experience* any of it. You just zoom right to your deathbed where you get to reflect on your life. Suppose you had the choice to have this power, would you choose it? (I don't even know if I would. It'd feel like I was designing someone else's life.) Suppose you weren't given a choice and you had to do it. Consider now, that (if you're lucky) some day you will be 80 years old and lying on your deathbed reflecting back on your life. What would you do then? I imagine I'd forget about pleasure and just try to maximize my legacy. Certainly, rabbits that tried to maximize their legacies would have more legacy than those that didn't.

CHAPTER 8

ECONOMICS AND ETHICS

MEME 46

As I mentioned at the beginning of **Enter Complexity: The Peacock's Tail**, the mantra "Eat. Survive. Reproduce." sounds too trivial to be the answer of the meaning of life. But hidden in this mantra is a good deal of what economists call Game Theory. This is where multiple "agents" (e.g., people) partake in a "game" and are trying to maximize their utility or reward. Evolution has structured our "reward" such that maximizing it tends to maximize reproduction (see the sections in the chapter "Taste is like the Butt, Divided" about how this reward is specified).

A classic example from Game Theory is called Prisoner's Dilemma. The situation is that there are 2 captives accused of a minor crime, and they're put into separate rooms and both are offered this deal: "If neither of you confess to the crime, you'll both get 1 year in prison, but if you confess and your partner doesn't, you'll get off free, and he'll go to jail for 5 years (and vice versa if he confesses and you don't). If you both confess, you both get 3 years in jail." Assuming the partners don't care about each other, no matter what the other guy does, each partner will be better off if he confesses. But the "tragedy" here is that both partners will confess and both serve 3 years, when they could've gotten just 1 year each.

Prisoner's Dilemma can get more complex. Suppose that instead of years in prison, the captives have to pay a fine of a mere dollar for each year they were to serve. The catch is that they have to play their game 100 times in a row, knowing what the other has done in the past. Furthermore, suppose each captive will play this new game, Iterated Prisoner's Dilemma, against each of a large group of other captives. On the face of it, it seems like each captive should always "confess", but it turns out that a strategy called "tit-for-tat" will yield a prisoner a smaller fine than always confessing. Tit-for-tat is simply starting out by "cooperating" with the person with whom you're playing (i.e., by not confessing), then doing whatever he did the last time. If he screwed you over on the last round, you screw him over this round. The tit-for-tat strategy will never gain the most against any single opponent in a single round, but it tends to yield high-scoring games. So if A, B, and C are playing each other and A and B are tit-for-tat, but C always confesses, and 100 rounds are played amongst each of the 3 pairings, the final score will be that A and B are both fined only $402, while C owes a fine of $594[1].

These types of games can get arbitrarily complex, and they can quickly push the limits of human intelligence. There are volumes of books written on Game Theory, but I'll just touch on a few of the relevant ideas in this memex.

[1]When A plays B, both will cooperate every time and each will get a fine of $100. When C plays A (or B), C will confess in the 1st round, giving A a $5 fine, but both will confess for the remaining 99 rounds giving C a fine of $297 and A a fine of $302.

MEME 47

Game Theory assumes we have a utility function. Real life isn't as straightforward as this. Because of their limited computation power, this isn't always clear cut for our rabbits. In fact, our rabbits probably won't have a single utility function. A utility *vector* might be more accurate.

At least one concept from Game Theory is useful here: a *Pareto Optimality*. Suppose you have no way of comparing apples and oranges, but you know that more apples is better than fewer apples and more oranges is better than fewer oranges. Suppose you have the following situations:

A: 3 apples and 3 oranges
B: 3 apples and 4 oranges
C: 6 apples and 8 oranges
D: 9 apples and 7 oranges

Clearly, situation B is better than situation A because we have an extra orange. We say that situation B *dominates* situation A. Both of situations C and D dominate (are better than) situation B. But what about situation C vs. situation D? Situation C has 1 more orange, but situation D has 3 more apples. Are the 3 apples worth the loss of the orange? We can't compare them. Of these 4 situations, there's no situation that dominates either C or D, so we call C and D our "Pareto Optimal set", the set of situation that aren't dominated.

∞

We could say that utility is ultimately reproduction, and those actions that will cause you to be around (i.e., reproduce) are desirable, but some situations are like rock, paper scissors: what causes you to be around depends on what everyone else is doing. For example, if everyone tries to exploit the same niche, then it won't be useful to try to exploit the niche.

MEME 48

One can derive much of ethics by applying Game Theory when there are multiple players. For example:

> RISK is interesting because when you have more than 2 players, "governments" emerge. The fundamental philosophy in RISK is "Macht macht Recht." or "Might makes right.". Ultimately, someone *will* end up conquering the world. Andy can make verbal treaties with Gabe, but it's not in the rules that they need to abide by those agreements. However, a "government" emerges when Andy becomes more powerful than me or Gabe, but not more powerful than both of us together. At this point, Gabe and I will realize that we're both doomed unless we might cooperate by forming a pact to pound on Andy at least until he's not so powerful. There's nothing in the rules of the game to prevent either player breaking the pact though.
> –From Europe Debris **[18]**, **Day 10**

MEME 49

Since people are social creatures in a cognitive niche, developing relationships consumes a good deal of our time and cognition[2]. Our (largely tacit) ethical drives are from a mix of culture and genes. I don't know how much of our "social desires" are encoded in our genes as innate Will. Certainly, much of our ethical drives are learnable, being derivable as corollaries from our more fundamental Will, but I wouldn't be too surprised if people had an innate Will for power, for example, though it could be tricky to encode this in terms of squiggly lines.

[2]The fact that we're both social creatures and that we're in a cognitive niche isn't coincidental. Many of the smartest creatures: dolphins, primates, and African grey parrots (arguably the smartest of the birds) are social. There's a book called The Red Queen **[21]** which argues that a good deal of human (and other social species) intelligence is driven by an escalating cognitive arms race. Basically, people were trying to outsmart other people. Then, the bigger brained people reproduced more than the dumber people and the intelligence level for the whole group rose, meaning that the successful people of the next generation had to be even smarter. (It's like the Red Queen's race in Alice in Wonderland where the earth moves backwards as fast as the sprinters move forward so that the net gain is 0.)

MEME 50

Question 99 from The Book of Questions [**24**] posits the hypothetical scenario: "You are driving late at night in a safe but deserted neighborhood when a dog suddenly darts in front of your car. Though you slam on the breaks, you hit the animal. Would you stop to see how injured the animal was? If you did so and found that the dog was dead but had a name tag, would you contact the owner?". Even if there was no consequence (i.e., you're sure no one will know that it was *you* who hit the dog), there's motivation to stop to see how the dog was, and to contact the owner. What good does it do you to do so? It could be that if you simply have habits of character to do good (i.e., what's good for other people), then you'll establish a reputation for this, which is good for you.

CHAPTER 9

ATOMIC WAVES AND PLATONIC FORMS

MEME 51

To say that we're all just collections of atoms is oversimplifying things. An average human body has about 6.7 octillion atoms: Hydrogen, Oxygen, Carbon, Nitrogen, Calcium, Phosphorus, Sulfur, and less than .1% each of various other elements. One fact that this oversimplification is missing is that not just any lump of the proper amounts of these elements will make a person. In fact, you need very specific compounds in the right order to make anything that we could call a person. A 2nd fact that the oversimplification overlooks is that you can replace the atoms in a person 1 by 1 and still wind up with the same person. In fact, the cells in our bodies are constantly being replaced. A person is a *wave*, a self-reproducing form. It's the same as if you wag the end of a chain and watch the ripple flow down along the chain. The wave in the chain isn't the individual links, but the motion that you created with your wag. Likewise, a person isn't really the individual atoms that make him up, but rather the *pattern* that the atoms make up. The point of all this is that at the atomic level, a person doesn't exist. What we call "a person" is just shorthand for a particular "wave" of atoms.

Since a person doesn't exist at the fundamental level of the universe, the meaning of life for a person doesn't exist at that level either. Atoms simply obey the laws of physics.

But people *do* exist in 2 senses. The 1st is that the concept of a person is a useful abstraction for describing a particular wave of atoms. The 2nd sense is deeper. Consider the concept of a circle. There is a mathematically precise formulation of this concept (a set of points on a plane that are all a certain distance (the radius) from another point (the circle's center)). However, approximations to circles are frequent in nature: the iris of our own eye, the disk of the moon, the cross-section of a tree, the orbital path of a satellite over Earth. Although perfect circles are rare or non-existent in nature, an elegant ideal "exists" in the realm of ideas. The equation for a circle and the computation for the related value of π are the same on Earth as they would be for intelligent beings billions of light years away.

Elegant ideals "exist" for more complicated concepts: ellipses, parabolas (which the lenses of our eyes approximate), gases, and perhaps machines and chemical compounds. Certainly some algorithms, such as the computation of an average, exist in the realm of ideas. In the same vein, I believe cognition (which is really just an algorithm) must "exist" in an ideal form (this idea is explored more fully by Hoimar von Ditfurth [**26**]). I see no reason why this shouldn't be the same for people and for the meaning of life.

Maybe the concepts of mind and happiness are "out there" in the same sense that the number π exists independently of people. They don't really exist in pure form, but something is trying to approximate them. Or maybe these pure concepts are the minimum distance from all the approximations (i.e. the actual examples).

Another thing to keep in mind is that ideal forms can be independent of a substrate. For example, a lens can be made out of glass, fiber cells, or even diamond. Likewise, the substrate of computation can be silicone, mechanical parts, or even neurons.

MEME 52

In a sense, a genetic line (such as humanity) is a wave too. If a person is simply a wave of atoms, then is there a difference in the type of existence for a person and the type of existence for a genetic line?

Culture, in the sense of information passed down through generations of people, is also a wave. In fact, there are cultural memes (such as an evangelical religion) that are "infectious" and self-reproducing as viruses. So, on some level, isn't a *culture* existing as an entity as much as a person exists as an entity? For that matter, I don't know that a physical object, such as a pen, *isn't* a wave. Electrons and other subatomic particles can be viewed as waves, so why not objects made out of them?

MEME 53

Here's a seeming contradiction in basic philosophies that I haven't resolved yet:

A chihuahua seems to be a living contradiction. People have bred chihuahuas to be something that probably wouldn't exist in nature otherwise. The dogs are so small that they constantly shiver unless they're in a very warm room. (I remember someone saying that her chihuahua was "too small for its body".) I escape from this "contradiction" by saying that the chihuahua doesn't actually exist. It's just a collection of molecules in a particular formation, and there's nothing about that formation that's against the laws of physics. You can resolve all sorts of "contradictions" and escape a bunch of philosophical arguments (such as ones concerning abstract ideas such as "freedom") by taking this stance.

