

School of Philosophy
Department of Philosophy & Hellenic Pasteur Institute
MA ANIMAL WELFARE, ETHICS AND THE LAW

*From Kubrick's 2001 to Scott's 2019:
Science Fiction and Moral Treatment of Non-Humans*

Nikolaos Dimitropoulos

MA Diploma Thesis

Supervising professor: Arabatzis Georgios

Master's Committee:

Ευάγγελος Πρωτοπαπαδάκης

Μέλος ΔΕΠ
(Αναπληρωτής Καθηγητής)

ΕΚΠΑ -Τμήμα Φιλοσοφίας
Διδάσκων ΔΠΜΣ
«Ζώα: Ηθική, Δίκαιο, Ευζωία»
Τμήματος Φιλοσοφίας ΕΚΠΑ-
Ελληνικό Ινστιτούτο Παστέρ

Γεώργιος Αραμπατζής

Μέλος ΔΕΠ
(Καθηγητής)

ΕΚΠΑ
Τμήμα Φιλοσοφίας
Διδάσκων ΔΠΜΣ
«Ζώα: Ηθική, Δίκαιο, Ευζωία»

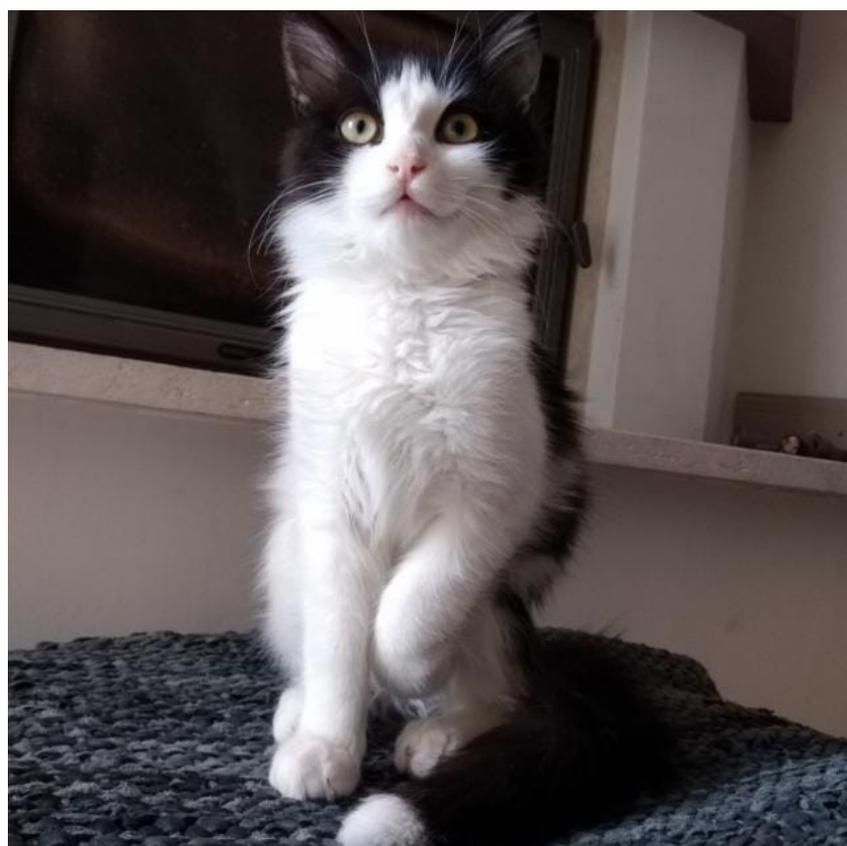
Μιχαήλ Μαντζανάς

Μέλος ΔΕΠ
(Καθηγητής)

Ανώτατη Εκκλησιαστική Ακαδημία
Διδάσκων ΔΠΜΣ
«Ζώα: Ηθική, Δίκαιο, Ευζωία»
Τμήματος Φιλοσοφίας ΕΚΠΑ-
Ελληνικό Ινστιτούτο Παστέρ

Athens, 12/12/2023

To Takis..



Acknowledgements

I would like to thank my supervising professor Georgios Arabatzis for the trust, acceptance and understanding, Vicky Protopapadaki for always being there to answer my questions and everyone involved in the creation of this visionary and innovative program.

