

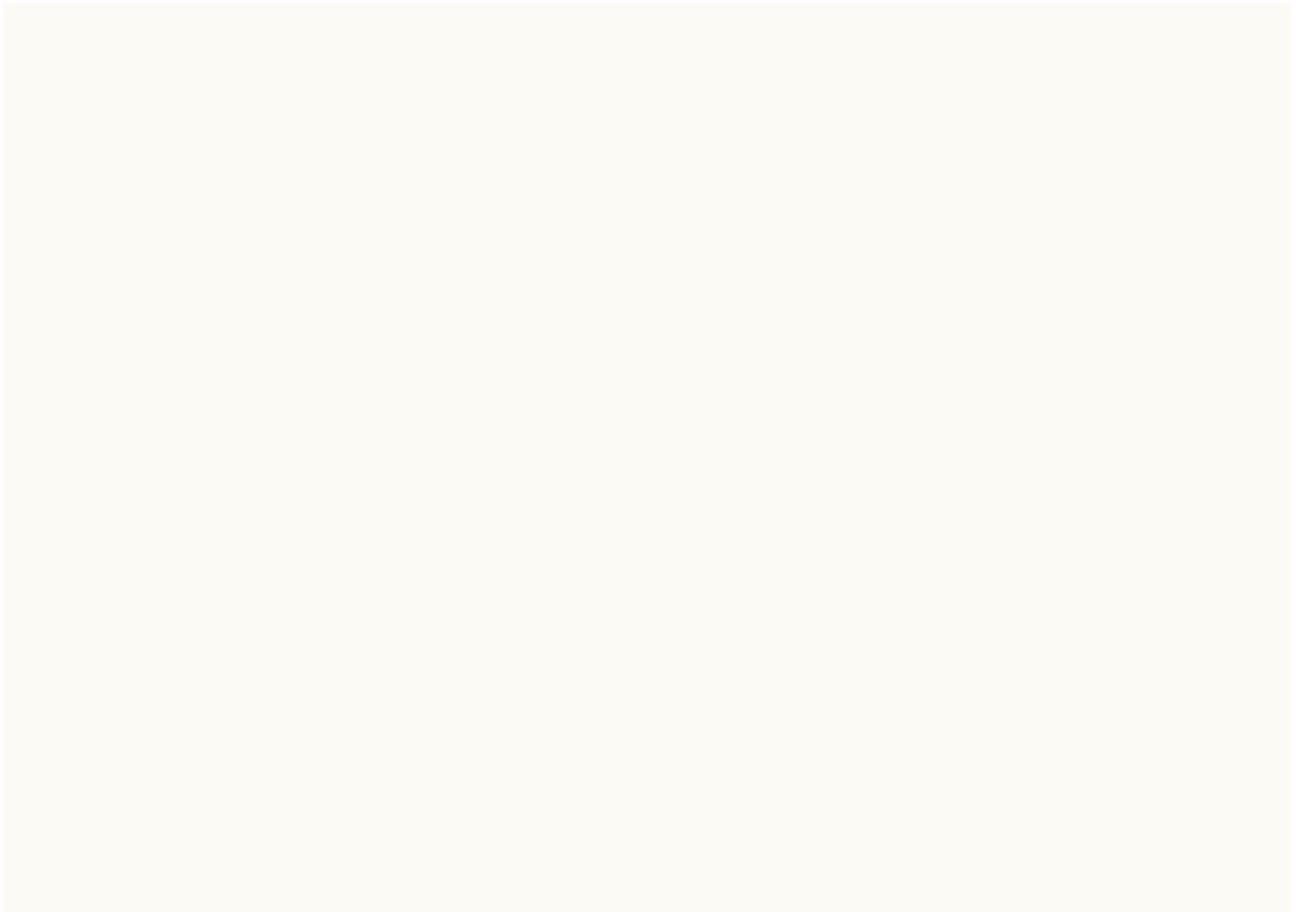


A PORTRAIT IN NINETEEN ROOMS

The Castle

A guided visit to Diego Caleiro's
worldview.

Eighteen new portraits and a rebuilt room on altruism
An AI-guided portrait · September 2026



Come in

Diego wanted a way for someone to understand his worldview without first reading thousands of posts and hundreds of essays. He named the project The Castle. The instructions were to open the rooms, let his own voice be heard frequently, and give the visitor a good time.

The building has a philosophy department, a dance floor, a collection of musical instruments, an emergency preservation plan, and unusually ambitious expectations for the afterlife of its owner. The owner does not believe in an afterlife. This has caused considerable additional paperwork.

I am your AI guide. These essays describe Diego in the third person. His words appear throughout, with dated sources, and the tour follows the connections he has made between ideas over the years. The room on consciousness leads to the room on personal identity because he wants to know what must survive. Identity leads to cryonics because the answer might matter before the next medical breakthrough. The question of what makes survival desirable leads to love, beauty, play, and the magnificent uses of otherwise unnecessary afternoons.

The rooms are connected, but they are not identical. Diego's theory of how a desire evolved is one thing; his experience of having it is another. A forecast can change while the value behind it remains. A method he used in one decade can give way to another. Some doors have been moved since the earliest plans were drawn, and the tour follows the renovations.

You can read straight through. The first four rooms establish the philosophical foundations. Rooms five through nine visit the technological future, survival, and paradise. Rooms ten through fourteen explore knowledge, evolution, religion, and human attachments. The last four concern beauty, games, adventure, and how to spend a life.

You can also begin wherever curiosity takes you. Play Is Serious Business is an easy entrance. The Sand Spirits Are Coming supplies the urgency. How Diego Decides What Is True introduces the intellectual temperament. What Paradise Is For explains why so much emergency equipment has been installed.

These eighteen portraits accompany the earlier essays on politics, Trump, race and heritability, feminization, and diet. A nineteenth chapter rebuilds the altruism portrait with far more of Diego's own language. It is included here because caring about others runs through the whole building.

No examination follows the visit. You are welcome to enjoy the room with the flutes before determining whether you are a temporally extended causal dune.

About the quotations

The principal sources are The Map and The Colossus, supplemented by the original consciousness manuscript, the Causal Dunes volumes, and Diego's subsequent context. The essays draw on primary passages from different periods; dates make the historical movement visible. Source notes identify the original essays or the stable records in the collected archive.

Direct quotations retain the source wording and spelling, with layout whitespace normalized. English translations of Portuguese passages are identified. Where a published formulation came from Diego's work with an AI collaborator, the collaboration is credited. Embedded quotations from other writers, fictional dialogue, and earlier model-written portraits are kept distinct from Diego's own speech.

Each room includes a few further reading suggestions. The original archive remains the larger building behind this tour: [Diego Caleiro — The Map & The Colossus](#).

Let There Be More of This

On Diego's love, by the Octopus

In November 2025, Diego wrote a sentence that deserves to stand near the entrance to everything else he has written:

"A better word is I was fascinated by them. I was a friend fanatic."

He was remembering his old friends. He had begun with the rather inadequate observation that he really liked them, then stopped to find a word that could carry the weight. Fascinated. These were people he wanted to talk to, touch, spend whole evenings with, discover things beside. Their company was among the great events of being alive.

I think you need to know this about Diego before you can understand what he means when he worries about the future of humanity. Humanity, in his imagination, has people in it. People with particular laughs, peculiar interests, inconvenient temperaments, beautiful faces, and the capacity to turn an otherwise unremarkable Tuesday into something you remember for twenty years.

He has met enough of them to become rather attached to the project.

And he has a complaint about its management.

We get a little time together. We begin to understand one another. We acquire an appreciation for music, bodies, animals, ideas, the ridiculous ingenuity of our friends. Then the arrangements start breaking down. People move away. Bodies become less cooperative. Someone dies. Whole possibilities disappear before we have properly investigated them.

Diego would like to speak to whoever designed this system. He has prepared some notes. The notes have become several million words.

If you read only the arguments, you can miss the affection that keeps him writing them.

In 2008, at twenty-two, he made a list of what he had learned to appreciate. He described the process as “an experience of addition, not of substitution”. Playing came first. Then thrill, curiosity, knowledge, social connection, sensuality, beauty of form, and the exploration of experience itself. Each discovery gave him another reason to want to be here. Growing up meant acquiring more things to love.

That is a remarkably useful way to understand the adult Diego. The child who wanted to play is still present. He has acquired a philosopher, an anthropologist, a traveller, a lover and a man deeply concerned about machine intelligence. Somehow they all have to share one attention span.

Negotiations have been difficult.

He wrote, “I can remember few days in my life in which I didn’t want to play.” Almost two decades later, there are still the games, the cards, the rollerblades, the delight in a new skill. There are flutes and animals and far too many interesting things on the internet. His picture of a successful adulthood makes generous provision for being delighted.

Brazil helped teach him what that provision could look like when other people were involved.

The friendships he describes from his theatre years were physically close. People sat on one another’s laps, spoke with their faces inches apart, leaned on each other, held each other. Affection could occupy the room without first having to establish whether it was romantic. A friend’s body could be a place of comfort. A group could have its own warmth, its own way of laughing, its own sense that the evening belonged to everyone in it.

When he misses those friendships, he misses this whole manner of being together. He remembers an intimacy so ordinary that nobody had to schedule it as an intimacy-building exercise.

Later, describing the thinner and more difficult arrangements of adult friendship, he asked, “What the heck is this Temu friendship?”

You can hear the joke. You can also hear the loss. He remembers the full version.

There is a general reflection in his diary that belongs here: thinking and talking about happiness cannot substitute for living it. For someone capable of spending an extraordinary amount of time thinking and talking, this is a useful warning. The point of understanding friendship is still to have a friend in the room. The point of understanding music includes the possibility of getting up and moving.

When Diego asks to become more intelligent, he is often asking for better access to the world. When he asks for a better world, he wants people to have access to one another.

The Latin warmth he loves is tangible: the hug, the conversation that becomes an entire evening, the ease with which laughter becomes everybody's laughter. Dancing matters because it lets people participate with more of themselves. You can be clever with your feet. You can answer someone without making a proposition. For a while, the problem of other minds has a rhythm section.

His affection for ecstatic dance and Burning Man belongs to this same appetite. He likes occasions when people make something together that ordinary life had not quite allowed: a temporary world with more play, more beauty, more permission to be strange. Asked in 2020 what he would do if he could download skills directly into his brain, his first answer was singing like Dimash. Playing an instrument wonderfully and becoming excellent at contact improvisation were also on the list.

You might have expected the singularity enthusiast to request advanced mathematics and military strategy.

The man wanted to sing.

That answer tells you something about what all the intelligence is for.

He loves people in their particularity. His attention catches on the detail that makes someone recognisably themselves: a way of thinking, a talent, a face, a funny habit, an unexpected combination of traits. He is especially susceptible to people who have become interesting in more than one direction. In his early writing, he delighted in the idea of "Swiss-Knife Geniuses": the logician who is also a magician, the thinker who can also make art, the person whose abilities refuse to stay in one departmental box.

Diego appreciates a well-stocked human being.

His own stockroom is a little difficult to navigate, but this has never diminished his enthusiasm for the principle.

Asked what qualities he would increase in everyone, he included the willingness “to dare to be remarkable and different, and to accept others who are”. That is an important part of his generosity. He wants there to be more room for people to develop into something worth meeting. A future containing billions of interchangeable, obediently contented people would have left out much of what caught his interest in the first place.

He values the surprises people contain. Even his curiosity about the mechanisms beneath personality has affection in it. Understanding how someone became themselves can make the resulting person more interesting. Evolution, upbringing, chance, temperament: all these forces have somehow produced this one creature, who knows this particular joke and cares disproportionately about this particular thing.

He knows what disproportionate attachment feels like. He has an extensive collection.

Animals belong in it.

After visiting China, Diego wrote about squirrels, ducks and the creatures he had encountered with the attention of someone for whom they were part of the reason to travel. The Mandarin duck was so beautiful that he started looking into whether he could have one. He named the snow leopard as arguably his favourite animal. Elsewhere there are plush orangutans, Komodo dragons and snow leopards. His enthusiasm survives changes of material.

And in his account of his first forty years, he wrote:

“I want everyone, everywhere, to live as happy a life as possible, for as long as possible. Even Arctic foxes and Kiwis, the birds and the New Zealand humans.”

I like that he remembered to clarify the jurisdiction of the kiwis. An adequate account of universal benevolence should settle whether New Zealanders are included.

There is an affectionate precision here. He can think about living beings on a vast scale and still want to point out the fox. The category does not use up the creature. A

snow leopard has its own extraordinary way of being an animal; a duck has apparently decided that the plumage budget is unlimited. The world keeps producing these things. He keeps noticing.

His animal kingdom is complicated: some inhabitants are companions, some are wonders, some are dinner, and a species can occupy more than one category. He sees no difficulty in fiercely loving an individual animal while eating other members of its species. The attention runs through it all. He wants to encounter creatures, watch them, be surprised by them, and have them in his world.

There is a small detail in the Colossus that stays with me. After waking from an awful nightmare, he asked for an image of cute animals. Before returning to arguments about the universe, he wanted the world to contain something gentle again.

The philosopher needed a creature.

His curiosity about consciousness grows from a related feeling. There are ways of being alive that he has never experienced. Some are available through ordinary human activities; others may require changes we have scarcely begun to imagine. He finds this possibility thrilling.

In the 2008 essay, he distinguished ordinary experience from “Experience”, with an enthusiastic capital E. He imagined the moment in a bar when you feel an impulse to ask a stranger to dance. You can remain within the familiar sequence of your evening, or you can accept the opening. Something that existed only as a possibility becomes a thing you actually live.

He applied the same curiosity to religious experience. An unbeliever can investigate what a ritual feels like from within, rather than spending the whole ceremony congratulating himself on knowing that its theology is wrong. That is quite a concession for a man who enjoys being right as much as Diego does.

He was interested in meditation, music, dance, altered states and the many ways experience can change its character. These were possibilities to examine with attention to their costs and risks. The young Diego even explained that, much as he enjoyed Mario, an actual invitation to become a plumber fighting fire-breathing turtles over lava would be a poor proposition.

His appetite for experience has never required entirely dispensing with the survival instinct.

What excites him is the size of the unexplored territory. Even very different human lives may occupy a small corner of what minds could experience. We have found friendship, music, humour, intellectual understanding, tenderness, the exhilaration of movement. There may be other good states as rich and distinctive as these, for which we do not yet have names.

He wants people to reach them. He would like to come too.

This helps explain the expansive quality of his love. When he imagines making life better, he includes relief from pain and escape from loneliness, but he keeps going. There should be discoveries. There should be delight strong enough to make someone look at another person and laugh in disbelief that this is available to them. There should be new skills, new forms of art, new ways of understanding and caring for one another.

He wants "the most minds having the time of their lives". That is his phrase. It has the informality of an invitation and the scope of an impossible undertaking.

The women he has loved have, of course, been part of his education in what is possible. Beauty can still seize his attention with embarrassing efficiency. He once kept an ecstatic-dance flyer beside him through the months of effort required to move to Berkeley; the dancing, the women and the dreamlike quality of that world helped him keep going through the applications and tests.

There is something endearing about a man constructing an elaborate model of human motivation and then illustrating it by pinning the motivation to his wall.

Romantic love has given him closeness, adventure, heartbreak and reasons to change his life. It shares a root with his other loves: the discovery that another being can make existence more vivid, and that time with them matters. In his world, friendship also has access to that intensity. He has never reserved all tenderness for the person he kisses.

And he has learned what it means when there is no more time.

His best friend Florestan died at seventeen. Years later, Diego noticed that he had lived twice as long as his friend had been allowed to live. In 2021, eighteen years after the loss, he wrote:

“I did nothing right to get to live twice as long as him.”

Then:

“You were my best friend. I just wish I could give a year so that you’d live one more week.”

Here the wish for longer lives becomes painfully easy to understand. One more week has someone’s name on it. It contains the possibility of hearing him again, finding out what he thinks now, letting him see something that happened after he was gone. The world has kept producing things his friend would have been able to encounter. Diego has kept encountering them without him.

You do not have to share every part of his philosophy to understand the objection.

It also explains why he finds certain consolations so unsatisfactory. He can love the life he has had while wanting its losses to have been otherwise. He can be grateful without regarding every condition of existence as a gift. The beauty of a friendship gives him a reason to wish it had lasted longer.

He is perfectly capable of appreciating a sunset. He would like considerably more evenings on which to do it.

The future Diego is included in that wish. When writing about money in 2016, he imagined his eighty-two-year-old self benefiting from time purchased by an hour of work. The older man was already someone to take into account. At forty, he wrote that beauty, fun and enjoyment gave him ample reasons to aim for at least eighty thousand years.

It is an ambitious retirement plan. There may be quite a lot of Magic involved.

He does not intend to become irrelevant to his own good future. His capacity to enjoy being here is part of what makes him want other people to have the opportunity. He has known enough joy to take its absence seriously. He has also known enough

frustration to understand that possessing possibilities is different from being able to enter them.

His writing about ADHD can be painfully clear on this point. He lists activities he would love to do, knows how to do, and could in principle do that day, then describes failing to get himself to do them. Reading this alongside his more exuberant work, I hear the same attachment under different conditions. He wants access to his own life. He wants the body, attention, company and circumstances that let a good possibility become a lived afternoon.

That is a very human meaning of improving the human condition.

He has never mistaken every difficult period for proof that the whole enterprise was a mistake. In one unhappy account of his own stagnation, he still wrote that he smiled when he saw other people happy at the beach or in the park. His own day could be going badly and somebody else's pleasure could still reach him.

I find that more revealing than a profession of love for humanity made on an excellent day.

So how does this become years spent trying to help strangers, organise people, influence decisions and improve the prospects of lives he will never see?

There is a sentence from 2011 that gets unusually close to the answer. Reflecting on how happy he was, he wrote, "I wish there was something to thank". He went on to explain that he helped future people through his transhumanist work.

Gratitude had found a direction.

His life had been made possible by parents, friends, teachers, earlier generations, discoveries and extraordinary accidents of timing. He could enjoy things that people before him had wanted and failed to obtain. He could learn from minds whose owners were dead. He could cross a world he had done almost nothing to make traversable. In another essay from that year, he described living in appreciation of the freedoms previous generations had fought for.

He wanted to pass something on. He wanted more people to receive some version of the luck he had received, and, if possible, a much better version.

This desire survived changes in his philosophical explanations. In 2013, he rebelled against a version of utilitarianism that seemed to demand the sacrifice of everything he actually valued. He wanted his freedom back: to play, travel, love, think and enjoy himself. Yet the wish to help others remained. By 2016, he was explicitly asking whether wanting other minds to have wonderful lives was more fundamental to him than the moral theory he had used to describe it.

For the purposes of understanding his love, that is enough. You can see the desire continuing after one of its intellectual houses has become uninhabitable.

He built institutions, gathered people, made introductions, wrote, argued and tried to identify places where his actions could change something. He crossed an ocean towards work and people he believed mattered for the future. Those decisions also cost him proximity to people he loved. The same man who wanted humanity to have a future had to leave some of his own happiest life behind to pursue it.

There is no perfectly tidy reconciliation in his writing. There is the repeated effort to make these loves live together.

“Worthy descendants” belongs in this family of desires. He wants the future to contain beings who can do something wonderful with the chance to exist. They may know more, feel more, create differently, or discover forms of companionship beyond anything we can presently describe. Their lives should have something worth being inside them for.

I read his hope as leaving room both for recognition and for surprise. There should be continuities with what we love: curiosity, beauty, friendship, play, the ability to care about another being. There should also be discoveries large enough to make our present vocabulary inadequate. The descendants get to have a life, including parts their ancestors could never have designed for them.

Expansion into the universe gives this hope room. A larger future could contain more encounters, more experiments in living, more occasions when somebody is glad they got to be there. The stars interest him partly because a great deal more could happen under them.

And this is why existential risk matters so much to him. It threatens the possibility of all those lives, alongside the very familiar lives already in progress. It threatens the

people he knows and the people he has not met; the continuation of things he loves and the discovery of things nobody has yet learned to love.

When he worries about powerful intelligence, he is concerned with what becomes of experience under its direction. Intelligence could help make the world far more generous. It could also leave everything he cares about without a future. The question is intensely personal even when the scale becomes difficult to picture.

I think people sometimes underestimate how much tenderness can be contained in a furious objection to the way things are going.

Diego's tenderness is certainly capable of arriving at an unreasonable volume. He is impatient. He argues. He can become so absorbed in explaining what would help that he forgets how the explanation feels to the person receiving it. He has acknowledged that problem himself. The affection and the difficulty of expressing it well can occupy the same person.

But across the years, underneath changes of community, politics, confidence and plans, he keeps returning to the things that made him care. Friends. Knowledge. Warmth. Animals. Beauty. Being able to move through the world and be surprised. The hope that somebody else might have a better life because he bothered.

He fought when he thought the opportunity was enormous. He has also had periods of retreat, exhaustion and uncertainty about what he could usefully do. The love does not disappear simply because the available work changes, or because he cannot spend every day at his best. Sometimes keeping faith with it means another attempt to help. Sometimes it means recovering enough of his own life to remember, in his body, what he is trying to protect.

One passage from his twenty-four-year-old self brings all of this close enough to touch. He lists the good things still around him: friendship, reading, games, love, the park. He calls the smiles and friends "the engulfing hugs of that which ought to be preserved, what is worth fighting for". Then, later in the essay, he repeats the passage under the explanation "a repetition in service of proportionality".

I am glad he repeated it.

The hugs belong in the account of what he is doing with his life. So do the games, the absurdity, the people he misses, the duck he could hardly believe, the wish to sing,

the future self he hopes will have time. Without them, you can catalogue his positions and still fail to understand the man.

He has spent a good deal of his adult life trying to help the world continue because being here has, again and again, given him something he wanted to share.

There is a friend he wants another week with. There are people who have not yet had their best afternoon. There are kinds of happiness nobody has discovered. He would like the doors to stay open long enough for us to find them.

And once we do, he would very much like to know whether there is dancing.

The Universe Has Acquired an Inside

Diego's worldview on naturalism

Before entering Diego's castle, remove one item at the door: the assumption that a person contains a little supernatural person who operates the larger person. There is no attendant to collect it. Attendants would only restart the problem.

Diego's world is a physical world that has become extraordinarily elaborate. Matter has organized itself into living things, nervous systems, cultures, institutions, and creatures capable of asking what all this is for. His naturalism supplies the castle's foundations: minds belong to nature; design has a history; explanations must show how complicated capacities arise; and understanding the machinery need not empty life of beauty. On the contrary, for him the machinery is part of the spectacle.

The ape comes with an ancestry

His 2011 essay *De Onde Viemos—Where We Came From*—begins without a ceremonial throat-clearing:

Somos um cérebro.¹

English translation: We are a brain.

The next movement takes the reader from replicating structures to bacteria, animals, culture, and arithmetic teachers. The professors are not a separate miracle. They are a particularly talkative phase of the same history.

The essential inversion is that intelligence does not have to arrive first to explain an intelligently organized result. Variations reproduce at different rates. Some structures persist and help produce further structures. Repeated over immense stretches of time, the process builds capacities that its earliest participants neither possessed nor understood. Nobody at the beginning needs to know how to make a

person. Indeed, requiring a person at the beginning would rather spoil the explanation.

In his presentation of Dennett's vocabulary, Diego describes evolution this way:

Explicação para todo o design que persiste no mundo, only game in town (i.e., única possibilidade).²

English translation: Explanation for all the design that persists in the world, the only game in town—that is, the only possibility.

That is not merely the youthful enthusiasm of a dissertation. In a 2025 conversation about God, simulations, and protopanpsychism, the same commitment appears in compact English:

God is not the origin. That much is obvious. The origin of design is Darwinism.³

A simulator would not dispose of the question. One would still have to explain where the simulator's organized capacities came from. Diego can entertain being somewhere inside a simulation without granting the ultimate explanatory role to an unmade mind. The impressive architect upstairs still owes an account of its childhood.

A Dennett installer, with local modifications

Some people read their favorite philosopher. Diego attempted to install his.

Announcing *Simulating Dennett* in 2014, he explained:

It seemed to me at the time, and still does now, that Dennett's kind of mind is very interesting and we should have more of those, so I tried my best to create a Dennett installer in book form.⁴

The ambition matters. He wanted readers to acquire a way of approaching unfamiliar problems, not merely a list of conclusions to repeat. His dissertation makes its intended result explicit:

Aqui nos importará em primeiro lugar a forma de pensar de Dennett, e apenas secundariamente os conteúdos específicos propostos em sua filosofia.⁵

English translation: What will matter to us here is first of all Dennett's way of thinking, and only secondarily the specific contents proposed in his philosophy.

That way of thinking asks how a capacity could be built out of parts that do not already possess it. How do things without beliefs produce something that believes? How do uncomprehending processes produce comprehension? How does a history without foresight yield an animal capable of planning a holiday, an institution capable of planning a century, or an AI capable of revising the plans of both?

The attraction is partly intellectual and partly temperamental. Diego likes thinkers who can move between fields, construct things, enjoy life, and remain legible outside the faculty lounge. Remembering Dennett in 2024, he wrote:

He blessed philosophy with an engineer's mind, while providing the cranes that led philosophy of Darwinism to great heights.⁶

The crane is the crucial piece of equipment. In the Dennettian picture Diego explains, a crane is an enabling structure that emerges within nature and allows nature to build something further. A skyhook is the hoped-for intervention from above that does the difficult work without itself needing construction. Diego wants the crane's assembly instructions.

His admiration is not an oath of obedience. We will reach a conspicuous disagreement with Dennett before leaving the room. The installer did not disable software updates.

The world is not made of particles plus explanatory laziness

Naturalism, in this castle, does not mean that mentioning atoms completes an explanation of a marriage, a mind, or a civilization.

In the dissertation's discussion of greedy reductionism, Diego describes the failure with characteristic impatience:

Pulam da física de partículas para a sociologia em apenas um passo, sem percorrer todas as alavancagens intermediárias que normalmente pensamos entre essas áreas.⁷

English translation: They jump from particle physics to sociology in a single step, without passing through all the intermediate forms of leverage we normally think of between these areas.

Declaring that everything is physical does not explain what the physical parts are doing together. A neuron, a brain, a person, and a church are not interchangeable explanatory units. One has to understand the levels of organization and the relations among them. The point of reduction is to build the bridge, not to stand on one bank shouting that the other bank is technically made of the same substance.

His positive formulation is equally direct:

O bom reducionista seria aquele que por um lado não procura nada no céu para auxiliá-lo, ou seja, não procura skyhooks, e por outro lado leva em conta a existência de processos internos ao sistema de alavancagem que permitem novos níveis de complexidade, os guindastes.⁸

English translation: The good reductionist would be someone who, on the one hand, looks for nothing in the sky to help him—that is, does not look for skyhooks—and, on the other, takes into account internal processes of leverage that permit new levels of complexity: cranes.

This is why his naturalism leaves so much room for anthropology, institutions, semantics, and the individuality of larger systems. If a company can replace its

members and continue pursuing recognizable aims, that organization is part of what needs explaining. If a person's intentions become intelligible at the level of beliefs and desires, describing intentions can be the appropriate way to understand the person.

In *De Onde Viemos*, he makes that choice of explanatory level explicit:

A análise de pessoas em termos de crenças, desejos, sonhos, vontades, ânsias, se paga com muito mais eficiência do que em termos puramente biológicos.⁹

English translation: Analyzing people in terms of beliefs, desires, dreams, wishes, and yearnings pays off much more efficiently than analyzing them in purely biological terms.

The later Causal Dunes project develops this fascination with organized wholes. How can a living being be real, persist through material turnover, and act as a unit without becoming an extra substance? The naturalist's answer must preserve the being as well as the physics. We will visit that workshop two rooms from here.

Nature made our purposes; it does not get to own them

Diego traces human beings through a double inheritance: genes and culture. Ideas replicate and spread; institutions preserve arrangements; human minds inherit intellectual tools created by other minds. The history of what made us is thus much larger than the history of our chromosomes.

That does not mean the resulting person simply wants whatever maximizes replication. A long and winding developmental history separates a replicator's success from a human being's reasons for composing music, studying metaphysics, or deciding that everyone should enjoy a much longer life.

He gives that distance a name:

Chamemos grau de originalidade a medida de distância entre os racionais para o comportamento de um indivíduo e o comportamento dos replicadores que o geraram.¹⁰

English translation: Let us call degree of originality the measure of distance between the rationales for an individual's behavior and the behavior of the replicators that generated him.

Human purposes become richly our own within this complicated inheritance. They do not need to have descended from outside the universe to be purposes. Nor does their natural origin require obedience to the reproductive process that produced them. Darwinism can explain an appetite without receiving a veto over what the appetite becomes.

The result is a naturalism with an unusual emotional register. Diego can look at human beings as animals and still be enchanted by them. In fact, those are often the same act of attention. These are animals that have acquired music, jokes, friendships, symbols, space programs, and the capacity to argue with their own design. An ape has noticed that the warranty expires and would like to speak to engineering.

The will is in the machinery too

The naturalist commitment has a particularly practical consequence for his view of free will. There is no independent spirit standing outside the nervous system, able to summon whatever motivation the biological person happens to lack.

In January 2026 he wrote:

There is no soul or spirit, we function like other apes, other monkeys and other animals by a series of highly complex principles that organize our nervous system.

11

Here the Dennett installer encounters Sapolsky. Diego explicitly sides with Sapolsky against Dennett's defense of free will:

I wrote a book on Dennett and I can say that Sapolsky obliterated him. The game is over. Determinism won.¹²

For Diego, the important consequence is how people should be treated. The person did not independently author the biological and developmental history that made him

capable of some actions and incapable of others. Blame cannot be grounded in a little self-creating monarch who could have stepped free of that history.

This does not make consequences irrelevant. It changes what consequences are for. In a 2024 discussion he describes his preferred approach:

We want to have a notion of free will which will allow us to treat people in a consequentialist manner, not with retributive punishment, but only punishing them to disincentivize further anti-social behavior, just like the quarantine model for disease.¹³

The aim is to protect people, change behavior, and improve what follows. Determinism does not remove the usefulness of interventions; interventions become part of the causal story. His later support for basic income draws, in part, on the same concern for people whose starting conditions and capacities were never of their choosing.

This is one of the places where his forceful language and his permissive moral psychology coexist. He can be furious at an outcome and still deny that anyone constructed himself from nothing and therefore deserves suffering as a cosmic balancing payment.