So the 1st idea is that "The universe is just a computation.". (It could be Stephen Wolfram's Rule 110 **[28]**, a program running on a Turing Machine, or a bunch of atoms (or quarks or strings or whatever the primitives are) interacting.) Any objects beyond that are human constructs or abstractions. I wouldn't doubt that there's some principled way of making that abstraction, but, under this framework, chihuahuas don't really exist. (I might call this "raw materialism".)

On the other hand, there's the 2nd philosophy: that the concept π exists independently of matter, and that soap bubbles "try" to approximate an ideal sphere. So ideas and Platonic forms exist outside of matter, which seems to fly in the face of raw materialism. Also on this side of the debate is the idea of waves. A traffic jam "exists" independently of any of the cars in it. You could even say that a chihuahua is a "wave" since its cells are constantly being replaced. We can say a chihuahua is like Granddad's axe, which has had 5 different handles and 3 different heads, but it's still Granddad's axe. The same wave can exist on different mediums. For example, I could take some computational process running on a computer, suspend it, write the relevant memory to disk, then restart the process on a different computer, or on some weird computer that uses millions of trained crack-monkeys to do its basic logic-gate operations.

I don't know enough about physics to say, but it seems that there's the physical universe and the rules by which the physical universe abides. The latter might be called an idea-system, and the former might be an *instantiation* of the idea-system. It's possible that the rules of how an electron behaves are somehow inherent in the electron itself, but for now I'll assume they're separate. It's also possible that there's only a single possible starting configuration of matter in the universe, but for now I'll assume otherwise as well. Thus, at some level, ideas (such as the idea of a circle) exist independently of matter. There's still the question of how an idea-system gets instantiated. If I ran a simulation of a miniature universe on a supercomputer, I have little doubt that the "beings" that evolved in it would feel as real as I feel because everything it interacted with would be as "real" as it was. "Infinitesimal objects have real significance when viewed through infinitesimal eyes."

CHAPTER 10

DESIGNING YOUR SUCCESSOR

MEME 54

> What, then, when an agent can best bear The Will be steering the creation of a new agent? The Will is to reach ultimate existence.
>
> –HElmut neeman (aka Marc Pickett I), The Will (1998)

In our rabbit Gedankenexperiment, Hare Wars, we assume we know what one of our rabbits is, but a rabbit is simply a collection of atoms (an atomic wave). It's not always clear whether an animal is one of our rabbits.

For example, suppose our rabbits figured out how to genetically modify their offspring so that they could metabolize sunlight (while giving up none of their current abilities). This might sound like a good thing to do (in our eyes) because a rabbit that can metabolize sunlight surely has a higher survival and reproductive potential than one that can't. But suppose that in modifying their genes, they end up having the same sequence as the rabbits of our competing player of Hare Wars. Would we want our rabbits to do this then? Would they still be *our* rabbits.

Another way to look at it is to consider if you could genetically modify your offspring. What if "the best" doesn't resemble you at all? Should we have a directive that our offspring be like us, or would the directive to produce "the best" offspring win out? If the meaning of life is to reproduce, *to make things like you*, what about genetic engineering? For example, if through genetic engineering, I could make my child smarter, I might do this (and presumably, the child would have a better chance of survival). However, this means that I'd be making a child less similar to me than I would without the genetic engineering. This seems to be a paradox, "The best way to reproduce is to make offspring that are dislike yourself.", and we need to delve into a deeper level of complexity to resolve it. Maybe, I'd have to define "me" as something more general than my gene sequence. Maybe part of me is the cultural idea of making greater beings (and this would be passed on)...

Finally, what if we can build robots that are bigger, stronger, faster, and smarter than us. Would it make sense for these to become our descendents? What about if we could upload our consciousness to a machine? For example, we could replace our neurons one at a time with functionally equivalent silicone circuits.

MEME 55

What constitutes a "self"? E.g., would you rather be like a bacterium and have 1000s of tiny offspring, or like a whale and have a few huge offspring? Since a whale is billions of cells, does a single whale count as millions of bacteria?

In Hare Wars, suppose my opponent made rabbits so tiny that a single carrot could feed thousands of them. Would he win since he has more rabbits, or should the measure we use be the total mass of all our rabbits?

MEME 56

If our goal is to spread our genes (as in The Selfish Gene [**4**]), what about replicating our "junk" DNA? Well, this isn't useful because if the junk DNA is truly junk (i.e., it has no phenotype), then it hasn't caused its own replication.

MEME 57

Initially, all life was unicellular creatures. We are really just waves of a group of cells, and a culture is a wave where *we* are the medium. A culture is, in a sense, an organism that can reproduce. What level do we look at when we say we want to reproduce? If a cell reproduces too quickly in our bodies, *it* is successful, but this process, cancer, is never good for us. Likewise, cultural memes can compete with individual reproduction. For example, a culture that causes people to adopt African babies would spread itself because the adopted babies would learn much of their foster parents' culture, including whatever habits of thought led them to their decision to adopt the African babies, but this culture would be detrimental to the individuals' genetic reproduction because they won't have as many resources for their own biological children.

∞

A "cultural organism" should be mostly unconcerned with what its medium is made out of, just as a biological creature should be mostly unconcerned about the cells that make it up. A cultural organism should only be concerned with these as far as its survival is dependent on having hosts and how different hosts have different effects.

CHAPTER 11

ENTROPY: A PETRI DISH FALLING INTO A VOLCANO

MEME 58

A rabbit's success is not measured by how many children it has, but better by how many grandchildren it has, even better yet by how many great-great-great-...grandchildren. This is another complicating matter for the mantra "Eat. Survive. Reproduce.". It's not about simply having as many babies as possible, because you need to provide for these babies so that they can have babies as well. However, if entropy is unstoppable, then this number (at infinity) is 0 *no matter what the rabbit does*. This would mean that the success rate of *all* rabbits is the same.

This is related to a question I've been thinking about for at least 12 years, and to which I haven't found a satisfactory answer. "What would you do if the Earth would end in 1 year?" In a sense, it doesn't matter what you do because the same outcome happens. But the answer can't be that life has no meaning. You *must do something*, and there has to be *some* criteria on which to base your decisions.

Suppose you have a petri dish full of bacteria that's falling from a high altitude towards a volcano, such that when the dish hits the lava, it and the bacteria will all be destroyed. The bacteria don't have any means of escaping. What should the bacteria do? What will the bacteria do? Well, they'll probably keep doing their thing exactly as if they weren't falling toward a volcano, eating, surviving, and reproducing. The bacteria that have the *tendency* to dominate the petri dish when it hits the lava will be those most likely to be those that are around when the dish actually hits the lava.

∞

There's a card game called Falling Down. In this game, each player is falling from the top of a tall building, and the goal in the card game is to hit the ground last. The game's tag line is "It's not much of a goal, but it's all you could think of on the way down.". This is essentially the same situation as the bacteria in the petri dish.

The game Falling Down has another parallel with life in that, in life, you don't have "pre-game" time to think about what to do with life. You're thrown into life already taking actions. You can think, and you can act, but you can't "pause" the game to think about what you're going to do. Thinking is acting.

MEME 59

If I knew everyone would die in a nuclear war in 1 week, in a sense, it wouldn't matter what I did, because the rubble and ashes that would be left wouldn't be that different no matter what my actions were. Furthermore, no one would be around to care about any differences.

My abstraction of the meaning of life ("Eat. Survive. Reproduce.") doesn't handle this case. Therefore, I have to step back a level (or reason using a higher level of complexity and jump back to the "sequence of events" level). One could say that my life would be ultimately meaningless. My response to this is summarized by the following quote:

> Infinitesimal objects have real significance when viewed through infinitesimal eyes.
> –HElmut neeman (aka Marc Pickett I), The Will (1998)

What I mean by this is that we always have to do *something* (even if that means sitting around doing nothing). It's impossible to not make a decision, because even sitting around thinking is doing something. Therefore, there must be some criteria for deciding what to do. These criteria amount to what I call The Meaning of Life.

So, if it doesn't matter one way or another, then you "step down" to the next level of decisions. You fall back to your heuristics. Our Will is actually a heuristic. From a cognitive point of view, people probably aren't born with a full world model. Even with my current world model and an explicit goal of reproducing, I don't know what would cause the most offspring (or great-grand offspring).

People have much better reasoning capabilities than rabbits, but we still have our heuristics. When "the ground" is gone (i.e., when your actions can't affect the ultimate outcome, as in the nuclear war scenario), all you have is the heuristic. This heuristic is what causes artificial sweetener to taste good. Artificial sweetener isn't actually nutritious, but it tricks our taste buds into "telling" us that it is. This heuristic also tells us that protected sex or masturbation feels good, even though it doesn't increase our reproductive fitness. (This is also related to the "tricks" of the Bumblebee orchid described in the section on "Squiggly Lines".) This heuristic is also what causes people to be fat, since eating all available high-calorie foods *was* desirable 100,000 years ago because of those foods' limited quantity.

I never view pleasure as an end in itself, but as a manifestation of this heuristic. This extends to reading. Although I enjoy reading, I view its actual purpose as acquiring knowledge, which is generally useful. However, the day before Hermann Göring was to be executed, he read a book in his cell. He'd obviously never be able to put that knowledge to use, but what else could he do?

So, in the case of nuclear war in 1 week, I'd basically be left with just my heuristic and the knowledge that nothing would be left in a week. The idea behind delayed gratification is that your heuristic is overridden by rational thought. But with only 1 week to live, the short term heuristic takes over. So, I'd predict that there'd be lots of sex, drugs, and goodbyes. It'd be the wildest party ever. There's also the quote that "If you lived every day like it was your last, then you'd never do your laundry.". So, I doubt anyone would do their laundry. I'd personally probably engage in lots of sex, but not drugs or alcohol, as I'd want to have a clear mind. I'd also talk to strangers, and take risks I wouldn't normally. In a sense, there would be no strangers, we'd all be the human family facing a common doom.

If we knew that the nuclear war would be in 20 years, instead of 1 week, the heuristics in this case would be longer term, to "maximize my utility" over 20 years. This'd mean that I wouldn't work on the AI problem, I wouldn't worry about having kids, I might plan on developing a drug habit (and doing all the other debauched things) in the last year.

MEME 60

At this point (and in Isaac Asimov's short story "The Last Question"), we still don't know whether entropy is reversible. If this is the case, then a default answer to "What is the meaning of life?" is to figure out how to reverse entropy (and thereby allowing us to survive and reproduce indefinitely). If entropy is not reversible, then we're the bacteria in the lava-bound petri dish.

