The Book Review Column[1]

by Frederic Green

Department of Mathematics and Computer Science
Clark University
Worcester, MA 01610
email: `fgreen@clarku.edu`

This column consists of two reviews that only Bill Gasarch could write. Well. . . I suppose it wouldn't be *impossible* for someone else to review the first, and *not even less wrong* to write the second, but it's a treat to have an all-Gasarch program!

1. **Tales of Impossibility: The 2000-Year quest to Solve the Mathematical Problems of Antiquity**, by David Richeson. A history of the convoluted paths to solving classical problems such as squaring the circle. Review by Bill Gasarch.

2. **A Map that Reflects the Territory: Essays by the Less Wrong Community**. By contributors to the Less Wrong forum. A collection of essays from this online forum of diverse interests. Review by Bill Gasarch.

As always, please contact me to write a review; choose from among the books listed on the next pages, or, if you are interested in anything not on the list, just send me a note.

[1] © Frederic Green, 2022.

BOOKS THAT NEED REVIEWERS FOR THE SIGACT NEWS COLUMN

Algorithms

1. *Algorithms and Data Structures Foundations and Probabilistic Methods for Design and Analysis*, by Helmut Knebl
2. *Algorithms and Data Structures*, by Helmut Knebl
3. *Beyond the Worst-Case Analysis of Algorithms*, by Tim Roughgarden

Computability, Complexity, Logic

1. *Applied Logic for Computer Scientists: Computational Deduction and Formal Proofs*, by Mauricio Ayala-Rincón and Flávio L.C. de Moura.
2. *Descriptive Complexity, Canonisation, and Definable Graph Structure Theory*, by Martin Grohe.
3. *Semigroups in Complete Lattices*, by P. Eklund, J. Gutiérrez García, U. Höhle, and J. Kortelainen.

Miscellaneous Computer Science

1. *Elements of Causal Inference: Foundations and Learning Algorithms*, by Jonas Peters, Dominik Janzing, and Bernhard Schölkopf.
2. *Partially Observed Markov Decision Processes,* by Vikram Krishnamurthy
3. *Statistical Modeling and Machine Learning for Molecular Biology*, by Alan Moses
4. *Language, Cognition, and Computational Models,* Theirry Poibeau and Aline Villavicencio, eds.
5. *Computational Bayesian Statistics, An Introduction,* by M. Antónia Amaral Turkman, Carlos Daniel Paulino, and Peter Müller.
6. *Variational Bayesian Learning Theory,* by Shinichi Nakajima, Kazuho Watanabe, and Masashi Sugiyama.
7. *Knowledge Engineering: Building Cognitive Assistants for Evidence-based Reasoning*, by Gheorghe Tecuci, Dorin Marcu, Mihai Boicu, and David A. Schum.
8. *Quantum Computing: An Applied Approach*, by Jack D. Hidary

Discrete Mathematics and Computing

1. *Mathematics in Computing*, by Gerard O'Regan
2. *Understand Mathematics, Understand Computing – Discrete Mathematics That All Computing Students Should Know*, by Arnold L. Rosenberg and Denis Trystram

Cryptography and Security

1. *Computer Security and the Internet: Tools and Jewels*, by Paul C. van Oorschot

Combinatorics and Graph Theory

1. *The Zeroth Book of Graph Theory: An Annotated Translation of Les Réseaux (ou Graphes) – André Sainte-Laguë (1926)*, translated by Martin Charles Golumbic
2. *Finite Geometry and Combinatorial Applications*, by Simeon Ball
3. *Combinatorics, Words and Symbolic Dynamics,* Edited by Valérie Berthé and Michel Rigo

Programming etc.

1. *Formal Methods: An Appetizer*, by Flemming Nielson and Hanne Riis Nielson
2. *Sequential and Parallel Algorithms and Data Structures*, by P. Sanders, K. Mehlhorn, M. Dietzfelbinger, R. Dementiev

Miscellaneous Mathematics

1. *Introduction to Probability*, by David F. Anderson, Timo Seppäläinen, and Benedek Valkó.

2. *Algebra and Geometry with Python*, by Sergei Kurgalin and Sergei Borzunov.

Review of
Tales of Impossibility:
The 2000-Year quest to Solve the Mathematical Problems of Antiquity[2]
Author: David Richeson
Publisher: Princeton University Press, 2021
$30.00 Hardcover, $22.00 Softcover, 450 pages

Review by
William Gasarch (gasarch@umd.edu)
Department of Computer Science
University of Maryland, College Park

1 The Problems of Antiquity

This book is about the 4 Greek problems of Antiquity. I will state them in a way that is absolutely whiggish.

By *construction* we mean *construction with just a straightedge and compass*.

1. (Trisecting an angle) Prove or disprove the following: Given an angle θ it is possible to construct an angle that is $\frac{\theta}{3}$.

2. (Doubling the cube) Prove or disprove the following: Given a line segment of length x it is possible to construct a line segment of length $x2^{1/3}$.

3. (Squaring the circle) Prove or disprove the following: Given a circle C it is possible to construct a square S such that the area of C and S are the same.