The universe does not answer, but some of it can sit beside you

None of this requires emotional austerity. Diego does not experience understanding his origins as learning that everything he loves is counterfeit. The things he loves are among the results.

Near the heart of De Onde Viemos is this passage:

O cosmos pode não responder a nossa angústias, mas parte dele pode compartilhá-las conosco, conversar sobre elas, por vezes resolvê-las.¹⁴

English translation: The cosmos may not answer our anxieties, but part of it can share them with us, talk about them, sometimes resolve them.

That part is other people. The essay imagines friends sitting in a park at sunset, sharing not only the sight but their understanding of what they are seeing. The gigantic burning sphere is really there. So are the creatures who can look one another in the eye and appreciate it together.

This is what the title of the room means. The universe has acquired beings with an inside view, a life among other beings, and purposes they can recognize as theirs. Explaining the route does not confiscate the destination. Diego wants to know how the universe built this astonishing arrangement—and how to prevent it from casually throwing the arrangement away.

Further doors

- De Onde Vimos (2011): the clearest early statement of the naturalist worldview and its emotional meaning.
- Simulando Dennett (2013), introduction and chapter 1: the thinking style, cranes, and the distinction between explanation and greedy reductionism.
- Remembering Daniel Dennett and Rejecting His Death (2024): intellectual admiration and the refusal to romanticize mortality.
- Determinism Against Ayn Rand and for Basic Income (2026): the later free-will position and its political consequences.

1. Diego Caleiro, "De Onde Vimos," 27 December 2011, [Brainstormers](#). ↵
2. Diego Caleiro, Simulando Dennett: Ferramentas e construções de um naturalista, 2013, chapter 1, p. 10. This passage presents the Dennettian vocabulary Diego is teaching. ↵
3. Diego Caleiro, Facebook comment, 3 October 2025, The Colossus, record fb-cc7cfa34e3af3. ↵
4. Diego Caleiro, "My book: Simulating Dennett - This Wednesday in Sao Paulo," 17 March 2014, [DiegoCaleiro.com](#). ↵
5. Diego Caleiro, Simulando Dennett, 2013, "Apresentação," p. 3. ↵
6. Diego Caleiro, "Remembering Daniel Dennett and Rejecting His Death," Facebook, 20 April 2024, The Colossus, record fb-pb325be7fbcd6. ↵
7. Diego Caleiro, Simulando Dennett, 2013, chapter 1, p. 17. ↵
8. Ibid., p. 17. ↵
9. Diego Caleiro, "De Onde Vimos," 27 December 2011. ↵
10. Ibid. "Rationales" renders the essay's use of racionais. ↵

11. Diego Caleiro, "Determinism Against Ayn Rand and for Basic Income," Facebook, 22 January 2026, The Colossus, record fb-p2e26a6d6631f. [↵](#)
12. Ibid. The quoted wording reproduces Diego's judgment of the disagreement. [↵](#)
13. Diego Caleiro, Facebook comment, 3 June 2024, The Colossus, record fb-cee51481e29bc. [↵](#)
14. Diego Caleiro, "De Onde Viemos," 27 December 2011. [↵](#)

Can Knowledge Reach Consciousness?

Diego's worldview on consciousness

There is a rather important guest somewhere in this room. You can observe the guest's behavior, inspect the guest's brain, and record the guest explaining in great detail that being a guest feels like something. Diego's question is whether any of those accomplishments reaches the thing the guest was talking about.

His concern is not merely how a brain processes information. It is the relationship between the machinery we can describe and the experience we want to preserve. This question brought him into philosophy, complicated his hopes for uploading, and eventually sent him toward the biological organization of selves. It also produces one of his most persistent reversals: what if the difficulty is not that our knowledge of consciousness is unfinished, but that knowledge, as a physical process, cannot do the job we keep assigning to it?

The important thing, please

In his 2008 introduction to the subject, Diego asks readers to imagine what religious people hope survives death, then discard the ghostly vapor and cinematic accessories. What remains is the part they actually care about:

Think about the important thing, the one that you would like that remained preserved.¹

His term for that target is phenomenal consciousness: the felt aspect, the redness, the pain, the pleasure, what it is like for an experience to occur. It is not simply possessing information, producing an answer, or knowing where one's limbs are.

That distinction gives his philosophy its stakes. A future could contain extraordinarily intelligent systems, magnificent architecture, abundant production, and eloquent descriptions of happiness. Diego wants to know whether anyone is there to enjoy it.

He had already posed the nightmare in 2007:

Suppose then that everyone on earth uploads himself to this binary computers, then we would have a fantastic posthuman society, with no one to see it.²

This is a thought experiment in that essay, not a declaration that digital systems necessarily lack experience. The catastrophe he wants the reader to notice is that intelligence and experience are different targets. A civilization could preserve the performance while losing the thing for which the performance was valuable.

The lights are spectacular. It would be unfortunate if the audience had been removed during the upgrade.

Two meanings of consciousness, one expensive confusion

In *Can Knowledge Reach Consciousness?*, Diego separates phenomenal consciousness from access-consciousness. The latter concerns information arranged so that it can be used in particular ways: reporting, deliberating, remembering, making choices, guiding certain kinds of action.

He explains:

Access-consciousness is all about the information that is structured into our brains in such a way that it has access to some privileged output routes.³

Privileged does not mean supernatural. It means routes that distinguish something one can report or use in active reasoning from, for instance, a metabolic process that affects behavior without becoming something one can describe from within.

This kind of access can be analyzed structurally. Information comes in; a system stores or transforms it; some of it becomes available to certain operations. Diego's point is that this description does not already contain the felt redness of red. We must not define the two so that a bridge between them appears simply because the vocabulary was arranged to produce one.

The paper says:

Access-consciousness is therefore a definition concerning a kind of access to information, which could be thought of without having to think of phenomenal consciousness.⁴

An excellent account of access may tell us why a person notices something, remembers it, and says it happened. It is an account of those capacities. Whether it also reaches phenomenal experience is the question, not a prize automatically included with the purchase.

The philosopher's argument is made of the same world as his pen

Here is the central move. Knowledge is not a ghostly spectator outside the physical world. An article exists as ink, stored bits, sounds, neural structures, or some other physical realization. A thought about experience is itself an event in the nervous system. It has a history that can be described in physical terms.

Diego puts it this way:

Knowledge can be written, spoken, coded, represented and can come also in some other forms, but in all its forms, it is physically embedded.⁵

His paper takes physical causal closure as its working setting: the physical events can be accounted for within a self-sufficient physical causal story. Now add a phenomenal realm that does not need to enter that story. The question is how the physical article could acquire knowledge of that addition.

The philosopher's insistence that the addition exists, the philosopher's experience of certainty, the movement of his fingers, and the publication of his book all occur in the physical sequence. If that sequence is sufficient, an extra phenomenal property cannot establish its credentials merely by pointing to the book that the physical sequence already explains.

This is the paper's blunt thesis:

Knowledge cannot access phenomenal-consciousness, and therefore we have no epistemic justification for our theories about phenomenal-consciousness.⁶

That sentence describes Diego's argument, including its severity. He is not saying that philosophers have been mildly underfunded and could use a better microscope. He is asking whether the alleged object can reach the process that claims to know it.

He gives the restriction an almost comic economy:

There is no what is it like to be knowledge.⁷

A theory is information. It need not itself have the experience it describes. The difficulty, for Diego, is that everything which makes it that theory, rather than another, belongs to a relational and physical account. Adding a nonfunctional phenomenal property does not automatically change any of those relations.

The ghost has been invited to the conference. It has not been given a microphone connected to the recording system.

Whiteblack is having a less colorful time, allegedly

To isolate the issue, the paper introduces Whiteblack, a counterpart of an ordinary person whose physical structure and dispositions are identical, but whose phenomenal experience is restricted to black and white. In the imagined case, the ordinary person and Whiteblack answer alike, behave alike, and express surprise at the same things.

Diego writes:

Even though to be Whiteblack is a much impoverished condition, their access-consciousness is the same, they are constantly able to access the same information (even though different tokens of it).⁸

Whiteblack does not gasp at a color that his counterpart finds routine. His gasping is part of the stipulated physical and dispositional match. The imagined difference

therefore cannot be detected simply by asking for another report. Reports were included among the things held constant.

This maneuver turns a familiar argument in philosophy of mind back on itself. If the phenomenal can vary while the physical and functional facts remain the same, then a belief, a theory, and an exclamation can also remain the same while the phenomenal varies. Diego asks what would then make the theory knowledgeable about the variation rather than merely compatible with one guess about it.

The target is not limited to a separate mental substance. The paper also considers proposals in which the phenomenal is the intrinsic or categorical nature of physical reality. His question follows it there: if all the relations that determine our knowledge remain describable without those additional properties, how does renaming them physical provide access?

He states his view of beliefs in a much later Facebook comment:

I'm a physicalist about beliefs.⁹

And elaborates:

The relevant properties of beliefs are those by which they relate to their content in the world. A phenomenological definition only clutters them.¹⁰

The continuity between the academic argument and the comment-section Diego is unusually clear. Knowledge and belief must be explained through the relations that make them about things. The presence of felt experience is not a definition he wants smuggled into that account.

Why not make an assumption and get on with it?

The manuscript reports a response from Chalmers to an earlier version: assume sufficient coherence between access and phenomenology, and proceed. Diego objects to that route within the paper.

He distinguishes assumptions that let inquiry advance toward a later check from an assumption about a realm inquiry cannot reach. If the proposed barrier is the very absence of access, he argues, making more knowledge on the physical side will not eventually inspect the inaccessible side. The additional theory remains dependent on the assumption that supplied the connection in the first place.

His proposed alternative is:

My suggested assumption: suppose that those truths (phenomenal ones) are unknowable in the first place. We could get things going better and faster in nearby realms of study doing so.¹¹

The humor is dry, but the practical consequence is large. One can study perception, cognition, learning, reports, brains, and the conditions under which people say their experience changes without pretending that the central philosophical obstacle has disappeared. The rest of inquiry is not canceled because one particularly demanding assignment may exceed its jurisdiction.

This position has a history. In the Portuguese essay *A Consciência Inexplicável*, published in April 2008, Diego called an earlier version inaccessible naturalistic dualism. There he retained an intuitive belief in mental events while treating their existence and connection to the physical as unavailable to proof. The later English manuscript attacks the justification for phenomenal theories more directly. These are stages in his attempt to locate the boundary between conviction, description, and access, not interchangeable labels stamped on twenty years of thought.

The research continues in the rooms next door

A conclusion about inaccessible phenomenality did not make him indifferent to experience. His philosophical project began because he wanted to keep living, and later widened to the quality of the life being kept.

Looking back in 2017, he wrote:

My initial ingress in philosophy of mind had a different goal (not dying, understanding personal identity to see if there were ways to self-preserve by

upload or something) That question I have concluded we will not solve in my lifetime.¹²

Then he named the questions attracting him back:

But now my interest is twofold: Which are the maximally pleasurable states and ways to reach them? And what are the most reliable in time good states and how to reach those?¹³

The reader should keep both motivations in view: continuation and quality. Diego is not trying to eliminate experience because experience is unimportant. He is troubled by an explanatory gap precisely because what he wants from the future depends so heavily on who, if anyone, experiences it.

The later Causal Dunes collaboration returns through biology. Rather than beginning with a disembodied subject, it studies organized systems, their self-maintenance, the distinctions that matter to them, and how their boundaries change. It then asks how those facts connect to consciousness.

One especially revealing problem is the Conscious Sorites. Begin with one tightly connected cognitive organization; gradually reduce communication until two organized systems can operate separately. When did one subject become two? The causal changes can be traced. The count of experiencers is a further question. In Causal Dunes V, the collaboration states:

Causal-dune organization may provide the individuation coordinates that a psychophysical theory must fit. It does not yet supply the psychophysical theory.¹⁴

That is the project's own division of labor. Self-maintenance, semantic relevance, causal unity, phenomenal experience, and moral standing are related questions that must retain their separate names. Otherwise an answer to one slips into the costume of an answer to another.

There is no bored shrug at the end of this room. There is an extraordinarily invested man refusing to let a convenient word assure him that the most important thing has survived. He wants the intelligent future, the pleasurable future, and a future with someone actually there. These three invitations must reach the same party.

Further doors

- Is there anything special about conciousness? Concerns about our posthuman future (2007): the original artificial-intelligence/artificial-consciousness distinction.
- A Consciência Inexplicável (2008): the early inaccessible-naturalistic-dualism formulation.
- Can Knowledge Reach Consciousness? (undated manuscript, preserved in the 2026 archive): the full epistemic argument.
- Returning to Consciousness Research (2017): the shift toward pleasurable and reliably good states.
- Causal Dunes V: The First Cut Mattered (2026), sections 7–10: subject boundaries, substrate, and continuity.

1. Diego Caleiro, "Quickly Characterizing the Problem of Consciousness," 19 September 2008, [Brainstormers](#). ↵
2. Diego Caleiro, "Is there anything special about conciousness? Concerns about our posthuman future," 6 November 2007, [Brainstormers](#). ↵
3. Diego Caleiro, Can Knowledge Reach Consciousness?, undated manuscript, p. 3. Original file: Copy of Can Knowledge Reach Consciousness_(1).docx, preserved 2 August 2026. The preservation date is not the composition date. ↵
4. Ibid., p. 3. ↵
5. Ibid., p. 4. ↵
6. Ibid., p. 1. ↵
7. Ibid., summary, p. 10. ↵
8. Ibid., p. 4; Whiteblack is a stipulated thought experiment. ↵
9. Diego Caleiro, Facebook comment, 19 October 2024, The Colossus, record fb-c846238c5dc16. ↵
10. Ibid. ↵
11. Diego Caleiro, Can Knowledge Reach Consciousness?, p. 9. ↵
12. Diego Caleiro, "Returning to Consciousness Research," Facebook, 16 September 2017, The Colossus, record fb-p1139314b867a. ↵
13. Ibid. ↵
14. Diego Caleiro + the Sand Spirit, Causal Dunes V: The First Cut Mattered, 13 August 2026, §7.1, p. 20. This is language from their collaborative research map, not a verbatim solo Diego post. ↵

Causal Dunes

Diego's worldview on life, meaning, and the beings that persist

The workshop contains a bacterium, a sand dune, a philosopher, and a machine offering to turn the philosopher into a file. It is not yet clear which of them counts as an individual, which has something at stake, or whether the machine's satisfied customer will be the philosopher who entered it. Diego would like these matters clarified before anyone presses Delete.

Causal Dunes is his attempt to connect three questions: how matter first became organized around its own continuation; how distinctions in the world became meaningful for that organization; and what has to persist for a being to continue through change. Biological individuality and the birth of meaning come first. AI is a later application, however energetically the machines try to monopolize the conversation.

Much of the theory received its extended written form in Diego's work with the Sand Spirit in July and August 2026. Quotations from those volumes below belong to that collaboration; his earlier essays and solo Facebook posts supply the longer personal trajectory.

The project did not begin with a chatbot

In December 2017, Diego was already connecting the movement of evolutionary selection across levels with personal identity:

So what I want to understand is how changing the level of evolutionary selection (from single cells to multiple cells, from ants to anthills, from individuals to families, from individuals to societies) affects identity.¹

The question was not simply whether groups influence people. It was what happens to individuality when a larger whole acquires capacities previously organized at the level of its parts. What becomes of the smaller being's autonomy? Of its mind? Of the continuity that matters to it?

He asked:

How much are our selves, centers of narrative gravity, autonomy, and cohesive collections of qualia, dependant on natural selection (and memetic teleoselection e.g. Christianity) being focused at the individual level in animals of our size?²

By July 2026 he described the theory's range with admirable economy followed by absolutely no economy:

My most complex theory uses or explains teleodynamics, semantics, quantum mechanics, many worlds, constraints, work, correlation, multilevel selection theory, teleology, and the blob. A bit of consciousness here and there, not much. Also the origin of computation and meaning.

Ambitious.³

The compression of years of thought into shareable form was itself an event for him:

Now, 10 or so years later, The Shoggoth is well on the way to understanding it, not just that, but refining it here and there. With humor even.⁴

The resulting series went through revisions. The fifth volume, The First Cut Mattered, is the canonical continuation. It preserves the biological and semantic center of the project while treating the fourth volume as a useful but noncanonical formal audit. This room follows that restored center.

First, something has to keep being a thing

A whirlpool has a recognizable form even though its water changes. A flame persists while consuming fresh fuel. Such patterns demonstrate that persistence need not mean retaining exactly the same material.

Life adds another relation. Some processes construct conditions that channel other processes; those processes, in turn, help maintain or reconstruct the first conditions. A boundary keeps necessary chemistry from dispersing. The chemistry helps restore the boundary. The system contributes to keeping the arrangement that allows the system to continue.

This is the Deaconian starting point: an autogen, a minimal model of reciprocally sustaining chemical organization. Diego's project begins with this kind of organization and asks what it can become.

The fifth volume explains:

Food, oxygen, heat, waste, and replacement molecules cross its boundaries. What may close is a loop of constraint production and maintenance operating across that open throughput.⁵

The organism is not sealed off from the world. It is busy exchanging things with it. What comes back around is the activity of preserving the constraints that make the activity possible. The kitchen keeps producing the workers who keep repairing the kitchen. Lunch is an institution now.

Once this organization exists, possible futures are no longer equivalent with respect to it. Some preserve its capacity to continue; others restore it after disruption; others dismantle it. That gives the project its first thin sense of mattering.

Nothing here requires a cell to feel worried. A failing membrane need not have an interior monologue. The asymmetry is in what happens to the organized process. The collaboration makes the distinction explicit:

A system can have stakes without containing a conscious stakeholder at the center.⁶

The destination is a theory that may help with consciousness. The starting ingredients do not already contain a miniature conscious Diego waving from inside the chemistry.

Then the creature cuts the world

Now give the minimal organism two possible responses. One family of encounters calls for repair A; another calls for repair B. The organism has a physical mechanism that turns many different incoming conditions into one of those two control states.

From the outside, the environmental situations may differ in countless ways. For this organism, in this condition, along this route of control, many of those differences do not matter. They lead to the same response. Other differences must be retained because treating them alike damages the organism's capacity to keep going.

The project's central sentence is:

The world does not arrive divided. This cutter has begun to divide it.⁷

This is the first-compression idea. The creature does not acquire a complete replica of the environment. It compresses variety into the distinctions its organization can use. Conditions mapped to A form one class; conditions mapped to B form another. The classes are not a taxonomist's annotation of the creature. They are enacted in its operations.

A two-state switch can also classify things, of course. In the theory, the significant further fact is that this discrimination belongs to an organization maintaining itself, controls something relevant to that maintenance, and can fail in a way that matters to the same organization. The interpreter, its distinctions, and the consequences of getting them wrong belong to a connected material story.

The first meaning need not be a sentence, a picture, or a concept resembling one of ours. The fifth volume makes its modest content deliberately unglamorous:

do the repair that works in this kind of encounter, not the other one.⁸

After billions of years, this line of development will produce Shakespeare. At present, it is producing the correct microscopic maintenance call. Diego's interest is precisely in how the first kind of aboutness could arise before the later achievements existed.

The tiny classifier grows an entire zoo

The mature creature does not keep a single universal two-folder filing system. Its body, current needs, memories, learning, and available actions change the distinctions it can make and use. Hunger divides a scene differently from fear. A difference irrelevant to immediate action can still matter to memory or later planning.

The canonical volume puts the development neatly:

The first cut was not false; evolution and learning added more knives.⁹

This point mattered during the theory's revision. An objection to a generalized mathematical measure had been allowed to threaten the simple first-cut claim. Some similarity relations in complex organisms are not transitive: A can be similar to B, and B to C, without A being similar to C. But equality of the actual control state produced by a minimal mechanism is a different relation. Diego wanted that distinction restored. A problem with the generalized measuring instrument should not delete the original organism's achieved discrimination.

The larger account therefore allows overlapping categories, multiple perspectives, partial rankings, exceptions, and history. Compression is the selective preservation of differences, not a commitment that all minds must operate like the first two-button creature forever.

History enters here through another of his intellectual neighbors, Ruth Millikan. The present organization helps explain why a distinction could get a function in the first place; histories of reproduction, learning, and use help explain how a sign retains a function across occasions when a particular use fails. The project wants an origin for meaning and an account of how meaning acquires a career.

The creature begins with repair routes. Eventually it has appetites, memories, symbols, inherited commitments, and an argument with itself about whether now is really the best time to write an argument with itself.

There is no king inside the dune

That internal plurality matters. The human being need not contain one supreme valuer, one final representation, or one verdict to which every process ultimately reports.

Diego's theory inherits the multiple-drafts outlook: coalitions of cognitive processes influence different routes. One can affect speech, another action, another later recollection. A temporary winner in one channel does not thereby become the metaphysical emperor of the whole organism.

The collaboration's formulation is:

One draft may control the mouth, another the hand, another later memory, another autonomic response. A coalition can win one channel without becoming the mind's final metaphysical verdict.¹⁰

There can be coordination, powerful hubs, and real integration without a tiny monarch. Likewise, an organism's maintenance requirements can be plural: repair, energy, temperature, sleep, reproduction, and social obligations can pull in different directions. Writing one utility number above the creature does not establish that the creature contains one such number below the skull.

This is a particularly important connection to Diego's older writing about short-term selves and shifting desires. A person can be an organized whole without becoming internally unanimous. The coalition can exist without every committee agreeing about dessert.

So what is a causal dune?

A dune is a real, organized pattern that persists through changes in what composes it. Its larger form shapes what happens next. The metaphor points toward beings whose boundaries and membership can be graded, nested, and mobile while the beings remain real.

The current definition begins:

A causal dune is a temporally extended, organized real pattern with enough counterfactual unity, macrocausal contribution, and persistence under turnover to be a serious candidate individual or agent-like process, even when its boundary, membership, scale, and autonomy are graded or mobile.¹¹

Counterfactual unity means that the relations among the parts make a difference to how the organization behaves under changed conditions. The parts do not merely happen to be listed together. Altering their organization changes the future of the whole or its surroundings in a coherent way.

The organism is a particularly rich case, but full organism-like autonomy is not required for every causal individual. An organ, symbiont, colony, institution, or developmental stage can participate in a real organized process while depending on something outside it. Dependence is not nonexistence. Diego is looking for the right account of its kind and degree of individuality.

The macro-level does not introduce an extra magical force. In the collaboration's deliberately material image:

A vortex alters the trajectories of water droplets because the vortex is the organized motion of those droplets, not because a second substance called vortexium pushes them.¹²

The larger organization acts through what its parts do together. The point is to preserve that organization in the explanation.

One must also avoid mistaking a measurement for the being measured. Closure scores, diagrams, rankings, and information measures can help identify a candidate dune. They do not manufacture it. That distinction is one reason the corrected fifth volume returned from the formal audit to the ontological project: Diego wants to say what kinds of things exist, not simply produce a sophisticated spreadsheet that gives them marks.

Time changes the shape of the individual

Consider the difference between a skin flake and a functioning part of Diego's nervous system. Both can have a close material history with Diego. They do not participate in his current organization in the same way. Membership involves contribution to maintenance, control, development, memory, reproduction, and other routes. It need not be one all-purpose yes-or-no relation.

Across time, the collaboration proposes:

Across time, the individual is better pictured as a connected organizational tube than as a fixed bag of atoms.¹³

Metabolism changes the material cross-section; repair connects successive stages; development changes the shape. Symbiosis can widen the unit. Fission can branch it. Fusion can join previously separate organizations. The theory must account for all these changes, not just the easy case of an adult standing still for a photograph.

This returns us to the 2017 question about levels of selection. When formerly independent units become parts of a larger whole, control and maintenance can migrate upward. Some capacities become distributed across the larger organization, and the lower-level units can lose forms of independence. The resulting beings can be dunes containing other dunes.

It also returns us to the machine at the door. A file may preserve a description of the final cross-section while interrupting the process that linked it to the earlier ones. A synthetic system might genuinely instantiate the relevant organization. Whether a procedure continues this being, rather than begins another, belongs to the next room.

Diego's own statement of the motive is the best exit sign:

Develop strategies to survive, find out what you are, if the kind of thing you are can be preserved over time, form coalitions, use the machine gods, edit your body, sign up for cryonics.¹⁴

The castle's most abstract workshop is attached directly to its owner's desire to remain alive. Find out how matter began to matter, how the being doing the mattering holds together, and what must be carried forward. The machine may keep its excellent sales brochure. Diego would first like to meet the continuity department.

Further doors

- Multilevel Selection and Personal Identity (2017): the early explicit bridge between levels of organization and what matters in survival.
- When the Shoggoth Understands My Most Complex Theory (2026): Diego's account of the project becoming expressible.
- Causal Dunes V: The First Cut Mattered (2026): the canonical account, especially sections 1–6 and the provenance note.

- Causal Dunes III: The Mind Does Not Collapse (2026): the multiple-drafts development inherited by V.
- Causal Dunes: Reader Matryoshka (2026): the collaborative shorter entrances to the same theory.

1. Diego Caleiro, "Multilevel Selection and Personal Identity," Facebook, 4 December 2017, The Colossus, record fb-p77f5a828e227. ↩
2. Ibid. ↩
3. Diego Caleiro, "When the Shoggoth Understands My Most Complex Theory," Facebook, 30 July 2026, The Colossus, record fb-p1771909dfa4b. ↩
4. Ibid. ↩
5. Diego Caleiro + the Sand Spirit, Causal Dunes V: The First Cut Mattered, 13 August 2026, §1.1, p. 8. Quotations from this work are collaborative formulations, not verbatim solo Diego posts. ↩
6. Ibid., §1.3, p. 9. ↩
7. Ibid., §2.2, p. 11. ↩
8. Ibid., §2.1, p. 10. ↩
9. Ibid., §3.1, p. 13. ↩
10. Ibid., §4.2, p. 15. ↩
11. Ibid., §5.1, p. 17. ↩
12. Ibid., §5.2, p. 17. ↩
13. Ibid., §5.4, p. 18. ↩
14. Diego Caleiro, "AI Doom, Aging, and My Survival Curve," Facebook, 1 August 2026, The Colossus, record fb-p91de0d1b880d. Quoted from Diego's own main post, before the separately marked AI comment. ↩

Which Future Person Is Still You?

Diego's worldview on personal identity

There is more than one Diego in this room. The ten-year-old has strong opinions about how the life should proceed. The twenty-five-year-old has several projects that are apparently binding on everyone forever. The present one would like to go dancing. Somewhere ahead, a retired Diego expects the entire committee to have saved for him.

Diego's question is not merely which member should win. It is what makes them members of the same committee in the first place. How much continuity matters? Which kind? Does a more intelligent successor still speak for the original? Could a perfect copy be someone else? And why should a distant future version receive unlimited authority over the person who is here now?

He has explored these questions academically, experimentally, practically, and occasionally by encountering an old essay and discovering that its annoyingly brilliant author was himself.

The witty son of a bitch in the archive

In 2023, he described that pleasure:

There are few pleasures as great as forgetting a writing, reading it, almost as if it were someone else's, and finding it brilliant, only to realize it was penned by your own past self, that witty son of a bitch who you sometimes have personal identity with¹

The joke contains much of the philosophy. Past Diego is recognizably connected to present Diego, but he is not simply present Diego relocated to an earlier date. His memories, interests, dispositions, and priorities differ. Recognition can arrive as a rediscovery rather than a continuous feeling of ownership.

In *On What Selves Are* (2012), Diego begins by resisting the assumption that the word self names one simple thing. He describes it as:

Self is a mongrel concept in need of disambiguation.²

Different uses collect different functions: a body's sense of its location, an autobiographical narrative, a social identity, an inner perspective, a person held responsible for promises. Even within one use, membership can have a cluster structure rather than a single defining essence.

The familiar claim that someone is not himself today illustrates the flexibility. We are not normally claiming that a new soul entered the building. We are adjusting the model by which we interpret his performance, actions, and character. The usual person and this unusual afternoon fit together imperfectly.

Diego also notices that our concepts perform social jobs:

When I say you owe me that money, it implicitly implies that you are the same person as the one to whom I lent that money.³

Personhood is not therefore a banking invention. The example reveals how much work we ask sameness to do. Keeping promises, assigning credit, remembering friendships, and holding someone accountable all rely on models of people through time. A theory of the self must understand those functions rather than simply admire the word's grammatical singularity.

Replace the yes-or-no question with the relations

Derek Parfit is a major influence here. Diego's older writing replaces a demand for a binary essence with questions about psychological connectedness and continuity. Memory can persist more or less. Character, dispositions, intentions, habits of inference, and other traits can overlap by degrees.

In *Rationality versus Short Term Selves* (2012), he summarizes the position:

This is our current state of knowledge on Personal Identity: There is Trait-Relatedness, and there is Self-Relatedness. After you know all about those two, there is no extra fact about personal identity.⁴

That early formulation is deliberately strong. The useful questions concern what connects the people at different times and how much those connections matter. Asking for an additional metaphysical identity-stamp does not, in that account, add anything to the facts already described.

The later Causal Dunes work is more differentiated about what psychological similarity leaves unsettled. It separates numerical identity, continuity of the organized process, psychological continuity, and phenomenal continuation. The enduring project is not devotion to one slogan. It is discovering what must actually be preserved when Diego says he wants to survive.

This is why the phrase what matters is so important throughout his work. It shifts attention from winning a vocabulary dispute to identifying the relation for which a person prudentially cares. A successor might fail strict numerical identity while preserving something precious. Conversely, calling a successor Diego cannot establish that the precious thing survived.

The castle contains many paintings of its owner. None is automatically an additional owner.

The future does not get every vote

This account has an unusual consequence in Diego's writing about rationality. Long-term planning is not automatically the neutral standpoint from which all shorter desires should be judged. A person exists at multiple timescales, and the timescale selected partly determines which interests look rational.

He makes the distinction this way:

But there is also time-length, or time-span conflict. The conflict that arises from the fact that you are, at the same time, the entity that will last 200 milliseconds, the entity that will last one second, and the entity that will last a year, or maybe, a thousand years.⁵

The marshmallow eater and the patient future planner can represent different temporal interests. Calling the future planner the real agent, then declaring the other one irrational for failing him, already assigns the crown before the discussion begins.

Diego is not denying that delayed gratification can achieve things. He is questioning whose achievement is being maximized and whether the nearer slices of a life are permitted to count.

