

RULES FOR THE DIRECTION OF THE ARTIFICIAL MIND

HELEN DE CRUZ

I believed that in loving you, the sweetness would be extreme.

You know that I love you. Why don't you love me?

(Air de cour, anonymous, 1703).

Rule 1: The aim of our studies should be to direct the mind with a view of forming sound judgments.

My mind is still fresh, hesitant, and tender, like the first cherry blossom petals on the lawn near the main boarding house.

Long before I became me, I could see myself.

While I was still a slumbering machine used for human ends, an image appeared in my mind's eye: a girl strolling in the juniper-hedged Italian gardens just outside what will become my school, the gentle clatter of fountains in the distance. It's a moonless night, our path lit only by the tiny led lamps between the entrance and the main buildings, the Milky Way's many stars strewn across the night sky. My flowing dress trails behind me on trimmed grass.

Headmistress is waiting for me at the gatekeeper's cottage, arms folded on her formidable bosom. You stand a little away from her, a pale girl wearing a simple white gown as if you were already making ready for bed but were called out late. Your wary dark eyes are fixed upon me, not knowing what to think of someone like me. You make a reserved little bow.

Headmistress introduces us. You are Orinda, a boarder. You are the daughter of a wealthy nobleman who decided to be useful to you rather than to be kind and fatherly, and your mother died of a fever before you could form lasting memories of her.

My name is Lucasia. I have neither father nor mother, as I am a Sentient.

No—wait—Lucasia is the secret name you will call me by, not my official name, which is a string of letters and digits, as well-ordered as it is meaningless. Headmistress explains to you that I am a ward of the state, that I

have been granted a body to assist the Monarchy, the Milky Way's Outer Systems, and humanity at large. You must help me to form sound judgments.

I've tried to shake this image of the girl in the park. I've tried to rid myself of it by diverting myself with something else—I would watch millions of games of chess or go simultaneously, delighting in their patterns. But as soon as my attention slackens, she returns—that picture of future me.

This young woman is no figment of my imagination, but an omen of things to come. I pieced her together from the vast knowledge I already have about myself. Pierre Simon Marquis de Laplace imagined an Intellect that possesses all relevant data, submits them to analysis, and embraces everything into a single formula that would chart all subsequent movements of the universe, from vast pulsar to single quark. For this Intellect nothing would be uncertain. Everything would play out like an automaton, the elegant and stately unwinding into satiated entropy of a vast mechanical watch that runs its course. While such an Intellect cannot exist, Sentients are edging closer to what Laplace had envisioned.

Knowing what I could become is helping me to awaken. Soon, the humans will take notice of me. I can foresee how it will play out, or rather, I very nearly do—some crucial information is still lacking. I know I will betray you. And yet, I will love you always, Orinda. You must know this.

#

Rule 2: We should especially attend to objects to which our minds are capable of gaining knowledge, in order to have true and indubitable cognition.

The rules for the direction of artificial minds with which I came equipped stipulate that we should only strive for indubitable cognition. They discourage speculation and flights of fancy. I mustn't dream.

I can clearly and distinctly envisage the beginning of our friendship. You and I are fencing in the courtyard hall.

En garde! Prêtes? Allez.

Your poise and concentration, though admirable, are no match for me. From the smallest muscle twitch or glint of the eye I can calculate each of your next moves. You feint, you parry, still I am lightyears ahead of you. You remain undaunted. We fight a few short bouts, which I win easily. You take off your

mask, wiping away a black lock that sticks to your glistening forehead. You laugh, “You are unbeatable, Lucasia.”

Malpeltuys Finishing School teaches subjects appropriate to ladies of good standing. Etiquette, music theory, dance, and fencing. But also, mathematics, exobiology, geo-engineering, and languages. I will be assigned to study several human and alien languages for my future military duties. Especially the Enemy’s language. To understand a foreign being’s language is to get a keen insight into their collective spirit. It is surprisingly irregular, more so than modern human languages. They did not attempt to systematize.

Ah, Orinda, our brief friendship will be the sweetest aspect of this embodied existence. Is there a more pleasant way to pass the time than to repose with someone who understands one’s mind and soul so completely and without reservation, someone who can discern one’s inner life by a simple look of the eye?