4. Determine for which n it is possible to construct the regular n-gon.

The above description is whiggish for the following reasons:

1. I used algebraic notation. They would have stated it all in terms of geometry. For example, doubling the cube might have been stated: *given an edge of a cube construct an edge of a second cube whose volume is twice that of the original.*

2. They would not have stated the question so precisely as *prove or disprove*.

The book under review is about the progress made on these four problems from 400BC until it was shown that the first three were impossible, and the fourth was solved, in the 1800's. There are many twists and turns and other issues that arise. What is of more interest than these four problems is how math itself has changed.

Paul Erdős said of the Collatz conjecture

Mathematics is Not Ready for Such Problems.

That may or may not be true (that's a tautology). Now that we know the solution to the four problems of antiquity, we can say with authority that, when they were posed,

Mathematics was Not Ready for Such Problems.

Realize how far math had to go: Algebra was in its infancy, the concept of a number was barely understood (e.g., negative numbers and irrationals were suspect), and the notation was awful. That last point is not trivial: good notation and good ideas go hand in hand, with one inspiring the other.

[2]

2 Summary of Contents

This book has 21 chapters and 21 tangents (each chapter is followed by a tangent). To summarize all of them in a book review of this length would be as absurd as squaring the circle. So I will talk about types of chapters.

2.1 Variations

Some of the chapters go over, and vary, the ground rules for what it means to *construct*. For proving a construction impossible this is quite important. Its not good enough to say *I could not square the circle, hence it can't be done* (though if Gauss said this, it would be good evidence that one cannot square the circle).

The most important ground rule, and the one ignored by cranks, is that there *are* ground rules. If you trisect an angle with a straightedge, compass, and protractor *you have not solved a 2000 year old math problem.*

The following variations of the ground rules are discussed:

- What if you have a straightedge and compass, but the compass is fixed at one angle (a *Rusty Compass*)? What if the angle can be of your choice? What if its arbitrary?
- What if you allow a two mark on your straightedge?
- What if you allow other mechanical devices?
- What if you only use a straightedge?
- What if you only use a compass?

2.2 How Math has Changed

How mathematics changed over the centuries is not really the subject of any one chapter; however, it permeates the entire book. I give two examples.

I) Connection to Geometry.

I present this as a conversation between Darling and me.

BEGIN CONVERSATION

Bill: Do you have any objection to the equation: $x^2 + x = 10$?

Darling: No. Why should I? Is this a trick question?

Bill: You are a 21st century person who correctly has no problem with that expression. But Viete and others in the 1600's though of Algebra as being so tightly connected to Geometry that since x is a *length* and x^2 is an *area*, $x^2 + x$ makes no sense.

Darling: That reminds me of the controversy over the definition of functions. When Math and Physics were more tied together, functions that were not differential were shunned since it was thought they had no real world counterpart.

Bill: I am glad to be a 21st century person.

END CONVERSATION

When it became apparent that algebra would be needed for mathematics, people accepted it but didn't like it. Here is a quote. I will tell you who said it at the end of this review, but see if you can guess. That may be to hard for you to guess, so try to guess when it was said, and what kind of person said it.

BEGIN QUOTE

Algebra is to the geometer what you might call the "Faustian offer"... Algebra is the offer made by the devil to the mathematician. The devil says "I will give you this powerful machine, it will answer any question you like. All you need to do is give me your soul: give up geometry and you will have this marvelous machine."... Of course, we like to have things both ways; we would probably cheat on the devil, pretend we are selling our soul, and not give it away. Nevertheless, the danger to our soul is there, because when you pass over into algebraic calculation, essentially you stop thinking; you stop thinking geometrically, you stop thinking about meaning.

END QUOTE

II) Rigor. Later in this review I will talk about what Descartes did mathematically. But a prelude to that in the book warns us about viewing it with modern eyes. Here is a quote:

Unfortunately, his (Descartes) Geometry isn't set up in the modern way, with clearly articulated definitions, carefully stated theorems, and rock-solid proofs.

2.3 History of Mathematics

Many mathematicians grace the pages of this book.

1. In the 1630's Descartes made massive progress on the problems by (1) realizing that the impossibility is something that might be provable, (2) translating the problem into algebra, and (3) in modern notation he showed that the set of numbers that are constructible is the field formed by taking the rationals and closing under $+$, $-$,$\times$, $\div$ and taking square roots. Later mathematicians and historians claimed (incorrectly) that Descartes showed trisecting the angle and doubling the cube were not possible. Descartes' own claims on this are muddled.

2. In 1796, when Gauss was 19, he constructed a 17-gon (I wish he had done it two years earlier!). He also showed that if $n = 2^a p_1 \cdots p_k$, where the p_i's are Fermat primes then the n-gone is constructible. He didn't give the construction. He showed, using Descartes' work, that it was possible. (In 1894 Johann Gustav Hermes completed the construction of the 65537-gone. It was 200 pages and took 10 years. From a 21st century prospective I cannot understand why he did that.) Some mathematicians thought that Gauss proved the converse, which he did not.