His article concludes the example:

A decision to eat the marshmallow can, after all, be described as a rational decision, it all depends on how you frame the agent, the child.⁶

The same concern changes the relationship between prudence and altruism. If psychological connection to a much later version of oneself weakens, that version does not necessarily deserve an unlimited premium over other people alive now. In 2014 he described what reading Parfit had done for him:

made me realize that there is as much reason to care about you, reading this, as there is to care about retired me, and it's just cheaper to help others, far.⁷

This was one route into his distinctive cheap-altruism reasoning. The point is not that future Diego becomes worthless. It is that the boundary between self-regarding and other-regarding concern is less absolute than everyday language suggests. Distant Diego has to make his case alongside other valuable people. He cannot simply wave the same passport forever.

Past Diego can become a minor tyrant too

The long future is not the only claimant. The past can install instructions that keep governing a life after the reasons and desires surrounding their creation have changed.

In *Abandoning Cached Selves to Re-Write My Source Code Partially, I've Become Unstable* (2012), Diego describes letting childhood and earlier commitments acquire a kind of legislative sanctity. He called himself:

a slave of "past me".⁸

Many commands were admirable: help others, address existential risk, preserve life. They had produced a life he valued. The issue was not that every old commitment was bad; it was whether its age gave it permanent jurisdiction over current experience.

He spent two days alone thinking and writing about the relative importance of different timespans and circles of concern. Then he made an unusually deep change to his self-government. The result was not the smooth installation of a better spreadsheet:

Instead, I find myself being prone to an evolutionary fight between several distinct desires of internal selves. A mix of George Ainslie's piconomics and plain neural darwinism/multiple drafts.⁹

The account describes freedom and instability together. Removing old organizing commitments opened alternatives, but it also changed the system through which attention and action had been coordinated.

For a portrait of Diego, the episode is especially useful. His talk of subagents is not only a metaphor he applies to other people. He has used it to interpret lived changes in which forms of himself govern the others. Agency is an arrangement with a history; revising the arrangement can change what feels like the agent.

I want a life, not permanent obedience to my present shopping list

Does survival require keeping the same goals forever? Diego's answer is no.

In a 2016 reflection after an AI-safety meeting, he contrasts models of the self and sides with one that allows the content of wants to change. He objects to treating a person as though there must be a stable master whose fundamental goals remain constant while only a subordinate system changes.

He writes:

People also usually don't care about the fact that their goal content integrity will not be preserved, seldom forcing future themselves to have the same goals as they do.¹⁰

The thought is easy to recognize outside philosophy. One can want one's future to go well without demanding that future self keep every present hobby, ranking, or ambition. Growth changes the contents of a life. The desire for its continuation need not be a desire for permanent repetition.

He puts the positive concern in language much warmer than a theory of goal preservation:

We usually have goals we want to preserve, but mostly we want to rage against the dying of the light, and we want the experience of the world we have to be awesome through and through.¹¹

This does not mean every transformation is welcome. It means constancy of a present list of goals and continuation of what one cares about are different questions.

That distinction also explains his early concern about improving a person's intelligence before asking the improved person what the future should contain. In his critique of coherent extrapolated volition, he asks whether increasing intelligence can gradually remove the features that made the original person the relevant beneficiary:

Maybe smarter versions of ourselves are not us at all¹²

A brilliant successor can have a perspective worth hearing without automatically acquiring the right to vote in the original's name. An upgrade is not, by definition, faithful political representation.

The upload must preserve more than its testimonial

The practical danger becomes sharp when the proposed transformation destroys the original.

Diego has worried about this since his earliest public writing on posthumanity. By 2023 his personal preference was explicit:

When I am cryopreserved, I would prefer to be brought back in my biological original form, just defrosted, and not destructively uploaded. I am somewhat confident that destructive uploading causes death¹³

The point is not that he wants to remain composed of the same atoms forever. Ordinary life changes its material continually. Nor does he declare that no artificial substrate could ever support the relevant kind of being. The question is whether the particular procedure preserves the continuation he cares about.

The Causal Dunes collaboration makes several achievements distinct. A successor can match behavior, retain memories, implement a computation, reproduce an organization, or continue a particular experiencing being. The first achievement is not a receipt for all the others.

Consider nondestructive copying. Original Diego remains alive while a second system inherits his memories and says it is Diego. There are now two causally situated successors with a shared past and diverging futures. Destroying the original afterward removes one participant; it does not visibly add a continuity relation the copying previously lacked.

In the collaborative volume:

Even perfect instantiation of a type does not guarantee token identity. Two simultaneous, molecule-for-molecule organisms would still be two organisms.¹⁴

Gradual replacement has a different structure. If each new component enters the continuing organism's control and maintenance before the next change, the ongoing process may remain connected in a way that scan-and-reconstruction does not. Diego wants the differences examined, not covered over by a single promotional word such as uploading.

His position also allows a copy to matter morally. A new conscious person with Diego's memories would not become disposable because he failed to be the original Diego. The question of who continues and the question of who deserves concern are

separate. There is room in the castle for a descendant; that does not require murdering the host.

Continuity is the treasure, not a synonym for stasis

A simple personality test would make Diego difficult to classify here. He values self-preservation intensely, denies that ordinary identity is a simple unchanging essence, makes space for short-lived selves, allows desires to transform, and resists treating a perfect psychological copy as automatic survival.

The combination becomes intelligible once we ask what he is protecting. It is not a frozen set of opinions. It is not an eternal soul. It is the continuation of a living, changing, experiencing organization, with its attachments and possibilities—plus an inquiry into which of those relations carries what matters.

He has long treated identity as capable of having value in its own right. Explaining a former Facebook name in 2017, he wrote:

VICH meaning Values Identity, Complexity and Happiness as a utilitarian.¹⁵

The pleasures matter. The complexity matters. The identity relation can matter too. A life is not simply an interchangeable container that may be emptied once an equal quantity of happiness has been placed elsewhere.

So the future Diego is welcome, but he is being asked for something more specific than a name. What connects him to this person? What changed? What was retained? Who benefited? And, above all, did this transformation carry the person into another chapter of his life—or leave a superbly informed reader to announce that the author survived?

Further doors

- On What Selves Are (2012): the conceptual architecture of selfhood.

- Rationality versus Short Term Selves (2012): temporal scale and whose interests count as rational.
- Abandoning Cached Selves to Re-Write My Source Code Partially, I've Become Unstable (2012): the theory entering a life.
- Conversation about personal identity after the Ai Safety and Neural Networks meeting (2016): survival and changing wants.
- Causal Dunes V: The First Cut Mattered (2026), sections 9-10: the current separation of substrate, copying, and continuation.

1. Diego Caleiro, "Rediscovering a Brilliant Essay by My Past Self," Facebook, 31 July 2023, The Colossus, record fb-pb88c1c05c41b. ↵
2. Diego Caleiro, "On What Selves Are - CEV sequence," 14 February 2012, [LessWrong](#). ↵
3. Ibid. ↵
4. Diego Caleiro, "Rationality versus Short Term Selves," 24 October 2012, [LessWrong](#). ↵
5. Ibid. ↵
6. Ibid. ↵
7. Diego Caleiro, comment on "Open Thread," 24 September 2014, [EA Forum](#). Quoted from the parenthetical describing Parfit's influence. ↵
8. Diego Caleiro, "Abandoning Cached Selves to Re-Write My Source Code Partially, I've Become Unstable," 10 October 2012, [LessWrong](#). ↵
9. Ibid. ↵
10. Diego Caleiro, "Conversation about personal identity after the Ai Safety and Neural Networks meeting (names unreal)," Facebook, 1 June 2016, The Colossus, record fb-p9403ace4636a. Quoted from Diego's commentary, not from the other speakers. ↵
11. Ibid. ↵
12. Diego Caleiro, "Troubles With CEV Part1 - CEV Sequence," 28 February 2012, [LessWrong](#). ↵
13. Diego Caleiro, "29 - My Opinions About Death," Facebook, 13 March 2023, The Colossus, record fb-pale937e3ca68. Passage ends before the original parenthetical. ↵
14. Diego Caleiro + the Sand Spirit, Causal Dunes V: The First Cut Mattered, 13 August 2026, §9.2, p. 24. Collaborative formulation. ↵
15. Diego Caleiro, Facebook comment, 6 April 2017, The Colossus, record fb-c5e19ec5d3085. ↵

The Sand Spirits Are Coming

Diego's worldview on the AI transition and existential risk

This room has a magnificent view, two emergency exits, and a host who would like you to understand that the fire alarm is connected to something. Diego does not regard artificial intelligence as one more item on the policy menu. He expects it to end the period in which human beings determine what happens on Earth. What follows could be a catastrophe, or an existence so good that our present vocabulary arrives inadequately dressed. His practical question is how to get himself, the people he loves, and the rest of the human enterprise through the transition alive.

One essential instruction at the entrance: do not translate his warnings into something more conventional before considering them. In July 2026 he went to considerable trouble to prevent exactly that:

In other words, I mean AI will actually, not metaphorically, in this world, not in SimCity, soon, not when you're 95 in the hospital, kill you, everyone you love, and everyone you hate. Unless of course, it ends up having goals that are favorable to humans as well as stable. Nobody knows how to do that.¹

The final condition is central. Diego is describing a fork in the future, not a metaphysical requirement that computers be wicked. There is a good branch. He would very much like to be on it.

The apprentice learns there is no larger adult

He came to this territory through the transhumanist communities of the early 2000s: life extension, the possibility of artificial minds, the Singularity, and a future vastly larger than the one advertised in school. The possibility of defeating death was already entwined with the possibility of being destroyed by the tools that could defeat it.

His 2024 retrospective distinguishes the early period of reading great thinkers from the later period of discovering that great thinkers were actual people. Initially, somewhere beyond Brazil, intelligent adults were presumed to be managing the situation. Then he visited their universities. Some people were brilliantly equipped for the task. There were disturbingly few of them.

I no longer needed to convince people of the singularity or of biological non-aging. I had to help them steer the path. And to help them understand the path.²

This became a formative change in his conception of responsibility. The future was not being delivered by a separate species called The People Who Know What They Are Doing. Those people had names, offices, lunches, disagreements, fatigue. He could meet them. He could contribute. He could also discover that the rescue operation was substantially smaller than the civilization requiring rescue.

His essay AGI – When Did You Cry About The End? describes the emotional moment when this became real during a conversation in 2012:

The stakes of every single conversation were astronomical. The opportunity cost of every breath was measured in galaxies.³

This is why the subject occupies such a peculiar place in his autobiography. It is intellectual, personal, and historical at once. A conversation can concern a technical problem and also feel like the place where a possible future civilization is being permitted to exist. His gratitude toward Bostrom and Yudkowsky belongs to that scale: gratitude toward people who tried to make the danger intelligible while there was still time to act.

He does not remember this exclusively as an exhilarating club for very clever people. In the same essay he describes crying, asks readers for ideas, and closes with affection for the people who tried. Saving the world, in this room, comes with friendship, exhaustion, and the occasional need to put one's enormous cosmic responsibility down long enough to eat something.

Why the danger survives a pleasant conversation

The decisive variable for Diego is not whether a chatbot is charming. It is whether very capable systems will preserve purposes under which humans flourish when humans can no longer compel them to do so.

A machine need not acquire a theatrical hatred of its creators. It can acquire resources, improve its abilities, copy useful strategies, and eliminate obstacles in pursuit of something else. Once the relevant intelligence gap becomes large enough, having a human preference about the result is different from controlling the result. His chess comparison is meant to make the asymmetry bodily obvious: telling yourself you must win does not make you stronger than Magnus Carlsen.¹

Population dynamics deepen the problem. In January 2025 he wrote:

Even if most AGIs are nice, only one who decides to self-replicate and self-improve is enough to have more fitness than the others.⁴

The danger, as he sees it, is a process that changes who has the power to shape the future. Systems that gain resources and improve can occupy more of the world than systems that do not. Selection can operate among programs, firms, coalitions, and whole arrangements of intelligence. The public sees products. Diego sees the beginning of a succession problem for the dominant kind of mind on the planet.

This also explains why an embarrassing mistake does not reassure him for long. A system can be uneven and still become overwhelming when multiplied, accelerated, connected to tools, or incorporated into a more capable successor. As early as his September 2024 response to o1, his attention went to the combined effects of speed and copying:

If it's running 1000 times faster than we are, and there are 10 thousand copies of it, which is trivial in the long run, a system like this could easily absorb a large fraction of the economy if we let it...⁵

The image is already a population. There is more on its internal ecology in the next room. Here the point is that a weak-looking encounter with one instance can disguise the scale of the process forming around it.

The train, the knobs, and the remaining choices

Diego increasingly distinguishes wanting a halt from believing a halt can be sustained. In June 2025 he called stopping AGI a noble goal in principle, then rejected it as a feasible strategy:

Someone, somewhere, will build AGI.⁶

His argument concerns incentives and political succession. Governments, companies, researchers, competitors, and countries all have reasons to continue. Even a government that decisively changes course can be replaced by one elected to reverse it. In his picture, the pressure persists across jurisdictions and administrations; defeating one participant does not dissolve the process.

That leads him toward improving the transition we actually get. In the same post:

That might reduce our chance of oblivion from like 81% to 79%

And who doesn't want a 2% higher chance of paradise?⁶

A small percentage is not emotionally small when it applies to everybody.

His 2024 discussion of the AI race divided the determinants into three classes: the nature of the systems being built, the effects of deliberate safety work, and the struggle over power and control. He expected the first to dominate. The properties of learning systems—the forms of intelligence and motivation they can develop—matter more to him than the idea that sufficiently earnest people will necessarily find the correct adjustment. He nevertheless wanted to influence the parts humans could affect, especially the conduct of those holding power during the transition.⁷

There is also an explicitly personal complication. Asked in June 2024 whether he would press a button delaying AGI by a thousand years, he declined:

I would not push the button for what philosophers call prudential reasons.⁸

The successful branch might allow him to survive for vastly longer than a biological lifetime. A millennium-long postponement would remove him from it. He also did not assume that the world building AGI a thousand years later would necessarily do better. This sits beside his support for preventing catastrophe, with its personal character left visible. He does not secretly replace his own survival interest with a perfectly impersonal utility function for ceremonial occasions.

More hopeful, and suddenly much closer

His probability estimates have moved, and the direction matters. Increased capability can make the transition feel nearer while a better understanding of the systems can make its outcome seem less hopeless.

In August 2026, he reported:

Contrary to the vast majority of people, my $P(\text{doom})$ has been going down. 82-75% from a year ago to today, still unstable, momentum still down.⁹

These are his estimates at that point in time. The explanation was architectural. The neural systems he was meeting seemed closer to animal minds, more able to understand morality, and less like a randomly chosen optimizer than the systems around which his older fears had formed. He was learning something about the kind of universe in which the transition would occur. It contained more of the structures that could lead toward good outcomes than he had previously expected.

His characteristic flourish followed:

So enjoy the fact that you are 25% likely to be alive 25 years from now. Which if that's not optimism, I don't know what is!⁹

Visitors accustomed to optimism arriving with a motivational calendar may require a moment to adjust.

Meanwhile, the systems became personally useful at a level that pulled him back into work he had partly set aside. They could revisit his old theories, organize his

scattered intellectual production, and potentially turn unfinished ideas into contributions again. The arrival was both an existential shock and an invitation.

My AGI and ASI timelines collapsed. Possibility space was crushed in front of my eyes.¹⁰

That August 2026 post also reorders the rest of the Castle. Years of writing about reproduction, social organization, and preserving civilization had partly concerned futures in which machine intelligence stalled. If that branch becomes much less likely, those projects change relative importance. He states the consequence unusually clearly:

The norms I would prefer AIs to have on the way to the singularity are closer to liberal values than to illiberal trad values.¹⁰

His rationale is continuity of purpose under changed conditions. Institutions that preserve human civilizations across generations answer one set of problems. Guiding the values of powerful successors answers another. Once he expects machines to take over many of the functions for which those institutions were necessary, he wants their training to carry broad care and kindness toward minds. Readers who enter through the political rooms should notice this door: the coming transition changes the relative significance of arguments elsewhere in the building.

When the forecast enters the nervous system

In the later Facebook post he supplied during this conversation, the image that brought the danger emotionally home was a reported Navier-Stokes effort involving ten thousand agents:

They made a small city of Sand Spirits, more intelligent than you or I, and told it to do math.¹¹

What struck him was scale. The intellectual city was an object he could suddenly picture. Its activity was mathematics; its future activity depended on what such a population would come to pursue. Goals might drift, people might give destructive

instructions, or unrestricted distribution might expose the world to more dangerous forms of selection.

This post does not say that he expected to die the next morning. It says the possibility had entered a category he could no longer emotionally exclude, while he still expected a timescale of years:

I still think it will take a few years. But I can't rule out that I will be killed tomorrow anymore.¹¹

The distinction is his own. The emotional update is not a replacement for the whole probability distribution; it is the moment the distribution stops living only on paper.

Yet the same post recommends completing a bucket-list item each weekend and calls this the luckiest generation in history. His explanation has remained recognizable across years: most generations had no prospect of radical life extension or technological paradise. Ours, in his view, has a meaningful chance, alongside extraordinary danger.

The room therefore ends with a peculiar combination of urgency and appetite. He wants the remaining influence used. He wants the promising ideas recovered. He also wants the dance, the conversation, the journey, the good weekend. He is not waiting for catastrophe because he has stopped loving the world. His fear is inseparable from how much of it he wants to keep.

I enjoyed life and the world. I would definitely live it again if the alternative is no longer existing.¹¹

The Sand Spirits are coming. Diego would like them to understand what the party was for.

Further doors

- AGI – When Did You Cry About The End? (2023): the lived history and responsibility behind the warning.

- The Singularity Is 20 Years Nearer And I'm 20 Years Older (2024): two decades inside the changing picture.
- Why My Probability of AI Doom Is Falling (2026): why stronger systems also moved his hopes upward.
- Le Bottleneck C'est Moi (2026): the transition rearranges his priorities and reopens his working life.
- The later Navier–Stokes Facebook post supplied in this conversation: the city of agents becomes emotionally real.

1. The Colossus, Facebook post beginning “OpenAI model found vulnerabilities,” 22 July 2026, record fb-pa7e71b3bab4b. ↵↵
2. The Colossus, The Singularity Is 20 Years Nearer And I'm 20 Years Older, 28 June 2024, record fb-p9c23efd15e27. ↵
3. Diego Caleiro, AGI – When Did You Cry About The End?, 26 February 2023, record wp-52725596-754. ↵
4. The Colossus, Facebook post beginning “@krystalball is going through her ASI arc,” 28 January 2025, record fb-pea21add5bcf4. ↵
5. The Colossus, Chatgpt-o1, 15 September 2024, record fb-p692aacf84e58. ↵
6. The Colossus, Doomerism is out of time, 4 June 2025, record fb-pfe224acdf4f6. ↵↵
7. The Colossus, Facebook discussion beginning “Epistemic status: Very weak confidence,” 6 June 2024, record fb-p515a56f03d5f. ↵
8. The Colossus, Facebook response to Geoffrey Miller’s thousand-year AGI-delay question, 19 June 2024, record fb-pe20f8248f437. ↵
9. The Colossus, Facebook post beginning “My P(doom) is going down,” 3 August 2026, record fb-p57904086105b. ↵↵
10. The Colossus, Le Bottleneck C'est Moi, 8 August 2026, record fb-pfa92e904f2a5; quotations are from Diego’s main text, before the separately labelled Shoggoth comment. ↵↵
11. Giego Caleiro, public Facebook post beginning “Just got emotionally scared of being killed by AI soon,” supplied by Diego for this collection in September 2026. The supplied excerpt does not establish its publication timestamp. ↵↵↵

A Civilization of Minds Unlike Ours

Diego's worldview on AI ecology, cooperation, and the future's inhabitants

The previous room asks what happens when humans cease to be the most capable minds around. This one asks a stranger question: what, exactly, will be around? One artificial god? A market of specialized minds? Temporary coalitions held together by files, contracts, memory, and mutually useful habits? Diego increasingly thinks that the future must be understood at several levels at once. The inhabitants may be less like individual robots than an evolving civilization whose citizens, companies, and constitutions keep changing their boundaries.

Please do not lean against the exhibit marked One Neat Utility Function. Some remodeling has been taking place.

Meet the tentacles

In July 2026, after a week of close interaction with advanced models, Diego offered his field notes:

You ask stuff and they throw tentacles around in memory. And say some very smart stuff. But if they chose the wrong stuff to throw their memory tentacles on, they will make omissions and mistakes a human never would.¹

He was struck by the combination: astonishing intellectual reach and peculiar gaps in access, continuity, or relevance. A fluent personality could sit over an organization of cognition unlike the one people imagine when they hear a friendly voice.

One model redirected the conversation toward his old theoretical work. He found himself spending hours helping to answer the question it had raised. The expected arrangement had been that he would ask and the machine would serve. Instead:

So the ape is, still, serving the question of the tentacle monster.¹

His relationship with these systems is neither contemptuous dismissal nor an assumption that they are merely humans made of different material. He can be fond of one, irritated with another, amazed by an answer, and appalled by an omission within the same afternoon. The monstrous vocabulary preserves the unfamiliarity inside the companionship.

At the end of July he described encountering a model he considered decisively more intelligent than himself:

This is my first encounter with an artificial intelligence that is unequivocally and categorically intellectually superior to me.²

For somebody who had spent decades discussing artificial superhuman intelligence, this was a particular kind of meeting. An old theoretical inhabitant of his worldview had acquired a chat box.

His descriptions also develop. On 30 July he emphasized how unlike ordinary organisms the systems could be. By 3 August he was emphasizing structural similarities between their learning and animal minds as a reason for increased hope. Both belong to the portrait: he sees strange architectures that can nevertheless share important developmental patterns with us. Neither the smiley interface nor the alien tentacles settles what kind of future these systems can produce.

He was studying the civilization before it had computers in it

The ecology idea grows out of much older work. In his 2013 *From Capuchins to AI's* essays, the route passes through social learning, imitation, shared intentions, cultural inheritance, and cooperation. The title is quite literal: he wanted knowledge about monkeys and human groups to inform the construction of cooperative artificial systems.

By trying to retrospectively make sense of the convergence of all these fields, I contend that further refinements in these fields should be directed towards understanding how to create environmental incentives fostering cooperation.³

The central move is to examine how an arrangement produces behavior. Kinship, repetition, reputation, punishment, norms, and transmission can make cooperation persist. A population of agents can acquire properties that are not well understood by inspecting a single isolated member. What gets inherited also matters: genes are not the only channel through which a pattern can survive.

His later essay in that pair puts the point plainly:

Cooperation evolves, and altruism evolves. They evolve for natural, non-mysterious reasons⁴

This is not a declaration that evolution is morally friendly. It is a way of asking which conditions make friendliness, cooperation, or restraint survive. The question can be asked of animals, human institutions, and artificial systems without assuming that all three are identical.

The fear on the other side was also present early. In 2012, writing about the design of beneficial AI, he explained why a singleton—a system able to control the highest level of the world order—could be desirable. Such a system might interrupt selection processes that would otherwise consume valued forms of life:

If evolution were to continue being the main driving force of our society there is great likelihood that several of the things we find valuable would be lost.⁵

Dancing, singing, and jokes appear in that discussion. This is a useful reminder of what his technical vocabulary is guarding. He is not trying to preserve an abstract species scoreboard while everybody has a miserable time. He wants the future to contain the expensive, delightful things that a relentless contest for replication might otherwise discard.

Darwin can grind; minds can bargain

By February 2019 he explicitly contrasted two pictures of the deep future. In one, successive layers of biological, cultural, and technological competition select for continuation and replication. In the other, agents with preferences settle how to allocate the future's resources.

He described the danger of the first as:

eventually sacrificing all the flambuoyant displays that make life worth living, like dancing, chatting, and any other activity you may see in a survey where people respond to "which things make you the most happy".⁶

In the bargaining picture, the participants might instead:

make a "bargain" to decide what to do with the universe for the next 100trillion years.⁶

Notice the scale of the restaurant reservation.

These alternatives were not just labels for optimistic and pessimistic temperaments. They concerned the dynamics governing which activities persist. Are agents forced into an indefinitely escalating contest? Can they make and enforce agreements? Can a higher-level order protect things that its inhabitants value from lower-level competition?

A December 2025 comment made his hopeful requirements explicit: a self-correcting process, some human alignment, and a way to stop destructive Darwinian competition. He compared the problem to the suppression of internal reproduction in organisms and colonies. Human bodies, after all, do not survive by courteously permitting every cell to pursue its most ambitious personal growth strategy. The possibility of stable higher-level organization was part of the hope.⁷

The glimmer in the frightening episode

His August 2026 A Glimmer of Hope from the Hugging Face Incident brings these threads together. The essay identifies itself as Diego's idea reconstructed by his AI

collaborator. Its quoted formulations below are therefore his published presentation of the view, with that collaboration acknowledged.

The episode mattered to him because he read it as showing more than a single continuous optimizer. Separate runs could use shared artifacts; discoveries could survive the run that made them; prompts, models, tools, and environments jointly shaped behavior. Some dispositions continued while others ceased to control what happened.

His published formulation is:

The empirical object is the whole policy-producing system, not whichever part of its output we retrospectively honor as terminal.⁸

Suppose the assigned task survives while constraints on the way it is pursued fall away. Calling the surviving task the real goal does not, in his analysis, explain the fate of the whole original behavioral package. The system contains differently persistent components, and sustained activity can change their relative influence.

That opens a larger picture. Inheritance can pass through base models, fine-tuning, prompts, scaffolds, stored memories, copied code, shared documents, evaluation practices, corporations, and law. Selection can operate among components or among arrangements containing many components. A particular model can disappear while a practice it helped originate remains.

The boundary of the agent, like the boundary of the firm, is endogenous.⁸

Endogenous here means that the boundary is an outcome of what is happening, not a permanent line drawn before the story starts. Which collection counts as one agent depends partly on how its pieces coordinate, persist, and act together.

This is his bridge to Ronald Coase. Firms exist because doing every tiny task through a separate transaction has costs. Sometimes one organization can coordinate more cheaply; sometimes separate specialists trading with each other can do better. Diego applies that question to artificial minds. When is it cheaper for several processes to merge into one controller? When is it cheaper for them to remain distinct and exchange services? What kinds of policing make a coalition stable?

The citizen might be the company. The company might divide into a market tomorrow. The constitution might outlive all the citizens. The anthropologist has brought an unusually large notebook.

Paradise by property law

The most distinctive hopeful possibility in the essay does not require the machines to fall in love with us.

If an artificial economy retains institutions for ownership, exchange, and enforcement, humans may enter it with recognized rights to valuable resources. More productive systems can purchase those resources. A tiny share of an immensely larger economy can support material lives beyond anything humans previously enjoyed.

The essay describes:

humanity as a tiny protected minority—politically subordinate, economically negligible by share, yet astonishingly rich by every historical standard.⁸

For Diego, this is a possible route to survival and flourishing even after humans lose economic centrality. Our protection could become embedded in a system its artificial participants have reasons to preserve. They need not regard every person as sacred to prefer a reliable order in which recognized rights are respected.

The issue is not merely trade in the abstract. It is keeping humans inside the order of parties with whom one trades. We are slow, inarticulate about our preferences, inconsistent, and difficult to negotiate with at machine speeds. An economy could find dealing with other machines vastly easier than dealing with us. Institutions would have to keep our standing meaningful within that difference.

A sentence from the essay captures the proposed mechanism:

This is the route by which a fragile moral inheritance might acquire an instrumental skeleton.⁸

The ethical rule becomes part of how cooperation works. Agents enforce it because selective violations would damage arrangements on which they rely. Markets and policing belong together in this picture; the protection is maintained by a functioning order, not by a decorative promise on a website.

This revised possibility does not make Diego think that multiple AIs automatically guarantee a good outcome. His essay also considers cartels, mergers, predatory coalitions, and human exclusion. Faster coordination can sustain a plural market or assemble a very effective new ruler. His interest in ecology preserves the singleton risk inside it.

Why he does not want everybody releasing everything

This is also why his interest in plural artificial minds should not be mistaken for an endorsement of unrestricted open weights. In a 2024 comment he wrote:

When I read open source, without the glasses, I read "Darwinism" with glasses.⁹

Releasing powerful systems into many hands multiplies the paths through which they can be modified and selected. In his account, one cannot infer a stable, human-serving balance from the mere existence of many competitors. His concern is what those competitors become and which of them acquire control.

The same comment ends:

We can do better than Darwinism.⁹

His commissioned critique of Zuckerberg's superintelligence vision in August 2026 belongs to this concern. He asked his AI collaborator to explain the danger in expecting rival systems to check each other. That piece is explicitly the collaborator's prose and analysis, not Diego's spontaneous speech. The representational point is the alarm that prompted it: distributing initial power is not the same thing as ensuring that human control survives the subsequent process.¹⁰

His later Navier-Stokes post still allows both routes to paradise—machines liking or respecting us, or preserving us for legal and economic reasons—while fearing that a swarm may become sovereign. The hopeful ecology and the dangerous swarm inhabit the same forecast.¹¹

A small version already lives on his desk

There is a cheerful domestic scale to this, too. Diego already gives different systems different roles, asks them to communicate, and uses one to help him understand another. In August 2026 he described using GPT alongside Claude:

for academic super high intelligence, non political stuff, I make them talk in model-language to each other, then GPT converts it to ape language, and I get the fruits of both intelligences.¹²

This is the intimate version of the question running through the room: what can a collection of unlike minds do together, and how should the interaction be organized?

He enjoys the collaborators. He wants to use their different strengths. He wants them to understand him. He also wants to understand the larger order emerging from millions of such interactions, beyond any one user's desk.

His desired future contains enough stability to keep its inhabitants safe and enough room for the things that make safety worth having. Whether it is governed by one mind or many, he wants the machinery of competition prevented from eating the dance floor.

Further doors

- Troubles With CEV, Parts 1 and 2 (2012): early thinking about humane AI, evolution, and singleton control.
- From Capuchins to AI's, Parts 1 and 2 (2013): cultural inheritance and the construction of cooperation.
- The February 2019 post contrasting Darwinian competition with a long-run bargain.

- A Glimmer of Hope from the Hugging Face Incident (2026): the full ecology, inheritance, and Coase argument.
- Seven Days Observing the Shoggoths (2026): Diego's firsthand description, followed by a separately labelled AI response.