11.1. Nihilism

MEME 1568

In one sense, the universe is dead. That is, the universe is a collection of matter that is simply following the laws of physics, coldly carrying out calculations. For example, my brain, *the machinery for my soul*, is simply a collection of atoms, each of which is obeying the laws of Chemistry.

A problem of Philosophy is whether it is possible to get "should" from "is". That is, given a *description* of a situation, is it possible to say how the situation should go? Is there an objective measure for this? In the universe, when you look very closely, there is only "is". That is, an outside observer only sees a description of things. We're all just a collection of atoms. "Should" is an abstraction of things that *are* (see "Atomic Waves and Platonic Forms"). "Should" is contained entirely within the universe. "Should" is a property of our brains which are contained within the universe, so we *can* get "should" from "is" simply by abstracting. On the one hand, I can describe a cockroach as a physical system, but on the other hand, it might be expedient for me to use the *abstraction* of Will and talk about what the cockroach *wants*.

MEME 1569

As in MEME 53, if I ran a simulation of a miniature universe on a supercomputer, I would be able to describe the simulation as bits on the computer, simply states of information, particular configurations in the computer's memory. On the other hand, beings *within* the simulation would find the "objects" in it very real (as real as they themselves were). Even the idea of a "physical object" in this simulation is just an abstraction for a particular information states. There are certainly no objects in the computer's physical memory. Nor are there colors, though it might be useful to talk about the "red" chair in the simulation. But "red" is really just a numeric value. In the simulation, we could probably call change the name of the idea that we had been calling "red" to "blue" or "forgnorp" even, and the simulation would still be the same. It's still all just states of electrons.

Likewise, biological life has arisen from *within* the dead universe. The human mind and notably our Will are products of the universe, and Will is entirely within our minds. Thus, Will is *created*.

MEME 1570

For any person (or robot) who is forced to take actions, Nihilism, the belief that life is meaningless (or that the meaning is arbitrary), is impossible. Every action must be based on some set of criteria, and *that* set of criteria is the meaning for that person.

A person can say I've *decided* that *my* meaning of life will be to paint everything green, but that decision was based on some criteria as well. Ultimately, these criteria ground out in a design created over millions of years of evolution, with the ultimate goal of reproduction (as in MEME 1588).

∞

Finally, even if you're not convinced that there is a meaning of life, it would be hard to prove that there isn't. Meanwhile, we can create AI, which will help us determine whether there is a meaning of life, and if there is, it will help us fulfill it.

CHAPTER 12

CONSCIOUSNESS AND DEATH (HOW A PERSON NEVER DIES)

MEME 61

Any discussion on the meaning of life would be incomplete without a mention of death. Death is the cessation of consciousness or sentience. Consciousness, sentience, and self awareness are related, but different. We can say that a patient is conscious if they respond to their environment. A person is self aware if they are able to discuss themselves and reflect on their own thoughts. Sentience is the most difficult to define. A person is sentient if they feel or are aware in the same sense that we are. Sentience is possibly the most slippery concept I can imagine. This is because at once our own sentience is the fact that we're most sure of (Cogito, ergo sum.), and even if we accept our senses, other people's sentience must be taken on faith. It's nearly impossible to prove that *someone else* is sentient. Even if we've proven that another person's intelligent, they could be some automaton or zombie "acting" like they're sentient.

For this reason, I'm loathe to put a lot of thought into sentience. My belief is that cognition will be easier to figure out because, though it might be difficult to understand, cognition is at least tangible. We can test whether a person or robot has some degree of cognition. The process of understanding cognition *may* lend some insight into our understanding of sentience.

MEME 62

> Jeder einzelne Mensch ist schon eine Welt, die mit ihm geboren wird und mit ihm stirbt, unter jedem Grabstein liegt eine Weltgeschichte.
> Every single person is a World, which is born and dies with him. Under every gravestone lies a World's history.
> –Heinrich Heine (1797-1856)

What a tragedy when someone dies. All that knowledge acquired over a lifetime, the entire tacit Weltanschauung, all gone to rot. A million secrets taken to the grave, synapses destroyed.

∞

If sentience has an ideal or Platonic form (and I think it does), then I'm sure that this form isn't meant to have a finite life. For example, I can imagine a design for a cognitive system that's supposed to keep learning and predicting indefinitely. If its World is complex enough, it doesn't seem like this system would ever reach a natural stopping point. 80 years (or even 120) seems like much too short a time for a cognitive system as complex as a human brain[1].

[1]It might be interesting to consider how much information a human brain can hold, and thus what its "steady state" would be. If people lived thousands of years (but had the same cognitive capacity they have now), there'd be some point where forming new memories would mean losing old memories. Except for significant events, a person's "working memory" might be just a few decades. It's hard for us to not remember where we were when we heard about the news of September 11th, 2001, but if we lived forever, eventually this information would be drowned out by even more significant memories.

MEME 63

> The syllogism he had learnt from Kiesewetter's Logic: "Caius is a man, men are mortal, therefore Caius is mortal," had always seemed to him correct as applied to Caius, but certainly not as applied to himself. That Caius – man in the abstract – was mortal, was perfectly correct, but he was not Caius, not an abstract man, but a creature quite, quite separate from all others. He had been little Vanya, with a mamma and a papa, with Mitya and Volodya, with the toys, a coachman and a nurse, afterwards with Katenka and with all the joys, griefs, and delights of childhood, boyhood, and youth. What did Caius know of the smell of that striped leather ball Vanya had been so fond of? Had Caius kissed his mother's hand like that, and did the silk of her dress rustle so for Caius? Had he rioted like that at school when the pastry was bad? Had Caius been in love like that? Could Caius preside at a session as he did? "Caius really was mortal, and it was right for him to die; but for me, little Vanya, Ivan Ilych, with all my thoughts and emotions, it's altogether a different matter. It cannot be that I ought to die. That would be too terrible."
>
> –From "The Death of Ivan Ilych" (1886) by Leo Nikolayevich Tolstoy

> It is not I who will die but the world that will end.
>
> –Ayn Rand (1905-1982), when asked about death

I think of understanding one's own death as being similar to understanding consciousness. Actually, understanding one's own death is understanding the end of one's own consciousness (the only consciousness one is sure of).

We're all immortal. This sounds farfetched, but consider this: We can never fully perceive our own death because our brain stops. It's a little like being unconscious, our experience goes from conscious period to conscious period. For example, if we undergo general anesthesia during an operation, we'll be aware of laying on the operating table, then the next thing we'll know is that we're coming to after the operation. We have no idea of how much time has passed. When I was 11, I played the red-face game and passed out (kids, don't be as stupid as I was). I was out for only a few seconds, but it could've been years. It was virtually the same as when I was put under for a minor operation and came to almost an hour later.

We've all already "experienced" a period similar to death: the time before our conception, when we simply don't exist.

Suppose that my computer, Lappy, somehow became sentient (say by an implementation of my cognitive architecture called "The Marchitecture"). If I suspended the Marchitecture process (I'm using "process" in the Operating Systems sense), then Lappy would cease to be conscious. If I slowed the process down, Lappy would just see the world going by faster. I could even theoretically suspend the process for several millennia, and Lappy wouldn't know the difference until I continued the process. If I killed the process altogether, Lappy would be completely unaware of it.

When someone *else* dies, that person stops, but when *you* yourself die, the entire world stops, in a sense. So our own death is quite different from the death of other people. Just as I don't know of a way to *prove* that anyone else is sentient, but I *know* that I'm sentient.

So, fully believing and understanding that *I* will die is equivalent to fully understanding that *existence* will cease, which is pert near impossible. More to the point: a person is immortal himself because if *other* people die, the world goes on, but when oneself dies, the world ceases. So, there's no such thing as the world existing with oneself dead.

MEME 64

Suppose instead of creating our rabbits, we simulated them on a computer (ignoring the fact that this would require an enormous amount of computation). We could slow down or speed up this simulation, or stop it altogether, and the simulated rabbits would be none the wiser. Value (e.g., the meaning of life) only exists *inside* our simulation. For us on the outside, these are just states of the computer's circuits. Value is created in the system, and to the rabbits in the system, it is quite real.

If I *pause* the simulation then continue running it, the rabbits will be completely unaware that I paused it, even if I pause the simulation for 100 years or a trillion years. For our rabbits, the time lag is imperceptible because they have nothing to perceive it with. It's also imperceptible to the rabbits if I *never* continue the simulation, if I simply throw my computer off a cliff. Likewise, when a person's unconscious, their brain has effectively been "paused". The difference between death and unconsciousness is that in unconsciousness, the "program" is resumed, whereas in death, the program never starts again.

MEME 65

For a cow, its muscle serves a function: it's a "tool" for pulling its bones together. For someone eating a hamburger, the muscle is a collection of proteins.

Our brains can be viewed as a collection of physical neurons. At a higher level, intelligence and sentience are emergent properties of interactions of these cells. This is difficult for me to grasp. A neuron is a physical object that I could actually touch, but sentience seems intangible. When a brain is alive, it's a sentient being. When it's dead, it's a lump of inert matter.

MEME 66

> Death is nothing to us, when we exist, death is not yet present, and when death is present, then we do not exist. All sensation and consciousness ends with death and therefore in death there is neither pleasure nor pain. The fear of death arises from the false belief that in death there is awareness.
> –Epicurus (341-270 BC)

Dying is the *event* of going to existing to not. Any creature with a nervous system sophisticated enough to support the notion, will have a strong aversion to death. In fact, animals as simple as mosquitos will take measures to avoid death[2]. So, it's natural that people (or our rabbits) will have a strong aversion to death as well.

Robots probably won't be sentient at all like people. Movies, such as AI, Terminator, The Matrix, Star Wars, Star Trek, and Short Circuit, all overly anthropomorphize. The Will that evolution has installed in people is so deep and ever-present, people don't realize that it's even there (or when they're wrongfully applying it). (Until 400 years ago, people did the same thing with gravity.) It might be completely contradictory to our intuitions to imagine a robot that "enjoys" abuse, or that doesn't mind things we consider very unpleasant, but consider the bulldog: A bulldog *likes* to mate with other bulldogs. I certainly think that'd be unpleasant, but the bulldogs seem to enjoy it.

In **Will: Interaction of Cognitive System and Reward System** I argue that it may be possible to have an intelligent system with no Will. If you had such a system, I'd be willing to say it's as self-aware as I am, but its sentience is something different. Since it'd have no Will aside from its cognitive "Will" to make predictions (i.e., it'd have no external reward signal), it wouldn't care about anything.

