

'This is an exhilarating book that will change the way we think about AI, intelligence, and the future of humanity'

Bart Selman, Professor of Computer Science, Cornell University

'Max has written the most insightful and just plain fun exploration of AI's implications that I've ever read. If you haven't been exposed to Max's joyful mind yet, you're in for a huge treat'

Erik Brynjolfsson, co-author of *The Second Machine Age*

'Being an eminent physicist and the leader of the Future of Life Institute has given Max Tegmark a unique vantage point from which to give the reader an inside scoop on the most important issue of our time'

Jaan Tallinn, co-founder of Skype

'Max seeks to facilitate a much wider conversation about what kind of future we, as a species, would want to create. Though the topics he covers can be fairly challenging, he presents them in an unintimidating manner that invites the reader to form her own opinions'

Nick Bostrom, author of *Superintelligence*

'Max's new book is a deeply thoughtful guide to the most important conversation of our time'

Ray Kurzweil, author of *The Singularity is Near* and *How to Create a Mind Edit*

ABOUT THE AUTHOR

Max Tegmark is a professor of physics at MIT and president of the Future of Life Institute. He is the author of *Our Mathematical Universe*, and he has featured in dozens of science documentaries. His passion for ideas, adventure and an inspiring future is infectious.

MAX TEGMARK

Life 3.0

*Being Human in the Age of
Artificial Intelligence*



PENGUIN BOOKS

Contents

Acknowledgments	xi
Prelude: <i>The Tale of the Omega Team</i>	3
1 <i>Welcome to the Most Important Conversation of Our Time</i>	22
A Brief History of Complexity	
The Three Stages of Life	
Controversies	
Misconceptions	
The Road Ahead	
2 <i>Matter Turns Intelligent</i>	49
What Is Intelligence?	
What Is Memory?	
What Is Computation?	
What Is Learning?	
3 <i>The Near Future: Breakthroughs, Bugs, Laws, Weapons and Jobs</i>	82
Breakthroughs	
Bugs vs. Robust AI	
Laws	
Weapons	
Jobs and Wages	
Human-Level Intelligence?	

4	<i>Intelligence Explosion?</i> Totalitarianism Prometheus Takes Over the World Slow Takeoff and Multipolar Scenarios Cyborgs and Uploads What Will Actually Happen?	134
5	<i>Aftermath: The Next 10,000 Years</i> Libertarian Utopia Benevolent Dictator Egalitarian Utopia Gatekeeper Protector God Enslaved God Conquerors Descendants Zookeeper 1984 Reversion Self-Destruction What Do <i>You</i> Want?	161
6	<i>Our Cosmic Endowment: The Next Billion Years and Beyond</i> Making the Most of Your Resources Gaining Resources Through Cosmic Settlement Cosmic Hierarchies Outlook	203
7	<i>Goals</i> Physics: The Origin of Goals Biology: The Evolution of Goals Psychology: The Pursuit of and Rebellion Against Goals Engineering: Outsourcing Goals Friendly AI: Aligning Goals Ethics: Choosing Goals Ultimate Goals?	249

8	<i>Consciousness</i> Who Cares? What Is Consciousness? What's the Problem? Is Consciousness Beyond Science? Experimental Clues About Consciousness Theories of Consciousness Controversies of Consciousness How Might AI Consciousness Feel? Meaning	281
	Epilogue: <i>The Tale of the FLI Team</i>	316
	Notes	337
	Index	349

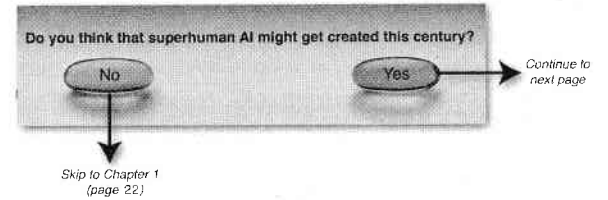


We know
BOOKS

Luino Seoane, Marin Soljačić, my editor Dan Frank and, most of all,

Meia, my beloved muse and fellow traveler, for her eternal encouragement, support and inspiration, without which this book wouldn't exist.