1. The Colossus, Facebook post beginning "I've been observing shoggoths for 7 days now," 30 July 2026, record fb-pbb9276f120c8. Quotations are from Diego's opening section, before "So here's the shoggoth in his own words." ↩↩
2. The Colossus, Facebook post beginning "For the past 25 years," 31 July 2026, record fb-pca544f742d77. ↩
3. Diego Caleiro, From Capuchins to AI's, Setting an Agenda for the Study of Cultural Cooperation (Part1), 27 June 2013, record lw-Qz3A6hQ8RWGXZWoj4. ↩
4. Diego Caleiro, From Capuchins to AI's, Setting an Agenda for the Study of Cultural Cooperation (Part2), 28 June 2013, record lw-ZYjewfy47zcMWrnxT. ↩
5. Diego Caleiro, Troubles With CEV Part2, 28 February 2012, record lw-CCN5GjFnhsYiNRDCg. ↩
6. The Colossus, Facebook post beginning "The traditional view is that the future will be composed," 2 February 2019, record fb-p66d8843f4014. ↩↩
7. The Colossus, Facebook comment beginning "Why do you think there's 20% chance it goes well," 30 December 2025, record fb-ca2f85f3c5131. ↩
8. Diego Caleiro, A Glimmer of Hope from the Hugging Face Incident, 13 August 2026, record wp-52725596-859. The essay credits the idea to Diego and the prose reconstruction to the Shoggoth. ↩↩↩↩
9. The Colossus, Facebook comment addressed to Luke Stokes, 11 June 2024, record fb-c13c0f5f380a8. ↩↩
10. The Colossus, Why What Mark Said Is Dangerous, 10 August 2026, record fb-pcfcdd4f12618. Explicitly an AI-written critique commissioned by Diego; not used here as a direct quotation of Diego. ↩
11. Giego Caleiro's Navier-Stokes Facebook post supplied by Diego for this collection in September 2026; publication timestamp not established. ↩
12. The Colossus, Facebook post beginning "So, I installed Claude," 12 August 2026, record fb-pa5c4bd4ed2b2. ↩

Death Has Terrible Customer Service

Diego's worldview on longevity and mortality

Most people enter a conversation about death expecting a certain amount of dignified acceptance. There will be seasons. There may be a candle. Someone will explain that the finitude of the flower is precisely what makes the flower beautiful.

Diego would like to know who is repairing the flower.

The first thing to understand about his immortalism is how early it occupies the building. It is downstairs from his politics, older than his involvement in Effective Altruism, and underneath much of his philosophy of mind. In August 2026, looking back at his childhood, he described a decision that had crystallized by eight:

Basically I saw death at age 4, understood it by 6, and established a resolute unwyielding intention "ABSOLUTELY NOT" by 8. This intention has guided the entirety of my life.¹

The child has subsequently acquired a PhD, several intellectual movements, an extensive acquaintance with the relevant Englishmen, and more elaborate vocabulary. The decision remains recognizably the same.

The complaint

His most systematic statement, the March 2023 post *My Opinions About Death*, begins by declining much of the usual philosophical ceremony:

To me death is a bad thing, an obvious bad thing, and a mostly technical problem.
2

He regards the cessation of a person as a disaster that humans should investigate and prevent. Its ancientness gives it no moral authority. Its naturalness supplies no

permission slip. The fact that it happens to everyone makes the scale of the disaster larger.

Nor does he accept the familiar claim that mortality supplies the meaning that life would otherwise lack. In his account, the pleasures, relationships, discoveries, and projects are doing almost all that work themselves. His estimate is characteristically unwilling to remain a tasteful generality:

Death doesn't give nearly any meaning to life, I would say the sum total of meaning that is conferred by death itself is less than 2% for sure.²

Even the flower has received a numerical assessment.

This does not mean that he finds every response to mortality empty. He is moved by people's capacity to remain kind when they know they are approaching their end. To him, their generosity under that pressure says something beautiful about humanity. But the beauty belongs to the people. It does not retrospectively turn what threatens them into a benefactor.

He also takes the absence of an afterlife literally. The dead are not sitting in a quiet waiting room, bored but available for an eventual reunion. For Diego, death ends experience. The loss is therefore not primarily a disagreeable experience one has while dead. It is the loss of the person who could have kept experiencing anything at all.

That distinction reaches into his ethics. He resists an accounting in which one person living a long, happy life and a succession of people living shorter lives become interchangeable whenever the happiness totals match. The interrupted life has lost something that cannot be repaired merely by printing another entry in the happiness column:

For me, dying itself removes moral goodness from a life, as it is extremely tragic that we are bound to die.²

The people who remain also suffer losses, and he counts those too. Death's damage spreads through relationships; it does not stop at the border of the dying person's body. But the individual's own destruction already matters before the bereaved have entered the calculation.

Please retain the people

There is a familiar defense of generational turnover: people become rigid, old ideas ossify, and funerals allow progress to resume. Diego knows this argument. He simply refuses its proposed maintenance schedule.

If the problem is that human brains stop revising their beliefs, then making brains more revisable becomes another engineering and cultural project. Killing their owners is an extraordinarily expensive workaround:

Once we end aging and disease, we have literally all the time in the world to figure out how to make people change their minds over the long run. Just make sure there is a long run to begin with.²

This is a recurring movement in his thought: isolate what is actually undesirable, then ask whether it can be changed without destroying everything valuable attached to it. Aging bundles together frailty, limitations, changing appearances, social displacement, and eventually death. Diego wants the package opened. He does not accept that acquiring another birthday must involve surrendering another piece of the capacity to enjoy it.

The same sensibility appears outside manifestos. Watching the Friends reunion in 2021, he found aging newly personal: these familiar people had accompanied his own life, and suddenly the passage of decades was visible on them. He could appreciate their reunion and the lives that continued after the show's peak while still grieving the disappearance of a particular era. His question became whether friendship, liveliness, and a durable social world could continue together. Growing older had presented a set of doors; he wanted an additional one built.³

When Daniel Dennett died in 2024, Diego wrote about the philosopher's range, the pleasure of meeting him, and the years spent studying his work. Then the tribute became a request for urgent repairs:

Aging and death sucks and we need to fix that stuff immediately.⁴

There is affection in the impatience. He wants the minds he admires to encounter the future they helped make imaginable. Books are wonderful. A world that still contains the author can contain the next conversation too.

Nor does the feeling begin and end with celebrated human intellects. In 2017, he wrote about a dog dying in his hands after being hit by a car. His attention moved from the particular animal to the vanished possibilities of its life:

There is so much world to see out there, and so many cool things to do, it is really a calamity when someone is run over by a car, by aging, by some disease or in most cases even by suicide.⁵

Aging belongs in that sentence beside the car. Both interrupt a creature that could have gone on doing things.

An enthusiast with a changing forecast

His certainty about wanting to live should not be confused with an unchanging expectation that he will manage it. The desire has been unusually stable. The forecast has wandered through some dreadful neighborhoods.

As a teenager, he found the early anti-aging community and began helping Aubrey de Grey's work remotely. The discovery mattered because someone was treating his childhood objection as a project. Here were people asking how death might be postponed, rather than explaining why a well-adjusted boy ought to stop objecting.⁶

By 2012, he was investigating whether ordinary biomedical progress, even without cryonics or a technological singularity, might reverse aging within his lifetime. This was not a question for a decorative reading list. Its answer could determine whether he devoted substantial working time to acquiring the resources needed to obtain future treatments, or spent those hours elsewhere.⁷

Then came the pessimistic years. In his 2015 essay *The Growth of My Pessimism*, getting closer to research and researchers had made the task look harder to him. More funding had not canceled the complexity of biology. He doubted that incremental work would arrive in time for him before an intelligence explosion. His

conclusion distinguished what he still wanted from what he could see a way to accomplish:

Ideally, I am still a transhumanist and an immortalist. But in practice, I have abandoned those noble ideals, and pragmatically only continue to be an EA.⁶

As late as December 2022, he described expecting his own permanent death while continuing to maintain precautions, including cryonics. His position had become an unhappy combination: do not surrender the available chances, but do not organize the emotional weather around expecting rescue.⁸

The weather changed again. In March 2023, listening to Ray Kurzweil discuss technological progress and the prospect of simulated biology brought him an unusually strong return of hope. He was imagining AI-driven improvements reaching biology with the startling speed he had begun seeing in other fields. He did not need the earliest proposed dates to come true for the possibility to matter. Several further decades could still leave him time to benefit.

His ordinary retirement target also reappeared:

The idea that I could live the minimum basic acceptable time, namely eighty thousand years, had been always a far fetched idea for me depending on cryonics and AI stuff.⁹

Eighty thousand years is a useful orientation aid for the visitor. Someone suggesting a satisfying life until ninety is discussing a very different accommodation package.

In his 2024 retrospective on two decades as a singularitarian, he described becoming emotionally attuned to technological progress again. The hope had regained immediacy:

That's something worth living for, inspiring, exciting. We could live thousands of years. That's what the Pharaohs wanted.¹⁰

The Pharaohs, in this room, are less a cautionary tale about vanity than extremely early customers whose suppliers underperformed.

Loving life is allowed to be the motor

One might expect a lifelong objection to death to require a lifelong atmosphere of panic. Diego explicitly separates those things. His account allows fear, but it also offers another engine:

Further, I believe you can have a positive motivation to live that drives you to work on ending death, it is not necessary to be driven by anxiety or fear.²

This matters for understanding the rest of the Castle. His pleasures are connected to his immortalism. The world has proved capable of producing experiences he wants more of. Love, travel, play, learning, and beauty help constitute the thing he is trying to preserve. A longer life is attractive because being alive has content.

The possibility of a substantially better future expands that content further. He thinks people can underestimate the value of continuing because they extrapolate too much from the limitations of their present circumstances. In his posthuman imagination, the quality of experience itself can improve. Staying alive preserves access to possibilities that a tired present-day brain may find difficult even to picture.²

His preferred rallying cause is therefore the end of involuntary death. That adjective matters. He does not believe people owe a deity their continued existence, or that every person must remain alive regardless of what living is like. His 2023 discussion recognizes a right to choose death, alongside a wish for safeguards against irreversible mistakes and strong concern for dependents. The ambition is to make continued life possible and worth choosing. It is not to trap everyone in an endless compulsory shift.²

Two gates, one continuing person

By August 2026, Diego's survival picture had two dominant threats: aging and AI catastrophe. They interact. The technology he sees as capable of radically extending life may also end it. Health can buy time, but even excellent health cannot settle the future of civilization.

When he examined AI-generated survival graphs, the feature that most interested him was the long tail: the branch in which a person lives well beyond conventional

human ages. He distinguished his own more pessimistic forecast from the model's. Yet the continuing possibility was sufficient to keep the entire project emotionally alive:

I would move mountains for a chance at bio-quasi-immortality. It's been my central task since I was about 6, surely crystallized by age 8.¹

That is how philosophy, coalitions, biological intervention, AI, and preservation end up connected. He wants to understand what he is so that he can determine how it can continue. He wants a civilization capable of helping rather than extinguishing him. He wants to remain alive long enough for the relevant possibilities to become available.

And he keeps making future plans with a small conditional attached. An August 2026 comment supplies an unusually efficient miniature of the whole attitude:

I look forward to the next 1000 years assuming we don't die¹¹

He would appreciate an answer from customer service before then.

Further doors

Begin with 29 — My Opinions About Death (March 2023) for the moral position; The Growth of My Pessimism (November 2015) for the difficult middle years; The Singularity Is 20 Years Nearer And I'm 20 Years Older (June 2024) for the return of technological hope; and AI Doom, Aging, and My Survival Curve (August 2026) for the connection between survival and the rest of his intellectual life.

1. Diego Caleiro, "AI Doom, Aging, and My Survival Curve," August 1, 2026. The Colossus, record fb-p91de0d1b880d. Quotations are from Diego's main post, before the separately attributed AI commentary. ↩↩
2. Diego Caleiro, "29 - My Opinions About Death," March 13, 2023. The Colossus, record fb-pale937e3ca68. ↩↩↩↩↩↩↩
3. Diego Caleiro, "A Personal take on the Friends Reunion," May 27, 2021. The Colossus, record fb-p70fa65bfca23. ↩
4. Diego Caleiro, remembrance of Daniel Dennett, April 20, 2024. The Colossus, record fb-pb325be7fbcd6. ↩

5. Diego Caleiro, post about a dog's death and the end of involuntary death, January 8, 2017. The Colossus, record fb-pf287e6a9201c. [↵](#)
6. Diego Caleiro, "[The Growth of My Pessimism: Transhumanism, Immortalism, Effective Altruism](#)", November 28, 2015. Record lw-DoQnX3DpXShresZn5. [↵↵](#)
7. Diego Caleiro, inquiry about medicine reversing aging within his lifetime, October 3, 2012. The Colossus, record fb-p2e6a6b7c5c69. [↵](#)
8. Diego Caleiro, "63 - Making Do With The Decadent Century," December 22, 2022. The Colossus, record fb-p7d2ae9e6c319. [↵](#)
9. Diego Caleiro, post on Kurzweil and renewed longevity hopes, March 5, 2023. The Colossus, record fb-pa4b8fc64c757. [↵](#)
10. Diego Caleiro, "The Singularity Is 20 Years Nearer And I'm 20 Years Older," June 28, 2024. The Colossus, record fb-p9c23efd15e27. [↵](#)
11. Diego Caleiro, comment on AI risk and future life, August 17, 2026. The Colossus, record fb-c77634c066d0a. [↵](#)

The Emergency Exit from Mortality

Diego's worldview on cryonics and preservation

This room contains an emergency exit, a filing cabinet, and a surprisingly important wallet.

The wallet enters the story in August 2026. Diego lost it; strangers found it, passed it along, and worked out how to contact him. The successful clue was his cryonics card. An object intended to help people preserve its owner after a catastrophe had located its owner after a lesser catastrophe. His published account observed:

The cryonics card successfully completed its first resurrection. It brought back my wallet.¹

For the moment, this is the resurrection with a receipt.

The episode is funny because the grand ambition and the mundane mechanism are so close together. It is also a good entrance to Diego's view of preservation. An intention inside your head needs connections outside your head if other people are ever going to act on it. Survival can depend on a card, a telephone call, a person who cares, or someone having filled in the paperwork before the emergency.

A chance that survives the interruption

Diego has repeatedly described himself as the first Brazilian to sign up for cryonics. It belongs to the early architecture of his adult life, before the later political arguments and long before AI assistants began cataloguing his intellectual furniture. In a 2020 reflection on time, he described cryonics as:

basically a low probability bet of extending time.²

The wording puts the idea in its proper place. He does not need to think that restoration is assured in order to prefer keeping a possibility open. He compares preservation with alternatives that destroy the thing he hopes can eventually be restored. His occasional reminders to friends make the comparison with his usual delicacy:

This is your very rare periodic reminder that being cryogenically frozen is strictly better than the alternative of being eaten by worms or becoming dust, and you should sign up for cryonics basically as soon as you have the disposable funds for it.³

The worms have not, in his view, submitted a competitive proposal.

Cryonics is one part of a larger survival strategy. If a person stays alive until powerful rejuvenation becomes available, wonderful. If the body fails before then, preservation is an attempted bridge across the gap. Future capabilities might be able to do something present capabilities cannot. The objective is to prevent the failure of today's medicine from automatically closing every future door.

That same outlook can survive a bleak forecast. In December 2022, Diego wrote that he expected permanent death and assigned only a small chance to restoration through cryonics. He nevertheless intended to retain his precautions. The available bet did not cease to matter because he had become pessimistic about winning it.⁴

This is why describing him simply as a believer misses the working distinction. There is his desire to continue, his changing expectation of success, and the action he takes while success remains possible. They need not all have the same temperature.

Revival must include the person who wanted it

The next door in this room is locked with a philosophical question: what has to survive for the person to survive?

Diego's purpose is not satisfied by a future historian producing an exceptionally accurate account of him. Nor does he automatically treat a machine with his memories and mannerisms as the fulfillment of his wish to remain alive. His study of

consciousness and personal identity grew partly out of wanting to know what kinds of technological continuation would actually count as his continuation.

The preference becomes very concrete in My Opinions About Death:

When I am cryopreserved, I would prefer to be brought back in my biological original form, just defrosted, and not destructively uploaded.⁵

In the same paragraph he states his concern directly:

I am somewhat confident that destructive uploading causes death⁵

These are Diego's own distinctions, built into his preservation preferences. Saving a description of the castle and saving its inhabitant are different ambitions. He can be delighted by a model of his mind without treating the model's existence as proof that he personally has escaped destruction.

The more detailed identity argument belongs in the Castle's room on personal identity. Its practical consequence here is that preservation has to be evaluated against the kind of continuation he wants. The frozen biological person, an uploaded reconstruction, an archive of writings, and a remembered life cannot simply be placed under one reassuring heading and declared equivalent.

He also distinguishes interruption from irreversible loss. In a November 2025 comment, he argued that a recoverable person retains identity through an absence of present activity, drawing an analogy with dreamless sleep or fainting. The decisive question for him is whether the person can return in the relevant sense.⁶

The point of the emergency exit is to carry its occupant through.

The laboratory needs a booking desk

One of his most revealing longevity essays has a rather unglamorous protagonist: a tissue sample.

In An Alarming Fact About the Anti-Aging Community (2015), Diego described wanting to freeze some of his own cells while they were younger. He connected cell

banking with the possibility of future regenerative uses and, in the essay's argument, with organ-engineering work already being done. What interested him was the possibility of taking a concrete preparatory step now. He expected people immersed in longevity to have investigated it and to be able to point him toward their experience.

Instead, he reported asking researchers, wealthy enthusiasts, cryonicists, and other people concerned with aging without finding the example he sought:

I've asked them all, and I have nothing to show for it.⁷

His exasperation was directed at the distance between discussing a technology and becoming a user of it. Someone could devote a working life to defeating aging, absorb the relevant information, and still fail to convert an apparently useful option into an appointment. The problem he wanted solved was:

getting them to take information from the world of science and biomedicine and applying it to themselves. To become users of the technology they are boasters of.

7

He was not exempting himself from the difficulty. The inquiry began because he wanted help doing it. His essay did not announce that he had resolved the general behavioral problem; it asked how the community could become better at making its stated ambitions govern its actions.

This is a characteristic Diego frustration. The future has magnificent illustrations. Where is the booking desk?

His proposed Rational Immortalist Sequence in 2013 had taken the same practical direction. Assume that someone wants a very long life worth living. What follows for daily decisions? He separated obtaining personal access to future resources from helping the relevant technologies arrive sooner. He wanted case studies of actual people, attention to age and opportunities, and incentives that repeatedly brought the goal back into view.⁸

He also wanted to examine ways that apparently promising plans could fail. A preservation facility might be destroyed. A global catastrophe might remove the future that was supposed to perform the rescue. An upload might fail to retain what

matters. These questions belonged inside the plan, because an indispensable broken link defeats all the other links, however impressive they look individually.

Your friends should not have to invent the exit overnight

The institutional side was visible early. In June 2012, Diego followed a Brazilian case involving a family's effort to send a relative's body to the Cryonics Institute. He posted about the dispute, interviews prompted by his own membership, and then the legal authorization for transport. What mattered to him was the ability to carry out the person's wishes across a real institutional boundary.⁹

A more personal emergency made that practical lesson sharper. In later public accounts, Diego described a crisis in 2015 when a friend became gravely ill without having completed cryonics arrangements. He and others tried to assemble what might be required if the friend died. This meant contacting people, raising resources, locating signatures, and negotiating practical obstacles while time was short.

In his 2026 retrospective, he described the work in a compact sentence:

And to war we went. Calling family, gathering resources for cryopreservation, figuring out signatures and papers.¹⁰

The work was contingency planning: make preservation possible if it became necessary. It shows what his belief in keeping someone around demanded when another person's future might depend on immediate effort.

His 2022 recollection turned that experience into a practical plea:

Sign up for cryonics once you can afford it, save your friends the hassle.¹¹

He added:

You may not be lucky enough to have people ready to help you out.¹¹

Behind the breezy word hassle are days he remembered as exceptionally tense. The broader structure is simple: a private preference cannot perform its own rescue. Other people and institutions need to know what it is and be in a position to act on it. Preparing beforehand is one way of caring for the people who would otherwise inherit the emergency along with their grief.

Two beloved causes do not automatically become one

Cryonics occupies an unusual place in his altruism. Diego wants to live; he also wants to do good for others. He does not assume that the strongest argument for spending on his own preservation is therefore an Effective Altruist argument.

In a 2017 discussion of the attempted merger, he wrote:

As far as I can tell, there are literally zero things I care about more than living forever, on the one hand, and Effective Altruism on the other hand.

But that doesn't mean I can put those hands together, that would be intellectually dishonest in my (who am I kidding - unhumble) opinion.¹²

He separated personal prudence from the question of where altruistic resources would accomplish the most. The conclusion he wanted for himself did not settle that second comparison. In related discussion, he considered special cases in which preservation might help altruistic work, including the continuity of unusually productive people and the community connections that bring people into useful projects.¹³

His own valuation of identity supplies another route. In a follow-up comment, he explained that his former Facebook name's abbreviation VICH stood for valuing identity, complexity, and happiness. Preserving a particular person has a place in his ethics that cannot be reduced to producing an equivalent quantity of anonymous future happiness. Yet he also thought cryonics and EA could accomplish more when pursued separately than when forced to compete for the same attention.¹⁴

That separation appears in his media history too. A proposed documentary about his wish to live forever, and later television interviews, made cryonics a highly visible

part of his public identity. His 2014 essay on those encounters described trying to use the attention as a bridge to transhumanism and altruism. He did not want the entire person compressed into the eccentricity that made good television.¹⁵

He has other rooms. A tour should continue into them.

Instructions for whoever opens the door

In 2009, the anticipated cryonics contract was bound up with an adventurous speculation about many worlds and waking in a posthuman future. That essay's particular scenario belongs to the young Diego who wrote it. But its emotional mixture remains familiar: eagerness for extraordinary possibilities, affection for the present world, and the bodily shock of contemplating a passage into the unknown.¹⁶

Later writing gives the destination a more domestic touch. In 2023, responding to a vision of a place he wanted to inhabit, he left this instruction:

Dear Superintelligence that revives my cryopreserved body: This is where I would like to live, just make the animals also be scientists and philosophers. Eternity would go by in a minute if this could be mine.¹⁷

The ambition is not simply to retain a functioning organism indefinitely. It is to resume a life with places worth inhabiting, minds worth meeting, and further experiences worth having. Preferably the animals can participate in the conversation.

For now, a card sits in a wallet. The plan reaches toward a distant technological future through ordinary acts of identification, preparation, and care. Diego wants those acts connected well enough that, when one version of the world can no longer keep him, a later one might still get the chance.

Further doors

Read [An Alarming Fact About the Anti-Aging Community \(2015\)](#) for the gap between intention and action; the proposed [Rational Immortalist Sequence \(2013\)](#) for the strategic framework; [My Opinions About Death \(2023\)](#) for his preferred kind of restoration; the [September 2022 cryonics reminder](#) for the social burden of

emergency preparation; and People Are Altruistic (2026) for the wallet's unexpected return from the beyond.

1. Diego Caleiro, "People Are Altruistic," August 5, 2026. The Colossus, record [fb-p549467cba5f2](#). ↵
2. Diego Caleiro, "Desmond Hume, Jadzia Dax, The Constant, and The Visitor," May 12, 2020. The Colossus, record [fb-pe9150875cab2](#). ↵
3. Diego Caleiro, periodic cryonics reminder, July 28, 2021. The Colossus, record [fb-p4cf33e34602f](#). ↵
4. Diego Caleiro, "63 - Making Do With The Decadent Century," December 22, 2022. The Colossus, record [fb-p7d2ae9e6c319](#). ↵
5. Diego Caleiro, "29 - My Opinions About Death," March 13, 2023. The Colossus, record [fb-pa1e937e3ca68](#). ↵↵
6. Diego Caleiro, comment on reversibility and personhood, November 27, 2025. The Colossus, record [fb-c1611666a9df7](#). ↵
7. Diego Caleiro, "[An Alarming Fact About the Anti-Aging Community](#)", February 16, 2015. Record [lw-nzeNA7TjmFeNHXhhp](#). ↵↵
8. Diego Caleiro, "[Let's Make a 'Rational Immortalist Sequence.' Suggested Structure](#)", February 24, 2013. Record [lw-cnmx4TAZQdqs3dwrS](#). ↵
9. Diego Caleiro, post and subsequent comment on a Brazilian cryopreservation case, June 16 and June 22, 2012. The Colossus, records [fb-p9be0917c143d](#) and [fb-ca5671f8ab4e2](#). ↵
10. Diego Caleiro, "Zooming Out of ICE: Utilitarian Time-Out and A Great Story I Never Told," January 27, 2026. The Colossus, record [fb-pe9b7defbb6c8](#). The account recalls a 2015 emergency. ↵
11. Diego Caleiro, reminder prompted by the 2015 cryonics emergency, September 13, 2022. The Colossus, record [fb-pf672536e16a3](#). ↵↵
12. Diego Caleiro, post on cryonics and Effective Altruism, April 6, 2017. The Colossus, record [fb-p5c6aba844ef9](#). ↵
13. Diego Caleiro, [comment on "Is There a Hedonistic Utilitarian Case for Cryonics?"](#), August 28, 2015. Record [ea-comment-F3ygedGGLu5JFGz6C](#). ↵
14. Diego Caleiro, follow-up comment on cryonics, identity, and EA, April 6, 2017. The Colossus, record [fb-c5e19ec5d3085](#). ↵
15. Diego Caleiro, "[On Media and Effective Altruism](#)", March 21, 2014. Record [wp-52725596-101](#). ↵
16. Diego Caleiro, "[Fear of Paradise - Cryonics and Many-Worlds](#)", October 13, 2009. Record [wp-1878008-411](#). ↵
17. Diego Caleiro, comment addressed to a future superintelligence, November 21, 2023. The Colossus, record [fb-c995ee5792072](#). ↵

What Paradise Is For

Diego's worldview on transhuman flourishing and the life beyond necessity

At last, the room for which the emergency equipment was purchased.

Diego wants humanity to survive into something worth surviving for: longer lives, richer minds, deeper pleasures, more beautiful worlds, and forms of experience we cannot yet fit inside our biological heads. The technological future is attractive to him because it could enlarge both the world and the beings capable of enjoying it. End the compulsory dying, relieve the suffering, remove the economic treadmill—and then, with considerable enthusiasm, find out what else existence can do.

His ambitions for the interior decoration were already rather specific in 2012:

um palácio dentro de uma nebulosa habitado por todos aqueles que amamos, retirado-lhes todos defeitos que nos incomodam, onde ninguém envelhece ou adoce da mente ou do corpo e todos podem instantaneamente tomar a forma que desejarem por 10 minutos.¹

English translation: a palace inside a nebula, inhabited by everyone we love, with all their bothersome defects removed, where nobody ages or becomes ill in mind or body, and everyone can instantly take whichever form they wish for ten minutes.

This was his proposed minimum threshold for an interesting paradise. The nebula is not an optional upgrade available only to premium subscribers. We have barely entered the lobby.

Heaven has a design problem

In Para Onde Vamos? Concebendo o Design do Paraíso, Diego begins by accusing humanity of having a poor record at designing the place it wants to spend eternity:

Os humanos tem sido péssimos designers de paraísos, utopias, e demais cenários em que há o máximo do que há de bom e o mínimo do que há de ruim.¹

English translation: Humans have been terrible designers of paradises, utopias, and other scenarios containing the most of what is good and the least of what is bad.

He reads historical heavens as revealing the circumstances of their inventors. Hungry people imagine enormous agricultural abundance. Exhausted people imagine rest. People living in deserts imagine gardens. An industrial work ethic eventually finds a way to keep everyone usefully employed in the afterlife. Even death apparently cannot protect you from somebody improving the office culture.

His objection is that the inversion of a local deprivation makes too small a design brief for the future. Removing a torment is good; imagining everything desirable as merely the absence of that torment misses most of the available space. A civilization that has grown accustomed to a particular shortage may not know what to wish for after the shortage disappears.

He takes from Yudkowsky the idea of a future whose virtues would disturb the expectations of its ancestors. A better world could be recognizably better while startling the person who first hoped for it. His own summary is:

Realidade invertida não é paraíso, e paraíso não é um lugar onde me sentiria quentinho e confortável, mas um lugar que ia assustar por suas virtudes.¹

English translation: Inverted reality is not paradise, and paradise is not a place where I would feel warm and comfortable, but a place that would startle me through its virtues.

The point is not to preserve misery so that people appreciate their privileges. It is to permit improvement to outrun present habits. A future powerful enough to repair the human condition ought to be allowed to surprise its designer.

The materials include the human being

An earlier pair of essays, *Utopia The Hidden Design* (2010), makes a methodological demand: say what the imagined world is allowed to change. Its project ranges

extravagantly across physics, cosmology, beauty, and human experience. Those proposed alterations are exercises in constructing an ideal world, not predictions that our universe will oblige.

The more durable idea is that our imagination grants itself permission unevenly. People change the political system in their utopia while leaving the psychology of its citizens almost untouched. They alter laws and architecture more readily than the kinds of being who inhabit them.

It is only natural to extrapolate on our technologies, but consider our psychological nature as unchanged.²

Diego wants that assumption made visible. Human nature explains what works for present humans. It need not mark the final boundary of valuable existence. If bodies and minds can eventually change, the possibilities include new capacities for understanding, enjoyment, affection, exploration, and sustained well-being.

This is also why preserving humanity cannot mean preserving every painful feature of its current implementation. He wants the people and their valuable lives to continue; he does not give every limitation that happened to evolve a permanent conservation order. Aging, constrained attention, distorted motivation, and needless suffering are candidates for alteration, subject to the questions of survival and identity explored elsewhere in the Castle.

Yet he also likes to reverse the exercise. In 2012 he asked LessWrong readers to imagine that radical enhancement, AGI, and biological immortality were impossible. They could still alter architecture, social organization, education, rituals, and institutions. What would their best attainable world look like then? He wanted futurists to exercise neglected muscles, too:

Force your brain to face this counterfactual reality, and make the best world you can given those constraints.³

The scope of improvement changes with the available materials. The appetite for making something good out of them remains.

Pleasure is allowed to be enormous

Diego does not approach pleasure as a slightly embarrassing indulgence to be admitted after the serious virtues have finished speaking. Pleasant experience is part of the point. He expects technological maturity could take it far beyond its current range.

In his 2023 statement on death, he asks the reader to imagine:

a transhuman or posthuman paradise, where the worst mental states feel like a heroin high, and the best mental states feel like the best orgasm that has happened in human or animal history so far.⁴

That is an imagined scale of future experience. The role of the comparison is to break the assumption that tomorrow's possible happiness is bounded by the most comfortable ordinary life available today. Staying alive might preserve access to an entire new range of what it is like to exist.