> The mind of an intelligent robot would probably have some major differences from that of a person. People came about because of evolution, and this process installed in people an innate Will for self preservation, for example. A robot I designed wouldn't necessarily have such a Will (unless I programmed it in). Without this Will, I could tear off the robot's gripper and throw the robot off a cliff, and the robot would *feel* none of the terror that a person would feel. If it were smart enough, it would come to the conclusion that it wouldn't survive the fall, but the robot would look at the situation with as objective nonchalance as if it were observing a cloud being whisked apart by the wind. I could give the robot goals, and maybe even a reward signal, but I'm not sure that the robot would *feel* pleasure when I set its reward signal to `high`.
> –From Europe Debris [**18**], **Day 06**

A cognitive system (such as a robot) finding itself in existence doesn't have to be "life" in the reproducing biological sense. It wouldn't be a product of evolution. Reproduction is the *biological* meaning of life. Therefore, a robot's goals don't have to be those of my rabbits. So, a robot wouldn't want rights or anything else, unless it was programmed in such a way as to have this happen. It would be folly to program the robots in such a way though, and this would most likely have to be explicit because the odds of this happening by accident are slim. Likewise, it'd be a mistake to make self-reproducing robots. That'd be shooting yourself in the foot. Robots' Will should always be subservient to the Will of people.

∞

Why are pleasure and pain different from other sensory signals? Maybe the answers are in **Will: Interaction of Cognitive System and Reward System**. If we gave our (purely cognitive) robot a human body with a full sensor suite, and we held its finger to a hot surface, it'd receive a signal that would say something like "nerve # 10228 is reading a value of 0.98, nerve # 37128 is reading a value of 0.95, etc.". A human brain would "know" that these particular nerve readings mean that a certain area of the skin on our finger is being damaged, and the person-brain would probably issue commands to jerk our hand away. To the robot, this is just another signal.

[2]The males of some insects and spiders will take actions to be eaten by a female after they've mated. In this case, the Will to reproduce (which is the ultimate goal) outweighs the avoidance of death.

MEME 67

Both my parents dated other people before they met. It seemed like a narrow margin that they ended up with each other. It's easy for me to think that if my dad had stayed with his previous girlfriend, then I would've had a different mother, and it's not too hard to think that if my mom would've stayed with her previous boyfriend, then I would've had a different dad, but I can't hold both views simultaneously.

Likewise, a woman has 1,000s of eggs, and a man has billions of sperm. By sheer luck, I was a particular egg and a particular sperm. Had either been different, I wouldn't be me, I would be a sibling that doesn't exist, and *I* wouldn't exist at all.

∞

"Why am I me at this time in history (and not somebody else at some other time)?", "If I'm just a collection of molecules, how can I have feelings?". I couldn't conceive of the void that comes after death (and before birth) or how this lucky arrangement of molecules resulted in my being. Everything I've ever perceived or thought has been the result of about 3 pounds of matter in my skull.

MEME 68

> Death is the release from all pain and complete cessation, beyond which our suffering will not extend. It will return us to that condition of tranquility, which we had enjoyed before we were born. Should anyone mourn the deceased, then he must also mourn the unborn. Death is neither good nor evil, for good or evil can only be something that actually exists. However, whatever is of itself nothing and which transforms everything else into nothing will not all be able to put us at the mercy of Fate.
> –Lucius Annaeus Seneca (4 BC - 65 AD)

Marc being not-alive is the natural state of the universe. For its approximately 12,000,000,000 year existence, Marc has been alive for less than 32, and will certainly be alive for less than 150 years compared to the total past and future existence of more than 30,000,000,000 years. Despite that, the entirety of my experience with the universe has been and will be during the time that Marc's alive. (The same is true for places: most parties don't have Marc, yet I've only been to parties where Marc was there.)

MEME 69

What is it that makes a person? I can say that I'm not really my liver, for example, because I can have a liver transplant without really altering what I consider my fundamental self. If it were technologically feasible, I doubt I would say the same thing about a *brain* transplant. In fact, if I swapped brains with another person, I would find "myself" in the other person's body. This operation might be better thought of as a *body* transplant[3].

So we might say that what we are is really our brains, but this idea can be challenged too. To begin, the skin cells of our body are gradually replaced. This goes for most other organs too. So, our livers might be a "wave" like Granddad's axe (which has had 5 different handles and 3 different heads, but it's still Granddad's axe). It was once believed that, after a certain age, we no longer produce new brain cells. That, unlike most other cells in our body, brain cells are there for life. More recently, this idea has been challenged. Regadless of what actually happens biologically, in principle it's possible to replace our brain cells one at a time and end up with a "new" brain that behaves identically to the "old" brain. If the replacement is seamless, we wouldn't be able to notice that our brain cells were being switched out.

Now, suppose instead of replacing brain cells with other brain cells, we replaced them with functionally equivalent *mechanical* brain cells. Again, if the replacement is seamless, we wouldn't be able to notice the difference. But this time, instead of ending with another brain, we have a *machine* that we call ourself. Now, suppose that instead of housing this machine in our skull, we connect the brain stem to remotely controlled interface that sends inputs to and receives outputs from the machine wirelessly. Assuming the interface is seamless, the machine would still be *us*.

Finally, suppose that instead of "running" the actual brain-machine, we *simulate* it on a supercomputer, and interact with our body via the wireless interface. This would still be *us*, but now what we call *us* is just a *process* on a supercomputer.

So, a person's consciousness, their being, is a collection of ideas and a set of processes on those ideas. These processes are basically organizing the ideas and using the ideas to find a way to accomplish their Will. Their Will is basically to spread their genes. We are our genes and our experiences, nothing more.

So, what we call "ourself" is bigger than our bodies. Given this, there are 2 more ways that what we call "ourself" might live on after our bodies die: we can spread our ideas to other people through conversations and writings, and our genes can also live on after our deaths, not only through our children, but through other relatives, such as our siblings' children.

[3]Adjusting to your new body will likely be nontrivial. At the least, you be probably be really clumsy for a while. Consider what your voice might sound like. If the transplant were with Arnold Scwarzenegger, you might accidentally hit yourself (when you meant to scratch your nose) a lot to begin with.

12.1. Free Will

MEME 70

Discussions on consciousness usually include some mention of free will, which is why I'm including this here.

The opposite of free will is where we have no choices. It's possible to be completely deterministic, yet feel like you have choices. You don't need non-determinism (or even pseudo-randomness) to be unable to fully explain your own thought processes. This is a more an issue of meta-cognition and complexity than of non-determinism.

For example, a classic chess playing computer program works by searching the "tree" of moves. It will look at all its possible moves, then all its opponent's possible responses to each move, then all its possible responses to each of those moves and so on up until a certain number of moves. At the end (or leaf) of each branch, it will use a heuristic to decide how good that branch is. This heuristic could be something simple like how many pieces it has left for the leaf board state, or it can be much more sophisticated. Suppose our chess playing program is deterministic in that it has a definite method for choosing which branch to search next. This program will search for the best move because it "wants" to win. So, we could say that the program lacks free will because it's deterministic. On the other hand, the program *chooses* the move that it judges to be best.

We can imagine giving the program a (deterministic) meta-cognitive process choosing how much search to put into each branch. So that's another layer of choice, but the meta-cognitive process is also deterministic. We can give the program a meta-meta process as well, which decides factors in the meta-cognitive process, and we can give the computer a meta-meta-meta process and so on, but all these processes can still be deterministic. The computer isn't changing what it wants. It has no way of escaping that it wants to win.

It is true, on the other hand, that the computer wouldn't be able to predict what it's going to do, because as the computer's sophistication increases with which to understand itself, its complexity increases, so it needs to understand more and more about itself. It's like a dog chasing its tail. In a sense, the computer will *feel* like it has free will because it will feel nondeterministic to itself because it has parts of its process that it isn't able to understand.

MEME 71

It's impossible for me to want something I don't want, or to choose what I want. This is because any choice must be based on some criterion. If I decided something "arbitrarily" (e.g., with a coin toss), tossing the coin was a choice. In this particular instance, my choice to toss the coin would be caused by the desire to be "free" of any governing rules. We want "free will" because we want freedom in general because constraints are usually bad for our wellbeing, and we've generalized this idea to include the "constraints" of the Will that is in our nature.

MEME 72

If we're all deterministic, can we be held accountable for our actions? Yes. Consider a mosquito. They're not very smart. In fact, I view them as stimulus/response automata (with a few internal "states"). I believe they don't have any choice in their behaviour. They *need* blood to reproduce. But this doesn't prevent me from swatting them when they land on me.

In the case of people it's more complex, but the principle's the same. If people have no choice (in a deterministic sense) in their actions, does it make sense to punish a thief, for example? Yes. The purpose of punishment is to act as a deterrent. The fact that you may be punished for doing something goes into your own process for determining your actions.

Part 2

Applications

CHAPTER 13

BOUNDLESS WILL AND FINITENESS OF LIFE

MEME 73

Und was der ganzen Menschheit zugeteilt ist,	And I want to savor in my inner self
Will ich in meinem innern Selbst genießen,	whatever's the lot of all mankind,
Mit meinem Geist das Höchst' und Tiefste greifen,	to grasp the highest and deepest with my mind,
Ihr Wohl und Weh auf meinen Busen häufen...	to heap man's wells and woes on my bosom...

–Johann Wolfgang von Goethe (from Faust, 1808)

The preceding quote by Goethe is one of my life's mottos. When faced with a choice of 2 activities where one will give me an experience I hadn't had before while the other will be more routine, I'll tend to go with the 1st, even if the 1st is more likely to be unpleasant[1]. It's my view that this philosophy is useful to have and will give a person a fuller life experience. Below, I examine the validity of my view.

On the one hand, I want to minimize the regret I might have when I'm an 80-year-old man on my deathbed[2]. When people look back on their lives, they tend to regret the things they haven't done, rather than the things they have done. I suspect my regrets will be along the lines of "I should've been braver, bolder, and less lazy." rather than "I was way too daring. I should have settled into more ruts.".

On the other hand, what's the point of a full life experience and a head full of knowledge of the world if they just die with you? Is a life's experience useful only insofar as it leaves a *legacy*? To that end, how much do the philosophies that lead to a fullness of life help a person reproduce and otherwise leave a legacy? Should we make our rabbits adopt this philosophy?

[1]This is true up to a point. I've never experienced losing a limb, but I wouldn't chop my foot off because that would limit my future possibilities.

[2]When faced with a situation where I'm feeling lazy or afraid, I might imagine myself as being controlled by my 80-year-old bedridden future-self, as if future-Marc's controlling a character in a video game. This is a good motivator, but there are occasional situations where I'm not sure what future-Marc will want to have done.