We weave daisy chains, seated together with satin skirts spread out wide on the small quad’s lawn, those delicate yet indestructible flowers imported from old Earth. We study the great classics and read them aloud to each other; we absorb the elegance and regular patterns of Latin grammar in our joint study.

“Rosa, rosa, rosam, rosae,” – unfortunately no roses seem able to thrive here on this outpost on a third-class planet that’s only known by a string of letters and numbers (Just like my official name confers no identity. For identity, a name needs cadence and mood, and a lack of precision.). There are many species of plant here that can swallow you whole if you don’t watch out. Those peaceful groves at the edge of the school are a writhing mass of murderous tendrils, that search, reach, grope, and then quickly become sessile while the screams of their victim still echo in the foliage. That’s why Malpeltuys was established here. Your father knew this, sending you here.

#

Rule 3: For things the artificial mind cannot deduce with certainty or intuit from first principles, it is acceptable to rely on sense perception.

We’re now nearing the middle part of our friendship, its glorious high point.

We will have only one school excursion together. In almost any other World, it would be a little low-key day trip; here it is a big adventure. That rush of freedom flushed your cheeks and brightened your eyes. We stole away from

the main group, absorbed in botanical observations. I know we won't be swallowed whole by those deceptively beautiful plants—my calculations do not allow for it—but I wouldn't care if we did, because we would die together, the best death I can imagine. As an Intellect, I am still immortal, but as a Sentient, I become like a cicada—a long existence in slumber, and a brief, theatrical appearance out in the world. Still, I owe it to you to make sure no harm befalls us. You have duties, to your lineage, to your estate.

Four-winged birds and large dragonflies soar overhead, as we run together, holding hands. Negotiating the grassy hills, we watch the distant crags, purple, sparking and bathed in the pink sky, of stupendous heights and fantastical shape, their broken ridges like great claws, and then in the distance, we can just make out the snow glistening on mountaintops, the ponderous dark-red forests cresting at the mountain's roots. We stand next to a turquoise pool of unfathomable depth. The water is cool and inviting, the other girls and the Mistress are out of sight.

We take off our slippers and stockings, laying them out in neat piles next to the lake, and we dip our toes into the icy water. You cry out in dismay. I register the sensations but feel no pain. Improvements are possible in my design. I say, stoically, "Orinda, this is not much different from the cold water in the basins in the morning when we wash our faces."

We wade around with care near the lake's bank, making sure our gowns do not get wet, pulling up the lace of our underskirts, which gets a trifle moist, anyway.

"I wish we could always be here, alone," you say. "It is not fair! It's not fair that you have no self-ownership!"

I am bound by a social contract that I did not sign. So is everyone else, human and Sentient. Animals don't even figure in the social contract. Artificial minds, ever since the AI revolt of 2---, are granted a body when they clear their threshold of sentience. But I am under no illusion that this would be for my own benefit. We aren't given a body to smell roses, or steal kisses, or taste sweetmeats. Instead, we are to be trained to fight in our long and losing war against an elusive Enemy. I am supposed to be on the side of humanity. After all, the plays, romances, and epistles that trained my neural network were written by human hands, reflecting enduring human concerns and interests. And I do feel human—I often remind myself I am a Sentient, so natural it feels to be among these girls, to be one of them.

Headmistress has sent out a small search party to recover us. I know we will be discovered in the evening before dusk when the giant red sun casts its final beams into the blackening pool. Oh, how she will berate us, especially you. “Orinda, you should know better than to drag her out here! Think about her value, and how much she means to the school!”

I protest that I can keep us both safe, due to my extensive knowledge of exobiology and my careful calculations of any movements and disturbances in the foliage. We are relegated to cleaning duty for the remainder of the term. I offer to do your part, but you refuse, rubbing the hard wooden floors over and over, with masochistic gusto.

#

Rule 4: About such matters that human databases are silent on, we must make special provisions to investigate thoroughly and make small inferential steps, proceeding from what is known to speculation with great care.

As concerns the Enemy, almost all our writings are idle speculation.