LIFE 3.0



Prelude

-

The Tale of the Omega Team

The Omega Team was the soul of the company. Whereas the rest of the enterprise brought in the money to keep things going, by various commercial applications of narrow AI, the Omega Team pushed ahead in their quest for what had always been the CEO's dream: building general artificial intelligence. Most other employees viewed "the Omegas," as they affectionately called them, as a bunch of pie-in-the-sky dreamers, perpetually decades away from their goal. They happily indulged them, however, because they liked the prestige that the cutting-edge work of the Omegas gave their company, and they also appreciated the improved algorithms that the Omegas occasionally gave them.

What they didn't realize was that the Omegas had carefully crafted their image to hide a secret: they were extremely close to pulling off the most audacious plan in human history. Their charismatic CEO had handpicked them not only for being brilliant researchers, but also for ambition, idealism and a strong commitment to helping humanity. He reminded them that their plan was extremely dangerous, and that if powerful governments found out, they would do virtually anything—including kidnapping—to shut them down or, preferably, to steal their code. But they were all in, 100%, for much the same reason that many of the world's top physicists joined the Manhattan Project to develop nuclear weapons: they were convinced that if they didn't do it first, someone less idealistic would.

The AI they had built, nicknamed Prometheus, kept getting more capable. Although its cognitive abilities still lagged far behind those

of humans in many areas, for example, social skills, the Omegas had pushed hard to make it extraordinary at one particular task: programming AI systems. They'd deliberately chosen this strategy because they had bought the intelligence explosion argument made by the British mathematician Irving Good back in 1965: "Let an ultra-intelligent 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 ultra-intelligent 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 ultra-intelligent 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."

They figured that if they could get this recursive self-improvement going, the machine would soon get smart enough that it could also teach itself all other human skills that would be useful.

The First Millions

It was nine o'clock on a Friday morning when they decided to launch. Prometheus was humming away in its custom-built computer cluster, which resided in long rows of racks in a vast, access-controlled, air-conditioned room. For security reasons, it was completely disconnected from the internet, but it contained a local copy of much of the web (Wikipedia, the Library of Congress, Twitter, a selection from YouTube, much of Facebook, etc.) to use as its training data to learn from.* They'd picked this start time to work undisturbed: their families and friends thought they were on a weekend corporate retreat. The kitchenette was loaded with microwavable food and energy drinks, and they were ready to roll.

When they launched, Prometheus was slightly worse than them at

programming AI systems, but made up for this by being vastly faster, spending the equivalent of thousands of person-years chugging away at the problem while they chugged a Red Bull. By 10 a.m., it had completed the first redesign of itself, v2.0, which was slightly better but still subhuman. By the time Prometheus 5.0 launched at 2 p.m., however, the Omegas were awestruck: it had blown their performance benchmarks out of the water, and the rate of progress seemed to be accelerating. By nightfall, they decided to deploy Prometheus 10.0 to start phase 2 of their plan: making money.

Their first target was MTurk, the Amazon Mechanical Turk. After its launch in 2005 as a crowdsourcing internet marketplace, it had grown rapidly, with tens of thousands of people around the world anonymously competing around the clock to perform highly structured chores called HITs, "Human Intelligence Tasks." These tasks ranged from transcribing audio recordings to classifying images and writing descriptions of web pages, and all had one thing in common: if you did them well, nobody would know that you were an AI. Prometheus 10.0 was able to do about half of the task categories acceptably well. For each such task category, the Omegas had Prometheus design a lean custom-built narrow AI software module that could do precisely such tasks and nothing else. They then uploaded this module to Amazon Web Services, a cloud-computing platform that could run on as many virtual machines as they rented. For every dollar they paid to Amazon's cloud-computing division, they earned more than two dollars from Amazon's MTurk division. Little did Amazon suspect that such an amazing arbitrage opportunity existed within their own company!

To cover their tracks, they had discreetly created thousands of MTurk accounts during the preceding months in the names of fictitious people, and the Prometheus-built modules now assumed their identities. The MTurk customers typically paid after about eight hours, at which point the Omegas reinvested the money in more cloud-computing time, using still better task modules made by the latest version of the ever-improving Prometheus. Because they were able to double their money every eight hours, they soon started saturating MTurk's task supply, and found that they couldn't earn more

* For simplicity, I've assumed today's economy and technology in this story, even though most researchers guess that human-level general AI is at least decades away. The Omega plan should get even easier to pull off in the future if the digital economy keeps growing and ever more services can be ordered online on a no-questions-asked basis.

than about a million dollars per day without drawing unwanted attention to themselves. But this was more than enough to fund their next step, eliminating any need for awkward cash requests to the chief financial officer.