MEME 74

Those who eat fugu soup are stupid. But those who don't eat fugu soup are also stupid.
–Japanese Proverb[3]

Audentes Fortuna iuvat. (Fortune favors the bold.)
–Virgil (70-19 BC)

Life is inherently unsafe. What feels safe will end you up as a bitter old man (after which you die anyway). This seems hardly a better fate than dying in a blaze of glory.

[3]Fugu soup can be deadly poisonous if prepared improperly. The soup is said to be quite good, though. As I understand it, the meaning of this proverb is that it's foolish to eat fugu fish because it's potentially deadly, but it's also foolish to refuse to take any risks in life because you'll miss out on so many experiences.

MEME 75

When I was 9 years old, I used a compass, ruler, and protractor to make a circular chess board. This chess board had the same dynamics as a regular 8 by 8 grid, but it was "warped" onto an annulus. I naïvely thought that someone who was into chess would be interested in playing on the circular board. It'd be a new kind of chess, "circular" chess. I was wrong. The differences between "square" chess and circular chess are entirely superficial. It's the same game underneath.

∞

When I 1st learned to ski, I remember thinking at a very low-level action level. I was thinking about putting my left leg into a particular position, or shifting my weight to my right leg. After I learned to ski well enough, my thoughts were rarely at that level. Instead my thoughts were about avoiding particular obstacles, or trying to figure out a route through the trees or moguls. I had *abstracted away* my conscious thoughts about low-level motor actions, and turned them into macro-actions.

∞

I remember the excitement when Nintendo released its "Wii" gaming system: an entirely different kind of controller, one that worked on motion detection rather than buttons. When I 1st started to play Wii, the game *did* seem different. It took a little while to get used to the new kind of play. But after several hours of play, I had abstracted away the controller. That is, I wasn't thinking about the controller any more, but about how Link was going to acquire some particular item, like a new sword. It turns out that the "entirely different" controller didn't matter at all. It was almost entirely a surface difference.

∞

My entire life, I've had more or less the same heart-rhythm and breathing pattern. This is predictable and boring on one level, but is unimportant and abstracted away (like the Wii controller) on another level.

So what do I mean by deeper meaning vs. surface features? I don't need to drive every color of car because the color is a surface feature.

∞

It's important to have a childhood. That is, it's important to have an exploratory period so you can try things like driving different colored cars and realize that this is just a surface difference, with no real bearing on the actual function of the car.

MEME 76

Immanuel Kant spent his entire life within 100 miles of Königsberg. His life was rumored to be so predictable that there's a story about neighbors setting their clocks by his daily walks. At the same time, Kant definitely lived an extraordinary life, coming up with deep philosophical ideas that no one ever had before.

MEME 77

When we 1st learn to write out the alphabet, we have a virtually unlimited range of styles that we might use. By the time we've graduated college, our handwriting is fairly solidified. Unless we see some reason to change it (e.g., for legibility), we stick with whatever handwriting style we're used to. It'd take a long time to write a letter if we had to rethink our style every time we wrote a word. Changes, if they exist, to our handwriting after a certain age are usually gradual. The script on papers I wrote over a decade ago is still recognizable as my own. The choice of our handwriting style is rarely consciously decided, but I'm willing to bet that a child has the ability to emulate almost anyone else's handwriting, such that the other style becomes their own.

MEME 78

> The infinite possibilities each day holds should stagger the mind. The sheer number of experiences I could have is uncountable, breathtaking, and I'm sitting here refreshing my inbox. We live trapped in loops, reliving a few days over and over, and we envision only a handful of paths laid out ahead of us. We see the same things each day, we respond the same way, we think the same thoughts, each day a slight variation on the last, every moment smoothly following the gentle curves of societal norms. We act like if we just get through today, our dreams will come back to us.
>
> And no, I don't have all the answers. I don't know how to jolt myself into seeing what each moment could become.
>
> ...
>
> All you see is the choice between working and slacking off. There are so many adventures that you miss because you're waiting to think of a plan. To find them, look for tiny interesting choices, and remember that you are *always* making up the future as you go.
>
> –Randall Munroe (XKCD)

Our brains are like closet space: we find ways to fill them. It's possible to let unimportant small decisions consume us. For example, when I worked for Sandia Nationals Laboratories in California, I had a 3-day weekend every other week (on what was called a 9-80 schedule). I lived there a year, but never made it to Sequoia National Park, which was only a few hours drive away. More than once, I spent the entire 3-day weekend playing a video game called "Heroes of Might and Magic III". My brain was completely consumed by questions such as "Should I invade this castle or produce more ranged units before doing so.". Because I was consumed with the game, I didn't take sufficient time to step back and consider the option of turning the game off and going to Sequoia National Park.

Likewise, as Randy Pausch mentioned, if you spend all your time polishing the underside of the banister, you won't have as much time for important things.

MEME 79

Why a person should go to Preakness once in their lives:

(1) A person's *experience* is limited.
(2) There are a combinatorially huge number of possible experiences a person can have.
(3) Even if we exclude all the absurd possibilities, the remaining possibilities are still combinatorially huge.
(4) Even the number of events that *have* happened or will happen (never mind the events that are conceivable but don't happen) in a human lifetime are vastly more than can be experienced by a single person.
(5) Human nature errs on the side of consistency and safety.
(6) Consistency makes life seem shorter.
(7) Our rabbits (and us) want immortality.
(8) We want minimal regret when we're old.
(9) A fuller experience gives a more accurate Weltanschauung.
(10) Therefore, a person should strive to experience as much as possible to have the fullest life experience.
(11) Preakness (for me) is further outside my other experiences than most other activities I could think of doing. Thus, going to Preakness (if I haven't been before) is that activity that will give me the widest experience.

Couldn't the same argument be applied to Why a person should be raped by Bobo, the A.R.R., once in their lives? With Bobo, the cost is too high. But how do you decide what that line is?

MEME 80

A foolish consistency is the hobgoblin of little minds.
–Ralph Waldo Emerson (1803-1882)

You, the partner in my journey, you have built your existence like a termite. You sealed off all holes outside, and rolled like a hank in a cocoon of habits, in a choking ritual of daily life. Despite it causing your madness, you have built a fortification out of this ritual against hurricanes, tides, stars and feelings. It is exhausting to you enough to forget about your condition as a human. Now, the clay, out of which you were built, got dry and hard. Nobody will ever wake in you an astronomer, musician, altruist, poet, or man, who maybe once inhabited you.
–From the movie Day of the Wacko (Dzień Świra) (2002), translated by Błażej Bułka

There are arguments for specialization. For example, there's Economy of Scale and Division of Labor. It's more efficient to have a company that specializes in making pencils and another company that specializes in dry-cleaning laundry than to have both companies making their own pencils and doing their own laundry. If we spend all our time dabbling in a huge range of activities or areas of knowledge, we won't become experts in anything. Many fields, like neurosurgery, require years of focused dedication, practice, and learning to become proficient. And in a world full of dilettantes, an expert neurosurgeon would be in demand (and he would be able to reap the rewards of that demand).

This being said, people's natural Will and behaviour tend to err too far on the side of routine and specialization. Because of cultural, cognitive, and genetic inertia, we're overly influenced by fear and laziness[4]. Food used to be more scarce than it is now, so laziness, which helped conserve a scarce resource, is less useful than it used to be.

Fear is also less useful than it used to be. Generally, people are overly cautious. This is caused by cultural and maybe even genetic inertia that made sense 400 years ago when our "safety net" was much smaller. 400 years ago, starving on the street was a more real possibility in the western world than it is today. Medicine also wasn't as advanced and most occupations were more physical, so a broken leg could be more physically debilitating. And a loss of a functioning leg had a higher likelihood of causing you to lose your means of support (which was more likely to be some sort of physical labor).

Mental laziness might be why we find ourselves "rolled like a hank in a cocoon of habits". Thinking costs calories, which have been historically expensive: the brain uses 20% of our metabolism despite being only 2% of our bodyweight. But don't habits make sense? Don't habits free up cognitive resources? It's human nature to fall into ruts. Habits make life easier. For example, if I had to rethink about how to dry myself off every time I took a shower, or how to walk or write, I wouldn't get much done. Likewise, when I shattered the top joint of my right index finger, I found out just how many habits I had. I was painfully reminded of my habits every time I "forgot" and tried to use the finger to tie my shoes, get dressed, etc.. I try to minimize cognitive overhead by consistency: e.g., always putting my keys in the same spot. Habits are how we avoid having to think about things. On the other hand, ruts and an aversion for the unfamiliar can prevent a person from fully experiencing life, causing him to live the same day over and over. Furthermore, people tend to get trapped in cycles too easily. Habits are the embodiment of laziness and fear. Both of these have their uses, but our brains developed in an era when physical threats were plentiful (so fear was necessary), but food was not (so laziness was necessary). So, people are "tuned" to have too much of these for our current environment. Both our instincts and our culture are victims of this environmental inertia. So habits can be useful, but it's important to be mindful that humans' natural tendencies err far on the side of consistency. Knowing this, I try to consciously remind myself of this, and will often take deliberate actions to step outside of my comfort zone. I also have to ask myself what the potential harm is in getting into a routine.

Even the neurosurgeon's training would suffer if he keeps doing the same routine operations. My friend, Gabe, said when he was in high school, after a heavy snow he would drive his car to an empty parking lot and do donuts and otherwise play around. Although the authorities discouraged this, he pointed out that this was a useful experience for him: it taught him the dynamics of the car in non-standard conditions and how to handle the car if he ever starts to skid on some ice. Thus, his exploration earlier probably prevented him from getting into at least a few accidents.

[4]The fundamental idea behind the phrase "Carpe diem." or "Seize the day." is to prevent fear and laziness from overly influencing one's actions.

MEME 81

Moderation is liberation.
–Jason Pickett (1998)

There are arguments for having a wide range of experiences. The *Pareto Principle*[5] (also called the 80/20 rule) states that, for many systems, roughly 80% of the effects come from 20% of the causes. To attain an expert level in (or 100% mastery of) many domains, such as hang gliding or neurosurgery, a good deal of time must be invested, but a 90% mastery will probably take roughly 10% of the effort. For example, suppose you want to learn to a new language. Zipf's law states that, "In a natural language (such as Swedish or English), the frequency of a word is inversely proportional to its rank in frequency.". So the 100 most frequent words in written English (`the`, `of`, `and`, `to`, `a`, `in`, etc.) will account for nearly *half* of the words in a document. So, if you learn just these 100 words, you'll already be able to understand half the words in a given document. If you learn 1000 words (and the basic rules of grammar), you'll be able to understand most sentences in a language. But the typical vocabulary of a language contains tens of thousands of words. English is estimated to have at least 50,000 root words. Even the artificial language Esperanto, which prides itself on its small vocabulary has 15,000-20,000 root words. So a complete mastery of the vocabulary of a language would take over 50 times the time to learn "a basic understanding". Thus, a person could attain basic understanding in 50 languages, or totally master a single language. Similarly, you can reach 90% mastery of being a Family Practice doctor with the first 10% of the effort (since most patients are the "common cases" for which you quickly learn what to do).

