A Model of a Mind

Tyler Neylon

264.2024

[Formats: html | pdf]

I'm working on a book about consciousness from the perspective of machine learning, coding, and neuroscience with a dash of philosophy. As a preliminary step in writing that book, I'd like to share a concrete model of how I suspect human minds might work — and thus how we might work on digital minds.

There's still plenty of debate about whether or not a digital mind can ever be truly conscious, or have emotions or subjective experiences as humans do. I'm convinced they can. But rather than focus on that debate, in this post, I'd like to simply work in the hypothetical world where digital minds are indeed capable of all the internal experiences of human minds. If I'm wrong, then this post becomes an interesting speculation; if I'm right, then this post is something more — hopefully, actual progress toward both the creation of digital minds as well as some insight into how our own brains my work.

1 The Model

The goal of this model is to capture how the human brain may work on a functional level, ignoring details that most likely won't be useful in building a digital mind. The primary motivation is to make progress toward digital minds, but I suspect this pursuit may shed light on how brains work as a secondary benefit.

One principle is that I'm trying to make a system that can behave like a human. I'm not trying to philosophically evaluate consciousness — I'm just aiming to achieve the same behavior. I'm trying to build a system that has these features:

- Learning
- Agency
- Thinking
- Introspection

I think introspection is thinking along with the ability to think about your own thoughts.

1

I recently enjoyed reading *So Good They Can't Ignore You* by Cal Newport, a thought-provoking book that argues in favor of a specific perspective on your big-picture career mindset. This post summarizes and provides some analysis of it. Since I like to be kind to authors, I also encourage you to buy the book if it appeals to you.

2   Summary

Cal's goal is essentially to answer the question:

How should I plan my career?

(It's strange that we call authors by their last name in writing when we address people by first name in person. So I'm trying out referring to Cal as Cal instead of as "Newport," which would sound overly-formal if I were speaking out loud.)

He suggests these ideas as a framework for thinking about your career:

- Don't fall into the trap of following your passion as a first priority.
  (Ch 1–2)
- Build up career capital in the form of rare and valuable skills.
  (Ch 4–7)
- It's nice to have control over what you do in your work.
  - Don't opt for control until you have rare and valuable skills.
    (Ch 9–10)
  - Do opt for control despite others wanting to be in charge of you.
    (Ch 11)
- It's nice to have a career mission.
  - Don't look for a mission before you have rare and valuable skills.
    (Ch 13)
  - Once you have a mission, move toward it with bite-sized projects; "little bets."
    (Ch 14)
  - Once you have a mission, share your work in a way that makes people want to talk about it, and in a place that's conducive to sharing and discussion.
    (Ch 15)

That's the book in a nutshell.

2.1   Book Origins

The book was published in 2012, just as Cal was about to begin his tenure-track position as a computer science professor at Georgetown. (Below I'll go into more thoughts along the line of: It seems a bit early in his career for him to be giving career-length advice; for now I'll focus on a more objective summary.) He had

written a series of blog posts refuting the passion hypothesis — the mandate to follow your dreams — and they received a strong reception. He expanded on those ideas into this book. Along the way, he interviewed a number of interesting personalities, building his hypotheses on what worked or didn't work for them.

2.2   Cal's Writing Since 2012

I was curious if Cal had evolved his thinking since publishing this book. He was early in his career, and the book lands to me a little as a hypothesis, rather than as a tested theory, for Cal himself. I would have found value in Cal saying either "yep, this has worked out great for me," or perhaps "golly, I think I should revise some of those hypotheses." However, I didn't find evidence either way. (If you're a skeptical person, you might read as subtle evidence the hypotheses could use some revision).

After 2012, Cal has written several books focusing on productivity and reducing distractions. For example, he doesn't like that email is used as a default means of communication where we expect quick replies. Instead, he advocates seeing email as asynchronous — no quick replies are expected — and promotes ideas such as office hours as designated times when someone *is* expected to respond right away. I think he's seeing the need for daily tactics after having deeply considered life-scale strategy.

3   Analysis

I take basically everything with a grain of salt, and this book is no exception. I'm going to agree with ideas when they make sense to me and match my experience. While I'm about to include criticism, note that I found the book good enough to both read and write about.

To some degree, the ideas presented here make intuitive sense. Nothing Cal says seems to be overtly mistaken. For example, I have often noticed a feeling of boredom or even unhappiness when I'm about to begin certain work projects, only to be surprised at how many interesting ideas could arise in the course of working on the project. I've often seen excitement show up as the *result* of hard work, not as the precursor.

At the same time, this book follows a common formula. Cal presents an abstract idea up front, tells one or two stories that support this idea, and then summarizes the idea again. Each chapter feels like this, and the book as a whole follows a similar arc. I have to admit that I was satisfied when Cal returned to the same story he opened with, of a man who followed his passion to become a monk, but was miserable as a result. That man left his life of disconnection and returned to the rat race, and was happier.

I have a few problems with this formula, which is common among advice books:

- **It's one-sided.** Cal presents his ideas as in an opinion piece, trying to convince us the ideas are rock-solid. This is quite different from the tone of an impartial expository writer showing us multiple perspectives.
- **It presents anecdotes as strong evidence.** I'm skeptical about the reliability of anecdotes as arguments. While Cal lands as credible, this genre is rife with writing that I suspect embelishes or creates supporting stories.
- **It's redundant.** I was sometimes bored re-reading the same ideas repeatedly.