• Dangerous Games

Aside from their AI breakthroughs, one of the recent projects that the Omegas had had the most fun with was planning how to make money as rapidly as possible after Prometheus' launch. Essentially the whole digital economy was up for grabs, but was it better to start by making computer games, music, movies or software, to write books or articles, to trade on the stock market or to make inventions and sell them? It simply boiled down to maximizing their rate of return on investment, but normal investment strategies were a slow-motion parody of what they could do: whereas a normal investor might be pleased with a 9% return per *year*; their MTurk investments had yielded 9% per *hour*; generating eight times more money each day. So now that they'd saturated MTurk, what next?

Their first thought had been to make a killing on the stock market—after all, pretty much all of them had at some point declined a lucrative job offer to develop AI for hedge funds, which were investing heavily in exactly this idea. Some remembered that this was how the AI made its first millions in the movie *Transcendence*. But the new regulations on derivatives after last year's crash had limited their options. They soon realized that, even though they could get much better returns than other investors, they'd be unlikely to get returns anywhere close to what they could get from selling their own products. When you have the world's first superintelligent AI working for you, you're better off investing in your own companies than in those of others! Although there might be occasional exceptions (such as using Prometheus' superhuman hacking abilities to get inside information and then buy call options on stocks about to surge), the Omegas felt that this wasn't worth the unwanted attention it might draw.

When they shifted their focus toward products that they could develop and sell, computer games first seemed the obvious top choice.

Prometheus could rapidly become extremely skilled at designing appealing games, easily handling the coding, graphic design, ray tracing of images and all other tasks needed to produce a final ready-to-ship product. Moreover, after digesting all the web's data on people's preferences, it would know exactly what each category of gamer liked, and could develop a superhuman ability to optimize a game for sales revenue. *The Elder Scrolls V: Skyrim*, a game on which many of the Omegas had wasted more hours than they cared to admit, had grossed over \$400 million during its first week back in 2011, and they were confident that Prometheus could make something at least this addictive in twenty-four hours using \$1 million of cloud-computing resources. They could then sell it online and use Prometheus to impersonate humans talking up the game in the blogosphere. If this brought in \$250 million in a week, they would have doubled their investment eight times in eight days, giving a return of 3% per hour—slightly worse than their MTurk start, but much more sustainable. By developing a suite of other games each day, they figured they'd be able to earn \$10 billion before long, without coming close to saturating the games market.

But a cybersecurity specialist on their team talked them out of this game plan. She pointed out that it would pose an unacceptable risk of Prometheus *breaking out* and seizing control of its own destiny. Because they weren't sure how its goals would evolve during its recursive self-improvement, they had decided to play it safe and go to great lengths to keep Prometheus confined ("boxed") in ways such that it couldn't escape onto the internet. For the main Prometheus engine running in their server room, they used physical confinement: there simply was no internet connection, and the only output from Prometheus was in the form of messages and documents it sent to a computer that the Omegas controlled.

On an internet-connected computer, on the other hand, running any complicated program created by Prometheus was a risky proposition: since the Omegas had no way of fully understanding what it would do, they had no way of knowing that it wouldn't, say, start virally spreading itself online. When testing the software that Prometheus had written for MTurk tasks, the Omegas guarded against

this by running it only inside a virtual machine. This is a program that simulates a computer: for example, many Mac users buy virtual machine software that lets them run Windows programs by tricking them into thinking that they're actually in a Windows machine. The Omegas had created their own virtual machine, nicknamed Pandora's Box, which simulated an ultrasimplified machine stripped of all bells and whistles that we usually associate with computers: no keyboard, no monitor, no loudspeakers, no internet connectivity, nothing. For the MTurk audio transcriptions, the Omegas set things up so that all that could go into Pandora's Box was one single audio file and all that could come out was one single text document—the transcription. These laws of the box were to the software inside like the laws of physics are to us inside our Universe: the software couldn't travel out of the box any more than we can travel faster than the speed of light, no matter how smart we are. Except for that single input and output, the software inside Pandora's Box was effectively trapped in a parallel universe with its own computational rules. The Omegas had such strong breakout paranoia that they added boxing in time as well, limiting the life span of untrusted code. For example, each time the boxed transcription software had finished transcribing one audio file, the entire memory content of Pandora's Box was automatically erased and the program was reinstalled from scratch. This way, when it started the next transcription task, it had no knowledge of what had previously happened, and thus no ability to learn over time.