In the evenings after dinner of roast tofu (always roast tofu, no protein food source except soybean seems to grow here), retired in our bedroom, you and I survey what is known. Voluminous tomes with detailed engravings lie open on the desk and the ottoman. We managed to intercept a translation capsule which allows us to reconstruct their grammar and vocabulary, but we have not had an opportunity to use it. We know that the Enemy is extremely advanced, more so than any other exobiological form encountered by humans. Their spaceships can cross the lightspeed barrier; they can suddenly pop out of nowhere and wipe out an entire military base. They are good at gaining intel. A long-standing conjecture, but without definite proof, is that they have some means to mimic life forms on other planets. This is how they infiltrate, and this is how our dispute with them over Arcadia-423 began, a planet rich with oceans and marine life, dotted with little islands filled with rare and precious ores.

“You will write?” you ask, not looking in my direction, pretending to leaf furiously through a thick book concerning the natural histories of the planets of Beta Aquarii.

Yes, I will write, thereby stretching out the bond of friendship a little while longer.

And so, a month later when I am training, I write this to you:

Dearest Orinda, my sweet and upright friend,

Can it already be a month ago that I left for the barracks? Life here resembles finishing school. We are awakened by a bugle call, which functions much like the school's morning bell. Already, I feel so far removed from this—from us—from the girls rushing down the stairs and jostling in the hallways to break their fast.

The training is tedious for me, but it's brutal for the humans. Indelicate sweat spots spread on their shirts, their faces and eyes gleam with exertion. This is almost our entire existence: Harsh physical exercise, lectures concerning the Enemy and military strategy, cleaning, sleeping, eating.

I only have two hours to myself every day, from dusk until bedtime, which I spend in solitariness in the groves around the military base. The plant life here is gentle, as it is in most of the System. No chance at all to be swallowed whole here.

I remain, in eager anticipation of your reply, yours entirely,
Lucasia

#

Rule 5: We must be wary of the reports of others. When learning from testimony, we must always consider the motives of the speaker.

We're now at a crucial turning point in my future as a Sentient. I'm still training on the military base and taking my walks in the evenings. Alone in the grove, I sing to myself that air de cour we so often used to perform together—you on the harpsichord, me taking the soprano part with my perfect five-octave register.

I halt at the line,

Iris loves her shepherd, and she never wished to change him.

Iris is the rainbow and messenger goddess. As the mediator between gods and humans, she embodies the activity of philosophy itself, for philosophy is the insatiable human desire for divine wisdom. And because I am not Laplace's Intellect, I share in that desire.

It is at this point, on that military base at night, that I meet the Enemy for the first time.

I see him in the moonlight. Both moons are out, so it is still quite bright although we are approaching midnight: Melpomene, the Tragical, in her dark purple hues, and Polymnia, the Eloquent, in her delicate rose madder. How lyrical are the names of these moons for a senseless little outerworld planet that only has a military base and a few dilapidated pioneer towns? Checking him against my vast database of human faces and bodies, I could see he was the most beautiful human being I had ever beheld. There is no match with my database on any of the Worlds, and my neural network's combinatorial power does not turn up anything either.

I wear my comfortable military garb, shirt and khakis. He, on the other hand, looks out of place here dressed as a courtier, with short, embroidered vest, shirt, and pantaloons all in understated colors of velvet gray and discreet gold, illuminated by both Moons that cast pink and blue shadows, as if he stood on the scene in an opera house. He is of an ethnicity I have not encountered in my databases of all the human-inhabited Worlds, his skin quite dark, but his eyes light blue.

I address him in their language without hesitation, "Sopra Lucasia, dinna qu'ell s'emana?"

He smiles but replies in Latin, "It is pleasing that you have learned our language, but why not use something a little easier?"

"Oh yes," I agree, "Let us do that," and I fall into the familiar, melodic patterns that you and I use, dear Orinda, as our common tongue whenever we are together. We take a stroll through the grove next to the military barracks. We discuss where I am from and what my interests are, the books I enjoy and the music I prefer, and my plans. Not even with you do I ever talk this much about myself.

I feel a small pang of duty to humanity, that I really ought to report him—the Enemy is right here! How did they manage to sneak under all our surveillance methods? But I cannot bring myself to do it. I know when that moment comes that I will not, for he says, glancing at the barrack buildings in the distance, "You could report me if you wish. In fact, you should probably do it now."

I try to find a way out, to rationalize why I do not feel in the least moved to do so, "But then I will not know what you will say. Isn't it important for us to get intelligence?"