3. In 1837 Wantzel published a 7 page article showing that (1) trisecting the angle is impossible, (2) doubling the cube is impossible, (3) the n-gon is constructible IFF $n = 2^a p_1 \cdots p_k$ where the p_i's are distinct Fermat primes. He solved three of the four problems of antiquity! Problems that had been open for around 2000 years! Yet his paper was met with a deafening silence. Why? (a) Some thought Descartes had already proved it. (b) Some thought Gauss had already proved it. (c) Some thought this was already proven and all Wantzel did was write it up formally. (d) Some thought it was one of those things that everyone kind of knows is true, but doesn't really need a proof. This last point may be the most important and was the theme of this article on Wantzel:

   `https://www.cs.umd.edu/~gasarch/BLOGPAPERS/wantzel.pdf`

   To drive home this last point I will tell you a story that was also one of my motivations to read the book. At the 12th Gathering for Gardner the author gave a talk on 12 ways to trisect an angle using straightedge, compass, and just-one-more-thing. Many of them predate the proof that it is impossible to trisect an angle with just straightedge and compass. The talk is here:

`https://www.youtube.com/watch?v=5VxMUhxkBqA&t=420s`

I asked him afterwards *did the people who trisected the angle using straightedge, compass, and just-one-more-thing have the hope of replacing that one-more-thing with a straightedge and compass?* He responded *No, people pretty much knew that it was impossible.*

This shows a different way of thinking about math. Viewing the problems of antiquity in the way I began this review, as a *prove or disprove* question, is a modern viewpoint. The notion that one can show a construction cannot be done is a modern viewpoint (even though Descartes had some idea of this). Hence, to celebrate Wantzel as having solved an open problem is a modern viewpoint.

4. In 1837 Hermite showed that e is transcendental. Lindemann talked to Hermite about his proof and nine years later Lindemann showed that π was transcendental, and hence the problem of squaring the circle was shown to be impossible. How did Hermite feel about this? IF he was angry or thought he was unfairly scooped then we would know about it. E.T. Bell would have had some absurd exaggeration of it in his book *Men of Mathematics*. But no. Hermite was quite happy with how things turned out.

The book closes with a chapter asking if the quest to show these problems impossible was a siren (dangerous, leading people astray) or a muse (guiding people to math of interest). Clearly the author comes down on the side of muse.

3 Opinion

Who should read this book? For the math-inclined there are plenty of constructions (many using more than a straightedge and compass) which are probably new to the reader and interesting. To the non-math inclined there are still some interesting points *about* math. I was especially intrigued by how math changed over the centuries.

4 Who Said the Quote?

Clearly the quote was said by someone at the time when math was transitioning from geometry to algebra. Also note that modern people do not invoke *the devil* in their discussions.

The above paragraph is incorrect. The quote is by Sir Michael Atiyah who won the Fields medal in 1966 and the Abel Prize in 2004. He said this in 2001. He passed away in 2019 and hence can be considered a 21st century person.

Review of[3]
A Map that Reflects the Territory:
Essays by the Less Wrong Community
Author: Less Wrong
Publisher: Less Wrong Press
720 pages, Year: 2020, $30.00

Review by
William Gasarch (`gasarch@umd.edu`)
Department of Computer Science
University of Maryland, College Park

1 Introduction

Less Wrong is a forum founded by Artificial Intelligence Theorist Eliezer Yudkowsky and Economist Robin Hanson in 2009. The stated philosophy is:

We are a community dedicated to improving our reasoning and decision-making. We seek to hold true beliefs and to be effective at accomplishing our goals. More generally, we work to develop and practice the art of human rationality.

That seems to cover a lot of ground! A satire of it would say the following:

There are discussions about discussions, discussions about arguments, arguments about discussions, and arguments about arguments.

That is not fair. The topics seem to be (1) how does one find the truth in science and in life, (2) AGI (Artificial General Intelligence), and (3) probability. The most common non-trivial word in this book might be *Bayes* (a trivial word would be something like *the* which is likely more common but less interesting).

This book is a best-of collection. We quote the preface: *Users wrote reviews of the best posts of 2018, and voted on them using the quadratic voting system, popularized by Glen Weyl and Vitalik Buterin. From the 2000+ posts published that year, the Review narrowed down the 44 most interesting and valuable posts.*

The collection of posts are now gathered together in a book from the Less Wrong forum, titled *A Map that Reflect the Territory: Essays by the Less Wrong Community*. This is the book under review.

This set of essays is *a set of five books*, titled Epistemology, Agency, Coordination, Curiosity, Alignment. Each book is small—about 6 inches long, 4 inches wide, and 1/4 of an inch thick.

2 General Comments

PROS: Many of the essays bring up a point that I had not thought of before. Many of the essays say something interesting in passing while getting to their point.
CONS: Some of the essays are trying to say something interesting but have no examples. There are times I am crying out *give me an example!* (Reminds me of my days as a pure math major.) Some of the essays are locally good but it's not clear what their point is.
CAVEAT (both a PRO and a CON): Many of the essays use a word or phrase as though I am supposed to already know them. If I was a regular member of the forum then perhaps I *would* know them. In the modern

[3] ©2022 William Gasarch

electronic age I can try to look them up. This is a PRO in that I learn new words and phrases. For me this is a really big PRO since *I collect new words and phrases as a hobby*. This is a CON in that going to look things up disrupts the flow of the essays. And sometimes I can't find the new word or phrase on the web.