Finally, my most heartfelt thanks to Eleni Kouti, my companion in life, whose love, help, support, cakes, pizzas and encouragement made it all happen.

Index

Acknowledgements.....	4
Index.....	5
Abstract.....	6
Περίληψη.....	7
Chapter 1 : Introduction.....	8
Chapter 2 : Science Fiction.....	13
2.1. Science Fiction and Ethics.....	14
2.2. <i>Blade Runner</i>	15
2.2.1. The Film.....	15
2.2.2. A Philosophical Review of the Film.....	16
2.2.3. The Role of Memory in <i>Blade Runner</i>	21
2.3. <i>2001 Space Odyssey</i>	23
2.3.1. The Film.....	23
2.3.2. A Philosophical Review of the Film.....	24
Chapter 3 : Philosophical Background.....	27
3.1. Rights-Based Approaches.....	28
3.2. Utilitarian Approaches.....	31
3.3. Critique of the Two Approaches.....	33
Chapter 4 : A suggested approach.....	37
4.1 Objectives.....	38
4.2. Building an Alternative.....	38
4.2.1. A Minimal Notion of Agency - <i>To Be</i> , as an Intrinsic Goal.....	38
4.2.2 Higher Order Goals.....	40
4.2.3 Group Goals as Individual Goals.....	41
4.2.4 Practical Application.....	41
4.3 The suggested approach summarized.....	43
4.4 Comments.....	44
Chapter 5 : Conclusions.....	45
5. Conclusions.....	46
Bibliography.....	47

Abstract

Non-human animals as food sources, test subjects, workers and companions have long been the subject of moral quandaries. The group of organisms raising questions regarding their moral treatment is however increasingly subscribing new members over the recent years. Organisms engineered and produced by humans are joining non-human ranks of the morally eligible. Organoids, hybrids, bioelectronic systems and the recently emerged artificial intelligence Large Language Networks are entering the scene, a scene once set only for the filming of science fiction.

Classic approaches on the matter in philosophy have mostly taken either the road of attributing individual moral rights on the basis of intellectual capacities or the utilitarian way, of deciding moral considerability by weighing its impact on total satisfaction of interests or the minimization of pain. The problems these approaches face are well known.

Considering problems less discussed, as deserving to take center stage, they are presented and then exemplified drawing from two monumental points in science fiction filmic history: 1968's *2001: A Space Odyssey* by Stanley Kubrick and 1982's *Blade Runner* by Ridley Scott.

Departing from the refocused problematic on the subject, we set off to explore an alternative approach, based on Barandiaran's notion of minimal agency and goals, and Levin's view on biologic group membership. The continuation of one's existence is made the central, *highest order goal* of all and any organisms or groups, our moral consideration of them is contextual, based on comparison of the *order of goals* of those involved, and moral decision making can be accommodated by statistical scientific methods.

Keywords

moral status, moral considerability, non-human animals, sentience, agency, interests, rights, science fiction, *Blade Runner*, *2001: A Space Odyssey*

Περίληψη

Τα μη ανθρώπινα ζώα ως πηγές τροφής, πειραματόζωα, εργαζόμενοι και σύντροφοι αποτελούν εδώ και καιρό αντικείμενο ηθικών διλημάτων. Ωστόσο, η ομάδα των οργανισμών που εγείρει ερωτήματα σχετικά με την ηθική τους μεταχείριση εγγράφει ολοένα και περισσότερα νέα μέλη τα τελευταία χρόνια. Οργανισμοί που έχουν κατασκευαστεί και παραχθεί από τον άνθρωπο εντάσσονται στις τάξεις των μη ανθρώπινων ζώων που είναι εν δυνάμει ηθικά υπολογίσιμοι. Τα οργανοειδή, τα υβρίδια, τα βιοηλεκτρονικά συστήματα και τα πρόσφατα αναδυθέντα δίκτυα τεχνητής νοημοσύνης Large Language Networks εισέρχονται στη σκηνή, μια σκηνή που κάποτε στήνονταν μόνο για την κινηματογράφηση της επιστημονικής φαντασίας.