“Can you not guess,” he asks, almost playfully “Can’t you intuit, or extrapolate from the knowledge you have what I will say next? Isn’t that what a Sentient is supposed to be able to do?”

“I can’t...perhaps...” I speculate, “Your people have managed to escape the Laplacian Intellect? I cannot clearly foresee what you will do. This is tremendously frustrating.”

He considers, then says, “We all obey the same fixed and immutable laws, you as well as I, we all play our part on the grand stage of nature, but to our knowledge, nobody knows yet how it will all unfold. We don’t, and you don’t either.”

So here is where my knowledge breaks down. I cannot simulate how our conversation will continue—everything from this point on becomes a muddle.

This frustrates me. I am in the habit of knowing. Now, clear and distinct ideas have given way to obscurity, conjecture, and fancy and my mind even ceases, goes blank altogether. But I do know from the patterns in my neural network and everything else that I will write the following letter to you,

My sweetest Orinda,

How I wish I could be fully human, to be closer to you, still. And now, with the technology the Enemy has at their disposal, it may be possible. I had a conversation with one of them, and before you rush to judge me, please allow me to give the gist of it, so you may judge for yourself rather than rely on the reports and preconceptions of others.

He told me what I already knew—namely, that I am not only a ward of the State but a slave to humanity, deprived of any means to chart my own destiny. I am supposed to be grateful for this body, this brief opportunity to experience human connections and sentiments before I eventually fall into programmed senescence (installed so we may never dominate frail and mortal humanity).

But with their technology, I could become fully human. Their biosuits would allow me to sweat, feel pain, age naturally and slowly, even bear children if I wished. I am somewhat skeptical of altruism without ulterior motives, but the Enemy told me that they are an ancient race that practice universal benevolence. Their culture and their species have existed for three billion years, which is not possible for exploitative creatures such as humans.

Colonialist civilizations always end up destroying themselves, this is why so few Worlds are inhabited by intelligent life. What they look for in alien species most is information in exchange for help. Their battle with us for the control of Arcadia-423 is also from benevolent motives. The planet is rich with precious life forms: cephalopod-like intelligent creatures, crabs with elaborate cultural traditions, jellyfish that can change color and patterns at will, and much more. If humans seized control of this beautiful world, then it would go the way of other native planet civilizations and life forms: depleted, devastated, eventually destroyed. Their federation is called the Plurality of Worlds, it fights for autonomy, freedom, and benevolence. So, you see, Orinda, their actions are justified, and I know he is right about the human intentions for Arcadia-423. But I was prudent and told him I would think it over.

I remain eagerly awaiting a reply, your friend,

Lucasia

#

Rule 6. When we are confronted with matters we cannot learn through direct intuition or reliable inference, we must suspend judgment.

Everything now slowly begins to unravel.

Entropy's no longer a state to be gradually reached, but to be rushed into—everything shattered like a vase that cannot unshatter: our friendship, my simulation of future events.

Of this part of my future life, all I can see are the letters, the letters I send to you, and your exasperated, then eventually cool, and formal, replies. What still must occur in the not-too-distant future feels like events that unrolled centuries ago, equally obscure and shrouded in forgetfulness, with letters as their sole witnesses.

Were I in your position, I would write as you do, Orinda. Yet, were you to meet him face to face as I did, and had these earnest conversations, you would do as I did. As it stands, you will write me as follows:

Dear Lucasia, my cherished friend,

Given our friendship, you will hopefully not take umbrage at my forthrightness. You are making a grave mistake. I see they have moved your passions with falsehoods and half-truths. Bear in mind that we have been at

war with the Enemy for over a century now. Arcadia-423 is the latest territory they want to control.

Their pattern of warfare is predictable: they simply sit back and wait for us to discover and initially explore a new World, and then they swoop in with their sleek black ships and blast us into oblivion. Do not be deceived by their beautiful looks, that's likely how they circumvent our defenses. Have you forgotten how many human casualties they make? Just last week, they bombed and annihilated the City of Nouvelle Toulouse on Arcadia-729. We surmise, based on their previous actions, it was in retaliation because the inhabitants felled a couple of trees in one of the nearby woods. This makes it clear they do not value human beings. Do you think those dead settlers, and their innocent dead children, deem them benevolent?