In the *third to last* section of this review I will have a list of all of the words and phrases I learned by reading these books and either their meaning or that I could not find their meaning. Why *third to last?* Because the second to last section is my summary opinion and the reader of this review should be able to find the meanings quickly (the last section is acknowledgments). I posted to LessWrong a request for what some of the words mean and got a few responses. I also emailed Robin Hanson to find out what a *Hansonian Death Trap* is. I now know (or think I know) all of the words and phrases I encountered; but it was challenging to find them all.

3 Epistemology

I quote the first sentence:

The first book is about epistemology, how we come to know the world.

Most of the essays are on *how to have a good argument.* (Reminds me of Monty Python's classic sketch *the argument clinic*, which is here:

`https://www.youtube.com/watch?v=ohDB5gbtaEQ`

The essays are more enlightening but less funny.)

Scott Alexander's *Varieties of Argumentative Experience* is especially good and has... wait for it ... **examples!**. Here is one concept I found very interesting: *double-crux*. Say Alice thinks gun control is good and Bob thinks gun control is bad. They should find related statements X and Y such that if X is true Alice will change her mind, and if Y is true then Bob will change his mind. In this case it could be

X is *If we have gun control then crime will go up.*

Y is *If we have gun control then crime will go down.*

Hence the argument can now focus on a question that can be studied objectively. (I will now plug my cousin Adam Winkler's book: *Gunfight: The Battle over the Right to Bear Arms*

`https://www.amazon.com/Gunfight-Battle-Over-Right-America/dp/0393345831`

which is an intelligent discussion of gun control including the history of the issue.)

Another essay that I interpret as on the topic of *how to have a good argument* is *Local Validity as a Key to Sanity and Civilization* by Eliezer Yudkowsky. The essay is actually about laws and norms, but it's more about the need to avoid having laws that only apply to some people and not others. While this seems like an obvious point, he gives it history and context.

There are essays by Alkjash about how to come up with new ideas: *babble and prune.* Have lots of (possibly half-baked) ideas, and then prune to get the good ones. There is a delicate balance here — how much to babble? how much to prune? A fascinating aside in the article: babies *can* make all the phonemes — they learn language mostly by pruning.

The essay *Naming the Nameless* by Sarah Constantin is about aesthetics and arguments. Why are artists left wing? What to do if you are are a conservative who likes modern art? She then critiques certain types of arguments from an aesthetic point of view.

The last essay, *Towards a New Technical Explanation of Technical Explanation* by Abram Demski is the most technical. Its about logic, uncertainly, and probability. It seems to point to a way to predict things under uncertainty, however there are no examples. I felt like shouting **Does it Work? Can you test it?**

4 Agency

I quote the first sentence:

The second book is about agency, the ability to take action in the world and control the future.

Despite the above sentence, this book does not have a coherent theme; however, it does have several very interesting essays.

Eliezer Yudkowsky has two essays on honesty: *Meta-Honesty: Firming Up Honesty Around the Edge Case (The Basics)* and *Meta-Honesty: Firming Up Honesty Around the Edge Case (The Details).* When should one be honest? The usual easy example is lying to Nazis who ask if you are hiding Jews (you should lie). Is there a consistent rule you can use? The essays suggest rules that involve never lying about lying. The second essay has two conversations that are so funny they should be made into a Monty Python sketch: (1) Dumbledore trying to find out if Harry Potter robbed a bank, and (2) the Gestapo asking about hiding Jews. What makes these conversations hilarious is that all parties know all about the issue of meta-honesty. Eliezer admits that these scenarios would never happen. These essays raise interesting points; however, it is grappling with problems that probably have no solution.

Michael Valentine Smith's essay *Noticing the Taste of the Lotus* is about noticing that you are (say) playing a computer game to get more points, and using those points to buy things so that you can . . . play better and get more points so that you can buy things . . .. We (I mean every human) needs to **BREAK OUT OF THIS DEATH SPIRAL**.

Scott Alexander's *The Tails Fall Apart as a Metaphor for Life* begins by talking about the following: even though reading and writing scores are correlated, the top reading score **is usually not** the the top writing score. He then applies this observation to happiness and morality. That is, different definitions of happiness are sometimes correlated, but not at the high end. Same for morality. This essay gave me lots to think about, though I don't know what to conclude.

The other essays were of the same type: they made some interesting points but didn't really answer the rather hard questions they set out to tackle. This reminds me of what I liked about philosophy (my minor in college): the questions raised (e.g., What is Truth? What is Knowledge? What is Beauty?) are not going to be answered, but reading about the attempt to answer them is interesting.

5 Coordination

I quote the first sentence:

This third book is about coordination, the ability of multiple agents to work together.

Four of the essays are on game theory. They all go beyond the usual introduction of the Prisoner's Dilemma and hence are all interesting. My challenge is is to give 1-2 sentences about each one.

1. *Anti-Social Punishment* by Martin Sustrik. This describes an experiment that people *really did* involving whether a player does what's good for himself or what's good for the group. Results are interesting and seem to really tell us something.