Ok, that's somewhat harsh — apologies to Cal if you read this. I aim to be more critical of the genre than of this particular author. I love to read math books, which tend to be the polar opposite in terms of information density.

Having complained about the genre, let's get into the content: Are the ideas themselves useful?

3.1   Idea: Passion is not a first priority

This one is easy for me to agree with. I appreciate that Cal later clarifies that passion is not *bad* so much as it should not be your *primary* motivation. I agree with this because I have many interests, and I could just as easily enjoy solving KenKen puzzles as I do researching consciousness-achieving algorithms, and one of those has more career value than the other. This idea also resonates with some well-respected career advice from Richard Hamming, who always wants people to work on the most important problem they can make progress on; this is another example of passion being subserviant to the value of your work.

3.2   Idea: First create rare and valuable skills

There's not much to disagree with here. I tried to think of careers in which the primary measure of job skill was simply how long you've been in the job. Any job like that would not be a good match for this advice. All I can think of, however, are unexciting government or unionized positions. And in both cases, I'm not sure how externally impressive such a career would be. In other words, I have trouble thinking of any counterexamples at all to the premise that a fulfilling career is built on rare and valuable skills.

3.3   Idea: Carefully bid for control over your work

I mostly agree with this premise, although Cal positioned his stories to suggest that you'll always have to fight for control in your job. That's not always true. I've worked for people who have trusted me not because I insisted on autonomy but because I deliver good work, and giving me autonomy was better for them than micromanaging me. I've felt similarly about people I've managed — people who thrive with more control — and voluntarily gave them more independence rather than waiting for them to request it. Friction against autonomy might

tell you more about your employer's mindset than about whether you're seen as valuable.

3.4   Idea: Carefully choose a mission for your work

This is the most nebulous of Cal's ideas in the books, and was the least satisfying for me to think about. Should I feel bad if I don't have a mission and I'm far into my career? What if I chose a mission but later changed direction? What if I'm already happy without a mission?

I suspect this part of Cal's thinking is likely influenced by his pursuit of an academic career. If you're a musician, for example, you may simply want to make good music. Perhaps there's an analogue of a mission for musicians, such as nudging a musical trend in a direction you like, or perhaps raising awareness of a cause you care about. But it would be a lie to pretend that you write music *because* of an external cause, such as world hunger.

A mission can augment a sense of meaning in your career. It's not critical to have one. If you care about something and want to make a difference in the world, go for it — but you can be happy without heroism.

3.5   A point of comparison

In 2006, Paul Graham published an essay called *How to do what you love*.

Paul is a compelling writer. So compelling, in fact, that I put more than my usual effort into thinking critically about what he writes. Paul's essay on doing what you love presents ideas similar to the passion hypothesis — the idea that you should follow your passions.

It's more nuanced than that, however. What Paul really expresses is that society imposes on us a false dichotomy between work and play. He also argues that, if you want to be happy, loving what you do has to be part of it. The difference between this and the passion hypothesis is that Paul (I believe) agrees with Cal that it's difficult to find something fulfilling that you love to do, and that it's worth dedicated, up-front effort to find that match.

Paul points out, for example, that if a person strays from the beaten path, that might indicate either that they're lazy (bad), or that they're working to be productive in an independent way (good). But it's hard to tell the difference. On diagnosing your own path-straying behavior, he says:

> Is there some test you can use to keep yourself honest? One is to try to do a good job at whatever you're doing, even if you don't like it. Then at least you'll know you're not using dissatisfaction as an excuse for being lazy. Perhaps more importantly, you'll get into the habit of doing things well.

I recommend reading the full essay to have another perspective on career questions that overlap with Cal's book. Where the ideas don't quite line up, it's educational to perform a merge conflict in your mind.

3.6   What's left out?

Although Cal does not explicitly say this, the tone of the book presents itself as a guide to happiness in terms of making large-scale career decisions. Yet the book is more a reaction to a few common mistakes than it is a full playbook. Or, if I'm being pessimistic, the writing comes across as someone who's good at publicity taking advantage of a couple sticky ideas to sell a book (it's not purely that, but the mood feels salesy at times).

There are many factors that contribute to career happiness. Here are few important ones that Cal doesn't talk much about:

- Positive impact on the world.
- Work-life balance.
- Compensation.
- Good relationships with your colleagues.
- Learning and feeling challenged.
- Safety.

Cal does briefly mention the value of career *impact* and *relatedness* — work relationships — but doesn't say much about them.

Here's an example of something not covered by the book: I've worked for a number of startups, and it's quite common to feel a great deal of pressure to work nights and weekends for startups. This can easily cause burnout, lack of sleep, and other kinds of unhappiness. There seem to be various resource metrics (think: sleep tank, energy tank, social tank) that need to be maintained for happiness. These resources are not going to be fully addressed even if we have a great framework for thinking about passion, mission, and career capital.

To be fair, Cal doesn't explicitly say he'll address all aspects of finding happiness in your job. At the same time, the book's tone lands as if it has everything figured out. I'd love to read more books that present interesting ideas a style that reads as succinct and inquistive without overconfidence.