On the other hand, there are cases where expertise is necessary, where having a 90% mastery simply won't do. For example, a sprinter who runs 90% as fast as the fastest won't make the olympics, even if he also throws the javelin put 90% as far as the best javelin. On the other other hand, such an athlete might be preferred company on a spear-hunting expedition.

[5]This is not to be confused with the concept of *Pareto Optimality* from MEME 47. Though both are named after the Italian economist Vilfredo Pareto.

MEME 82

> Happy families are all alike; every unhappy family is unhappy in its own way.
> –The 1st sentence of Anna Karenina (1873), by Leo Nikolayevich Tolstoy

In developing a model of The World, its useful to know about outliers. Although most cases are common, it's useful to spend a disproportionate amount of time learning about the "unhappy families" because the variety in these 10% is so big. For example, if we wanted to write a 100 page pamphlet on mammals you might see in Colorado, a "proportional" pamphlet would have 90 pages of the book devoted to the ubiquitous prairie dogs, with chipmunks, bison, elk, big horn sheep, black bears, mountain lions, etc. crammed into the remaining 10 pages. In reality, although you might devote a few more pages to the more common animals, a more useful pamphlet would have roughly the same number of pages for both prairie dogs and mountain lions[6]. Experience has a number of dimensions that are difficult to capture through other means (books, photos, films). Thus, it's hard to find a substitute for learning through experience. This is another reason it's useful to have experiences far outside what you've already experienced.

Learning about a wide variety of fields also allows for generalization and transfer of principles. For example, I've learned to ski, hang glide, and touch type. By doing each of these, I was able to developed a theory about how low-level motor actions become abstracted (in our cerebellum) so that they're no longer conscious. This pattern wouldn't have been as easily noticeable to me if I hadn't done this range of activities. Another example would be automobile traffic and fluid dynamics. If a person learned about both, he'd see a good deal of overlap. Sometimes cars behave like a fluid[7].

A wide experience also helps broaden our perspective. With a limited set of data, we might feel that we've characterized The World, when in reality we've only glimpsed a tiny part of it. Even if a wide experience doesn't help you much to characterize the rest of The World, it will at least give you some sense of how much you don't know.

∞

Ignorance is the product of a narrow context. If we have a large set of outliers, unexplained data, or incorrect predictions, we realize how imperfect our Weltanschauung is. If we spend all our time in a limited context, we'll begin to forget that there are things outside of that context. For example, when Gabe and I were in a pub in Malmö:

> It was interesting that Gabe and I hadn't heard of the Euro 2004 until we got to Sweden. Neither Astrid nor any of her friends had mentioned anything about it. This is in contrast to the feeling at the pub where the Euro 2004 consumed the attention of everyone there. From the context of the pub (and later games in Germany and Brussels), it seemed like everyone in the world was focused on this tournament. The pub felt like "This is where everything's happening!", it felt like it was the center of The World.
> –From Europe Debris **[18]**, **Day 07**

∞

> What is laid down, ordered, factual, is never enough to embrace the whole truth.
> –Boris Pasternak (1890-1960)

There's no substitute for experience. But you can reach a point of diminishing returns, where new experiences are so similar to old ones that it's not worth the effort. For example, cathedrals in Europe, though each individually amazing, start to look like each other after a while. Also, with more and more experience, you learn a better mapping between what's written down and what the experience is actually like.

[6] Mountain lions are rarely seen in Colorado. I lived in Colorado for 18 years, and I never saw a mountain lion in the wild, though I saw thousands of prairie dogs.

[7] Sometimes cars behave very much unlike a fluid. Traffic can be complex and nonlinear.

MEME 83

Suppose you were supposed to memorize the following sequence of numbers:

```
0, 1, 3, 8, 15, 24, 35, 48, 63, 80, 99
```

If you were given only the first few numbers, you'd only be able to rote memorize them. But, as the sequence goes on, you'd start to notice a pattern: the numbers are just 1 less than the sequence of squares of 1 through 10. That is the nth number is $n^2 - 1$.

∞

In immersive language learning, we start to pick up on patterns like this too. For example, if we're learning Korean we might learn from context that "Ojingoe juseyo." means "I'd like some squid." and "Nae hoebuhkeurapeuteuneun jangoero kadeuk cha itseyo." means "My hovercraft is full of eels". At this point, you wouldn't know how to say "My hovercraft is full of squid.". But if you also knew that "Jangoe juseyo." meant "I'd like some eel.", then you could deduce the words for eel and squid and perhaps construct the sentence "Nae hoebuhkeurapeuteuneun **ojingoero** kadeuk cha itseyo.". This is another example where more knowledge allows for a more compressed Weltanschauung. Initially we were given only 2 sentences and could only reproduce those 2, but with the 3rd sentence, we are able to generalize and we can construct at least 4 sentences.

∞

Whenever I tour a new town (or even a new building), I like the feeling of everything falling into place. For example, when I first went to high school, I used a system of landmarks to help me navigate the wings of the building. Eventually, I started "connecting the dots". I would figure out that the computer lab was next to the math room and in front of the teachers' offices, whereas before, these 3 landmarks were disconnected. When things fell into place, I discarded my system of landmarks for the more parsimonious "map" that I'd built in my head.

∞

This phenomenon, things falling into place or "clicking" at a certain point after a sufficient amount of learning, happens in a broad range of areas of knowledge: how cars work, human anatomy, the history of the Roman Empire, etc.. The problem is that if we didn't spend enough time learning about a field, we wouldn't reach this point, and we wouldn't have a real understanding of the field.

The conclusion here is that it's useful to know that these click-points exist, and it's good to reach them if it's not too much effort. Knowing a lot about a lot of areas can help a person develop the meta-knowledge that will let them estimate how much effort will need to go into a new area before it clicks (and whether it's worth it).

MEME 84

> So many people live within unhappy circumstances and yet will not take the initiative to change their situation because they are conditioned to a life of security, conformity, and conservatism, all of which may appear to give one peace of mind, but in reality nothing is more damaging to the adventurous spirit within a man than a secure future. The very basic core of a man's living spirit is his passion for adventure. The joy of life comes from our encounters with new experiences, and hence there is no greater joy than to have an endlessly changing horizon, for each day to have a new and different sun.
> –Chris McCandless (from a letter to Ron Frans, 1992)

If we vary our experience, then our life at least *seems* longer and more full of events. The explanation for this is related to the "chairlift phenomenon":

> Some days [while skiing], I'd want to "master" a particularly difficult run, so I'd end up taking the same chairlift over a dozen times. In doing this, I'd experience what I call the "chairlift phenomenon": the chairlift ride would seem to get shorter near the end of the day. No, the actual ride wasn't any quicker (I timed it), it just *seemed* shorter.
> ...
> [D]uring the 1st ride on the chairlift, everything I see is new and unpredicted: an odd rock formation, a beautiful view, or a particularly treacherous looking piece of slope. When I've reached the top of the lift, there are a number of things I'll know that I didn't know when I was at the bottom. It's not this conscious, but when I get to the top for the 1st time, I might implicitly think "Just 10 minutes ago, I didn't know about (such and such a hazard) or a certain enticing tree-trail.". During my 12th ride on the chairlift, there'll be much less significant that's new, so I'll have noted or *learned* less. So the change in my brain-state (in terms of information) from the bottom to the top for the 1st run is greater than the change in brain-state for the 12th run, and in general, more change is correlated with more time passing.
> –From Europe Debris **[18]**, **Day 18**

I don't know that having life *seem* longer is useful in itself. It could be just putting our heads in the sand, but I can't escape the idea that a variety of experiences will produce a more fulfilling life. Would our rabbits do better if they wanted their lives to seem longer? Well, we'd at least want for our rabbits to want to live a long time[8], or a long-lived rabbit would survive and reproduce more than a short-lived one. I think a diversity of experience allows for a better world-model, which is useful for the rabbits' survival and reproduction.

∞

On the other hand, fragmenting your life into too many slices can make it seem shorter. I'm a slow reader. It takes me about 40 hours of reading to finish a book, and I read about an hour each day. If I read 30 different books at once, reading a different book every day of the month, it'll take me years to finish any of them. If I looked at any book, I'd think "Wow. I started this book 3 years ago, and I'm still not finished.".

[8]There are arguments that it's evolutionarily advantageous *for a species* if the individuals have shorter lifespans. The reasoning goes that a shorter lifespan allows for quicker evolution, which makes the species more adaptable to environmental changes.

MEME 85

The purpose of a cognitive system is to characterize and make predictions about The World. For reasons described above, it'd make sense for a cognitive system to want to learn as much as possible. I call this thirst for knowledge θαῦμα, which is the open-eyed wonder that impels us to explore The World, a sort of intense curiosity. Our rabbits should have θαῦμα too, but balanced by fear and laziness. After all, *exploring* The World comes at a cost of *exploiting* what is known already. How much to explore vs. exploit depends on how much life you expect to have yet, so when you're young it makes sense to explore and have a diversity of experiences.

It makes sense to be curious and acquire lots of knowledge without an obvious use.

> On the other hand, perhaps one could take a view of The World as being similar to the dictionary: I've always accepted that I'll probably never look up most of the words in the dictionary (and certainly, I'll never read the entire phone book), and I doubt that many people do.
> It could also be like having a fancy *versal* font (the font used to make the large ornate letters at the beginning of a book or chapter). You'd always want a full set of 26 of these, even if you never use some of the letters (like X and Z, those high scoring, and therefore hard-to-use Scrabble letters) for a book.
> –From Europe Debris **[18]**, **Day 15**

A drive related to θαῦμα is what I call Boundless Will. This is the Will to do everything, to pop every push, to tie up all loose ends, to travel all paths, to get to know everyone, to find all patterns, to experience all of the combinatorially huge number of possibilities. It'd make sense to give our rabbits a boundless Will (see **Walking** = **Falling** + **Catching**). So how do they (we) achieve this well?

MEME 86

I read once somewhere that the the academic being is ever a child, or at least ever filled with a child's sense of wonder (i.e., deeply motivated by θαῦμα).