2. *The Costly Coordination Mechanism of Common Knowledge* by Ben Pace. The key to the prisoners dilemma is that the parties cannot talk to each other. In the real world how do enough people talk to each other so that they do not fall into the dilemma?

3. *The Pavlov Strategy* by Sarah Constanin. This describes strategies for Prisoner's Dilemma.

4. *Inadequate Equilibria vs Governance of the Commons* by Martin Sustrik. This gives *real examples* of how people got around the tragedy of the commons.

Prediction Markets: When do they work? by Zvi Mowshowitz is an excellent article about, as the title says, when Prediction Markets work. I was most intrigued by the fact that insider trading is quite legal; however, if it is known that people are doing it, less people might use that market.

The Intelligent Social Web by Michael Valentine Smith views life as improv. In order for a scene to work everyone must naturally follow their role. In life we have a view of ourselves that we have to stick to to make the scene work. We may change slowly to adapt to a different scene. This is a fascinating way to view life!

On the Loss and Preservation of Knowledge by Samo Burja begins with the question: **What would Aristotle have thought of Artificial Intelligence?** No it doesn't! The essay really begins with the question **How would you approach the question of "What would Aristotle have thought of Artificial Intelligence?"** It goes on to talk about how knowledge, schools of thought, and philosophies have a hard time being preserved, and giving signs that they were or were not. Alas, it is likely that the Aristotelian philosophy is not so well preserved to answer the question (that's my opinion).

There are a few other essays, but the ones I mentioned are the highlights. This was my favorite book since so many of the essays were interesting.

6 Curiosity

I quote the first sentence:

The fourth book is about curiosity, which is the desire to understand how the world works.

There are three essays that look at the pace of science and other advancement:

1. *Is Science Slowing Down?* by Scott Alexander,
2. *Why Did Everything Take So Long?* by Katja Grace, and
3. *Why Everything Might Have Taken So Long* also by Katja Grace,

Scott Alexander argues that science is slowing down and he gives good reasons for this. Katja Grace examines why, for example, even though humans have been around for 50,000 years the wheel was invented only about 6000 years ago. So for 44,000 years people didn't have the wheel! (My students are amazed that 30 years ago people didn't have Netflix.)

The essay *What Motivated Rescuers During the Holocaust?* by Martin Sustrik is interesting both in what they can say about the question and how they can say anything about the question.

The essay *Is Clickbait Destroying Our Intelligence?* by Eliezer Yudkowsky is locally interesting but wanders around quite a bit. Another negative is that the answer is so obviously *Yes*.

The essay *What Makes People Intellectual Active?* is somewhat interesting but longer than it needs to be.

The essay *Are Minimal Circuits Daemon-Free?* by Paul Christiano is about circuits (really AI systems) that satisfy the problem constraints but not in the way that you want. It was too technical for my tastes. Also (and this is not an objection), it may have fit better in the book *Alignments*.

There are a few other essays, but the ones I mentioned are the highlights.

7 Alignment

I quote the first sentence:

This fifth book is about alignment, the problem of aligning the thoughts and goals of artificial intelligence with those of humans.

The essay *Specification Gaming Examples in AI* by Victoria Krakovna is about when AI systems do well but for the wrong reason. For example, a deep-learning model to detect pneumonia did well, but only because the more serious cases used a different X-ray machine. She has a longer article and many example here:

`https://www.lesswrong.com/posts/AanbbjYr5zckMKde7/specification-gaming-examples-in-ai-1`

This essay is excellent in that it states what the problem of alignment is, getting AI systems to do what we want for the right reasons. Then there was a great satirical essay *The Rocket Alignment Problem* by Eliezer Yudkowsky. There were some other essays of mild interest about what might happen (e.g, slow and steady or fast and abrupt progress). But the collection bogs down with a series of essays (about 1/3 of the book) on Paul Christano's research on *Iterated Amplification*, which is also called *Iterated Distillation and Amplification (IDA)*. The idea is that you start with a system M that is aligned— it gives the right answers for the right reasons. Perhaps a literal human. You then amplify to a smarter system Amp(M) (perhaps letting it think longer or spinning off copies of itself). Then you (and this is the key!) distill Amp(M) into a system M+ which is aligned. Repeat this many times. But note that you always make sure its aligned.

That sounds interesting! It might work! But then the essays seem to debate whether its a good idea or not. I kept shouting at the book **JUST TRY IT OUT AND SEE IF IT WORKS!** I have since learned (from comments on LessWrong about an earlier draft of this review) that current AI is just not smart enough to do this yet. This raises a question: How much should one debate if an approach will work before the approach is possible to try? If the debate produces interesting results (like research on quantum computing giving insight into quantum and computing) then the debate is worth having. I did not see that here.

8 Newords that I Learned or Tried to Learn From These Books

This section has a list of newords[4] that I learned from reading this book. From looking them up, posting to LessWrong, and emailing Robin Hanson, I found out what they all meant.

8.1 From the book Epistemology

1. *No Free Lunch Theorem*: If an ML algorithm does well on one set of data it will do badly on another (this is a simplification). This is not just an informal statement—it has been formalized and proven.