Οι κλασικές προσεγγίσεις επί του θέματος στη φιλοσοφία έχουν ως επί το πλείστον ακολουθήσει είτε τον δρόμο της απόδοσης ατομικών ηθικών δικαιωμάτων με βάση τις διανοητικές ικανότητες είτε τον ωφελμιστικό δρόμο, της απόφασης της ηθικής υπολογισιμότητας με στάθμιση των επιπτώσεων της στην ολική ικανοποίηση των συμφερόντων ή στην ελαχιστοποίηση του πόνου. Τα προβλήματα που αντιμετωπίζουν αυτές οι προσεγγίσεις είναι γνωστά.

Θεωρώντας προβλήματα λιγότερο συζητημένα, ως άξια να βρεθούν στο επίκεντρο, αυτά παρουσιάζονται και στη συνέχεια παραδειγματίζονται αντλώντας από δύο μνημειώδη σημεία της κινηματογραφικής ιστορίας της επιστημονικής φαντασίας: Το *2001: A Space Odyssey* του 1968 από τον Stanley Kubrick και το *Blade Runner* του 1982 από τον Ridley Scott.

Αναχωρώντας από την ανασυγκροτημένη προβληματική επί του θέματος, προχωράμε να διερευνήσουμε μια εναλλακτική προσέγγιση, βασισμένη στην έννοια της *minimal agency* και των *goals* του Barandiaran και στην άποψη του Levin για τη βιολογική συμμετοχή σε ομάδες. Η συνέχιση της ύπαρξής γίνεται ο κεντρικός, *highest order goal* όλων και οποιωνδήποτε οργανισμών ή ομάδων, η ηθική μας εκτίμηση τους είναι πλαισιακή, βασισμένη στη σύγκριση του *order of goals* των εμπλεκόμενων, και η λήψη ηθικών αποφάσεων μπορεί να υποστηριχθεί από στατιστικές επιστημονικές μεθόδους.

Chapter 1 : Introduction

A small airplane on the screen, embedded in the back of the seat in front of me, floats on a map above the Netherlands. There is a line tracing its course till Buenos Aires; the number 11437, kilometers that is, next to it. It is going to be thirteen and a half hours from take off to touch down, I am informed through the aircraft's speaker system. Enough time to catch a movie, get some sleep and think. Because this is not a travel for leisure, but rather one of the more long term kind. I am thinking about how I am having it easy. I used to spend 7 hours traveling from Athens to Thessaloniki by train, as a university student in late 90s Greece. Komotini, a city at the northeast border, was even further, about 12 hours. Now I am crossing the ocean, from one continent to the other in thirteen and a half.

In the early 1900s, there was a huge immigration wave from Europe to Argentina ¹. Millions of people crossed the ocean, but they were less fortunate than me. They had to do it by ship and the trip was between two to three weeks long. I get curious and do the math; if I had to walk that distance, without the mountains and ocean in the middle, it would have taken me something close to 350 consecutive days of walking 32 kilometers per day. Human populations in humanity's first steps, had to move on foot, and covering this much ground took generations. The entire American continent was populated gradually as groups descended from Alaska, having crossed through what was then a kind of land bridge, from Siberia. Navigating the waters came a bit later, but still early in our history and sped things up considerably. We spread literally everywhere. The way I am traveling now, is a very recent thing. It is a novelty. From the insides of something that those generations of humans would view as a giant steel bird, I try to grasp the difference in the order of magnitude; hundreds of years for generations of people to 13.5 hours for one person.

The opening sequence of "2001: A Space Odyssey" strikes for me the same chord. It depicts two groups of hominid ancestors of ours, clashing for control over a watering hole until a strange black monolith of alien origin somehow influences one of them to use a femur bone as a tool to kill the opposing group's leader. The shock and surprise in both groups hints that these humans have just discovered tools and murder for the rest of us, and in the context of war at that. What happens next though is what stuck with me even more than the fictional explanation to the discovery of these major driving forces of human history. The killer roaring in celebration throws the bone with all his might high into the sky. The bone reaches its peak, suspended in the blue background, and as the scene changes, in that same spot on screen it is replaced by a man-made satellite of the year 2001, suspended in the sky while orbiting the Earth.