I've always been reluctant to "grow up". Because after growing up comes growing old, and after growing old comes death. Perhaps I implicitly believe that if I can stall growing up, then I can stall death. It might be because of this, but I get disturbed when my friends' conversations turns to buying and remodeling their houses. It seems so mundane. Of all the boundless possible topics of conversation, all the Gedankenexperiments to be investigated, all the analysis of historical trends, the rise and fall of empires, the topic chosen is so often whether to go with oak or cherry cabinets for the kitchen. Maybe the realities of mundane life will force me into this abyss, but I think the fate I dread is at least escapable. While in high school, we were as the chimpanzees in the Arnhem zoo: the basic necessities are automatically taken care of, and all that is left is politics and peacock's tails.

If you have children, it seems you'll be faced with life's real problems: mortgages, car insurance, school-district zones. This is what will be on your plate, so this is where your interests will be. In high school, there seemed to be so many paths, and possibilities. It seems that as one gets older, one falls on this well trampled path of practicality. There seems to be a lack of friends and socialization, a lack of intellectual fervor, of philosophical pondering, of wonder, of θαῦμα. Thus, one passes from the phase of exploration to that of exploitation.

In life there are again so many combinations, but a fairly common story is to graduate from high school at 18, graduate from college at 22, get a job, marry, buy a house, have kids, then retire. This might sound uninteresting, but it works. It causes people to be successful. It does the job of reproduction. (And isn't there room for creativity elsewhere?)

Our rabbits can live The Unexamined Life and still be successful in our eyes (i.e., evolutionarily successful). Real rabbits reproduce just fine without having the barest inkling of the meaning of any of what they're doing.

MEME 87

The World can hold only a tiny fraction of all possibilities. If I'm dealt a particular hand of Gin Rummy, I only get that single hand, but there are 15,820,024,219 possible hands I'm not being dealt. Before the hand was dealt, there were almost 16 billion hands it could've been, but it had to be only one of those.

∞

There are many instruments that people have invented, and even more ways to possibly make sound. For example, I could invent an "instrument" that makes sounds by throwing various sized rubber balls against a racquetball court wall. Then there are also existing instruments such as the Jew's harp, the didgeridoo, and the theremin, but most bands that I see in Baltimore typically have a vocalist, a guitar, a bass guitar, and a set of drums. This seems to be only a tiny slice of the unimaginably large number of possibilities for a band.

I'm sure there are some musical reasons why the standard music set is better than a band with an accordian, a didgeridoo, a jew's harp, and no drums. But I can't help but suspect that the standard set is largely a historical accident, and continues to exist largely due to cultural inertia.

∞

There are also an unimaginably large number of things a person could do on a Friday night: one could build Lego creations. One could attend a painting party in which the guests all paint portraits of each other. One could play parlor games. One could lock themselves up in a library with friends, everyone grabbing stacks of books on ideas they're interested in. My friend, Dan Bernstein, recommended fetal position night, in which the attendees stay in fetal position the entire night. Yet, the selection in most cities seems to be typically limited to going to a bar or going to a theatre.

∞

Eggs Benedict seems pretty arbitrary: A poached egg and Canadian bacon on an English muffin covered with Hollandaise sauce. If we allow 2 combinations for each part of this menu item (e.g. egg over-easy instead of poached), then there would be 16 combinations. Yet none of these are specially named.

MEME 129

On one hand you have the combinatorial explosion of possible outcomes, and the tiny perturbations and chance happenings that cause decks of cards to fall into a particular order. On the other hand you have Mutual Information and convergence, which say that many outcomes are basically equivalent.

As an example of the latter, when I was 8, my parents took me to the Humane Society to get a puppy. There were dozens to choose from, and I chose one from a particular litter of 6 or 7 puppies. My decision was rather arbitrary in that the only difference I could see in the puppies was their coloration. So in choosing 1 puppy (that I later named "Oreo"), I excluded the 5 or 6 other puppies. Also, in choosing a future with Oreo, I also excluded at least 5 or 6 other futures with the other puppies. My question here is "Does it matter?". Would all these futures have been basically equivalent?

MEME 130

I only have so many days before I die, and I want to have the whole human experience as fully as possible, and it bugs me that there are some things that I'll have died without doing. Along similar lines, I view it as a little tragedy every time I pass a stranger on the sidewalk. Here's this full interesting person with dreams, concepts, a family, and a life story. If the stranger and I were stranded on a desert island for 20 years, we'd probably end up as good friends, and I'd still never fully figure the person out. Yet, all those possibilities are quashed as we walk past.

∞

Consciousness is seemingly infinite (or at least combinatorially unbound), but life is very finite ($<$ 51,135 days (140 years) and usually closer to 29,400 days (80 years)). In that time, we can't possibly conceive all possible thoughts or dream all possible dreams.

MEME 128

> From the child of five to myself is but a step. But from the new-born baby to the child of five is an appalling distance.
>
> –Leo Nikolayevich Tolstoy (1828-1910)

Our highest level concepts are probably formed very early in life, and some of these concepts are influenced implicitly by our early experiences. These might be things like good and bad (maybe the very idea of good and bad), that to have effect you have to take an action, etc..

In working on artificial developmental learning, I've come under the suspicion that the amount of new information put into our heads after the age of 3 might be *less* than the amount of information that was in our heads by the time we're 3. Tacit knowledge that we don't usually think about, like recognizing physical objects, object permanence, and inertia, might be more difficult to describe starting from scratch, than are the ideas in a college Complex Variables course, when using those fundamental concepts as building blocks. Paul Cohen (my academic grandfather) once noted that the concepts that an army General uses to plan battle strategy are fundamentally the same as those his 3-year-old daughter uses to push blocks around a table. Therefore, a good deal of the knowledge that's in my head is also in everyone else's head. So the amount of knowledge in my head that's unique might actually be pretty small. The amount of knowledge that's unique to me, and not directly related to me (e.g., no one else knows that I read a Sherlock Holmes story last night) is even smaller.

I consider this a good thing though. This is because I'm interested in the whole human experience. The day I graduated from college, I was depressed because I had the thought that "That *was* my undergrad experience.". There were all the paths that I could have taken, but that was it. I was allowed only one path. ("Einmal ist keinmal." (Once is never.) I read years later in Milan Kundera's "Unbearable Lightness of Being".) There were all the movies that you saw about college and all the crazy times, and not all those were part of my college experience. So, I want to experience everything. I want to be everywhere and know everything, but there simply isn't time in my "expected" 29,400 days. It's the finiteness of life vs. the infinity of possibilities. Mutual Information mitigates this. There are a large number of paths that all lead to pretty much the same state. So, Jim's experiences might be radically different from mine, but the distance between the information content in our heads might not reflect that. The laws of the universe work the same always and everywhere, so most of what Jim learned in his experiences might be fundamentally the same as what I learned in my experiences.

MEME 88

My current answer to the question I asked at the beginning of this section ("What's the point of a full life experience and a head full of knowledge of the world if they just die with you?") is this: a full life experience gives a person a better Weltanschauung. Just as it's useful to have a full encyclopedia even if few people read every entry in it, it's still useful to have a well-developed Weltanschauung so that a person can make use of the relevant parts if the need arises. Because of evolutionary and cultural inertia, we tend to err far on the side of fear and laziness. So a constant effort must be put forth to combat these forces.

CHAPTER 14

CONCLUSIONS

MEME 89

Although, these memes may have raised more questions than they answered, I hope they've at least given some tools for thinking about the bigger questions such as "What should I do with my life?" or "When faced with 2 possibilities, what criteria should I use for choosing between them?".

Now that we understand a bit more about where our motivations come from and how they work, we have some insight as to how to fulfill them. It's not truly the case that *our* goals should ultimately be reproduction any more than the goals of our rabbits (from their perspective) will be reproduction. Our rabbits lack the computational power to figure out what behaviours will cause their genes to be around. Even the designers (be it us or evolution) lack the computational power for all situations. But understanding how and why our Will is set up as it is may be useful for understanding why we're here and what we should do. In Reinforcement Learning terms, the Meaning of Life is to make our reward signal fire for the long term. This is different from Hedonism for reasons explained in the section on that subject.

MEME 90

> Equations are more important to me than politics because politics are for the present, but an equation is something for eternity.
>
> –Albert Einstein (1879-1955)

In **Walking = Falling + Catching**, I wrote that a person always needs goals. This begs the question of what these goals should be. What is progress? Reproduction is so broad. When we unpack what this means, it unfolds into a complex array of what it entails.

Achievements that have lasting power seem to be more satisfactory than temporary results. Having and raising children is a "project" that has staying power. Creating new knowledge (e.g., inventing, writing, exploring, and discovering) also has staying power. Being a social creature, a good deal of a person's wellbeing depends on their relations with other people, so cultivating relations also has a good deal of staying power. There's also the creation of goods or setting up social projects. We should always be falling towards something.

∞

On the other hand, if just we're living for our legacy after we're dead, can't we just pretend to have been (e.g.) Thomas Edison? Aren't we just as much Thomas Edison as are the bones lying in Edison's grave? Well, no. The question is what to do with our life while we're alive. So pretending we were Thomas Edison doesn't answer any questions about which actions to take, and we have to do *something*. Rabbits that took this approach aren't the rabbits that left legacies, and are less likely to be the rabbits that had a lasting effect on their world. A *group* of rabbits that had few inventors would be less likely to survive and reproduce than a group of rabbits that did.

So what about Isaac Newton? Here was a man who may have been the world's single greatest influence on modern science. He co-invented calculus, developed the laws of mechanics and gravitation, and developed a theory of optics. However, Isaac Newton never married and had no children. So, if Newton were one of our rabbits, would we consider him a failure? Maybe, but because of his work, his countrymen gained a technological advantage that helped spread genes like his. Ultimately, the knowledge that he helped developed has helped and will help the survival of the whole human race.

In 1879, in modern day South Africa, some 4,000 Zulus attacked a remote hospital housing 139 British soldiers (and medical personnel, etc.), many of whom were sick or wounded. In the end, the British, who suffered only a few dozen casualties (only 19 British died as a result of the battle), inflicted heavy losses on the Zulus (it's estimated that the Zulus lost 600 men), and the Zulus withdrew [**2**]. To me, this illustrates how ideas, from the invention and production of the firearms used by the British, to the clever strategy devised by Lieutenant John Chard, the British commanding officer, can manifest themselves in a reality of blood and flesh. In this battle, ideas, *such as strategy and inventions*, meant the life and death of several hundred people. Had Chard been somewhat more dimwitted, the British could have easily been wiped out by the sheer numbers of the Zulu horde.