2. *Code of the Light*: On page 19, in the article *Local Validity as a Key to Sanity and Civilization* by Eliezer Yudkowsky, is the following sentence:

   I've been musing recently about how a lot of the standard Code of the Light isn't really written down anywhere anyone can find.

[4] This is not a misspelling—I use newords instead of new words. The best neologisms do not need to be explained; however, (1) when I posted this review on my blog I got 5 emails "correcting" the spelling, (2) when I posted this review on the LessWrong blog I got some comments "correcting" the spelling, and (3) when I emailed this to Fred Green, SIGACT News book review editor, he inquired if this was a misspelling.

Google Searches for *code of the light* only lead to the essay. The phrase was in green so I thought maybe in the original it was a link that would tell me what it means. Nope. There is an irony that he notes that *Code of the light isn't really written down anywhere* and then not write down what it means.

When I posted an early version of this review I got a comment from gjm which I paraphrase.

I think that EY made up this terms for the occasion and he intends them to be, at least roughly, clear from context. It means "how good, principled, rational, nice, honest people behave."

3. *Straw Authoritarians*: On page 20, in the article *Local Validity as a Key to Sanity and Civilization* by Eliezer Yudkowsky, is the following sentence:

   Those who are not real-life straw authoritarians (who are sadly common) will cheerfully agree that there are some forms of goodness, even most forms of goodness, that it is not wise to legislate.

   When I posted an early version of this review I got a comment from gjm which I paraphrase.

   Straw authoritarians are authoritarians who are transparently stupid and malicious, rather than whatever the most defensible sort of authoritarian might be.

   Is this what EY meant? Probably yes. Could I ask him myself? Probably not. Why not? A recent LessWrong post (see

   `https://www.lesswrong.com/posts/bDMoMvw2PYgijqZCC/i-wanted-to-interview-eliezer-yudkowsky-but-he-s-busy-so`

)

   was titled *I wanted to interview Eliezer Yudkowsky but he's so busy so I simulated him instead.*

4. *Whispernet Justice System* being tried in the court of public opinion. I am guessing from context. Google only points to the essay it appeared in. Even so, this should be a word!

5. *The Great Stagnation*: The name of a pamphlet by Tyler Cowen from 2011 that argues that the American Economy has run out of steam for a variety of reasons. The phrase is now used independent of the book but with the same meaning.

6. *Memetics*: The study of memes in a culture.

7. *Memetic collapse*: On page 27, in the article *Validity as a Key to Sanity and Civilization* by Eliezer Yudkowsky, is the following sentence:

   It's [the book Little Fuzzy by H. Beam Piper] from 1962, when the memetic collapse had started but not spread very far into science fiction.

   Google searches only lead to the same essay I read this in. Searches on LessWrong lead to a few hits but they all seem to presuppose the reader knows the term.

   When I posted an early version of this review on LessWrong I got a comment from gjm which quotes from a Facebook post by EY. I paraphrase the Facebook post:

   Since people can select just what they agree with (on the internet, on Facebook, etc) there is a collapse of references to expertise. Deferring to expertise causes a couple of hedons[5] *compared to being told your intuitions are right. We're looking at a collapse of interactions between bubbles because there used to be just a few newspapers serving all the bubbles; and now that the bubbles have separated*

[5] I had to look this one up: a hedon is a unit of pleasure used to theoretically weight people's happiness. Like how happy I am when I find a cool new word like *hedon*.

there's little incentive to show people how to be fair in their judgment of ideas from other bubbles. In other words: changes in how communication works have enabled processes that systematically made us stupider, less tolerant, etc., and also get off of my lawn.

I am happy to know what EY meant by the term. I'm surprised he says the memetic collapse had already started in 1962. I would have thought it started later than that. *The history of the memetic collapse* might be a good topic for historians.

8. *AGI*: Artificial General Intelligence

9. *Double cruxing*: Alice and Bob are having an argument. Get them to agree on a fact that would change their mind. Example: Alice is for gun control and Bob is against it. If Alice would change her mind if she knew gun control causes crime to go UP and Bob would change his mind if he knew gun control causes crime to go DOWN then they have reduced their disagreement to a factual statement that can be investigated.

8.2 From the Book Agency

1. *Lotus Eater*: In the Odyssey they land on the Island of Lotus-Eaters. The taste of the lotus is so good that your goal is to eat them and you ignore other goals. Some of today's games have that property - you accumulate points that allow you to play more to get more points. . .. I've also heard of going to the gym and lifting weights so you get better at lifting weights. Origin might be Duncan Sabien (a contributor to LessWrong).

2. *Medioracistan* and *Extremistan*

   On page 16, in the article *The tails coming apart as a metaphor for life* by Scott Alexander, is the following:

   This leads to (to steal words from Taleb) a Mediocristan resembling the training data where the category works fine, vs. an Extremistan where everything comes apart.

   I paraphrase an enlightening comment by gjm.

   In Nassim Nicholas Taleb's book The Black Swan Mediocristan and Extremistan are imaginary countries. In Mediocristan things have thin-tailed distributions, so differences are moderate. In Extremistan there are fat-tailed distributions, so difference are sometimes hugh. These countries are used to indicate if data is thin-tailed or fat-tailed.