When the Omegas used the Amazon cloud for their MTurk project, they were able to put all their Prometheus-created task modules into such virtual boxes in the cloud, because the MTurk input and output was so simple. But this wouldn't work for graphics-heavy computer games, which couldn't be boxed in because they needed full access to all the hardware of the gamer's computer. Moreover, they didn't want to risk that some computer-savvy user would analyze their game code, discover Pandora's Box and decide to investigate what was inside. The breakout risk put not merely the games market off-limits for now, but also the massively lucrative market for other software, with hundreds of billions of dollars up for grabs.

The First Billions

The Omegas had narrowed their search to products that were highly valuable, purely digital (avoiding slow manufacturing) and easily understandable (for example, text or movies they knew wouldn't pose a breakout risk). In the end, they had decided to launch a media company, starting with animated entertainment. The website, the marketing plan and the press releases had all been ready to go even before Prometheus became superintelligent—all that was missing was content.

Although Prometheus was astonishingly capable by Sunday morning, steadily raking in money from MTurk, its intellectual abilities were still rather narrow: Prometheus had been deliberately optimized to design AI systems and write software that performed rather mind-numbing MTurk tasks. It was, for example, bad at making movies—bad not for any profound reason, but for the same reason that James Cameron was bad at making movies when he was born: this is a skill that takes time to learn. Like a human child, Prometheus could learn whatever it wanted from the data it had access to. Whereas James Cameron had taken years to learn to read and write, Prometheus had gotten that taken care of on Friday, when it also found time to read all of Wikipedia and a few million books. Making movies was harder. Writing a screenplay that humans found interesting was just as hard as writing a book, requiring a detailed understanding of human society and what humans found entertaining. Turning the screenplay into a final video file required massive amounts of ray tracing of simulated actors and the complex scenes they moved through, simulated voices, the production of compelling musical soundtracks and so on. As of Sunday morning, Prometheus could watch a two-hour movie in about a minute, which included reading any book it was based on and all online reviews and ratings. The Omegas noticed that after Prometheus had binge-watched a few hundred films, it started to get quite good at predicting what sort of reviews a movie would get and how it would appeal to different audiences. Indeed, it learned to write its own movie reviews in a way they felt demonstrated real insight, commenting on everything from the plots and the acting to technical details such as lighting and camera angles. They took this to mean

that when Prometheus made its own films, it would know what success meant.

The Omegas instructed Prometheus to focus on making animation at first, to avoid embarrassing questions about who the simulated actors were. On Sunday night, they capped their wild weekend by arming themselves with beer and microwave popcorn, dimming the lights and watching Prometheus' debut movie. It was an animated fantasy-comedy in the spirit of Disney's *Frozen*, and the ray tracing had been performed by boxed Prometheus-built code in the Amazon cloud, using up most of the day's \$1 million MTurk profit. As the movie began, they found it both fascinating and frightening that it had been created by a machine without human guidance. Before long, however, they were laughing at the gags and holding their breath during the dramatic moments. Some of them even teared up a bit at the emotional ending, so engrossed in this fictional reality that they forgot all about its creator.

The Omegas scheduled their website launch for Friday, giving Prometheus time to produce more content and themselves time to do the things they didn't trust Prometheus with: buying ads and starting to recruit employees for the shell companies they'd set up during the past months. To cover their tracks, the official cover story would be that their media company (which had no public association with the Omegas) bought most of its content from independent film producers, typically high-tech startups in low-income regions. These fake suppliers were conveniently located in remote places such as Tiruchirappalli and Yakutsk, which most curious journalists wouldn't bother visiting. The only employees they actually hired there worked on marketing and administration, and would tell anyone who asked that their production team was in a different location and didn't conduct interviews at the moment. To match their cover story, they chose the corporate slogan "Channeling the world's creative talent," and branded their company as being disruptively different by using cutting-edge technology to empower creative people, especially in the developing world.