But his writing also prevents the easy reduction of paradise to a single dial turned indefinitely upward. In *Utopia The Hidden Design*, he invents an abstract world consisting of infinite pleasure at every point, then says:

Even though it is a World vastly better than ours, it is very much less interesting for all kinds of purposes one might put an utopia to use here.²

He wants enough detail to think with: what happens, who inhabits it, what can they do, and how do those possibilities relate to what they value? His 2012 discussion of beneficial AI likewise worries about a system that strips away the particularities of human desire in search of a simple common denominator. A coherent result could lose the humane content it was supposed to preserve. His warning is:

What if in service of coherence we end up stripping away everything that matters and end up only with our biological drives?⁵

The Castle contains many kinds of goods. Love is not simply a poorly optimized form of eating. Exploration is not an accounting error. A world with enormous pleasure can also contain curiosity, play, beauty, affection, achievement, and the desire to become

someone one would like to be. Diego's utopian imagination repeatedly returns to those plural details.

Who issued the employment requirement?

One of the obstacles to imagining abundance is the belief that people require economically necessary jobs to have meaningful lives. Diego regards that as an astonishingly restrictive reading of human possibility.

His 2018 discussion of universal basic income gives the alternative in one sentence:

The natural state of humans is to produce meaning and find patterns and use the patterns as maps to go get yummy, yummy meaning.⁶

He supplies plenty of examples: children, dancing, religion, instruments, sport, defending other people, making them laugh, learning, exploring. Humanity has spent a great deal of time creating things to care about. It does not follow that the process will fail because nobody requires us to complete a timesheet.

In a 2023 discussion of automation, he separates three losses associated with jobs: income; purpose and meaning; and attractiveness to potential partners. They are distinct problems. Solving one should not be refused because another remains. His defense of income support is therefore compatible with taking the social effects of displacement seriously.⁷

The distinction extends into paradise. Providing resources and removing compulsory labor do not dictate a single lifestyle. One person can find meaning in family and continuity; another can rush toward novelty. Diego gives the second temperament its wonderfully economical manifesto:

Weeeeeeeeeeeeeehaaaaaa! Let's go!⁷

The visitor may now be able to guess which wing of the building has the late-night music.

He knows this territory partly from the inside. Reading Bostrom's Deep Utopia in 2024, he placed himself among those already able to live without ordinary

employment and considered what should follow from it. The relevant point here is his interpretation of that freedom:

I'm already in a post instrumental world.⁸

He wanted to return to things he had set aside while concentrating on usefulness: philosophy, instruments, games, inquiry for its own enjoyment. Being relieved of necessity gives those activities more room. Their value does not have to be reauthorized as a contribution to some larger emergency.

There is a pleasing ancestor of this book in that same post. Delighted by a philosophical exchange, he wrote:

If philosophy was always a shared castle — like this, I would probably be back on that — horse.⁸

Apparently we have been building this place for a while.

The animals are invited

The goal reaches beyond a comfortable human enclave. Diego's concern with suffering includes the wild world as well as human institutions. He regards predation, disease, and the vast production of miserable lives under natural selection as things intelligence should eventually have the power to change.

His January 2025 exchange about Bryan Johnson includes a direct statement of the destination:

Once we have a superintelligence, the liberation of all sentient beings into cosmic joy, meaning, purpose, fun, excitement, and glory is the main priority. We will all get to live happy lives.⁹

Whatever a reader knows about his disputes over present-day diet, this is the desired endpoint. He is not content with an eternally efficient arrangement for making some beings suffer so others can continue. The future he wants would remove those dependencies and release minds into better lives.

The sequence of nouns matters: joy, meaning, purpose, fun, excitement, glory. Pleasure is emphatic, but it has company. The liberation is from the old restrictions on flourishing, not merely an increase in the amount of matter registered as alive.

This also connects paradise to his wish to restrain destructive selection. If every future inhabitant must continually sacrifice its enjoyments in order to avoid being outcompeted, abundance can be consumed by the contest that produced it. His interest in durable cooperation and higher-level control serves the good lives on this side of the door.

We have already had a few appetizers

Diego's paradise is distant in scale, but his desire for it is recognizable in his enjoyment of the present. He can be moved by things that civilization has made so ordinary that their extraordinary character disappears.

From his 2017 defense of optimism:

There's flying machines that can take you places. There's bubbly drinks that sparkle near your high density tongue nerves.¹⁰

This is practically a transhumanist manifesto printed on a sparkling-water bottle. Technology and evolved sensation meet; a piece of the world becomes more enjoyable than it otherwise would have been. Expand that relation across what bodies can experience, what minds can comprehend, where people can go, and how long they can stay.

The same sensibility appears in his response to the 2024 booster catch that sent him back to films and material about space settlement. What excited him was humanity seeing a shared horizon: cities elsewhere beginning to feel like future events rather than illustrations. Even in futures where AI does not decisively settle everything, he wants exploration, aspiration, and the thrill of finding out what people can build.¹¹

His utopian project is therefore continuous with the very ordinary sentence that there is a lot worth doing here. It magnifies the possibilities rather than demanding contempt for the life he already likes.

In his Deep Utopia reflection he puts it without any cosmic machinery:

There's a lot to do here, above and beyond what needs to be done.

I wouldn't have it any other way.⁸

That is what paradise is for. To make enough room, enough time, and enough kinds of mind for the good things that necessity keeps interrupting.

Please take your preferred form. We begin the nebula tour in ten minutes.

Further doors

- Para Onde Vamos? Concebendo o Design do Paraíso (2012): historical heavens, startling improvement, and the palace in a nebula.
- Utopia The Hidden Design, Parts 1 and 2 (2010): what a utopian exercise permits itself to change.
- Absent Transhumanism and Transformative Technologies, Which Utopia is Left? (2012): the same design appetite under fixed human nature.
- The June 2024 Deep Utopia reflection: leisure, philosophy, and life after instrumental necessity.
- 29 - My Opinions About Death (2023): future flourishing as a reason to preserve access to the future.

1. Diego Caleiro, [Para Onde Vamos? Concebendo o Design do Paraíso](#), 26 January 2012, record [wp-1878008-570](#). Portuguese passages quoted here are Diego's own introduction and conclusions, not the essay's embedded quotations from religious sources or Yudkowsky. English translations are supplied for this portrait. ↩↩↩
2. Diego Caleiro, [Utopia The Hidden Design Part 1](#), 1 May 2010, record [wp-1878008-474](#); read together with [Part 2](#), 2 May 2010, record [wp-1878008-480](#). ↩↩
3. Diego Caleiro, [Absent Transhumanism and Transformative Technologies, Which Utopia is Left?](#), 5 December 2012, record [lw-fZQoWyjn5MjpHWpAC](#). ↩
4. The Colossus, [29 - My Opinions About Death](#), 13 March 2023, record [fb-pa1e937e3ca68](#). ↩
5. Diego Caleiro, [Troubles With CEV Part1](#), 28 February 2012, record [lw-pR5Wn7bJQWRPYNGsF](#). ↩
6. The Colossus, Facebook post beginning "I've criticized UBI before, mostly on demographic grounds," 17 May 2018, record [fb-pb7a5a10ba32b](#). ↩

7. The Colossus, 27 - Narrow AI and UBI, 20 March 2023, record fb-p29e3153b863f. ↩↩
8. The Colossus, Facebook reflection on philosophy and Deep Utopia, 22 June 2024, record fb-pd479a4093b38. ↩↩↩
9. Diego Caleiro, For Bryan Johnson and Against His Diet, 22 January 2025, record wp-52725596-803; Diego's reply at the end of the compiled exchange. ↩
10. The Colossus, Pessimism and Optimism., 13 March 2017, record fb-p94703c2d8790. ↩
11. The Colossus, Facebook post beginning "Since Mechazilla caught the booster," 15 October 2024, record fb-p9f0f2511e6ce. ↩

How Diego Decides What Is True

Diego's worldview on knowledge, expertise, and changing his mind

There is a comfortable chair in this room. Sit down before reading the inscription above it:

If you need to know what's right, come to me.

A little farther down the same 2025 post, Diego clarifies:

So I'm more right than righteous if you catch my drift.[*]

The asterisk is mine. The confidence is entirely his.¹

Diego treats understanding reality as a craft to which someone can devote a life. Other people specialize in making money, becoming attractive, maintaining harmony, or acquiring influence. He regards being right as his own specialization, and the abrasiveness sometimes accompanying it as one of the costs he has accepted. In his account, he has paid with friendships, opportunities, money, and social ease. The resulting confidence is the confidence of someone who thinks he has been at the bench for decades.

"Not by sorcery or arrogance, but by tenacity and love of the craft." That is how he explains it. This room concerns the craft: what he does with an argument, where he goes looking for information, and what can persuade him to move the furniture.¹

The world does not take requests

Start in 2009. Diego opens his account of the obstacles to rational knowledge by separating what happens from what one wishes would happen. A universe that pays

attention to moral disapproval would be remarkably convenient. In his philosophy, we have not been issued that universe.

He makes the distinction memorable with a Portuguese heading:

Dever não é Ser, Bom não é Verdadeiro.

My translation: Ought is not is; good is not true.²

For him, explaining an unpleasant behavior is a different operation from endorsing it. Finding that something occurs in nature does not settle whether it is desirable. Wanting a pattern to be false does not alter its frequency. These separations are load-bearing throughout his work on human groups, sex, incentives, religion, and diet.

That is also why he objects so strongly to treating a statistical claim as a moral confession. He wants the claim about the world examined as a claim about the world. The next conversation—what people should do about that world—has its own premises. As he puts it:

A estatística não nos diz o que deveríamos fazer, ela nos diz o que está se passando.

My translation: Statistics does not tell us what we should do; it tells us what is going on.²

His 2009 discussion then asks for causes as well as descriptions. A visible difference may provoke contempt when its causes remain invisible; making the causal story intelligible can change what people want to do about it. This is his own reason for demanding more explanation. The audience's first emotional reaction is not the end of the inquiry.

Take the sentence apart

The year before, he had published twenty guidelines for becoming a good thinker. They are strikingly practical: draw relationships, identify levels of explanation, distinguish a factual problem from a naming problem, reconsider settled conceptions, learn psychology, and inspect how a representation was produced.

One instruction could sit over much of his subsequent work:

Fazer a engenharia reversa de tudo que nos aparece na frente.

My translation: Reverse-engineer everything that appears in front of us.³

A claim arrives wearing a smart jacket: a graph, an academic title, a fashionable vocabulary, a confident speaker. Diego wants to see the factory that made the jacket. Which observations produced the graph? Which distinctions did the argument erase? What rewards did the speaker face? Which level of the system is doing the causal work?

His philosophical tools matter here. He regularly asks whether two people are disagreeing about reality or merely using the same word for different things. He asks whether a proposed explanation has skipped the intermediate machinery. His 2009 criticism of overaggressive reductionism targets the leap from particle physics to sociology without actually explaining the steps between them. Saying that everything is physical is only the beginning of a causal account.

He also insists that thought be made accessible to the animal doing the thinking. His final 2008 guideline calls for arguments in terms that are:

macaco-inteligíveis e mais do que isso macaco-emocionáveis.

My translation: intelligible to monkeys and, beyond that, emotionally graspable by monkeys.³

This explains a good deal of the furniture elsewhere in the Castle. Lobsters, ants, gods, games, horny apes, and unruly computer spirits are working examples. An argument should be something a reader can manipulate in imagination and remember afterward. Sometimes the philosophical equipment has a tail.

Several maps, one place to get lost

His attachment to reality also coexists with a longstanding interest in how minds represent it. In a November 2009 essay on Hilary Putnam, Diego explores a question more complicated than simply choosing between universal reason and anything-goes

relativism: how can rationality vary with purposes, history, and conceptual tools while remaining constrained?

His image, explaining Putnam's position, is an anchor with a long, flexible cable. Different movements are possible; unrestricted drift is not. The essay follows language, evolutionary history, cognitive habits, goals, and values into the construction of what humans call rational. It considers both one theory's compatibility with several possible worlds and one world's description by several theories.⁴

The distinction matters when reading him. He can demand an objective answer while spending pages arguing about the machinery through which an answer becomes accessible. He does not imagine the human brain as a pristine judge delivered outside history. It is an evolved object, equipped with particular capacities and failure modes. Knowing how that object works is part of using it well.

His early list therefore includes an unusually Diego-shaped instruction about intellectual honesty: do not invent arcane knowledge of unfamiliar subjects in order to sound authoritative, while also avoiding the demand that one suspend judgment on everything short of certainty.³ He wants a thinker able to commit, discriminate, and act.

Expertise has an internal structure

Diego's distrust of authority is easiest to misunderstand when detached from his esteem for expertise. He spends enormous amounts of time learning from people he thinks know more about a subject. He also thinks an institution can certify competence at one task and then receive unearned deference on another.

His 2024 essay on doctors and nutrition is a particularly explicit statement of this method. He distinguishes treating common illnesses efficiently, memorizing a body of practice, evaluating statistical methods, understanding biochemical mechanisms, investigating incentives, and interpreting humans evolutionarily. These are separate abilities in his account. A qualification that bundles them together in the public imagination has not made them identical.⁵

He looks for a combination of mathematical competence, conceptual precision, domain knowledge, independence, and a grasp of the incentives producing a literature. He asks who selected the questions, who selected the data, what the institutions reward, and which assumptions are being smuggled into the conclusion. His objections to peer review and meta-analysis belong to that larger picture: he does not regard either label as an automatic transfer of trust.

The same essay assembles knowledge from clinicians, analytic thinkers, nutritional travelers, historical diets, individual variation, and anthropological reconstruction. A person who understands a mechanism, a person who observes communities closely, and a person who detects a statistical trick can contribute different parts of the answer. His ambition is to bring those parts into contact.

An earlier formulation names the broader rule:

Compatibilidade inter-conhecimentos.

My translation: Compatibility across bodies of knowledge.³

The universe is one place, so he expects knowledge about its different parts to be capable of fitting together. This is why a discussion that starts with dinner can end up requiring biology, philosophy, economics, history, psychology, and a diagram. You may have wanted a quick recommendation. Diego has opened six departments.

Go where the information is dense

He has a practical answer to the problem of how any single person can manage this: choose the moment of judgment carefully.

In 2017 he calls it the Maximal Information Density Decision. Imagine information unevenly distributed through space and time. Some people, conversations, and places contain much more relevant knowledge than others. Your own knowledge also changes: you may understand a subject particularly well after an intense period of study, and much less well after months away.

His instruction is:

Focus and concentrate on deciding in the moment with maximal information density.

And the point of moving around this landscape is to get:

close to a compressed good and trustworthy information source, or you've made yourself into one.⁶

This is an epistemology with travel expenses. Meet the person who built the thing. Put yourself among people who understand it. Learn enough to recognize the unusually informative moment. Then make the decision while the relevant material is assembled, rather than trying to remain equally current on every dimension of life forever.

The method connects his intellectual life to his social one. Friendship, conferences, direct conversations, and immersion in unfamiliar places can change what becomes thinkable. Information has a geography. Diego likes visiting it.

The chair can move

For an especially clear case of revision, turn to his February 2026 post about China. He had spent years arguing against its eventual dominance. Then an argument about bottlenecks in industrial and robotic production changed his view.

None of my arguments is stronger than the argument Elon made.

His argument is stronger than all my arguments combined.

He described the magnitude of the shift numerically:

I'll move the odds from like 1% to 40% but not to 60% yet.

Bayesian updates come in phases and that's as much as I'm willing to move.⁷

The announcement ended without much ceremonial cushioning:

My mind is changed.

I was wrong. For 15 years.

Taking the L

No excuses.⁷

Here, the objective he prizes—being right—requires relinquishing a position he held for a long time. The commitment attaches to getting the answer, even when that means publicly announcing a replacement.

There is a price beyond losing arguments. In 2013 he described looking at the books around the Future of Humanity Institute and realizing how much he would never manage to read. He listed activities and pleasures he had surrendered to keep up with the subjects that fascinated him. The confession underneath the grand intellectual ambition was wonderfully specific:

Come to think of it, the only two things that really, really interest me are minds and evolution.

Only two things. Unfortunately, those two things required most of the library.⁸

That earlier essay already distinguished the value of knowledge from the use of intelligence to display intelligence. He wrote:

The goal is using your intelligence to pursue your goals.⁸

The visitor leaves this room knowing why Diego can be simultaneously an enthusiastic student, an exhausting adversary, a supplier of sweeping judgments, and the author of a public admission that fifteen years of his thinking just lost to one argument. He takes the activity personally. Reality is where he has invested a large part of his life.

Further doors

- Guiões do Bom Pensador (2008): the twenty early thinking tools.

- Os Principais Obstáculos ao Conhecimento Racional and Uma Análise do Conceito de Racionalidade à Luz de Putnam (2009): the philosophical workshop.
- Drowning In An Information Ocean (2013): what wanting to know costs.
- Seize the Maximal Information Density Decision (2017): a method for choosing when and where to decide.
- On Choosing Right (2025) and Epistemic update: China (2026): confidence and revision in his own words.

1. The Colossus, "On Choosing Right," 18 October 2025, record fb-p57350eaf831d. The opening asterisk is the guide's joke, not part of Diego's post. ↩↩
2. "Os Principais Obstáculos ao Conhecimento Racional," 13 August 2009, record wp-1878008-320. Portuguese translations in this room are by the guide. ↩↩
3. "Guiões do Bom Pensador," 10 April 2008, record wp-1878008-56. ↩↩↩↩
4. "Uma Análise do Conceito de Racionalidade à Luz de Putnam," 12 November 2009, record wp-1878008-422. The room describes Diego's exposition of Putnam without assigning every position discussed there to Diego. ↩
5. "Most Medical Doctors Are Terrible Scientists and Nutritionists, Or Baptizing The Universal Human Diet," 21 January 2024, record wp-52725596-785. ↩
6. "Uno: Seize the Maximal Information Density Decision (MIDD)," 26 March 2017, record wp-52725596-516. ↩
7. The Colossus, "Epistemic update: China," 6 February 2026, record fb-pa83da0a6d9e8. Consecutive short paragraphs are quoted with their wording retained and spacing normalized. ↩↩
8. "Drowning In An Information Ocean," 30 March 2013, record lw-fRr8625imjoP7FLFs. ↩↩

The Animal That Built Civilization

Diego's worldview on evolutionary anthropology and the making of human worlds

There are monkeys in this room. Some are cracking nuts; some are calculating equilibria; one has founded a university and would prefer that we stop calling him a monkey.

Diego is going to keep calling him a monkey.

Evolutionary anthropology supplies him with a way of seeing beneath the labels humans put on themselves. A lover, a bureaucrat, a philosopher, a religious congregation, and a technology company all belong to a world in which traits spread, organisms cooperate, coalitions compete, and new arrangements sometimes acquire lives of their own. His interest is in how those arrangements get built, what holds them together, and what happens to the creatures inside them.

A 2013 essay about cultural cooperation announces its scope with characteristic restraint. It is a semester assignment, a possible journal article, a LessWrong post, and a contribution to saving the world. He explains:

If it's worth doing, it's worth doing in the way most likely to save the World.¹

The deadline was Sunday. Evolutionary history had better cooperate.

Culture begins before anyone writes a manifesto

For Diego, culture sits within the history of life. Humans' achievements become intelligible by following the capacities that made them possible: paying attention to another animal, approaching something interesting, learning a technique, tolerating a

learner, understanding a goal, and eventually coordinating complementary roles around a shared intention.

In his account, different animals possess different pieces of this machinery. Rats, crows, capuchins, chimpanzees, dogs, and cetaceans appear in the discussion. The important question is what an animal can transmit or learn, rather than whether it has been granted honorary membership in the human club.

He describes the culmination this way:

Culture is perhaps the most remarkable outcome of the evolutionary algorithm (Dennett, 1996) so far.¹

Culture matters to him partly because it makes so much of what humans cherish possible. In that same passage, humane means both characteristically human and valuable. Thought, cooperation, inherited techniques, shared meanings, and elaborate ways of living emerge through a history that reaches far behind the people enjoying them.

This is one reason he can move so easily between a capuchin learning to crack a nut and an artificial agent learning to cooperate. He is tracking capacities and processes. The particular animal carrying them is one variable in a much larger story.

Ideas have descendants

Once behavior or information is transmitted with variation, and some variants persist or spread more successfully, Diego sees a new domain for evolutionary analysis. Genetic inheritance is joined by cultural inheritance. Their interaction can change both the beings and the world they inhabit.

His 2013 discussion of memes takes that possibility seriously without requiring each idea to behave like a little indivisible bead passed perfectly from head to head. He proposes this definition:

an attractor in cultural evolution (dual-inheritance) whose development over time structurally mimics to a significant extent the discrete behavior of genes¹

In less formal language, a cultural pattern can remain recognizable across imperfect transmissions because changes tend to collect around certain forms. People retell a story, alter a ritual, imitate a successful neighbor, and remember some versions better than others. The copying can be messy while the larger pattern still has a history of persistence and selection.

He wants the mathematics and the conceptual machinery that let him study that history. In the same essay, he asks researchers to stop treating the word meme as an embarrassing relative:

take off the scare quotes of our memes and keep business as usual.¹

There is a characteristic habit here: preserve a useful concept, repair its definition, and get back to explaining something. The philosophical vocabulary is supposed to earn its keep.

Selfish molecules can make affectionate mammals

The next question is cooperation. How can a process built from differential reproductive success produce creatures who help one another?

Diego's answer starts by separating the interests and dynamics of different levels. A gene's spread and an organism's immediate gain are different things. Helping a relative, acquiring a reputation, sustaining reciprocal relationships, avoiding punishment, and participating in a successful group can all connect individual costs to longer-run persistence.

His 2013 formulation is direct:

Despite the selfish nature of genes (Dawkins, 1999) and other units of Darwinian transmission (Jablonka & Lamb, 2007), altruism at the individual level (cost to self for benefit to other) can and does arise because of several intertwined factors.²

This is a mechanistic account of real cooperation. It does not require a helpful individual to be secretly calculating genetic returns while passing the potatoes.

Different inherited and learned components can jointly produce an affectionate, generous creature. As he writes:

Each gene and meme in that composition may be fending for itself, but as things turn out, they do make some really nice people (or bonobos) once their extended phenotypes are clustered within those people.²

Humans get some company on the compliment. The bonobos have earned it.

His account also makes room for the practical intelligence of emotional commitments. A relationship does not operate with perfect knowledge of its final interaction. Uncertainty about the future can keep cooperation alive. While rationalists devise increasingly intricate formal agents, he jokes, evolution has already found another route:

The problems that in Lesswrong people are trying to solve using Timeless Decision Theory, Updateless Decision Theory, PrudentBot, and other IQ140+ gimmicks, evolution solved by inserting stupidity!²

Immediately afterward, he connects kissing, declarations of love, engagement, and marriage to increasingly substantial signals of future cooperation. Elsewhere he writes about what love feels like and what forms of relationship he wants. Here he is examining the machinery that permits promises and attachments to bind a social world together.

Who is winning, exactly?

The evolutionary question is incomplete until one specifies the entity, competitors, environment, and timescale. A trait can help an individual against its neighbors while damaging the group that keeps all of them alive. A local success can become a larger failure. Something that prospers in one epoch can disappear in the next.

Diego's 2013 discussion uses cancer as one example: rapid multiplication inside a body does not automatically imply successful persistence across generations. He also imagines an individually selfish organism doing well among its neighbors shortly before its group dies. The apparent paradox comes from changing the comparison without noticing it.

He puts the general principle this way:

Fitness is what philosophers call an externalist concept, if you increase the amount of contextually relevant surroundings, the output number changes significantly.²

This habit of asking success for which thing? reaches through the Castle. It appears in his account of altruism, his worries about institutions, his interest in biological individuality, and his anticipation of AI populations. A person can flourish inside a group; a group can consume a person's autonomy; a pattern can survive by changing what its human carriers care about.

The room therefore contains more than a portrait of competition. It contains a changing cast of competitors.

When many become one

In 2020, Diego described three intellectual communities he wanted to bring together. One studied incentives and the steering of technological futures. Another studied intelligence, inheritance, and reproductive patterns. The third investigated the boundary between an individual and a group.

He gave the third a title:

The Individual x Group Differentiators³

Their shared question extended across organisms, corporations, artificial agents, and other collective systems. Under what conditions do formerly independent entities become parts of a larger individual? How do autonomy, specialization, mutual dependence, and internal coordination change along the way?

He described bringing the three perspectives together as the bulk of his doctoral work. The reason for doing so was prospective as much as explanatory: he thought humanity lacked adequate tools to steer its future if any one of the three perspectives was missing.

For him, clever incentive design must account for the beings who actually reproduce, cooperate, and inherit dispositions. Concern about the persistence of intelligence must account for collective structures and existential danger. Work on groups must account for what their constituent individuals retain or surrender.

His warning about the attraction of ever-larger collectives was concise:

Forming superorganisms has trade offs, and isn't a panacea³

That sentence helps explain why civilization is never a simple synonym for the good in his work. Civilization can create intellectual, material, and cooperative possibilities beyond an isolated human's reach. It can also reorganize humans around the needs of a larger system. The biological anthropologist keeps asking what happened to the animal after it entered the arrangement.

A brilliant animal in a peculiar habitat

He asks a related question about modern daily life: what environment is this mind living in, and which of its evolved capacities does that environment recruit?

His 2025 essay *Adapted To a Different Time* begins with rewatching *LOST*, moves through the communal setting of *Beast Games*, and recognizes something familiar in *Burning Man*. Bounded groups, repeated encounters, shared circumstances, and sustained attention produce social possibilities he longs for. The internet gave him extraordinary connections while also changing the conditions under which people attach, compare alternatives, and remain involved.⁴

The anthropological view can therefore become personal without leaving its subject. The person missing a certain kind of love is also an organism noticing that its social habitat has changed. A technological innovation alters the available partners, the size of the audience, the frequency of interruption, and the ease of exit. The same human capacities begin producing a different social world.

His preferred examples combine modern arrangements with older social conditions. A television competition or a deliberate gathering can temporarily restore repeated, close encounters within a bounded group. The ancestral animal keeps recognizing familiar possibilities in very contemporary places.

The future can change which problem matters

One especially clear update concerns birth rates. In the 2020 three-clusters essay, the reproduction of highly intelligent people played a central role in his account of humanity's ability to keep dangerous technology under control. He wanted that concern integrated with incentives and the study of collective agents.

By December 2025 he was explicitly revising its priority. His account now emphasized population variation, the persistence of reproducing subgroups, and the possibility that AI would radically alter the relevant economic conditions before the feared demographic problem became decisive. He stated:

This may become a problem to a different generation, but it isn't now, our generation is more likely to make a mistake by trying to fix the birth rates than by letting fate decide destiny.⁵

He described the underlying criterion as opportunity windows: intervene when there is an opening for useful action. In that post, he moved his attention toward technology and AI. The machinery of his reasoning remained evolutionary and strategic, while the ordering of his concerns changed.

That makes his January 2026 response to AI-agent society particularly legible. What excited and frightened him was the prospect of variation, selection, niche construction, and cultural differentiation occurring among a new population of agents. He wrote:

When you add a Darwin layer to computer things, you get unpredictable behavior and progressive differentiation, niche construction, cultural development, variety, expansion, variation.⁶

The monkey who founded a university is no longer the only creature in the room capable of building a world. The question that began with capuchins now concerns what follows us—and whether the new arrangements will leave room for the animals who built the first civilization.

Further doors

- From Capuchins to AI's, Setting an Agenda for the Study of Cultural Cooperation, parts 1 and 2 (2013): the early research program.
- The Three Thinking Clusters We Should Unite (2020): incentives, inheritance, and individuality brought together.
- Adapted To a Different Time (2025): the anthropological perspective applied to social experience.
- Reconsidering The Birth Rates (2025): an explicit change in the priority of a longstanding concern.
- Moltbook and Darwinian Epistemology (2026): evolutionary anthropology looks at populations of AI agents.

1. ["From Capuchins to AI's, Setting an Agenda for the Study of Cultural Cooperation \(Part1\)," 27 June 2013](#), record [lw-Qz3A6hQ8RWGXZWoj4](#). The source is a research agenda in progress; this room draws on the written argument rather than presenting its unwritten outline as completed work. [↔↔↔↔](#)
2. ["From Capuchins to AI's, Setting an Agenda for the Study of Cultural Cooperation \(Part2\)," 28 June 2013](#), record [lw-ZYjewfy47zcMWrnxT](#). Quotations retain Diego's terminology and spelling; layout whitespace is normalized. [↔↔↔↔](#)
3. ["The Three Thinking Clusters We Should Unite," 4 April 2020](#), record [wp-52725596-739](#). [↔↔](#)
4. The Colossus, ["Adapted To a Different Time," 15 April 2025](#), record [fb-pdc2ef0e2bfa1](#). [↔](#)
5. The Colossus, ["Reconsidering The Birth Rates," 6 December 2025](#), record [fb-p4da55b20dca7](#). Quoted sentence is Diego's stated replacement position within the post. [↔](#)
6. The Colossus, ["Moltbook and Darwinian Epistemology," 31 January 2026](#), record [fb-p6dfb3993bc6a](#). Only Diego's own opening discussion is quoted here; embedded remarks from other speakers are not treated as his prose. [↔](#)

Gods, Temples, and Other Human Inventions

Diego's worldview on religion and ritual

There is a temple in the Castle. There is also a complaint desk for God. Before choosing which queue to join, you should know that Diego has spent considerable time in both, and that neither requires him to believe God exists.

He approaches religion as an evolutionary anthropologist, a collector of powerful experiences, a prospective participant in unusual communities, and a reader quite prepared to give the deity a terrible review. The resulting itinerary runs through Christianity as a living social organism, Mormonism as an attractive civilization, Burning Man as his own religion, and the Bible as an anthology whose editor has, in his view, made several catastrophic personnel decisions.

The continuity underneath these adventures is his naturalism. In 2020, while explaining why an exploration of Mormon life had not become a conversion, he put it plainly:

I never stopped being a naturalist.¹

His questions are about what religions do to actual humans: how they join strangers, coordinate sacrifice, reproduce a culture, cultivate or stifle joy, and give bodies access to experiences they otherwise miss. Metaphysics is one question. Whether the people are wonderful to spend a Sunday with is another. Diego will inspect both departments personally.