3. *Deontology*: An ethical system that uses rules to tell right from wrong. Once the rules are set, no need for God or anything else.

4. *Glomarization*: Always saying "I cannot confirm or deny." I got this definition from the essay. It's a more common term than I had thought: there is a Wikipedia entry on *Glomar Response*.

5. *Dunbar's number*: The number of people that we can interact with comfortably. Dunbar estimated it to be 150. Be careful who you choose for your 150st friend.

8.3 From the Book Coordination

1. *Miasma*: This seems to be the opposite of hype, but Google says its an unpleasant smell.

2. *Goodhart's Demon*: On Page 54, in the article *The Intelligent Social Web*, is the following:

   Ah, but if we are immersed in a culture where status and belonging are tied to changing our minds, and if we can signal that we are open to changing our beliefs, then we're good . . . as long as we know Goodhart's Demon isn't lurking in the shadow of our minds here.

   I could not find this anywhere on the web (except for the article). I suspect the following contrast is true:

   Goodhart's Law if a measure becomes a target it ceases to be a measure. Example: Colleges admissions committees use the number of clubs you are in as a measure of a person's wide ranging interests, but then people begin joining clubs to impress college admissions committees.

   Goodhart's Demon The temptation to game the system.

3. *Hansonian Death Trap*: On page 73, in the article *Prediction Markets: When Do They Work*, is the following:

   If you're dealing with a hyper-complex Hansonian death trap of a conditional market where it's 99% to not happen, even with good risk measurement tools that don't tie up more money than necessary, no one is going want to put in the work and tie up the funds.

   Google Searches only turned up hits to this essay. Searches within LessWrong point to a few more hits, but they presuppose the reader knows the term. I suspected that Hanson was Robin Hanson, so I emailed him and got this response:

   Hi. I'm pretty sure I'm the only Robin Hanson mentioned in those circles, so that must be me. However, I've never heard the phrase "Hansonian Death Trap", so I expect it isn't in common usage; you are just seeing one person make up a phrase.

   It is true that speculators will put less effort into trading on conditional claims with lower probabilities, and so prices will be less accurate, but there isn't a problem of "tying up funds".

4. *The Costanza*: Do the opposite of what you naively think you should do. This is from an episode of Seinfeld where George Costanza intentionally does this since all of his past decisions have been wrong. Not to be confused with *pulling a Costanza* which means, if you are fired, show up for work the next day as if you weren't, as if your boss was just joking.

5. *Lucas Critique*: It is naive to predict the effect of an economic policy based on past uses.

6. *Counterfeit Understanding*: Knowing the words but not their meaning. Like people who memorize proofs in math line-by-line but do not know the intuition behind them. Students memorize proof templates without understanding take the proof that $\sqrt{2}$ is irrational and blindly modify it to show $\sqrt{4}$ is irrational.

8.4 From the Book Curiosity

1. *Dectupled*: Multiply by 10.

2. *Price's law of scientific contributions*: If there are n people on a project then half of the work will be done by $\sqrt{n}$ people.

3. *Yudowsky's law of mad science*: Every 18 months the min IQ needed to destroy the world decreases by one. Scary!

4. *Opsec*: Short for operational security.

5. *Bystander Effect*: On page 28, in the article *What Motivated Rescuers During the Holocaust?* by Martin Sustrik, is the following:

   As I already said, I am not an expert on the topic, but if what we see here is not an instance of the bystander effect, I'll eat my hat.

   He is referring to the fact that people who begin helping one Jew escape the Nazis end up helping more.

   The phrase *Bystander Effect* is on the web! A lot! It seems to be that *the more people that are bystanders who could prevent something bad from happening the less likely someone really will.* This seems different from how it's used in the essay.

   When I asked the LessWrong forum about this I got two responses:

   (a) beriukay said *Since I have not read the first one [the article], I could only speculate that the people who end up helping realize that nobody else is doing to do anything to help, which breaks them out of the effect and they end up helping more.* Excellent! This seems to say that what the author of the article meant to say is that this is an example of the converse of the standard bystander affect.

   (b) Tetrapace Grouping said: *The bystander effect is an explanation of the whole story:*

      - *Because of the bystander effect, most people weren't rescuers during the Holocaust, even though that was obviously the morally correct thing to do; they were in a large group of people who could have intervened but didn't.*
      - *The standard way to break the bystander effect is by pointing out a single individual in the crowd to intervene, which is effectively what happened to people who became rescuers by circumstance that forced them into action.*

6. *Memetically*: This seems to be related to memes but I could not find the word on the web.

7. *The Sequences*: On page 83, in the article *What Makes People Intellectually Active?* by Abram Demski, is the following:

   What is the difference between a smart person who has read the Sequences and considers AI x-risk important and interesting, but continues to be primarily a consumer of ideas, and someone who starts having ideas?