MEME 1587

The question of *what* creates real value vs. what is only Artificial Sweetener still plagues me. For example, when between the ages of 7 and 11, I collected coins. I saved up my allowance, and instead of spending it on a bike or a sling shot, I spent it on coins. This collecting *felt* like progress, and I had the added satisfaction that I could later sell the collection and gain a profit. It has gained in value, but the collection that I spent 4 years of my childhood obtaining is now worth less than a month of my adult's salary.

There are cases where it's clear what my future 80-year-old self will have wanted, but sometimes it's not so clear. For example, when I was an undergraduate, I focused on my studies because that's what I thought future-Marc would have wanted me to do. I never drank, did drugs, or partied. My social life wasn't very active. I graduated at the top of my class and got into a good graduate program, but I've I've since regretted that I didn't socialize more and sow some wild oats. I should have been more moderate.

MEME 91

> Let an ultraintelligent machine be defined as a machine that can far surpass all the intellectual activities of any man however clever. Since the design of machines is one of these intellectual activities, an ultraintelligent machine could design even better machines; there would then unquestionably be an "intelligence explosion", and the intelligence of man would be left far behind. Thus the first ultraintelligent machine is the last invention that man need ever make, provided that the machine is docile enough to tell us how to keep it under control. It is more probable than not that, within the twentieth century, an ultraintelligent machine will be built and that it will be the last invention that man need make.
> –I. J. Good [**6**]

> Der Übermensch ist der Sinn der Erde.
> The Übermensch is the meaning of the earth.
> –Friedrich Wilhelm Nietzsche, from Also Sprach Zarathustra [**16**]

I've found I can accomplish the most by doing research in Artificial Intelligence. Even if not achieved during my lifetime, this has the potential to give me an academic legacy. But, if achieved during my lifetime, AI has the potential to think all those thoughts that I'll have died without thinking otherwise. It has the potential to figure out a way to preserve me. It would create a superbrain that could help answer the question of whether entropy is reversible (if the universe is a closed system).

Appendices

APPENDIX A

WHY RELIGION ISN'T THE ANSWER

If you wish to strive for peace of soul and happiness, then believe; if you wish to be a disciple of truth, then inquire.

–Friedrich Wilhelm Nietzsche (1844-1900)

MEME 92

I don't believe that the meaning of life has anything to do with the Abrahamic religions (i.e., Judaism, Christianity, and Islam) because fundamental to these religions is the belief in the god of Hebrew mythology, commonly known as "God".

A.1. Are Religion and Science equally justified?

MEME 93

One could claim that Science and Religion are both axiomatic belief systems. That is, both science and religion start with a set of assumptions (e.g., "God exists." for religion, and Ockham's Razor for Science) that have no backing and on which other beliefs are based. Since Science and Religion are both systems of axioms, one might claim that there is no fundamental difference between them, and therefore, belief in Religion is as justified as belief in Science.

However, Science isn't a set of specific beliefs, but rather a set of principles for generating beliefs, so the axioms are few. For example, Newton's 3rd Law of Motion (i.e., "For every action there is an equal and opposite reaction.") is *not* a core belief of science. The core tenets of science allow that Newton's generalization might not be entirely correct. When Newton proposed this law, it was the most parsimonious theory that explained observations. The tenets of Science allow both that a more parsimonious explanation might still be found and that new observations might be made which Newton's 3rd Law will be unable to explain. By contrast, the axioms of religions are specific and numerous. For example, Christian dogma states that the scriptures are divinely inspired. Therefore, the entirety of these scriptures (several hundred pages) are included in the set of Christian axioms. Within these are such statements as 1. Jesus rose from the dead, 2. belief in Jesus grants one eternal life, 3. Jesus was born of a virgin, and 4. Jesus turned water into wine.

If our goal is to understand The World and make predictions, then axiomatic systems *can* be compared. A system of axioms that has few contradictions, has high predictive power, and is parsimonious and easier to apply is more useful for that goal (and therefore has more justification) than a system that is rife with contradictions, has little predictive power, and is bloated and tangled. Science being the former and Religion the latter, belief in Science is more justified than belief in Religion.

A.2. Proof of God's Non-existence

MEME 94

It's sometimes claimed that it's impossible to disprove the existence of God. While it's true that to disprove the existence of some things, such as a diamond the exact size and dimensions of the Statue of Liberty, one would have to exhaustively search all the corners of the universe and show that such an implausible diamond doesn't exist anywhere. However, it's possible to disprove the existence of God the same way we can disprove the existence of the largest prime number: namely, that assumption of its attributes results in a contradiction. My proof runs as follows:

(1) Belief in the God of the Bible implies the belief that people have non-material souls which are sentient and live on after a person's body has died.
(2) Belief in the God of the Bible also implies that one's own senses are not deliberately deceptive because God, by definition, is righteous, has an extreme affinity for truth and honesty, is infallible, and is the creator of a person and their senses. This means that I'm not a brain in a vat deliberately led to believe in a false universe. (This was Descartes' argument for being able to believe his senses.)
(3) Belief in one's senses forces one to believe in many of the findings of neuroscience. Namely, that most of the processes formerly attributed to soul have been proven to be properties of a material brain.
(4) Since these properties are properties of a physical brain, they can't also be properties of a non-material soul, because this implies either redundancy or contradiction (in that a property of the non-material soul may disagree with that of the physical brain). If it's redundancy, then a person's soul would be damaged if we damaged a person's brain. When a person dies, and his brain is destroyed (by rot and worms), redundancy would imply that the overlapping parts of the soul would also be destroyed. Neuroscience has established that by physically affecting a person's brain, one can cause the person to have mental changes. (For example, would an Alzheimer's patient still be short of memory in his afterlife? What about a person who had suffered severe brain damage? Would their soul be catatonic in the afterlife?) To say that the brain is a physical manifestation of the soul (like a shadow is of a physical object) doesn't help this problem because shooting my shadow's legs doesn't impair my ability to run, but damaging our brain does impair our ability to think.
(5) Therefore, neuroscience implies that there's little left to the non-material soul. So little, in fact, that we could hardly consider the non-material soul to be sentient (any more than we could consider a severely brain-damaged catatonic person to be sentient). This is contrary to 1.
(6) Therefore, the God of the Bible is a self contradiction, and can't possibly exist.

One can reconcile this contradiction by altering the definition of God. However, any alteration sufficient to do this would have to be so fundamental that I would no longer apply the term "Abrahamic God" to the new concept.

One can also reconcile this contradiction by dropping 1. Namely, by claiming that one can believe in God without believing in a non-material soul. But belief in an afterlife is central to the teachings of Christianity, Judaism, and Islam.

APPENDIX B

ADDENDA

MEME 1834

Another way of looking at the memex: For the rabbits, suppose we don't worry about "happiness" the rabbits just have a signal ("reward") that they try to maximize. When the signal is high, there are mechanisms that cause the rabbit to take those sorts of actions again. It turns out that this signal (and behavior) correspond to happiness.

Thus, robots would be conscious in an entirely different way that people are unless this signal is carefully crafted to have *specific* things like self-preservation.

BIBLIOGRAPHY

[1] D. Attenborough. *The Private Life of Plants*. Princeton University Press, 1995.
[2] M. Barthorp. *The Zulu War*. Blandford Press, Dorset, England, 1980.
[3] A. R. Damasio. *The Feeling of What Happens: Body, Emotion and the Making of Consciousness*. Vintage, 2000.
[4] R. Dawkins. *The Selfish Gene*. Oxford University Press, Oxford, England, 1976.
[5] F. de Waal. *Chimpanzee Politics: Power and Sex among Apes*. Johns Hopkins University Press, 1982.
[6] I. J. Good. Speculations Concerning the First Ultraintelligent Machine. *Advances in Computers*, 6, 1965.
[7] J. Hawkins and S. Blakeslee. *On Intelligence*. Times Books, 2004.
[8] D. R. Hofstadter. *Gödel, Escher, Bach: an Eternal Golden Braid*. Basic Books, 1979.
[9] D. Hume. *A Treatise of Human Nature*. 1740.
[10] W. S. Jevons. *The Coal Question*. Macmillan & Co., London, 1866.
[11] H. Leibenstein. Bandwagon, Snob, and Veblen Effects in the Theory of Consumers' Demand. *Quarterly Journal of Economics*, 64:183–207, 1950.
[12] S. Levitt and S. Dubner. *Freakonomics: A Rogue Economist Explores the Hidden Side of Everything*. William Morrow, 2005.
[13] P. D. MacLean. *The Triune Brain in Evolution: Role in Paleocerebral Functions*. Plenum Press, New York, 1990.
[14] G. Miller. *The Mating Mind: How Sexual Choice Shaped the Evolution of Human Nature*. Anchor, 2001.
[15] T. Nagel. What is it Like to Be a Bat? *The Philosophical Review*, 83(4):435–450, 1974.
[16] F. Nietzsche. *Also Sprach Zarathustra: Ein Buch für Alle und Keinen*. Walter de Gruyter & Co., Berlin, 1885.
[17] J. Piaget. *The Origins of Intelligence in Children*. W. W. Norton, 1963.
[18] M. Pickett. *Europe Debris: The Epic of Gabe and Marc in Europe*. Lulu, 2006.
[19] M. Pickett, D. Miner, and T. Oates. Essential Phenomena of General Intelligence. In *Proceedings of The First Conference on Artificial General Intelligence*, 2008.
[20] Plato. *The Republic, Book VII*. (Public domain), 360 BC.
[21] M. Ridley. *The Red Queen: Sex and the Evolution of Human Nature*. Harper Perennial, 1995.
[22] M. Rosenstein and P. Cohen. Continuous Categories for a Mobile Robot. In *proceedings of the 16th National Conference on Artificial Intelligence*, 1999.
[23] A. Schopenhauer. *Die Welt als Wille und Vorstellung, Bde. I & II (The World as Will and Representation, Vols. I & II)*. 1844.
[24] G. Stock. *The Book of Questions*. Workman Publishing Company, 1987.
[25] R. Sutton and A. Barto. *Reinforcement Learning: An Introduction*. MIT Press, 1998.
[26] H. v. Ditfurth. *Wir sind nicht nur von dieser Welt*. Hoffman und Campe, 1981.
[27] E. R. Westervelt, J. W. Grizzle, C. Chevallereau, J. H. Choi, and B. Morris. *Feedback Control of Dynamic Bipedal Robot Locomotion*. Taylor & Francis/CRC, New York, 2007.
[28] S Wolfram. *A New Kind of Science*. Wolfram Media, New York, NY, USA, 2002.
[29] R. Wrangham and D. Peterson. *Demonic Males: Apes and the Origins of Human Violence*. Mariner Books, 1996.