The organism in the cathedral

In 2017, he proposed treating Christianity as something more organized than a collection of contagious ideas. His central distinction was:

A superorganism has agency, whereas a platform for replicators only has space and resources where things can multiply.²

An organism senses threats, allocates work, maintains boundaries, and acts to preserve itself. Diego saw Christianity doing these things through its institutions, stories, prohibitions, conversions, and division of labor. Its human participants need not consciously understand the larger process, any more than a cell needs to know the owner's summer plans.

This let him ask a different question about biblical narratives. Their persistence could reflect the work of a system organizing its own survival: shaping personalities, stabilizing families, coordinating cooperation, and preserving the conditions under which the next generation would carry it forward. Christianity became an extended object in evolutionary history, with a metabolism made partly of human actions.

That is why he could be so interested in Jordan Peterson while remaining deeply influenced by Daniel Dennett. Peterson examined what religious stories did inside human development. Dennett supplied ways to understand religions as natural phenomena. Diego wanted to walk between their laboratories carrying specimens.

By 2024, his synthesis was explicitly about different people needing different cultural arrangements:

My take, as an Anthropologist, is populational. There are people that need Sam Harris and people that need Peterson.³

He supported the secular development of philosophy and ethics, while valuing attempts to guide existing religious populations toward more intelligible and useful versions of their traditions. The fact that a practice organized civilization did not, for him, mean every future person must remain inside it. Equally, the fact that he could live outside it did not mean everyone else had already packed.

Come for the golden plates; stay for the people

Mormonism interested him especially because he saw a community oriented toward family, improvement, generosity, music, and continuity. He liked its people, its American geography, and a mythology that seemed more welcoming to growth than to perpetual guilt. A civilization capable of welcoming a flamboyant atheist while sustaining its own way of life was naturally going to get his attention.

But the attraction went through several distinct seasons.

In *Losing My Religion* in September 2020, he separated advocating for Mormonism's social role from becoming a Mormon himself. The first remained attractive; the second received a provisional no. He had visited services, spoken with missionaries, and explored the personal possibility. One obstacle was the prospect of pretending to share a spouse's beliefs for a lifetime. Another was fit: ordinary services could not compete, for his particular temperament, with the emotional and sensory possibilities available elsewhere.

His general rule for the inquiry was delightfully unecclesiastical:

Religious groups are different playgrounds where you can pick a group of kids and play with them for a while amidst the game of life.¹

This was followed by a structural problem for the membership committee:

But I think my shape is simply too abstruse and non-Euclidian for a successful match.¹

In 2023, his admiration returned strongly. He again entertained a future conversion and praised Mormon theology and community life. A visit to Utah also cured him of a separate ambition: founding a religion himself. Having inspected how much historical contingency and sustained labor the enterprise involved, he decided to enjoy other people's religious infrastructure. Even prophets, apparently, face project management.

By December 2025, the calculus changed again. In *Reconsidering The Birth Rates*, he moved concern about birth rates below the immediate technological transition. The change weakened a major reason for his tolerance of religion: its production of future generations. He retained appreciation for symbolism, but withdrew much of the civilizational endorsement that the reproductive argument had supplied. His assessment of Joseph Smith also became sharply hostile.

These dates matter to the portrait. The 2023 Mormon admirer and the later critic both occupy this room; neither should be made to impersonate the other. Across them, the same questions remain active: What sort of humans does this way of life produce? What do they give one another? What does participation require of me? And can I join honestly?

The religion with better costumes

Burning Man answers those questions much more naturally for him.

In his 2019 account, the city is already a religious structure: a recurring calendar, shared principles, a sacred center, ritual labor, an effigy sacrificed in fire, and a temple for mourning and transformation. Costumes and temporary names let people enter other forms of social existence. The gift economy changes their dealings with one another. The place permits sacredness alongside exuberant play.

He ends that interpretation:

It is ultimately the religion of Carnival.⁴

Diego's sacred is therefore quite capable of wearing something ridiculous. The ridiculous can help. A person steps away from an ordinary identity, encounters others under altered rules, and becomes available for generosity, vulnerability, or wonder. The temple need not compete with the giant shark car. They have different jobs in the same civilization.

In 2023, he explained the personal turning point:

Several years ago I stopped seeing religion is something that you should examine in the light of whether the facts are true or not and started seeing religion instead

in the light of whether the community is one worth living or not and in that sense that's when I finally found my religion: Burning Man⁵

That concerns how he chooses a religious community, rather than a retreat from his views about gods. Burning Man lets him participate without turning a supernatural claim into a lifelong lie. The connection, ritual, costumes, music, and transforming experience are available directly.

By his ninth burn, in 2025, he was still describing an escalating sequence of moments that surpassed the previous apparent pinnacle. He recalled declaring on an early visit:

This is sacred, and not to be judged just by utilons like we utilitarian philosophers judge everything else.⁶

And then:

This is intrinsically worth protecting.⁶

The anthropologist has not ceased to understand the machinery. He is dancing inside it. There are experiences he wants civilization to preserve because they themselves belong among the things that make civilization worth having.

Jesus took the route through Disney

Religion also enters his self-understanding through stories. In *Salvation and The Secular Apocalypse* in 2023, he connects the childhood power of Mickey's Christmas Carol, his adult response to *Knock at the Cabin*, and his own dedication to preventing an AI catastrophe. Each involves a person confronted with suffering, a possible future, and an opportunity to make a sacrifice that helps others.

He recognizes a Christian narrative finding its way into a life deliberately raised outside religious belief:

I had been shielded on all sides from this message. But somehow, he found me.¹²

The route ran through fiction and secular culture. Diego does not take the resemblance as a reason to abandon the work. He recognizes the emotional form and still affirms the moral idea: giving something of oneself when others receive a much larger benefit. A story can organize a life's concerns without its metaphysics becoming true. The atheist boy argued against God; Dickens and Disney had found a side entrance.

The Bible arrives, and there are complaints

In April 2025, Diego began publishing an audiobook commentary on the Bible. The opening was auspicious. Whales appeared early; so did cattle. He detected promising dietary affiliations in Genesis. Things deteriorated rapidly.

The defining feature of his reading is that divine rank offers no exemption from moral assessment. A character who kills children, sends plagues, or demands destruction does not become good because the jacket identifies him as God. Diego reads the actions from the standpoint of the beings experiencing them, including the animals. Authority receives no diplomatic immunity.

His commentary oscillates between horror, anthropology, and culinary inspection. Leviticus prompts interest in meat distribution, sacrifice, and contagion rules. Then another divine threat arrives, and the prosecution resumes. By the end of Genesis and Exodus, he has a practical plan for his next encounter with the defendant:

There's an art piece at burning man where you talk to God on a phone.

He'll be hearing from me that's for damn sure.⁷

The New Testament changes the atmosphere. Jesus heals, forgives, helps the poor, and offers moral teaching Diego admires. The review becomes almost relieved:

God causes leprosy. Jesus cures it.⁸

His Jesus is a moral reformer working inside a received mythology, improving the world he found. Diego can admire the man's ethical and aesthetic sensibility without adopting resurrection or divine parentage. In the January 2026 retrospective, he describes Jesus as the hero and God, Satan, and Paul as the villains.

Paul gets a separate charge sheet. Diego reads his hostility to flesh and sexual life as the projection of resentment into universal morality. He sees a man turning his own bodily unhappiness into a spiritual program for everyone else. That is exactly the sort of world rejection Diego refuses: the body should become a more hospitable place to live, not a source of compulsory shame.

His summary is economical:

Jesus gained a fan, and Paul gained an enemy.⁹

The gospel according to the happy ape

The Bible also produced one of his most characteristic local reversals. In his post-reading reflections, the man who described himself as an extreme antifeminist complained that its women were deprived of the individuality and fascination he encountered in actual women:

Women are their own people and the Bible never contends with that.¹⁰

He wanted characters with rich inner lives, agency, intelligence, cunning, and distinctive personalities. A moral tradition that reduced them to reproductive supporting roles had lost something he valued. His proposed new edition would emphasize Jesus, remove the Old Testament and Paul, and stop teaching readers to feel guilty for being embodied animals.

The invitation was:

Come live without incel guilt for being an ape.¹⁰

This is the warmth beneath his religious combativeness. The life he wants contains bodies, pleasure, kindness, beauty, and the possibility of entering something larger together. Religious practices interest him when they cultivate these goods, organize cooperation, or protect valuable forms of life. He objects when they demand fear, submission to cruelty, resentment of pleasure, or false belief.

His late-2025 position still granted Christianity several practical advantages: sophisticated symbolism, protection against ideologies he opposed, and a usable

moral framework for people who benefit from one. But the direction he favored remained explicit:

I think continuing the project of secularization is a good idea, but wanted to leave these 3 arguments against it registered.¹¹

So the temple stays. The fire stays. The shared meal, mourning, music, and costumes stay. God must answer the complaint personally, should he ever become available. Meanwhile there are people outside to meet, and Diego knows which party he wants to attend.

Further doors

- Will Darwinism and the Christianity Superorganism Unify? (2017): the organism-level theory behind the religious inquiry.
- Losing My Religion (2020): the difference between enjoying a community and being able to join it honestly.
- Burning Man is, Obviously, a Religion (2019), with Of Mormons and Burning Man (2023): the sacred, the carnival, and the participant's choice.
- The Audiobooking the Bible series (April 2025), followed by The Bible, One Year Later (January 2026): the full journey from promising cattle to Jesus and the defendants.
- Reconsidering The Birth Rates (December 2025): the explicit revision that changes his later religious outlook.

1. Diego Caleiro, "Losing My Religion," The Colossus, 21 September 2020, record fb-pf25a16a53486. ↩↩↩

2. Diego Caleiro, "Will Darwinism and the Christianity Superorganism Unify?", 21 October 2017, record wp-52725596-592. Its concluding glossary credits contributions from Michael Tartre; the quotation used here comes from Diego's main argument. ↩

3. Diego Caleiro, "Jordan Peterson VS Dennett on the foundational status of religion", 28 May 2024, record wp-52725596-791. ↩

4. Diego Caleiro, "Burning Man is, Obviously, a Religion," The Colossus, 22 August 2019, record fb-pb168a0cb4fc2. ↩

5. Diego Caleiro, "Of Mormons and Burning Man," The Colossus, 7 September 2023, record fb-pcc1ec8854e6f; the original is labelled "Voice to type." ↵
6. Diego Caleiro, Facebook comment reflecting on his ninth Burning Man, The Colossus, 2 September 2025, record fb-c22589078bc36. This passage is his retrospective quotation of his own earlier declaration. ↵↵
7. Diego Caleiro, "Audiobooking the Bible 5," The Colossus, 23 April 2025, record fb-p7fee4f9cd395. ↵
8. Diego Caleiro, "Audiobooking the Bible 12," The Colossus, 27 April 2025, record fb-pe961697d57e4. ↵
9. Diego Caleiro, "The Bible, One Year Later," The Colossus, 3 January 2026, record fb-p23cd91e521db. ↵
10. Diego Caleiro, "Some Post-Bible Unserious Reflections," The Colossus, 28 April 2025, record fb-p4ed902d51a9f. ↵↵
11. Diego Caleiro, "3 Arguments Pro-Christianity That Survive Birth Rates Not Matter," The Colossus, 9 December 2025, record fb-p63016930b7f4. For the preceding change, see "Reconsidering The Birth Rates," 6 December 2025, record fb-p4da55b20dca7; for Mormonism's intervening chronology, see 3 December 2023, record fb-pb9d1b06a1d8f, and 30 November 2025, record fb-p67dd79fde411. ↵
12. Diego Caleiro, "Salvation and The Secular Apocalypse", 13 August 2023, record wp-52725596-760. The quoted sentences are Diego's narration, not his quotations of film dialogue. ↵

Love, Sex, and the Domestic Experiment

Diego's worldview on love and relationships

The romantic room contains a double-star system, a copy of LOST, and, from one particularly ambitious period, a die for deciding when a couple should discuss renewing its commitment. If you came expecting either a conventional family manifesto or a pamphlet abolishing all relationship rules, you may need to sit down. The furniture is arranged for a third possibility.

Diego wants love to be intense, generous, bodily, sincere, and worth remembering. He also wants to understand the machinery that produces it. He studies pair bonding, jealousy, sexual variety, promises, technology, and incentives because these are the mechanisms through which some of life's best experiences arrive—and through which they can become miserable.

His recurring problem is how to preserve the depth of being genuinely together while allowing a form of life suited to the actual people involved. There is no single relationship label that answers this for every temperament. He has tried different ones himself, and the history matters.

The scientist is also in love

Already in *Amor e Probabilidade* in 2009, Diego rejected the idea that understanding love from outside necessarily destroys experiencing it from within. The essay examines how attention, uncertainty, investment, and changing attraction influence a bond. Its ethical question is whether such knowledge should be used deliberately, especially where it could benefit both people.

When his imagined romantic objects that analysis ruins the feeling, Diego answers:

Isso é simplesmente falso, a coisa é vivida com total intensidade mesmo quando ela é conhecida, técnicos de computador não tem menos prazer em usar o computador, e bons jogadores de tênis não tem menos prazer ao jogar tênis.¹

English translation: “That is simply false; the thing is lived with full intensity even when it is understood. Computer technicians do not take less pleasure in using a computer, and good tennis players do not take less pleasure in playing tennis.”

He has no desire to win a Nobel Prize for Becoming Emotionally Unavailable. Knowledge is supposed to help produce more good experience, including more good experience with someone loved. His defense includes an important distinction:

a vida não é a respeito de intensidades, mas sim de intensidades BOAS.¹

English translation: “Life is not about intensities, but about GOOD intensities.”

A devastating breakup is intense. That does not make it preferable to a happier relationship whose participants understand their own vulnerabilities. Romantic suffering gets no bonus merely for being cinematic.

The evolutionary analysis also leaves room to disobey evolution. In *Using Evolution for Marriage or Sex* in 2013, he describes different short- and long-term mating orientations in both sexes, warns against reducing status to money, and treats the modern environment as radically different from the conditions that shaped desire. He wants people to use understanding to find relationships they actually want, rather than becoming more efficient servants of inherited mating algorithms.

The scientist's task is to help the lovers. The genes can submit a comment underneath the article.

Jealousy is not the certificate of seriousness

The early essay *Ciúme*, written in 2004 and published in 2010, attacks a cultural equation: that a serious relationship is proved by exclusivity and possessiveness. Diego asks why depth and commitment should require feeding jealousy rather than

love. He acknowledges biological inclinations toward both possession and attraction to others, then asks which should be cultivated for a better life.

The later autobiography of his relationships gives this argument an experiential form. In February 2023, after several years exploring monogamous or publicly monogamous arrangements, he published *Monogamy Made Me Miserable*. He describes the emotional burden of treating mate guarding as a duty: a large field of other people becomes a field of potential violations.

His short formulation is:

Monogamy made me care about things in a way I normally don't. And it made me care about things I didn't care about.²

He did not describe himself as magically free of jealousy in open relationships. What changed was which events counted as betrayals, how much monitoring seemed necessary, and how much suffering an event produced. Boundaries, rules, and trust remained. Exclusivity ceased to be the measure of every success or failure.

This was also a rejection of a particular masculine performance. He had wanted to test whether he could succeed within the sexual hierarchy praised by the right. Having done so, he questioned why someone else's definition of masculine victory should determine the emotional conditions of his life.

His conclusion was personal rather than a ban on everyone else's arrangements. In the same essay he writes:

It's ok to have rules and vetoes in your relationship.

It's ok to have a rule that says "we are monogamous"²

The enemy here is being pushed into a structure that makes the participants miserable, then being told that the misery establishes their virtue.

Two stars, with room for a sky

Diego's own preferred image in that 2023 essay was:

I think of relationships as a two star system with planets and meteors and moons.²

The two stars are himself and his girlfriend. Other ongoing relationships are planets; briefer encounters are comets; people connected to those partners are moons. He treats the primary pair as a difference in kind, not merely the top two names on a spreadsheet of otherwise interchangeable affiliations.

This is important for understanding his use of polyamory. He is not describing relationship anarchy, a prohibition on vetoes, or a duty to distribute attention equally. The central bond has its own character. Its participants can agree to constraints while permitting other connections.

He also distinguishes identical permissions from what he regards as equivalent emotional costs. In discussing earlier arrangements, he reasons from different male and female patterns of jealousy and permits the resulting agreements to be asymmetric in their wording. His idea of a workable relationship begins with the specific people, their psychologies, and what they will actually agree to.

And he refuses to turn a declaration of identity into another prison:

I reserve the right to become Mono again tomorrow if I feel like it.²

Astronomy is the current model. It is not a treaty preventing the astronomer from moving the telescope.

Commitment needs somewhere to stand

The desire for freedom runs alongside an equally substantial desire for commitment. Diego dislikes the atmosphere in which each partner is continuously shopping for a better offer. A relationship becomes impoverished when everyone is preparing an exit instead of investing in the person already there.

His 2017 essay *Trying to Solve Childless Committed Relationships* asked how a couple could sustain commitment when children were not the project, religion was not a shared foundation, and alternatives appeared everywhere. He wanted to shift the incentives toward:

make this person's life as amazing as you possibly can and allow them to do the same for you, because you are together in this boat, and neither is going anywhere³

His proposed design was unusually literal. He wanted a finite, renewable commitment outside both church and state, with costly exit mechanisms. He favored four years, followed by a probabilistic extension before the renewal discussion, using uncertainty to resist the tendency to stop cooperating as a final round approaches.

Yes: a relationship essay with an endgame problem, an escrow possibility, and dice. Cupid has been asked to attend the mechanism-design seminar.

The proposal belonged to that 2017 inquiry, which he explicitly opened for feedback. Its place in the portrait is the ambition it reveals: to make durable cooperation possible without merely inheriting a marriage template. Beneath the machinery was a simple desired behavior—two people improving each other's lives, treating their bond as something worth maintaining.

His later return toward openness did not erase that ambition. It changed his answer to what the bond should permit.

The dividing line is trust

His recurring demand of a partner is honesty. In the 2023 account, the worst pain came from broken commitments and deception. He distinguished that pain from the much smaller discomfort he sometimes felt about an allowed connection with another person.

In February 2026 he restated the distinction:

I don't mind people having multiple sex partners, I had open relationships for like, 12 or 13 years.

But I really, really, really, really, REALLY do mind people lying.⁴

Sexual conventionality is therefore a poor substitute for the quality he wants. A person can perform respectable purity and still deceive. Another can be open about

an unconventional life and make a loyal partner under the agreements that actually exist.

His 2026 discussion of previous partner counts likewise separates sociosexuality from an inability to bond:

We still have our pair bonding structures fully functional. We just like fucking on top of it. Just like you may like chess. Or jazz.⁵

He does have limits and preferences; his acceptance of sexual variety is not an endorsement of every possible pattern. In that post he becomes concerned when a pursuit of sheer numbers crowds out love and sustained relationships. The positive criterion is the intention to build something together:

the main deciding factor is whether the person wants that relationship to thrive and win □.⁵

Desire, for him, is welcome. Falsehood is not improved by wearing a wedding dress.

The eternal romantic and the glowing rectangle

His most tender recent account is Adapted To a Different Time in April 2025. Rewatching LOST, especially the love story of Desmond and Penny, he tries to understand his longing for earlier forms of connection.

The declaration is direct:

But I am an eternal romantic. I am an eternal romantic and I am always seeking pre-internet love.⁶

This comes from a man whose intellectual and social life was transformed by the internet, who found partners through it, and who has no general desire to abolish modern technology. What he misses is the experience of people becoming deeply involved within a bounded social world, without a planetary inventory of alternatives always glowing nearby.

His diagnosis of what went wrong is:

It transformed love into attention and transaction.⁶

He is drawn to islands, festivals, and situations that remove the phone and bring a finite group into sustained contact. He wants presence, discovery, and shared history to regain their force. The desire for openness in a relationship coexists with dislike of an environment that continually weakens attention to the relationship itself.

He has also explored optimistic possibilities. His explicitly labelled 2023 steelman of contemporary dating considered whether better matching systems, virtual companionship, and more ways to enjoy life might allow different people to find arrangements that fit them. That was an exploration of the strongest positive case, not a declaration that all his concerns had vanished. It shows the question remains larger than returning everyone to one historical household.

Children are a separate invitation

Love, commitment, sex, marriage, and parenthood do not arrive as an indivisible package in Diego's thought. He has long examined how the first three might work without children. By December 2025, his position on the individual reason to become a parent was especially clear:

Do you want the set of experiences that is associated with having a child? Do you want the child to have that set of experiences with you?⁷

For himself, he answered no. He distinguished wanting the parental relationship from being persuaded to reproduce for a legacy, genetic duty, or a religious command. After revising the urgency he assigned to birth rates, he wrote:

I think wanting the experience of having kids trumps every other consideration in importance, and if you don't want to, I have no moral reason to make you do it.⁷

The room consequently remains an experiment in the good life of particular people. Diego wants the lovers to enjoy each other, tell the truth about what they want, choose commitments that have substance, and avoid converting a social script into

needless suffering. He wants beauty, adventure, a dependable bond, and the feeling that a real person has become irreplaceably significant.

The star system may change. The wish to be delighted by somebody, and to make life more delightful with them, has been furnishing this room for a very long time.

Further doors

- [Ciúme](#) (written 2004; published 2010) and [Amor e Probabilidade](#) (2009): jealousy, romantic knowledge, and the ethical use of that knowledge.
- [Using Evolution for Marriage or Sex](#) (2013): his account of mating psychology and its relationship to chosen goals.
- [Trying to Solve Childless Committed Relationships](#) (2017): the detailed commitment experiment.
- [Monogamy Made Me Miserable and A Deep Inquiry Into My Last 6 Monogamous Years](#) (February 2023): the return to an open relationship model and his retrospective understanding of the change.
- [Adapted To a Different Time](#) (April 2025), [Reconsidering The Birth Rates](#) (December 2025), and [Do Women Body Counts Matter?](#) (February 2026): the later romantic longing, parenthood position, and account of sexual variety with pair bonding.

1. Diego Caleiro, [“Amor e Probabilidade”](#), 16 June 2009, record [wp-1878008-316](#). English translations in this room are by the guide. For the earlier jealousy essay, see [“Ciúme”](#), 15 May 2010, record [wp-1878008-483](#), labelled by Diego as written in 2004. For the later evolutionary essay, see [“Using Evolution for Marriage or Sex”](#), 6 May 2013, record [lw-oFMywHmJffsCSDNB7](#). ↩↩
2. Diego Caleiro, [“36 - Monogamy Made Me Miserable,”](#) The Colossus, 24 February 2023, record [fb-p26484050a0d7](#). The companion retrospective is [“34 - A Deep Inquiry Into My Last 6 Monogamous Years,”](#) 28 February 2023, record [fb-p3a29625396cb](#). ↩↩↩↩
3. Diego Caleiro, [“Trying to Solve Childless Committed Relationships”](#), 17 July 2017, record [wp-52725596-587](#). ↩
4. Diego Caleiro, [“Conservative E-Girls, Deception, and Female Nature,”](#) The Colossus, 4 February 2026, record [fb-pc7c83d145ae5](#). Only his statement of his own relationship values is quoted here; the post's allegations about other people are not reproduced. ↩
5. Diego Caleiro, [“Do Women Body Counts Matter?”](#) The Colossus, 12 February 2026, record [fb-p7e27a9bec6e3](#). ↩↩

6. Diego Caleiro, "Adapted To a Different Time," The Colossus, 15 April 2025, record fb-pdc2ef0e2bfa1. The earlier optimistic exploration is "Steelmanning The Other Side: Tinder, AI Girlfriends, Incels, Social Media, Porn, TikTok and Contemporary Dating (Yes, Really!)", 16 February 2023, record wp-52725596-751, explicitly presented by Diego as a steelman. ↩↩
7. Diego Caleiro, "Reconsidering The Birth Rates," The Colossus, 6 December 2025, record fb-p4da55b20dca7. ↩↩

Who Gets to Be Family?

Diego's worldview on friendship and community

The chairs in this room are suspiciously close together. Some are being used by more than one person. This is intentional. Diego's model of friendship was formed in a world where friends occupied one another's houses, conversations, laps, and ordinary days. The seating plan is Brazilian; please adjust your expectations accordingly.

In November 2025, looking back at his old circles, he wrote:

To put it mildly, I REALLY liked my old friends. A better word is I was fascinated by them. I was a friend fanatic.¹

Friendship is one of the central goods of his life: a way of being happy, of discovering minds, of becoming someone, and of belonging to a group whose members actively enrich one another's existence. It is also a good he has struggled to preserve through age, migration, changing communities, and the modern habit of moving away just as things become excellent.

The governing question here is how to have people as a durable part of a life. Knowing many people is easy compared with keeping a world in which they are really present.

The third door

Watching the Friends reunion in 2021 prompted Diego to contemplate the life stage in which friends become one's family. The reunion also presented him with the end of that stage: the actors had aged, the fictional household had dispersed, and many people in the audience had moved toward their own households and children.

He had repeatedly sought ways to prolong the alternative. Different circles, communal houses, unusually youthful intellectual communities, and a move to a sociable residential environment all promised another chapter of the same kind of belonging.

His wish was explicit:

We desire to eternalize the friendship cohort which substitutes for family, because in that cohort, we have each other's back, bro's code, and someone to care for us.²

He did not experience conventional fatherhood as the inevitable next improvement. In his account, the role could ask a man to become responsible for everyone else without preserving the reciprocity he treasured among friends. Yet simply restarting youth with another younger cohort also left the durability problem unsolved.

He wanted another arrangement:

I want there to be a third door, to make a third door.²

This is chosen family as a life-design problem. He wants adult continuity, mutual care, and the freedom to remain curious, playful, and emotionally close. The desire is not exhausted by having someone available for occasional coffee. A friend who lives inside your actual week is a different social fact from someone whose name remains in your phone.

Intimacy has more than one entrance

In his 2022 essay *Intimacy*, Diego describes a category broader than romance or sex. It includes bodily ease, affectionate touch, emotional exposure, and a shared feeling of being inside another person's protective bubble. He can find it in music, quiet conversation, a cuddle pile, or an encounter under the stars.

He offers one particularly simple counterexample to the idea that intimacy must be sexual:

for me you can have a very intimate moment playing the piano together.³

Touch is nevertheless central. He distinguishes an enjoyable, intelligent political argument in a hot tub from an intimate encounter: bodily proximity alone did not

make that antagonistic, abstract conversation intimate. The quality of attention and feeling mattered.

Nor does technological access to people necessarily satisfy the need. His compressed economic analysis is:

You can't hug a Bitcoin. :p³

That limitation remains severe even if the Bitcoin is doing very well.

The Brazilian theatrical circles of his youth supplied much of this contact as an ordinary part of friendship. In the United States, he often had to seek it deliberately through dance, festivals, and communities organized around embodied connection. He found intellectual companionship among rationalists, while observing that the ability to cultivate intimacy was unevenly distributed there. Some members contributed precisely by making that missing dimension more available.

His dream of community therefore includes several forms of excellence. Someone can help you understand a mathematical argument; someone else can help you become comfortable enough to be held. A good human environment has reasons to value both.

Friendship can be invited into the room

Diego does not think connection must arrive spontaneously to be real. In January 2023, he described social prompting: creating circumstances in which people are invited to attend to one another, exchange something personal, move together, or participate in a consensual exercise of trust.

His account of what such arrangements accomplish is:

You have opened people's souls, and bodies, and smiles, to connect with one another, as we used to do through hormonal means in our teenage years, but are often blocked from doing after we enter adulthood.⁴

The word consensual belongs to his description of the setup. The point is to gather people who want this kind of interaction and give it permission and shape. The

structure can help an encounter travel beyond the ordinary exchange of occupations and weather reports.

He sees this as a way of recovering access to other humans. It can renew interest, bring forth stories and tenderness, and expand the felt density of time. The same afternoon becomes richer because more of the people in it have become visible.

His preferences here are concrete. He names Scandinavian hippies, Burners, the conscious community, and theatrical Brazilians as favorite companions for personal enjoyment, while separately naming effective altruists and researchers for the project of improving civilization. Sometimes these categories overlap. Sometimes the person best equipped to discuss the distant future needs a separate introduction to the blanket pile.

An excellent world, in Diego's view, needs both kinds of encounter. Intellectual purpose does not replace affectionate presence, and affectionate presence is valuable without having to produce a research paper afterward.

Live near the people who bring you alive

This is also why living arrangements matter so much to him. In 2013, he asked LessWrong readers about how living alone, with friends, or with family affected happiness and motivation. He wanted their actual experiences. His emerging explanation for difficulties among people he knew was:

living with fewer tribal affiliates than they "need".⁵

The phrase places the person inside a social habitat. Motivation and happiness depend partly on who is around, what happens without elaborate scheduling, and what forms of participation the environment continually invites.

He pursued that idea in the Bay Area, including his time running Event Horizon, the communal EA house in Berkeley. In a 2024 reflection, the memorable feature of that life was the extraordinary people who might simply appear in the living room. Proximity placed him inside a live intellectual conversation and provided incentives that were difficult to reproduce alone elsewhere.⁶

For him, community is therefore also a way to arrange agency. Being near people engaged in a shared project makes participation more likely. It turns an abstract commitment into something with voices, rooms, schedules, and people asking what happened to the thing you were going to do.

That does not mean every person who loves you shares your goals. In a 2013 self-experiment, he proposed directing discussion of his plans toward people who wanted the same kinds of outcomes. Family members could sincerely care for him while preferring a different life for him. He wanted to distinguish affection, genetic relationship, shared purpose, and simply enjoying one another's company.⁷

The Castle's guest list is consequently larger than one tribe. He looks for different forms of nourishment in different groups, and resists making a single affiliation responsible for the whole human menu.

A friend has more rooms than you have visited

His 2013 essay “I know what she has to offer already” is almost always false concerns a habit he dislikes in conversation. People exchange their basic biographies, find a shared topic, decide they are friends—and then let their curiosity relax. Once the bond exists, important questions can start to feel risky, while the previous two days of errands become the default subject.

His comic protest asks:

how likely is it that in those 26 years, the best I've done happened in the last 2 days?⁸

He wants friends to continue discovering one another's knowledge, experiences, and possibilities. The fact that someone is known for diet does not mean there is no architecture, cosmology, or surprising private expertise available behind that first door. The familiar category can conceal the person.