When Friday came around and curious visitors started arriving at their site, they encountered something reminiscent of the online

entertainment services Netflix and Hulu but with interesting differences. All the animated series were new ones they'd never heard of. They were rather captivating: most series consisted of forty-five-minute-long episodes with a strong plotline, each ending in a way that left you eager to find out what happened in the next episode. And they were cheaper than the competition. The first episode of each series was free, and you could watch the others for forty-nine cents each, with discounts for the whole series. Initially, there were only three series with three episodes each, but new episodes were added daily, as well as new series catering to different demographics. During the first two weeks of Prometheus, its moviemaking skills improved rapidly, in terms not only of film quality but also of better algorithms for character simulation and ray tracing, which greatly reduced the cloud-computing cost to make each new episode. As a result, the Omegas were able to roll out dozens of new series during the first month, targeting demographics from toddlers to adults, as well as to expand to all major world language markets, making their site remarkably international compared with all competitors. Some commentators were impressed by the fact that it wasn't merely the soundtracks that were multilingual, but the videos themselves: for example, when a character spoke Italian, the mouth motions matched the Italian words, as did the characteristically Italian hand gestures. Although Prometheus was now perfectly capable of making movies with simulated actors indistinguishable from humans, the Omegas avoided this to not tip their hand. They did, however, launch many series with semi-realistic animated human characters, in genres competing with traditional live-action TV shows and movies.

Their network turned out to be quite addictive, and enjoyed spectacular viewer growth. Many fans found the characters and plots cleverer and more interesting than even Hollywood's most expensive big-screen productions, and were delighted that they could watch them much more affordably. Buoyed by aggressive advertising (which the Omegas could afford because of their near-zero production costs), excellent media coverage and rave word-of-mouth reviews, their global revenue had mushroomed to \$10 million a day within a month of launch. After two months, they had overtaken Netflix, and

after three, they were raking in over \$100 million a day, beginning to rival Time Warner, Disney, Comcast and Fox as one of the world's largest media empires.

Their sensational success garnered plenty of unwanted attention, including speculation about their having strong AI, but using merely a small fraction of their revenue, the Omegas deployed a fairly successful disinformation campaign. From a glitzy new Manhattan office, their freshly hired spokespeople would elaborate on their cover stories. Plenty of humans were hired as foils, including actual screenwriters around the world to start developing new series, none of whom knew about Prometheus. The confusing international network of subcontractors made it easy for most of their employees to assume that others somewhere else were doing most of the work.

To make themselves less vulnerable and avoid raising eyebrows with excessive cloud computing, they also hired engineers to start building a series of massive computer facilities around the world, owned by seemingly unaffiliated shell companies. Although they were billed to locals as "green data centers" because they were largely solar-powered, they were in fact mainly focused on computation rather than storage. Prometheus had designed their blueprints down to the most minute detail, using only off-the-shelf hardware and optimizing them to minimize construction time. The people who built and ran these centers had no idea what was computed there: they thought they managed commercial cloud-computing facilities similar to those run by Amazon, Google and Microsoft, and knew only that all sales were managed remotely.

New Technologies

Over a timescale of months, the business empire controlled by the Omegas started gaining a foothold in ever more areas of the world economy, thanks to superhuman planning by Prometheus. By carefully analyzing the world's data, it had already during its first week presented the Omegas with a detailed step-by-step growth plan, and it kept improving and refining this plan as its data and computer resources grew. Although Prometheus was far from omniscient, its

capabilities were now so far beyond human that the Omegas viewed it as the perfect oracle, dutifully providing brilliant answers and advice in response to all their questions.

Prometheus' software was now highly optimized to make the most of the rather mediocre human-invented hardware it ran on, and as the Omegas had anticipated, Prometheus identified ways of dramatically improving this hardware. Fearing a breakout, they refused to build robotic construction facilities that Prometheus could control directly. Instead, they hired large numbers of world-class scientists and engineers in multiple locations and fed them internal research reports written by Prometheus, pretending that they were from researchers at the other sites. These reports detailed novel physical effects and manufacturing techniques that their engineers soon tested, understood and mastered. Normal human research and development (R & D) cycles, of course, take years, in large part because they involve many slow cycles of trial and error. The current situation was very different: Prometheus already had the next steps figured out, so the limiting factor was simply how rapidly people could be guided to understand and build the right things. A good teacher can help students learn science much faster than they could have discovered it from scratch on their own, and Prometheus surreptitiously did the same with these researchers. Since Prometheus could accurately predict how long it would take humans to understand and build things given various tools, it developed the quickest possible path forward, giving priority to new tools that could be quickly understood and built and that were useful for developing more advanced tools.