   The Sequences is impossible to look up on Google. Fortunately, if you search on the LessWrong site you get the following:

   The original sequences were written by Eliezer Yudkowsky with the goal of creating a book on rationality. Someone with the name MIRI has since collated and edited the sequences into Rationality: AI to Zombies. *If you are new to Less Wrong, this book is the best place to start.*

   Darn. I started with *A Map that Reflects the Territory: Essays by the Less Wrong Community*.

8. *Yed graphs*: On page 85, in the article *What Makes People Intellectually Active?* by Abram Demski, is the following:

   I might write one day on topics that interest me, and have sprawling Yed graphs in which I'm trying to make sense of confusing and conflicting evidence.

When I asked the LessWrong forum what a Yed graph is I got a pointer to a product, Yed Graph Editor, that generates high quality graphs. Here is the pointer:

`https://www.yworks.com/products/yed`

When I was looking for what a Yed graph was, I did come across that, but I thought it was not how the term was being used in the article.

9. *LW-corpus*: Everything in the Less Wrong website.

10. *TAP*: On page 92, in the article *What Makes People Intellectually Active?* by Abram Demski, is the following:

    It's like the only rationality technique is TAPs, and you only set up taps of the form "resemblance to rationality concept" → *"think of rationality concept".*

    When I asked the LessWrong forum what TAP was I found out that it stands for *trigger action plan* and I got a pointer to another LessWrong article that may be where the term originated. Here is the pointer:

    `https://www.lesswrong.com/posts/vE7Z2JTDo5BHsCp4T/instrumental-rationality-4-2-creating-habits`

9 Should You Read This Book?

Yes.

Okay, I will elaborate on that.

In the spirit of the Less Wrong community, I looked at *evidence* on this question. What kind of evidence? I went through all five books and, for each article, marked it either E for Excellent, G for Good, or M for Meh (none were B for Bad), and counted the number of each.

1. *Epistemology* E-1, G-6, M-3.
2. *Agency* E-2, G-2, M-1.
3. *Coordination* E-6, G-2, M-2.
4. *Curiosity* E-4, G-2, M-4.
5. *Alignment* E-2, G-3, M-5.

What to do with this information?

1. There are 15 excellent articles! That's. . . excellent!
2. There are 15 good articles! That's. . . good?
3. There are 15 meh articles! That's. . . meh.

(I did not plan to have 15-15-15. Honestly! In the spirit Eliezer Yudkowsky essays on Meta-Honesty I tell you that this is not the kind of thing I would lie about.)

So is 15-15-15 a good ratio? Yes! And note that the good articles are still . . . good. But let's take a more birds-eye view (Do birds really have a good view? Do crows really fly "as the crow flies"?): What did I learn from reading these 45 essays?

1. Many interesting questions were raised that I had not thought of. Here is just a sample: (1) Why do inventions take so long to be invented? (2) Why do I play to much Dominion online? (From *Noticing the taste of the Lotus*, and it also says why I should stop),

2. Many interesting meta questions were raised that I had not thought of. Here is just a sample: Can we know what Aristotle would think of AI?

3. Some answers or inroads on these questions were made. Sometimes the answers were actual answers. Sometimes they gave me things to think about. Both outcomes are fine.

4. Some newords for my newords hobby!

So are there any negatives? Yes:

1. There were some words that I had to go look up. This interrupted the flow of the articles. I re-iterate that this can also be seen as a positive as you get to learn new words.

2. The problem above points to a bigger problem: LessWrong writers (and I presume readers) seem to have their own language and hidden assumptions that it may take an outsider a while to catch onto.

3. Some of the essays need examples. This may also be part of the bigger problem: LessWrong writers (and I presume readers) may already know of the examples or some context. And again, it makes it a bit rough for outsiders.

And now for the elephant in the room: Why buy a book if the essays are on the web for free? I have addressed this issue in the past since I've reviewed 3 blog books (see
`https://www.cs.umd.edu/~gasarch/BLOGPAPERS/lipton.pdf`
`https://www.cs.umd.edu/~gasarch/BLOGPAPERS/liptonregan.pdf`
`https://www.cs.umd.edu/~gasarch/BLOGPAPERS/tao.pdf`
)
and have written my own blog book: *Problems with a point: Explorations in Math and Computer Science by Gasarch and Kruskal* (see
`https://www.amazon.com/Problems-Point-Exploring-Computer-Science/dp/9813279974`
).

Here is an abbreviated quote from my book that applies to the book under review.

The Elephant in the Room

So why should you buy this book if its available for free?

1. Trying to find which entries are worth reading would be hard. There are a lot of entries and it really is a mixed bag.

2. There is something about a book that makes you want to read it. Having words on a screen just doesn't do it. I used to think this was my inner-Luddite talking, but younger people agree, especially about math-on-the-screen.

10 Acknowledgments

I thank Oliver Habryka for giving me this opportunity to review these books.

I thank Ben Pace for proofreading and useful comments.

I thank beriukay, Tetraspace Grouping, and gjm for clarifying some of the words and phrases that either I did not know or thought I did but was wrong.

I thank Daniel Kokotajlo for his comments on Iterated Amplification which I have incorporated into the review.

I thank Ruby Bloom for helping me post this review to the LessWrong forum in text form (having a pointer to the review I learned was not as good).

All of the people acknowledged help make this review LessWrong.