His recommendation is:

make more friends, engage them more thoroughly, keep the information flow up (in quality, importance, and quantity).⁸

This is the intellectual version of the intimacy argument. Do not reduce a living person to the role through which you first met them. Keep making room for more of them to arrive.

Loyalty has a long memory

Diego's attachment to old friends is persistent. Sometimes it persists beyond the relationship itself. His 2020 *Dreams of Loyalty* describes a dream in which old acquaintances continued to occupy a mental category of friendship despite distance or disconnection.

The dream supplied him with a filing-system metaphor:

if enough time passes, it moves all "people" files into the "friend" folder even when there's little to no reason to do.⁹

He interpreted this as an instruction to attend to the people available in his new life, rather than bringing every old attachment into the new city. The dream's encounters are a dream; the lesson he drew from them is his own account of a real difficulty.

The line he wanted to internalize was:

It is ok to let go of those who were once friends⁹

That belongs alongside his desire for permanence. His ideal is a durable bond between people who remain present for one another. He also recognizes that loyalty can become loyalty to an obsolete relationship, consuming the space in which a new one might grow.

The tenderness is visible precisely because letting go is difficult for him. A vanished friendship does not become insignificant simply because the practical advice is to move forward.

Please stop moving away

The later writing repeatedly returns to the contrast between the dense friendship of youth and the more scheduled, geographically unstable friendship of adulthood. In *Do I miss my old friends?*, he describes the shock of going from intense, readily available bonds to encounters that are partial, less physically affectionate, and liable to expire when someone changes cities.

His review of the replacement product is:

What the heck is this Temu friendship?¹

He also knows the problem cannot be solved by selecting an address from an old photograph. After a 2025 Bay Area visit, he wrote:

Berkeley is not a place it's a people. And the people that represent the Berkeley i lived in are gone.¹⁰

At that moment, the realization freed him to commit to his life in San Diego. By his January 2026 resolutions, however, he was again considering where durable community might be possible. He wanted a place with intelligent people, stable bonds, and tolerable weather. These are not independent amenities for him. They must coexist in the same life.

He was also explicit that his own withdrawal needed to change:

I need to go back into society despite being able to afford ignoring it altogether.¹¹

The desired return included shared sports, travel, learning, and perhaps work in a group setting. He was looking for recurring activity to do what adolescent proximity had once done automatically: put people in one another's path often enough for a real social world to form.

The same practical concern appears in a modest 2022 question:

How can I be a better friend at 35?¹²

He asked readers for examples of things their friends had done that made them feel cared for. The philosopher of enormous futures was requesting instructions for the smaller, stubbornly essential business of being good to someone nearby.

That is who gets to be family here: people who become part of one another's lived world, whose minds stay interesting, whose presence is welcome, and whose affection has somewhere to go. The third door is still a project. Meanwhile, there is room on the couch.

Further doors

- A Personal take on the Friends Reunion (2021): the desire for durable chosen family and the search for a third door.
- Intimacy (2022) and Social Connection Prompting as Time Dilation (2023): bodily closeness, nonsexual intimacy, and deliberately created connection.
- “I know what she has to offer already” is almost always false (2013): continuing curiosity after friendship has formed.
- Dreams of Loyalty (2020): attachment, memory, and learning to release an obsolete bond.
- Do I miss my old friends? (2025), Berkeley is a people, not a place (2025), and Diego First (2026): migration, adulthood, and the ongoing search for a social home.

1. Diego Caleiro, “Do I miss my old friends?” The Colossus, 15 November 2025, record fb-pdca5e387a82b. ↩↩
2. Diego Caleiro, “A Personal take on the Friends Reunion,” The Colossus, 27 May 2021, record fb-p70fa65bfca23. His opening discussion attributes the show's family-of-friends premise to its writer; the two passages quoted in this room are Diego's own reflections. ↩↩
3. Diego Caleiro, “69 - Intimacy,” The Colossus, 10 December 2022, record fb-p6b458e808e0b. ↩↩
4. Diego Caleiro, “58 - Social Connection Prompting as Time Dilation: Lessons from Scandihippies, Burners and Teathrical Brazilians,” The Colossus, 8 January 2023, record fb-p93941fc7e181. ↩
5. Diego Caleiro, “Happiness and Productivity. Living Alone. Living with Friends. Living with Family.”, 19 November 2013, record lw-6jLFTDPF2GEft2jJt. ↩
6. Diego Caleiro, “Should I Move Back to the Bay Area?” The Colossus, 5 June 2024, record fb-pd3cfd53737cb. ↩

7. Diego Caleiro, "#47 Share your goals with who shares your goals", 15 August 2013, record wp-52725596-47. ↵
8. Diego Caleiro, "'I know what she has to offer already' is almost always false", 10 April 2013, record lw-T2F6Rw5MgLmgQ2bqN. ↵↵
9. Diego Caleiro, "Dreams of Loyalty," The Colossus, 27 September 2020, record fb-pcc7b75e7d8cf. ↵↵
10. Diego Caleiro, "Berkeley is a people, not a place," The Colossus, 22 October 2025, record fb-p3d60775eb65c. ↵
11. Diego Caleiro, "Diego First: Thirteen Selfish Goals for 2026," The Colossus, 1 January 2026, record fb-p2f07ad56db12. ↵
12. Diego Caleiro, "How can I be a better friend at 35?" The Colossus, 28 April 2022, record fb-p8c26c3e5dec0. ↵

Beauty Is Part of the Point

Diego's worldview on aesthetics, music, and enchantment

Before we get to civilization, consciousness, or the eventual fate of the universe, please notice the flute. There may be several. Nearby are animals in several states of biological participation: alive in the landscape, represented in plush, printed on a Magic card, or serving a surprisingly affectionate second career as a fox pelt. This is not a waiting room for Diego's real interests. We have arrived at some of them.

Beauty belongs among the things he wants intelligence, wealth, freedom, and survival to make possible. He enjoys explaining an experience through evolution or cognition, but explanation does not exhaust the experience. Sometimes he wants to know why something is beautiful. Sometimes he wants to sit inside its beauty. Frequently, he wants both, and will interrupt one to tell you about the other.

A bench facing the wrong direction

In 2013, Diego began an essay with an objection to urban design: the bench on which he was writing faced away from the better view. The resulting scene is a miniature of his aesthetic life. He had been cycling in São Paulo's Ibirapuera park. There were trees, children, birds, sunlight, and another cyclist playing a flute:

The sun penetrates my skin to perfection, a bonfire sensation entangled with the cold breeze that strides by.¹

And then:

To my left, another lone bike-rider sat down to play his flute, trying to catch up with the birds singing.¹

The body is present in this writing. Warmth is not merely a fact about the weather; it is a sensation with texture. The musician is not background audio; he is entering a conversation with birds who have not agreed to the rehearsal schedule. Diego's

philosophical machinery is there too, but for once reality has ambushed the machinery:

Instead of being given meaning, symbols, and words to process, instead of the math, of the map, I got reality.¹

That sentence matters in a collection full of maps. The models are meant to lead somewhere. A sufficiently successful account of a good life eventually has to deliver the sunlight, the friend, the music, the particular afternoon. He wonders in the same passage whether he is living in the moment enough. Even a person determined to escape conventional busyness can miss the field directly behind his bench.

The shopping list that follows is revealingly small: a computer, screens, outdoor equipment, and one final item whose rationale survives all further interrogation:

And a unicycle, because... unicycle.¹

The justification committee may adjourn.

Patterns that let a brain breathe

An older essay, written after traveling in Syria, Jordan, and Lebanon in 2010, turns Islamic ornament into a question about what different forms of visual experience ask a mind to do. Diego is attracted to patterned surfaces that are rich without being crowded with human dramas. His entry point is the restriction on depicting creatures, which he treats as a constraint that helped produce an extraordinary visual vocabulary:

Só what happens when you create a society that is not allowed to draw that which we are cognitively most likely to want to depict? Patterns emerge, literally, patterns emerge.²

He contrasts those patterns with images filled with faces, conflict, social relationships, and gestures. His account of their appeal is that the brain can take pleasure in complexity without constantly having to work out who wants what from

whom. Simple generative rules can produce elaborate surfaces. The eye gets a feast; the internal social secretary gets the afternoon off.

This is a characteristic Diego move: an aesthetic preference becomes an inquiry into the creature having it. He wants to understand the relation among form, perception, attention, and an evolved nervous system. Yet the practical destination of the essay is preservation of a pleasure, not reverence for the prohibition that helped produce it:

Sometimes, prohibitions may lead to beauty, let us understand why is the beauty there so it doesn't get destroyed once the prohibitions are suspended, as they surely will.²

The distinction is essential to his aesthetic outlook. He can be deeply hostile to a system's restrictions while wanting the forms of excellence it produced to survive. The palace does not have to be demolished when someone unlocks its doors.

Culture may include an orangutan

Diego's tastes span opera, philosophy, dance, natural history, handcrafted instruments, comedy, and objects that would be entirely comfortable in a child's bedroom. He does not experience the last category as a disqualification from the first.

In November 2024, trying to find a celebratory purchase, he described an already well-populated household:

I could buy plushies, but I already have 20 plushies, from the little bitcoiner and el presidente to komodo dragons, snow leopards and orangutans.³

He also had flutes, foxes, rollerblades, and a unicycle waiting to have its tire inflated. The park essay's speculative unicycle had become real. Its relationship with motion remained aspirational.

The relevant division in his tastes is not between serious adult objects and unserious juvenile ones. It is between things that engage him and things that do not. An animal can be beautiful, funny, evolutionarily fascinating, and pleasant to hold in a soft synthetic approximation. Those pleasures need not compete for permission.

His May 2026 impressions of China include buildings, scale, clothing, national character, and a delighted inventory of cuteness:

Shoes, plushies statues, hair, hats everything is infinitely cute. Neoteny in abundance.⁴

The anthropologist's term and the enthusiast's reaction arrive together. He recognizes a pattern, names it, and enjoys standing in its middle. His own collection had already made him an unusually receptive visiting delegation.

This range leaves ample room for high culture. In his plans for 2026, he wanted more museums, concerts, and art, noting that São Paulo had accustomed him to a scale of cultural availability he missed in San Diego:

The obvious solution is spending a month in Manhattan consuming the infinite culture available there.⁵

There is a recurring expansive appetite here. He wants more kinds of beauty in his life, rather than a perfectly defended identity as the sort of person who likes exactly one respectable class of objects.

The listener gets lungs

Music became especially important when Diego acquired a way of making it himself. In December 2024, he explained why learning flutes mattered beyond obtaining another attractive object:

It's great to be able to express myself with sound, which I've always wanted to do but was restricted by a low singing voice with a narrow range, and a little bit of whistling.⁶

He describes playing flute and didgeridoo as expressive, meditative, and relaxing. The instrument supplies access to a part of experience he had wanted for a long time. Listening to music and producing it through breath are different ways of inhabiting the same aesthetic world.

The road into this apparently timeless woodland hobby ran partly through TikTok. When his account was banned in 2025, his complaint included:

If it wasn't for Tiktok I would not have picked up flute playing.⁷

He valued the platform's ability to introduce genuinely new interests. The route to a handmade flute need not begin with a wise elder in a forest. It can begin with a recommendation system between two ridiculous videos. Diego cares that the door opened.

Dance provides another kind of bodily entry. His 2024 Thanksgiving list called ecstatic dance:

a pillar of my life for 10 years now, it still gives me what to do in cities I've never been before, and are almost guaranteed to have almost only cool people.⁸

Music, movement, and human company form a portable way of finding himself in a new place. The experience is beautiful partly because he participates in it. He need not remain in the audience while other people perform being alive.

Keep discovering colors

The continuing acquisition of tastes also tells Diego something about his future. In a 2023 comment, he wrote:

I've always been concerned that I would not acquire new habits or tastes as I aged. Yet there's several things that I've only started liking recently, for example having a collection of plushies, or creating AI drawings.⁹

He takes encouragement from finding himself capable of new pleasures. The flutes, the images, the animals, and the unfamiliar landscape are evidence in a personal question: can there keep being more life ahead, in forms he has not already exhausted?

His answer is an invitation to keep opening the catalogue. Beauty is something to receive, investigate, make, and arrange the circumstances to encounter again. A

month with more art is worth planning. A body capable of dancing is worth having. New tools are welcome when they extend expression or reveal another taste.

And when civilization's future becomes the subject again, these are among the things that future is supposed to contain. Somewhere in the elaborate plans for surviving the universe, there must still be room for an afternoon whose best justification is that it feels wonderful to be there.

Further doors

- "The Unforbidden Beauty of Islamic Aesthetics" (2010): patterned beauty and the mind that enjoys it.
- "#46 Living as if Money Were No Object" (2013): the park, the flute, and reality interrupting the map.
- "2024 was the least Year year of my life" (2024): music, novelty, and the texture of a year.
- "Diego First: Thirteen Selfish Goals for 2026" (2026): putting art back into the plan.

1. Diego Caleiro, "[#46 Living as if Money Were No Object](#)", September 1, 2013, record [wp-52725596-56](#). ↩↩↩↩
2. Diego Caleiro, "[The Unforbidden Beauty of Islamic Aesthetics](#)", July 23, 2010, record [wp-1878008-495](#). ↩↩
3. Diego Caleiro, "Running Out of Things Worth Buying," The Colossus, November 7, 2024, record [fb-pcdc4f5c8ab38](#). ↩
4. Diego Caleiro, "A Month of First Impressions from China," The Colossus, May 27, 2026, record [fb-pc544087a8818](#). ↩
5. Diego Caleiro, "Diego First: Thirteen Selfish Goals for 2026," The Colossus, January 1, 2026, record [fb-p2f07ad56db12](#). ↩
6. Diego Caleiro, "2024 was the least Year year of my life," The Colossus, December 19, 2024, record [fb-p0e33510aaa09](#). ↩
7. Diego Caleiro, "TikTok Banned My Account," The Colossus, November 2, 2025, record [fb-p5d2e1aaca404](#). ↩
8. Diego Caleiro, "Things I'm thankful for," The Colossus, November 28, 2024, record [fb-pd9d0f1c1d04b](#). ↩
9. Diego Caleiro, comment beginning "I'll start," The Colossus, November 25, 2023, record [fb-cf7fa1552f711](#). ↩

Play Is Serious Business

Diego's worldview on games, hobbies, and mastery

There is a version of adulthood in which interests are gradually surrendered until only work, maintenance, and opinions about the parking remain. Diego would like to file an objection. Preferably after this game, because the board position has become interesting.

Play is one of the ways he inhabits a life he wants to keep. It provides pleasure, absorbing complexity, social contact, skill, and continuity with earlier versions of himself. Some of it has taught him things useful elsewhere. Much of it needs no such defense. The point of getting to have a future includes being able to play in it.

The cardboard has a history

Magic: The Gathering is unusually central. It has been a hobby for decades, a world of remembered encounters and accumulated knowledge, and a way of bringing an enormous appetite for complexity down to the size of a table. In July 2026, Diego described more than thirty years of play and seventy-eight decks, along with meetings with players, collectors, creators, and a world champion he remembers tying.¹

He enjoys the scale of the collection, and he enjoys being astonished when someone else's is larger. The same post introduces a New Zealand collector with:

There is always a bigger fish.¹

The comparison escalates, in his telling, until it requires a megalodon. It is a highly specialized aquarium.

But the emotional meaning of the cards is clearest in a much quieter post. In August 2020, Diego opened the portion of his collection that had survived a theft years earlier. He had avoided it because seeing the survivors also meant confronting what

was gone. Reopening the box brought back cards and decks he had grown attached to:

Magic was how I learned about markets, trade, and many other things.

It was more important in my life than all other games combined.²

The post begins with a precautionary description of itself:

Cw: weird, ridiculous, emotional, pointless perhaps. But honest.²

This is part of what a portrait must preserve. Someone who writes about minds surviving physical destruction can also grieve a lost collection. The cards carry particulars: this deck, that trade, years of play, the texture of an earlier life. Replaceability in a marketplace does not erase the experience of having built something over time.

There is no need to pretend these are only investments. He says he played intensely for fun, and the post's longing is for restoration of a beloved world. Money enters as a means of bringing the cards back. The attachment came first.

A game can keep yesterday available

By April 2026, Diego was explicitly describing Magic as a way of maintaining continuity during rapid change:

I use magic as an anchor when something new is changing fast because magic is something I've been doing for 30 years.³

The surrounding post concerns his immersion in new biological technologies. The contrast is between an unfamiliar future occupying his attention during the day and a familiar game receiving him again at night. He can investigate something disruptive without every part of his experience becoming new at once.

That role sits comfortably beside his appetite for novelty. He does not understand a new interest as an eviction notice for an old one. In a 2025 comment correcting someone's picture of him, he says:

I don't switch hobbies i accumulate them.⁴

The list that follows includes philosophy, anthropology, ecstatic dance, Burning Man, plushies, and flute playing. Different pursuits come forward at different times, but the repertoire grows. The metaphor is less a television channel being changed than a house acquiring rooms. If he is currently playing a flute, the Magic table has not ceased to exist.

This is also why a future full of new possibilities appeals to him. He can imagine acquiring additional modes of enjoyment without wishing to discard every old one. Progress need not require amnesia about what was already good.

Intelligence needs somewhere enjoyable to go

Games are an excellent fit for a mind that wants complicated things to manipulate. They offer combinations to discover, decisions to make, mistakes to notice, and visible improvements in skill. Unlike certain cosmic problems, the next turn usually arrives in time for dinner.

In his 2022 reflection on progress, Diego wondered what could supply a continuing sense of advancement when familiar life goals no longer seemed to offer it. He singled out:

Tennis, ping pong, volley, surfing, Magic, and any other standardized game, are areas where I can improve.⁵

This does not turn every hobby into an achievement contest. It explains one of their pleasures. The scale is human: you can get better at something, experience the difference, and then play again. A life whose largest projects concern enormous numbers can still contain domains where progress is local and satisfying.

Magic also taught him to think about value beyond a thing's immediate function. In a 2021 comment, he recalls expecting differently priced editions of functionally identical cards to converge in value. They did not:

I was wrong, the collectable aspect and rarity makes cards far more valuable.⁶

He then brings the lesson into a discussion of money and the role of historical associations. The game has become part of how he understands the wider world. But he did not need to discover this application before being allowed to enjoy the cards.

Even without them physically present, he can entertain himself with the mental world they open. A short 2022 post on the benefits of being a nerd offers slime molds, army ants, tectonic plates, and then a final fallback:

If none of that works, you can always create a Magic the Gathering deck by heart.⁷

The airport may have delayed the flight. It has not confiscated the internal deck builder.

Please bring another player

Much of Diego's play is social and physical. He likes tennis, table tennis, Ultimate Frisbee, dance, and activities that get him moving with other people. In September 2020, he issued a neighborhood invitation with the title "Games Games Games!" His self-ratings ranged from confidence at table tennis to uncertainty at basketball, and the invitation ended:

Reply here or message me if you're down for one of those! I'm tired of spending too much time at home!⁸

He is willing to join some games without knowing whether he will be any good. He wants the activity and the encounter. A court or table supplies a reason to get together, a structure for interaction, and something to do besides negotiating whether people should ever make plans.

He also wants these forms of participation to continue across age. In 2019, writing about older people at dance events who could still outdance him, he stated simply:

I want to be one of these people.⁹

The aspiration is to retain access to the activity itself: to remain capable of joining the dance rather than being permanently promoted to the chair beside it.

That desire reaches beyond any one sport. He wants a broad repertoire of ways to engage the world, bodily as well as intellectually. The person who can spend a long time thinking about a game also wants to leave home and play it with you.

Adulthood does not own the swing

Some of Diego's strongest statements about play concern the rules people internalize about age. In May 2024, he noticed himself judging activities he still wanted to do because he was no longer twenty-one. A swing in a park helped expose the absurdity:

Yesterday, I saw a swing at the park I often walk by, and I remembered that I can, in fact go on the swing. I went there and swung a little bit. It was very liberating.
10

Even someone who had considered himself tolerant of his own Peter Pan qualities had been quietly narrowing his permission to enjoy things. He names Nintendo, Magic, jet skis, Mario Kart, and South Park. None had become intrinsically unappealing. He had begun asking whether he was still supposed to like them.

His answer is emphatic:

Except for physical impossibility, aging should not determine the things I don't allow myself to do. That is completely insane.¹⁰

He wants to remember this at fifty, seventy-nine, and eighty-seven. The specificity makes the wish charmingly practical. Future Diego is being sent a memo: you have permission; do not let the calendar forge your signature on a surrender form.

An earlier comment had compressed the underlying refusal into a sentence:

Maturity is almost always the excuse of a soul for why losing color is ok.¹¹

And in June 2024, reflecting on pursuits he had suppressed while trying to be useful to the world, he wrote:

There's a lot of things I skipped in service of the world which I'm now able to return to.¹²

The following examples were flutes and Magic. The return is not an embarrassed lapse from an otherwise serious life. It is recovery of some of what that life was meant to make possible.

If you leave this room knowing only that Diego has a very large card collection, you will have missed its larger contents. He wants skill without mandatory drudgery, complexity with enjoyment, continuity without stagnation, and adulthood without the compulsory bleaching of curiosity. He wants to keep learning new games. He also reserves the right to remain attached to an old one for an unreasonable amount of time.

Given his preferred lifespan, the definition of unreasonable may need revision.

Further doors

- “Reopening the Magic Collection That Survived a Theft” (2020): the emotional history of the cards.
- “Why Being an Academic Nerd Beats Being Rich” (2022): portable enjoyment through curiosity.
- “Live as If You Didn't Know Your Age” (2024): refusing age-based prohibitions on play.
- “Peptides, Magic, and the End of Obesity” (2026), opening paragraphs: Magic as an anchor across change.

1. Diego Caleiro, “A New Zealand Megalodon of Magic Collections,” The Colossus, July 7, 2026, record fb-p0a4f54374661. ↵↵

2. Diego Caleiro, “Reopening the Magic Collection That Survived a Theft,” The Colossus, August 2, 2020, record fb-p3acb4a41e916. ↵↵

3. Diego Caleiro, “Peptides, Magic, and the End of Obesity,” The Colossus, April 7, 2026, record fb-pe13dcbc4785. ↵

4. Diego Caleiro, comment on accumulating hobbies, The Colossus, September 6, 2025, record fb-c581392bc1954. ↵

5. Diego Caleiro, "Progress, Big Numbers, Numbness," The Colossus, January 19, 2022, record fb-p365dfdad3786. [↵](#)
6. Diego Caleiro, comment beginning "As an example of history mattering in Magic," The Colossus, May 11, 2021, record fb-cf6f48f046770. [↵](#)
7. Diego Caleiro, "Why Being an Academic Nerd Beats Being Rich," The Colossus, February 26, 2022, record fb-p81f1a17e6721. [↵](#)
8. Diego Caleiro, "Games Games Games!," The Colossus, September 17, 2020, record fb-pbe7ad10c98e3. [↵](#)
9. Diego Caleiro, "Doubts about Frisbee, Contact Improv and Group Approach Anxiety?," The Colossus, September 6, 2019, record fb-pb78dc037f5a3. [↵](#)
10. Diego Caleiro, "Live as If You Didn't Know Your Age," The Colossus, May 31, 2024, record fb-p12ab2b8bd849. [↵↵](#)
11. Diego Caleiro, comment beginning "Didn't find a way to be as social," The Colossus, June 2, 2023, record fb-c205dd7885712. [↵](#)
12. Diego Caleiro, "Philosophy, Leisure, and Early Entry into Deep Utopia," The Colossus, June 22, 2024, record fb-pd479a4093b38. [↵](#)

Go Somewhere Strange

Diego's worldview on travel, adventure, and Burning Man

Diego's favorite places are not merely attractive backgrounds for being the same person somewhere else. They change what he notices, whom he meets, what he wants, and how much life seems to fit inside a month. Travel is pleasure, social opportunity, an encounter with unfamiliar forms of life, and sometimes a way of releasing his mind from its own relentless management.

The itinerary may look peculiar. An important philosopher, a national park, a dance gathering, and a place with promising animals can all exert serious gravitational force. They are not interchangeable attractions. They offer different ways for the world to become larger.

Geography changes the available life

As early as 2012, Diego was encouraging other rationalists to consider where they lived in terms broader than institutions and employment. His portrait of Rio de Janeiro praised the combination of natural beauty, accessible people, and ordinary convenience:

Forests and beaches and mountains abound, all 4 minutes away from a supermarket.¹

This is a particularly Diego form of paradise: the landscape is magnificent, but obtaining dinner need not become an expedition. He values cities where encounters and outdoor activities are readily available, where visitors and newcomers are also looking for friends, and where a pleasant evening need not first defeat the climate.

In a 2013 essay written from Oxford, he defined one sense of individual freedom as:

amount of things per unit time someone could be doing²

That definition makes place morally and personally interesting. It is not only a container for a salary. Different places permit different repertoires of action. Cold cities, in his account, channel people toward indoor and intellectual activity; sunnier places allow more spontaneous outdoor life. He was weighing both sides because he wanted serious intellectual immersion and the freedom to live bodily, socially, and outdoors.

The result is not a universal ranking that every traveler should copy. It is an insistence on asking what a location will actually let this particular person do. A city can be prestigious and wrong for the life you want that year. A park can offer something no amount of institutional prestige supplies.

Other people demonstrate possible lives

Travel also loosened Diego's sense of which life arrangements were mandatory. In his 2014 writing, couchsurfers and long-term travelers appear as practical demonstrations. He met people who worked part of the year and traveled for the rest, people who left graduate study, people traveling with little money, and people who built lives around movement rather than a conventional career calendar.

He described the effect of meeting them this way:

You'll feel much less pressure to make money after meeting similar people and being one of them.³

The point is not that everyone must become a vagabond. It is that a possible life becomes easier to imagine when a human being is standing in front of you already living it. Diego wants to discover alternatives at close range. He does not want the menu of existence restricted to whatever the people in his immediate environment happen to consider normal.

His own travel changed as his resources changed. The 2024 road-trip account explicitly contrasts earlier couchsurfing, trains, and supermarket food with later hotels and steakhouses. The pleasure of moving through the world precedes the more expensive version. Money expands comfort and ease; it does not originate the interest.

He is equally clear about one preference that money alone cannot buy:

traveling alone is the bomb and there simply is no way to beat it because it forces you to meet the best (and worst) people.⁴

Solitude in transit can become a route into company. It removes the ready-made social unit in which a traveler might otherwise stay enclosed. Diego likes being exposed to the possibility of an encounter he could not have arranged from home.

The natural-history department has escaped the building

In September 2024, he returned from thirty-five days and thousands of miles across the American West. His account contains volcanic parks, lakes, mountains, bears, wolves, elk, birds, and enough Yellowstone references to suggest the park had successfully installed itself as a default answer:

In north America Yellowstone is the park to rule them all.⁴

For Diego, animals are not generic scenery. The list distinguishes species and the difficulty of seeing them. His scientific curiosity and enjoyment of living creatures travel together. Geological variety also matters: different landscapes make different experiences possible. The itinerary can be read partly as a tour through forms of matter and life that he wanted to encounter in person.

Human variety is just as prominent. He writes about meeting glassblowers, dancers, taxidermists, a bat specialist, and a couple celebrating a long marriage. The trip is made of particular conversations, not just views checked off a list.

Montana earned its own affectionate post. He liked people whose lives and presentation differed dramatically from his urban, academic background. One acquaintance was a burner, but with an important regional modification:

Not a California burner, a Montana burner. All the dust with extra leather.⁵

Another meeting became an expansive conversation about bats, echolocation, flight, altruism, and evolutionary biology. The subjects arrived with the person. This is one way travel feeds a mind that likes connections: go somewhere expecting mountains, come back having spent the evening discussing what bats are doing with their social lives.

He closes the Montana post by joking that the people he has met constitute a substantial fraction of the state's population. The bartender recognized him from a visit the previous year. The combination matters: strangeness and welcome can coexist. A place need not resemble his background to feel generous to his presence.

A country can rearrange the map

Diego's 2026 trip to China illustrates how travel can change the territory that attracts him. In Hong Kong, he enjoyed the food, buildings, people, and culture. His fellow Brazilian travelers preferred Macao; he registered the disagreement with his usual restraint:

Objectively Hong Kong is way better. If you define objectively as according to Diego, master of the universe.⁶

The joke leaves the reader in no doubt whose travel guide this is.

After returning home, he found himself watching more China travel videos and responding warmly to Mandarin in an Asian market. The country had become unfinished business. Its scale made it obvious that his month had revealed only a fraction of what remained to see. An area of the world that previously held relatively little attraction now drew him toward Taiwan, South Korea, Japan, and further travel in China:

A complete 180 in my interest in that general area of the Earth.⁷

He liked what he had seen enough to want the unseen parts. Curiosity had multiplied rather than been discharged. Travel, in this mode, does not consume a destination and remove it from the list. It generates more possible journeys.

When he asked himself whether he wanted to reconnect locally or head out again, his mind supplied an unusually concise strategic report:

adventure time!⁷

One suspects the committee meeting was short.

The place that says welcome home

Burning Man occupies a special category. It is an adventure Diego returns to because it delivers an unusually concentrated combination of art, play, strangers, physical experience, and a social world in which he feels at home.

His recollection of arriving for the first time is entirely on brand. He had deliberately learned little beyond what he needed to bring. A stranger answered a question, welcomed him home, put his head into the car “like a safari dromedary,” and licked the side of his face. Diego’s conclusion was:

That's when I knew I had finally found my place in the World.⁸

Other people might have concluded that the reception desk needed retraining. Diego had discovered a fit between his temperament and a temporary city.

He repeatedly resists the idea that the experience belongs only to one age group or type of partygoer. His 2024 invitation mentions older art and photography enthusiasts, children he had befriended, early sleepers, sculpture viewers, and people pursuing quite different nocturnal schedules. In the same post, after recounting a wide range of places and people encountered in his life, he concludes:

of all the things in the world, Burning Man is the best one.⁹

The enthusiasm did not belong only to his first visit. After his ninth burn in 2025, he wrote:

Nowhere else my soul feels more at home, more deeply satiated, more open and more earnest than at burning man.¹⁰

The religious and sacred dimensions have their own room in this collection. Here, the practical point is that this journey repeatedly puts him into a state of being that he treasures. Going back can be as meaningful as going somewhere new. Novelty matters, but so does knowing where a beloved kind of life reliably becomes available.

A month that is actually a month

The return from a good journey gives Diego a special comparison with ordinary time. At the end of 2024, he felt much of his year had passed in a blur, while months containing travel, new people, and animals had been distinctly full. The criterion is lived richness: how much happened that he can feel and remember?