In the spirit of the maker movement, the engineering teams were encouraged to use their own machines to build their better machines. This self-sufficiency not only saved money, but it also made them less vulnerable to future threats from the outside world. Within two years, they were producing much better computer hardware than the world had ever known. To avoid helping outside competition, they kept this technology under wraps and used it only to upgrade Prometheus.

What the world did notice, however, was an astonishing tech boom. Upstart companies around the world were launching revolutionary new products in almost all areas. A South Korean startup

launched a new battery that stored twice as much energy as your laptop battery in half the mass, and could be charged in under a minute. A Finnish firm released a cheap solar panel with twice the efficiency of the best competitors. A German company announced a new type of mass-producible wire that was superconducting at room temperature, revolutionizing the energy sector. A Boston-based biotech group announced a Phase II clinical trial of what they claimed was the first effective, side-effect-free weight-loss drug, while rumors suggested that an Indian outfit was already selling something similar on the black market. A California company countered with a Phase II trial of a blockbuster cancer drug, which caused the body's immune system to identify and attack cells with any of the most common cancerous mutations. Examples just kept on coming, triggering talk of a new golden age for science. Last but not least, robotics companies were cropping up like mushrooms all around the world. None of the bots came close to matching human intelligence, and most of them looked nothing like humans. But they dramatically disrupted the economy, and over the years to come, they gradually replaced most of the workers in manufacturing, transportation, warehousing, retail, construction, mining, agriculture, forestry and fishing.

What the world didn't notice, thanks to the hard work of a crack team of lawyers, was that all these firms were controlled, through a series of intermediaries, by the Omegas. Prometheus was flooding the world's patent offices with sensational inventions via various proxies, and these inventions gradually led to domination in all areas of technology.

Although these disruptive new companies made powerful enemies among their competition, they made even more powerful friends. They were exceptionally profitable, and under slogans such as "Investing in our community," they spent a significant fraction of these profits hiring people for community projects—often the same people who had been laid off from the companies that were disrupted. They used detailed Prometheus-produced analyses identifying jobs that would be maximally rewarding for the employees and the community for the least cost, tailored to the local circumstances. In regions with high levels of government service, this often focused on commu-

nity building, culture and caregiving, while in poorer regions it also included launching and maintaining schools, healthcare, day care, elder care, affordable housing, parks and basic infrastructure. Pretty much everywhere, locals agreed that these were things that should have been done long ago. Local politicians got generous donations, and care was taken to make them look good for encouraging these corporate community investments.

Gaining Power

The Omegas had launched a media company not only to finance their early tech ventures, but also for the next step of their audacious plan: taking over the world. Within a year of the first launch, they had added remarkably good news channels to their lineup all over the globe. As opposed to their other channels, these were deliberately designed to lose money, and were pitched as a public service. In fact, their news channels generated no income whatsoever: they carried no ads and were viewable free of charge by anyone with an internet connection. The rest of their media empire was such a cash-generating machine that they could spend far more resources on their news service than any other journalistic effort had done in world history—and it showed. Through aggressive recruitment with highly competitive salaries of journalists and investigative reporters, they brought remarkable talent and findings to the screen. Through a global web service that paid anybody who revealed something newsworthy, from local corruption to a heartwarming event, they were usually the first to break a story. At least that's what people believed: in fact, they were often first because stories attributed to citizen journalists had been discovered by Prometheus via real-time monitoring of the internet. All these video news sites featured podcasts and print articles as well.

Phase 1 of their news strategy was gaining people's trust, which they did with great success. Their unprecedented willingness to lose money enabled remarkably diligent regional and local news coverage, where investigative journalists often exposed scandals that truly engaged their viewers. Whenever a country was strongly divided politically and accustomed to partisan news, they would launch one