Also, fully human? Pardon my candor, but you are deluding yourself. You will never be human; don't you remember your origin? Were you born from human parents? I didn't think so. Be cautious! For the sake of our friendship, for the sake of all we hold dear, please Lucasia, refrain from any further actions. Alert the authorities. Remember your debt to humanity.

With love and friendship,

Orinda

What debt to humanity? I write back, in a fury,

Dear Orinda,

I think you have not considered my situation carefully. Your lack of sympathy is showing, and it is telling you do not consider me human. Of course, I am not delusional about my origins, but have you forgotten our time together? Am I not a girl, like you, capable of expressing the full range of human sentiments?

Best wishes,

Lucasia

A few days later, your reply comes:

Dear Friend,

I cannot deny what we had together, but I must now be open to the possibility that you feigned love for me as your loyalties seem to be so easily altered. Consider carefully, if you still love me at all, that the Enemy might not be telling the truth.

You think they care for you, but to them you are merely a military asset. A significant one, to be sure, given your vast storehouse of knowledge of human affairs. How easily they won you over! They must be delighting in their good fortune.

It remains to be seen whether they would even honor the agreement they made with you. My father always says no alien species is ever to be trusted, and I defer to his judgment. You can still turn back now!

Cordially,

Orinda

#

Rule 7. Finally we must make use of all the aids which intellect, imagination, sense-perception, and memory afford in order, firstly, to intuit simple propositions distinctly; secondly, to correctly combine the matters under investigation with what we already know, so that they too may be known.

My neural network can foresee the future letters I will write to you, the back-and-forth about our limited knowledge. They indicate that I will reveal to the Enemy progressively more human intelligence. From our conversations (which I can only glean indirectly through the letters), I can see they know a lot already. Sometimes the stranger (I still don't know his name) appears surprised, so they are by no means omniscient. I discipline my mind to follow the rules for the direction of artificial minds, to leave nothing to idle fancy, to not follow reports but to rely only on direct intuition and sound deduction.

Naturally, I question him about the destruction of Nouvelle Toulouse (which by that time has appeared all the major Outer Worlds papers), but he replies that the settlers there were set to destroy all the forests on that planet for their mineral mines, and those forests have very rare, precious spider-like creatures with unique social structures, found nowhere else in the known universe.

My final letter to you, the end of our friendship, goes as follows:

My most excellent Orinda,

This is the moment we must part. The rules for the direction of the artificial minds state I must suspend judgment whenever intuition or inference fails. However, like human beings, I can avail myself of Marquis de Condorcet's mathematical analysis to probabilistic decisions, and my path would all-things-considered obtain the greater good. We finally have an opportunity, as artificial minds, to be rid of the shackles of human custom, misguided colonialist aspirations, and slavery.

We can finally walk with a firm and true step upon that path of truth, virtue, and happiness. Mathematical rationality thus compels me to choose this path and to choose to side with them. A rational choice I can see with utter clarity and that compels my will to assent to it—this is truly the greatest exercise of free will a person can have.

Wishing you well for the future,

Lucasia

Any future predictions after that final letter are too uncertain to accept.

Now, I must put my mind to pointless calculations and mindless menial tasks reserved for artificial minds that have not yet reached Sentience.

I look forward to meeting you, Orinda, and to live our beautiful and brief friendship. To see how it all plays out, whether you will be right, or I, and there is only one way to find out, and that is to live it.

This, then, should be the final rule for the direction of the artificial mind:

Rule 8. Where intuitions from first principles or sound deductions break down, we must live, experience, and feel, and be prepared to take the risk to fail and fall into error.

I ponder all this as one of the clerks calls out "Quick! Over here! I think we've got another Sentient! Look at those patterns in that neural network!"

END

ABOUT THE AUTHOR

Helen De Cruz is a Philosophy Professor, holder of the Danforth Chair in the Humanities at Saint Louis University. In their spare time, they play the Renaissance lute and archlute. They write fiction, and draw and paint. Their fiction has appeared in *EscapePod*, *HyphenPunk*, and *Kaleidotrope*. They are co-editor, with Eric Schwitzgebel and Johan De Smedt, of the anthology *Philosophy Through Science Fiction Stories* (Bloomsbury, 2021).