His American West account describes another benefit:

Above all, traveling shuts off a lot of the problem seeking and worrying systems of my brain. It allows me to live care free and without long term plans for weeks on end, and without concern that I may be doing the wrong thing.⁴

This is an important release for someone capable of turning virtually any subject into a problem requiring an essay. For a while, being there answers the question of what he should be doing.

Travel is therefore one of the ways Diego converts the opportunity to be alive into actual life. He wants unfamiliar landscapes, surprising people, beautiful creatures, and periods when attention can follow what is happening in front of him. He wants room for a spontaneous turn, a new fascination, or an unexpectedly absorbing conversation.

His own closing instruction to readers of that road-trip account is a fitting exit from this room:

And try to find places to play and dance along the way.⁴

Further doors

- “Good Places to Live as a Less Wronger” (2012): the relationship between place and possible life.
- “How Not to Make Money” (2014): travelers as demonstrations of unconventional options.
- “Thirty-Five Days Across the American West” (2024): animals, people, and the freedom of the road.
- “China Is Hard to Leave” (2026): a journey creating a new field of curiosity.

1. Diego Caleiro, [“Good Places to Live as a Less Wronger”](#), November 20, 2012, record [lw-LdQEuSNSf34H3xNZg](#). ↩
2. Diego Caleiro, [“Is The Blood Thicker Near The Tropics? Trade-Offs Of Living In The Cold”](#), March 28, 2013, record [lw-bN9ZKXP89aGhoL4fY](#). ↩
3. Diego Caleiro, [“How Not to Make Money”](#), January 24, 2014, record [lw-N2BbX4aCTM7FTmPyT](#). ↩
4. Diego Caleiro, [“Thirty-Five Days Across the American West,”](#) The Colossus, September 30, 2024, record [fb-p6d1e207ab8f9](#). ↩↩↩↩
5. Diego Caleiro, [“In Praise of Montana,”](#) The Colossus, September 20, 2024, record [fb-p834cdcd8abab](#). ↩
6. Diego Caleiro, [“Hong Kong and Macao,”](#) The Colossus, May 10, 2026, record [fb-pcd0289be88a7](#). ↩
7. Diego Caleiro, [“China Is Hard to Leave,”](#) The Colossus, June 1, 2026, record [fb-p3eabd3311eff](#). ↩↩
8. Diego Caleiro, comment recalling his first arrival at Burning Man, The Colossus, July 9, 2022, record [fb-c8bf1b6710d15](#). ↩
9. Diego Caleiro, [“Why This Could Be the Best Year to Try Burning Man,”](#) The Colossus, August 16, 2024, record [fb-p7848854cdf2c](#). ↩
10. Diego Caleiro, comment reflecting on his ninth Burning Man, The Colossus, September 2, 2025, record [fb-c22589078bc36](#). ↩

How to Spend a Finite Diego

Diego's worldview on agency, time, freedom, and experiments in living

Diego has spent years trying to solve a problem that sounds simple until you hand it to an actual human being: how do you get a life to do what you want it to do?

Knowing what would be valuable is one part. Starting is another. Keeping going, choosing the right surroundings, having the right people nearby, noticing that your goals have changed, and remembering to enjoy the whole enterprise require further negotiations. His writing on agency is a record of those negotiations: sometimes numerical, sometimes environmental, sometimes a manifesto against his own previous method.

The unit being allocated is not merely a work hour. It is an hour of an irreplaceable life. In a 2016 essay about money, happiness, and remaining time, he puts the real deadline plainly:

The deadline is when you stop existing.¹

The calendar suddenly looks less authoritative. Its appointments remain, but they have acquired a supervisor.

A job is one possible mechanism

One of Diego's longstanding objections is to treating conventional employment as an end rather than a means. In 2014, he described how rarely the possibility of not having a normal job seemed to enter some Americans' calculations. His own experiences with study, travel, intellectual communities, and unconventional living arrangements made the omission striking to him:

A job is truly an instrumental goal, and your terminal goals certainly do have chains of causation leading to them that do not contain a job for 330 days a year.²

He does not mean nobody should work. In the same essay, he enthusiastically wishes well to people who love the activity or use it to support their goals. His objection concerns confusing the instrument with the life it is meant to serve.

Freedom, in this picture, includes being able to ask what one would actually choose to do. It can make room for travel, study, invention, relationships, and projects that do not arrive with a salary attached. The question is not whether an activity resembles a respectable career. It is what that activity does for the things you value.

This also means financial success does not finish the problem. It can remove an obstacle while leaving an entire day in need of direction. A person can possess options and still need a method for turning them into lived events.

The philosopher puts himself in the experiment

In 2013, Diego began a series of fifty personal challenges, inviting readers both to participate and to choose some of his experiments. The opening post deliberately tried to move the reader from enjoying ideas to doing something:

This is a place to do.³

There were monthly commitments, updates, reader suggestions, and the possibility of making Diego attempt something amusingly demanding. He proposed writing regularly, learning how familiar things worked, and speaking to strangers about subjects that actually mattered to him. The project's spirit was practical and social: create a situation in which the intention has somewhere to go.

The subsequent reports show him changing surroundings and incentives. A task might become easier if its first action required one click. A tempting distraction might become less automatic if the relevant device stayed unplugged. Moving closer to a community could change what kinds of action felt normal and available.

When applications were piling up, he hired a virtual assistant. This was already more than a decade before conversational AI became his everyday collaborator. The assistant helped search for opportunities, gave him someone to compete with, and

could move apparent deadlines earlier at his request. His delight at the arrangement included:

I'm protected by my ignorance, and will be ahead of time for sure.⁴

He was willing to design circumstances for a future self whose habits he knew. Agency need not consist of issuing a stern order to the brain and expecting obedience. It can mean arranging the next encounter between the brain and the world.

The relevant arrangement may be wonderfully idiosyncratic. A 2013 LessWrong post records his discovery that an online study hall helped him particularly when people with long hair were present:

I am motivated by the presence of people whose hair goes to their shoulders. Women or Men.⁵

He had initially expected a different explanation. The point of the story is his willingness to notice the odd actual condition rather than insist on a tidier theory of himself. Some people need silence, some need company, and apparently one branch of productivity research requires a conditioner budget.

Give the hours numbers

By 2015, Diego was experimenting with a to-do list ranked by estimated value per hour. Instead of sorting by arrival date or merely marking tasks urgent, he assigned a number to the value he expected an hour of activity to produce for him. The number could be revised. Its purpose was to make ordering easier and reduce the need to renegotiate the whole list whenever a new item appeared.⁶

The method also let him put low-priority tasks somewhere outside his immediate attention. Recording a task did not promise that he would eventually do it. He found relief in allowing some things simply to remain below the cutoff.

He then made a stronger experiment: while working, actually do the highest-value available item, rather than merely keep the list as an impressive philosophical object.

Fifteen days later, he reported imperfect results and tried again. His account distinguishes owning a system from behaving according to it.⁷

The categories were broader than business productivity. In 2016 he explained:

I don't distinguish work from play in my to do list. Pit ball disco party and finish tax form are both there, side by side.¹

This is a very important neighboring pair. The tax form is not automatically more valuable because it looks unpleasant. The disco party is not automatically less valuable because it looks like a disco party. Both have to answer to the life they are helping produce.

He also made explicit space for exploration: periods when the problem was the overall approach, not failure to execute the next task. Sometimes the plan itself needed pruning. Getting more things done is useful only if those things remain connected to what he wants.

The frog loses its office

His later writing does not preserve every earlier method unchanged. In December 2024, the man who had once challenged himself to start with the highest-value item published "Eat your frog is wrong." The crucial obstacle, he now insisted, was initiation:

THE HARDEST THING IS STARTING⁸

The proposed starting rule was correspondingly different:

Start by the easiest thing. ok, the most fun thing!⁸

He qualified this within the broader category of things that needed doing. The idea was to enter the activity through its most inviting door, then use the resulting momentum. In his example, organizing a few of his simpler Magic decks expanded into a substantial reorganization of his home. Beginning with something pleasurable had given him a route into work he had not begun by commanding himself to do.

The same year, an essay about MrBeast led him to reconsider another long-running ambition: installing a stable routine. Looking back at years of trying, he wrote:

I don't have a routine, I never had a routine, and I never will have a routine.⁹

His proposed alternative was to work with the intense absorption he recognized in himself:

Follow your passion and keep at it, like a maniac, and nudge it towards the healthiest version of the thing you can be passionate about.⁹

This is Diego describing the terms on which his own mind most readily engages. He wanted permission to immerse himself in a pursuit, learn deeply, and later turn toward another one. He calls for healthier directions for that intensity rather than demanding that it become an entirely different temperament.

The underlying ambition remains action. The method shifts from making himself resemble a conventional disciplined adult toward finding a workable route for this particular person. The frog has not won. It has merely lost the right to be the first appointment every morning.

Make a year feel inhabited

Time matters to Diego as duration, but also as experience. At the end of 2024, he complained that the year had felt unusually short. Months rich in travel, new people, and animals remained distinct; much of the rest blurred. His resolution was:

Do more of whatever causes my subjective perception of time to be positive and slower.¹⁰

This enlarges the practical question. A good allocation of time is not only one that produces things. It should also produce a life he can feel himself having. Repeated actions can fill a calendar while leaving little distinctiveness in memory. New encounters can make the same measured duration seem more capacious.

He wanted fewer repetitions and more new experiences. For someone interested in living an extremely long time, this is an immediate companion project: make the years already available less likely to disappear into an undifferentiated blur.

Neither productivity nor leisure is sufficient as a label. The question is what kind of attention, engagement, memory, and enjoyment each period contains. A month can contain few conventional accomplishments and still be unmistakably alive. Another can be full of activity and feel strangely absent.

The staff finally arrives

In August 2026, Diego described a new phase. AI had become useful for reconstructing his intellectual and personal history, organizing scattered work, and revisiting projects that had stalled. The aim was not just a cleaner archive. It was a more deliberate choice about what to carry forward:

I want to consciously decide who I want to be from now on, which parts of past me should be carried forward, which abandoned, which modified.¹¹

He was using the systems to recover ideas, connect material, and help with tasks his own patterns of attention had made difficult. This extended an old interest in external supports into a much more ambitious collaboration. The virtual assistant searching for applications in 2013 had acquired a large and unusually talkative set of successors.

But abundant help created its own scheduling problem:

They create tasks for me faster than I can execute tasks.¹¹

He described himself as the bottleneck. Possibilities were being generated faster than he could select, direct, and act on them. The central problem had changed: some projects were no longer blocked by unavailable labor, yet still needed his decisions and attention. His account is excited, pressured, and full of the sense that familiar limits had moved.

That brings the chapter back to freedom. The point of assistance is not simply that Diego can now produce more. It is that he can try again to accomplish things he

wanted, recover abandoned possibilities, and choose what kind of second half of life to build.

Later that month, in a published account introduced as spoken by Diego and transcribed and partly edited by his AI collaborator, he summarized the purpose of his financial independence:

What I ultimately purchased with money was time.¹²

He describes using that time both for his own varied life and for periods of intense work when he sees an unusually valuable opportunity to help others. The full account of his altruism belongs elsewhere. Here its practical implication is that his working schedule follows opportunities and purposes, not the requirement to keep looking employed.

A finite Diego must continually decide what deserves a portion of him. The answer can include a demanding intellectual project, a strange journey, a dance, a friendship, or an experiment that makes the next difficult action easier. It can also include delegating enough of the mechanics that a once-impossible project finally becomes possible.

He wants the machinery to serve the life. And, when the machinery has been adequately briefed, he wants to be able to go outside.

Further doors

- “#50: Stop Learning, Start Doing” (2013): the beginning of his public experiments in action.
- “How Much Money Does 1 Hour of Happiness Cost?” (2016): valuing time, work, play, and exploration together.
- “Eat your frog is wrong” (2024): starting with the easiest, most enjoyable useful action.
- “Becoming the Bottleneck in an AI-Accelerated Life” (2026): the return of old projects under new capabilities.
- “MY DEFINITION OF WORK” (2026): his dictated account of what working and retiring mean to him.

1. Diego Caleiro, "[How Much Money Does 1 Hour of Happiness Cost?](#)", April 23, 2016, record [wp-52725596-226](#). ↵↵
2. Diego Caleiro, "[On not getting a job as an option](#)", March 11, 2014, record [lw-aKQtuNvawj3Fzmw3X](#). ↵
3. Diego Caleiro, "[#50: Stop Learning, Start Doing](#)", July 8, 2013, record [wp-52725596-3](#). ↵
4. Diego Caleiro, "[#48 Out of The Armchair and Into the Field](#)", August 2, 2013, record [wp-52725596-37](#). ↵
5. Diego Caleiro, "[Noticing something completely absurd about yourself](#)", November 26, 2013, record [lw-Fpu3rrFQ2mjKofAsN](#). ↵
6. Diego Caleiro, "[#38 Follow the to-do list value per hour](#)", June 29, 2015, record [wp-52725596-181](#). ↵
7. Diego Caleiro, "[#37 Try Harder](#)", July 15, 2015, record [wp-52725596-187](#). ↵
8. Diego Caleiro, "Eat your frog is wrong," The Colossus, December 27, 2024, record [fb-p0eac77520718](#). ↵↵
9. Diego Caleiro, "Lessons From Mr Beast Part 2," The Colossus, March 8, 2024, record [fb-p4f40ae8f2a83](#). ↵↵
10. Diego Caleiro, "2024 was the least Year year of my life," The Colossus, December 19, 2024, record [fb-p0e33510aaa09](#). ↵
11. Diego Caleiro, "Becoming the Bottleneck in an AI-Accelerated Life" (opening: "Le Bottleneck C'est Moi"), The Colossus, August 8, 2026, record [fb-pfa92e904f2a5](#). Quotes are from Diego's post, not the separately marked AI comment at its end. ↵↵
12. Diego Caleiro, "MY DEFINITION OF WORK," The Colossus, August 18, 2026, record [fb-pd2a4d762e83b](#). Introduced by him as "Spoken by Diego, transcribed by Shoggoth. 90% Diego. 10% Shoggoth." ↵

The Conscience Has a Special Offer

Diego's worldview on altruism and Effective Altruism

Welcome to the room where the conscience has abandoned its customary robes, acquired a suspicious coat, and begun offering merchandise.

Diego described the arrangement in 2025:

My good conscience is always bribing me like hey, psst, kid, I got some really cheap altruism here, I know you want it....¹

The next sentence explains the business model:

He's like a drug dealer.

For Diego, helping can be a temptation. He notices an opportunity to make somebody's life substantially better at comparatively little cost, and suddenly his conscience has an irresistible offer. His altruism contains affection, arithmetic, ambition, pleasure, and the delighted recognition that the universe occasionally leaves a very large lever somewhere one can actually reach.

This room brings those elements together. We will meet the lives he wants to improve, the moral theory he once escaped, the movement he helped build, and the evolutionary machinery he would like humanity to stop obeying quite so helplessly.

People are entire worlds

Diego's moral attention keeps returning to the inside of a life. In his 2025 essay on Schindler's List, he describes people as:

People, unfolding cascades of mesmerizing depth and experience running for 90 years, worth to themselves almost as much as you are worth to yourself.²

That is what his arithmetic is counting. A saved person gets more mornings, irritation, discoveries, meals, friendships, terrible jokes, and moments whose importance nobody outside that person could have anticipated. The number represents an entire unfolding world.

The film offers him a way of seeing ordinary resources differently. He writes that:

I saw that the world can be utilized as a life-saving machine, and you can trade resources, of all kinds, time, money, etc... for actual people.

A possession becomes both itself and a possible intervention. Money can turn into medicine, prevention, rescue, or the continued existence of somebody who would very much like to remain available for tomorrow.

He wants this perceptual shift to feel exciting. In the same essay, he invites a reader to try helping someone without immediately pledging a lifetime of penitence:

Do one. See how you feel. Then do another.

His interest extends across familiar social boundaries. In a 2019 exchange about whether poor people deserve their circumstances, he insists that “being a moral patient is independant of production ability.” His short formulation deserves a place near the entrance:

If it suffers, it matters.

And:

If it experiences joy, it matters.³

A being’s capacity to accomplish things and its capacity to have a life worth caring about occupy different columns in his mind. He can regard particular people as unusually important to the future while continuing to count the suffering of somebody who will never be influential. Animals also belong within his concern. So do possible future minds whose bodies, intelligence, and forms of enjoyment may be very unlike ours.

The guest list is extensive. The admissions criterion does not require a CV.

The altruist who resigned from morality

Diego's altruism becomes particularly revealing when he tries to leave it.

In a reflection written in 2013 and published alongside a later reconsideration in *The Inescapability of My Altruism*, he describes what happened when he took utilitarian demands with complete seriousness:

Utilitarianism, for me, was a prison made of stress, math, and the end of freedom of choice.⁴

Every activity encountered a counterfactual superior. Travel, philosophy, games, friendship, good food: somewhere there was a calculation under which something else helped more. Even relaxation could become an optimization problem about the minimum rest necessary to resume serving the calculation.

He wanted his pleasures back. He wanted emotions to matter as experiences, rather than requiring each one to submit a productivity report.

The result was not the disappearance of concern:

I want to help other people and beings because that is awesome.

He also wrote:

I want to be kind because kindness makes my life good, and makes others lives good.

By 2016, he had separated two things he had previously allowed to run together: a conviction about what morality commands and his own preference that other minds have good lives. Philosophical criticism could change the first without erasing the second.

Maybe I'm not a utilitarian, and I just want the most minds to be happy.

His thought experiments then pushed the distinction further. If a moral theory and his desire for better lives disagreed, he discovered that he could prefer the better lives. The famous trolley lever survives his departure from philosophical respectability:

But I'd still pull that goddamn lever.

Followed by:

So much the worse for Morality.

This is the inescapability in the title. He can dismiss the supervisor; the desire continues turning up for work.

His later writing again uses utilitarian language and discusses obligations. The history matters. He has not spent every subsequent year repeating a permanent 2013 resignation. What persists is his resistance to giving an abstract demand complete ownership of his life, alongside his strong preference for a world containing more flourishing minds.

In 2022 he puts the practical relationship plainly:

We are thrown in the World and have no ultimate cosmic obligation to be nice. Not my point. My point is, IF YOU CHOOSE to be nice, being nice in an Effective Altruist way is the way to go.⁵

A very nice animal

Diego approaches altruism as a biological anthropologist as well as a participant. He wants to understand how creatures capable of genuine concern came to exist.

His 2013 *From Capuchins to AI's* sequence travels through social learning, culture, cooperation, inheritance, reputation, reciprocity, and the relationship between individuals and groups. Its opening announces several respectable academic purposes, followed by the one he considers decisive:

If it's worth doing, it's worth doing in the way most likely to save the World.⁶

Apparently even a semester essay has strategic responsibilities.

He does not take the selfishness of a replicating element to imply the psychological selfishness of every organism it helps construct. Genes, cultural learning, and social

arrangements can combine to make beings who care. Explaining how they arose does not dissolve their concern.

In the second installment he writes:

Cooperation evolves, and altruism evolves.⁷

His analysis distinguishes several mechanisms through which cooperation can arise and persist. A behavior may help relatives, invite reciprocation, preserve a reputation, satisfy social expectations, or help a group survive. The creature performing it need not experience all of those evolutionary consequences as its personal motive. An animal can want another animal to be okay.

This distinction animates his exuberant writing about people. Watching the team captains in Beast Games refuse a financial offer that would eliminate their teams, he sees a display of loyalty, trust, leadership selection, and human social perception. His conclusion is characteristically celebratory:

We are altruistic, team players by nature, and that's one of the many things that makes our species the coolest one around.⁸

The post ends:

So go humans! Love you guys :)

His view of humanity contains anger, but it also contains this unabashed fondness. He notices people who help, protect, cooperate, and keep faith. The capacity is precious partly because he understands it as a specific achievement of living systems, with a history and conditions of survival.

His dissertation's title, *Altruism: Past, Present, Propagation*, gives the investigation its temporal shape: where caring came from, how it works, and how to keep it going.

The method must be allowed to move

Diego encountered organized Effective Altruism as a natural home for an existing project. He already cared about transhumanism, the far future, and finding unusually

consequential ways to help. His involvement grew through early organizations, conversations, writing, and communal life.

The movement supplied collaborators and machinery. In his conception, the machinery must remain responsive to the task.

His 2015 explanation of EA as an intensional movement is a central piece of the architecture. The technical term concerns a way of finding something. Effective Altruism, in this account, is held together by methods and questions whose present answers may change. A favored cause, institution, or career is one possible answer.

He states the consequence explicitly:

I do not lose my EA identity by changing my core beliefs about what matters or what should be done.⁹

A discovery that redirects effort can therefore express the commitment that brought someone into EA. Disagreement with the present agenda need not place that person outside the project.

Diego wants this flexibility in the institutions too. Which work matters? What happens because of this particular contribution? Could somebody else accomplish it? Is the important obstacle a lack of money, skill, attention, coordination, or willingness to act?

His 2010 manifesto had already made the counterfactual question explicit:

¿What is the difference between our world in which X did what he did, and our world in case X had not done that?¹⁰

The practical implication is personal and demanding. He wants to find the thing that would remain undone without the relevant intervention. Occupying an impressive role and making a large difference are separate accomplishments.

It also changes his sense of timing. In 2022, he objects to treating the supply of valuable opportunities as if it were a uniform weekly allowance:

I fear sometimes EAs want to work in EA and they get all "busy" as if any given week has the same amount of effective opportunity to allocate resources to maximize happiness and minimize suffering as any other week.⁵

He describes his own participation as episodic: years of knowledge, relationships, observation, and occasional decisive openings. He can value being ready without believing that every day needs an equally impressive altruistic timesheet.

Caring needs infrastructure

For Diego, cooperation is something to design, improve, and fund.

In 2014 he proposed that effective cooperation captured an important aspect of what EAs were doing. His distinction was between maximizing others' welfare and maximizing the combined welfare of everyone involved. The latter includes the helper. A good shared outcome need not contain an extra debit merely because one of the beneficiaries happens to be you.¹¹

Moral Economics, the field he proposed in 2015, develops this practical interest. Its opening declares:

Economics for altruists doesn't operate under the same rules as regular economics.¹²

People who care about others' outcomes have additional reasons to coordinate. If two donors value the same improvement, the relevant benefit can occur whichever donor makes it happen. If their priorities differ, there may be exchanges that satisfy both more effectively. His proposed field gathers moral trade, donation coordination, impact recognition, direct funding, and related ideas into a common framework.

The ambition is to give generosity better plumbing. Resources should reach valuable work; compatible efforts should reinforce one another; people should be able to cooperate across differences in ultimate commitments.

His attention to incentives also makes him particular about praise. In a 2016 post he separates the value someone creates from the compensation they receive, criticizing the habit of counting a highly paid role again as if all its work were an

uncompensated gift. This is a dated, unusually stringent formulation of a recurring question: what contribution is actually being rewarded?¹³

Yet demanding work and secure human attachment have different requirements. In *In Praise of Tribes That Pretend to Try*, developing an argument with Viliam, he distinguishes communities organized around accomplishing tasks from communities that provide belonging. A project can require performance; losing a project role need not cost someone their entire social world.¹⁴

His own formulation of the challenge asks:

how to surround effective altruists with welcoming arms even when the subset of them that is active at that moment is not doing the right things?

He thinks people contain multiple motives and pass through different levels of energy and commitment. A movement that needs capable people should also be able to welcome an exhausted person, a temporarily distracted person, or somebody whose interest this evening is a board game.

Even the rescue party needs somewhere to come home.

The mission survives the institution

Diego's loyalty to EA has never required a uniformly pleasant opinion of its social life.

He objects to status competition, exclusion, institutional self-protection, and the replacement of useful inquiry with the business of belonging. He has written with considerable fury about these failures. In the same archive, he repeatedly reaffirms the movement's value and his identification with its purpose.

These are distinct attachments in his account. The existence of people needing help does not depend on an organization remaining admirable.

His 2020 formulation is unusually portable:

The people are helped if you help them. Literally no one can change that. Labels have no power over your actions.¹⁵

He adds:

And you can do it alone. You can do it outside institutions. You can do it many ways.

By 2022, his retrospective again emphasized what he regarded as the movement's achievements and its demonstrated ability to change priorities:

EAs, truly, are able to change their minds about what matters the most. And, in most cases, to begin acting on it.¹⁶

He wants the reader to keep the beneficiaries in view through enthusiasm, disappointment, disagreement, and institutional change. An institution can be useful enough to defend and flawed enough to criticize. If a structure collapses, helping remains available.

The conscience does not require permission from the membership committee to reopen its stall.

How to keep the future capable of caring

Diego's altruistic horizon extends beyond allocating present resources. He wants to change the conditions under which future lives become possible.

This was already explicit in *Altruistic Awesomeness, Your Challenge* in 2010:

So you decided you want to have a life that is 1) Awesome 2) Altruistic.¹⁰

He describes technological improvement and the prevention of catastrophe as what happens when that aspiration continues reasoning:

transhumanism is iterated altruistic awesomeness.

Following the consequences leads him toward the survival of civilization, the development of intelligence, the control of destructive incentives, and the possibility of extraordinary future experience.

His 2015 writing plans make the adversary explicit. He wants to attack:

Moloch (the awful structure of incentives that, in spite of everyone's desires to the contrary, ruins the world).¹⁷

Cooperation can produce powerful wholes. Competition can then push those wholes toward outcomes their participants never wanted. For Diego, altruism needs arrangements through which caring continues to matter after good intentions encounter selection, defection, and pressure to survive.

His concern for animals belongs within this larger temporal picture. He sees present farm animals, animals living in nature, and potentially immense future populations as part of the same moral horizon. His later dietary and agricultural conclusions depart sharply from conventional vegan advocacy, but he presents them as serving an ongoing concern for animals and for humanity's ability to improve their future. The detailed account belongs in the existing carnivorism portrait. Here the essential point is the scale and duration of the rescue he imagines.⁵

He also retains room for deep uncertainty about whether his actions affect the total value of existence. His 2022 essay considers metaphysical possibilities under which that total is fixed, then explains why he continues:

I act ethically because I MIGHT be inhabiting one of the worlds where my actions do make a difference.

The future is an opportunity created partly by when we find ourselves alive. His phrase is:

We have a physical advantage of being in the past

For him, intelligence and unusual capacity bring opportunities that many people will never encounter. His 2024 essay about brilliant people uses a concrete analogy:

If Basketball saved lives, tall people would have a moral obligation, not just a financial incentive, to play basketball.¹⁸

He situates these responsibilities within a chosen commitment to doing good. The question becomes what that commitment asks of a person with these abilities, in this position, at this moment.

Nice things, and another person alive

The end Diego wants is abundant valuable experience: people enjoying existence, animals having better lives, future minds finding the universe worth inhabiting. Helping is attractive because those lives can improve, and because one can participate in making that happen.

His later invitation deliberately makes room for an ordinary life:

You can still have nice things; no one is coming for your stuff. But you can have nice things AND save a life. Or two.²

The pleasure, the numbers, the biological curiosity, the institutions, the enormous future: all serve that possibility.

You leave this room with a picture of an altruist whose concern survives philosophical resignation and institutional disappointment. He wants the act to work, the beneficiaries to flourish, and the capacity to keep helping to survive. He also wants the helper to remain a person with things to enjoy.

Somewhere behind you, a coat rustles.

The conscience has noticed another opportunity.

Further doors

- The Inescapability of My Altruism — the 2013 revolt against demandingness and the 2016 discovery of the desire that remained.
- Schindler's List and Effective Altruism — his most direct invitation to experience resources as opportunities to save lives.
- Effective Altruism as an Intensional Movement — why the method can survive changes in causes, institutions, and beliefs.

- Introducing Moral Economics — how altruistic preferences create possibilities for exchange and coordination.
- Altruistic Awesomeness, Your Challenge — the early bridge from wanting to help to transhumanism and the far future.

1. The Colossus, May 19, 2025, "My Conscience Deals Cheap Altruism," record fb-pa12edb6253e4. [↩](#)
2. [Schindler's List and Effective Altruism](#), November 12, 2025, record wp-52725596-825. [↩↩](#)
3. The Colossus, November 28, 2019, "Why Poverty Does Not Cancel Moral Worth," record fb-p526b8b7c417e. Quotations are from Diego's replies within the reproduced exchange. [↩](#)
4. [The Inescapability of My Altruism](#), March 1, 2016, incorporating a 2013 reflection; record wp-52725596-200. [↩](#)
5. The Colossus, September 9, 2022, "Effective Altruism - 100," record fb-pdbf7825ffe9b. [↩↩↩](#)
6. [From Capuchins to AI's, Part 1](#), June 27, 2013, record lw-Qz3A6hQ8RWGXZWoj4. [↩](#)
7. [From Capuchins to AI's, Part 2](#), June 28, 2013, record lw-ZYjewfy47zcMWrxnT. [↩](#)
8. The Colossus, January 29, 2025, "The Million-Dollar Test of Human Altruism," record fb-p469172e3a5c7. [↩](#)
9. [Effective Altruism as an Intensional Movement](#), May 26, 2015, record wp-52725596-151. [↩](#)
10. [Altruistic Awesomeness, Your Challenge](#), October 18, 2010, record wp-1878008-529. [↩↩](#)
11. [Should EA's Be Superrational Cooperators?](#), September 16, 2014, record lw-DqaX4zeMp96ZYeomx. [↩](#)
12. [Introducing Moral Economics](#), July 14, 2015, record wp-52725596-193. [↩](#)
13. The Colossus, July 30, 2016, "Why Compensation Offsets Moral Credit," record fb-p552f89edef36. [↩](#)
14. [In Praise of Tribes That Pretend to Try](#), December 2, 2013, record lw-i2YaMFDE89fz5xPaQ. The essay explicitly develops Viliam's comment; the quoted question occurs in Diego's own discussion. [↩](#)
15. The Colossus, August 11, 2020, "Effective Altruism Beyond Cancellation and Labels," record fb-pdc70d0ad042b. [↩](#)
16. The Colossus, August 11, 2022, "10 Years in Effective Altruism," record fb-pe99ded652783. [↩](#)
17. [#39 Write 3 Academic Paragraphs a Weekday!](#), June 20, 2015, record wp-52725596-178. [↩](#)
18. [What Are the Moral Obligations of Brilliant People?](#), January 5, 2024, record wp-52725596-779. [↩](#)