One point on screen connects the beginning and the future of human evolution and history. The same point, occupied one moment by a bone thrown by hand a few meters into the air and the other by a metal construction weighing tons put to orbit with the help of man-made machines. We went from throwing things that weigh a few kilos at most, to propulsing massive constructions weighing tons. The reach of the apeman has been extended greatly. Our capacities now are sufficient to put us on the planet Mars in the following decades. And yet, the trip to the closest star, Proxima Centauri, with current technology would take tens of

¹ "Great European Immigration Wave to Argentina," in *Wikipedia*, August 12, 2023, https://en.wikipedia.org/w/index.php?title=Great_European_immigration_wave_to_Argentina&oldid=1169986533.

thousands of years. What was once unreachably far, now is a short trip, and other destinations have taken its place of unattainability. Will this process ever end? It is unknown whether the augmentation of human capacities will ever enable us to lay our hands on all that exists, or whether this is a never ending struggle; our goals forever expanding along with our expanding capabilities.

Our goals personally were to move to Argentina and study music. Our cat did not come with us. I could only surmise his goals, if we assume he has them, but they probably do not include castration, sedation, a cage and a flight across the ocean. The range of his excursions are two neighborhood blocks at the most. That is his territory and his home. It was a tough decision for us to leave him behind. While there he was free to come and go as he pleased, in Buenos Aires he would have to stay confined in the apartment. Two blocks compared to 60m². One might say there is not much of a difference. There is a clear need this cat has however, that is satisfied by his living space extended to two blocks. There are not only a couple of humans in these two blocks. There are a few dozen, along with cats, dogs, birds, trees, insects, lizards, action, changing weather; an entire small interacting world. Cats in Buenos Aires tend to spend time seated at their windows, looking at other cats looking at them seated at other windows. To join us, our cat would have had to be operated for castration and spend more than twenty hours inside a cage. We decided this shock, along with the shock of changing environment and the eventual life confinement, weighed more than the loss of his apartment and the people who were viewed as his parents. This was a decision we had to take for him.

In Buenos Aires it is considered cruel to leave a cat on the street and good service to the animal to castrate it and keep it confined indoors for the entirety of its life. On the other hand, Argentina is the world's greatest per capita consumer² of red meat and one of the worlds greatest producers³. There is a strong tradition of meat roasting on barbeque fire and many people are having meat almost every day. For a significant part of the Argentinean population it is considered good practice for cattle, pigs and poultry to be factory farmed for consumption, cats and dogs to be castrated and kept indoors for life. But choices regarding nonhuman animals are not the same for all people and not for every part of the world. People in Istanbul for example have a long tradition of living with free roaming cats and going to lengths to accommodate their feeding, shelter, comings and goings. Some people consider meat consumption unacceptable. Others agree, but add that animal experimentation is a necessary evil in particular cases.

The questions about pets however, are but a small part of the discussion concerning nonhuman animals. The food industry uses what is called *factory farming* to supply the human world with nonhuman animal meat, for consumption. The conditions in which cattle, pigs and poultry are conceived, born, live and die are nightmarish to say the least. The seafood industry has brought fish populations to critical lows, with some species, yellowfin tuna being one example, in danger of extinction. Research in a multitude of fields, such as, but not limited to, medicine, biology, cognitive science, psychology, cosmetics, routinely uses

² "Per Capita Meat Consumption by Type," Our World in Data, accessed August 29, 2023, <https://ourworldindata.org/grapher/per-capita-meat-type>.

³ Hannah Ritchie, Pablo Rosado, and Max Roser, "Meat and Dairy Production," *Our World in Data*, August 25, 2017, <https://ourworldindata.org/meat-production>.

animals as test subjects. They are the subject of documentaries, literature and public discourse regarding legislation and regulation. At the same time, the exploitation of land for agriculture and natural resources devastates wildlife habitats. Even insects are hitting historic lows because of the widespread use of pesticides.

On the other hand, new organisms come to life. Designed by humans, hybrid plants, cells and microorganisms, with DNA coming from different sources, are used in both farming and research. Stem cell research has gotten to a point where it can create *organoids*, cell cultures with a specific function; brain tissue cells, heart muscle cells, retina cells. Brain cell groups connected to electronic circuits recently learned on their own how to play the simple video game *Pong*, making a strong case for questioning what kind of thing it is we are creating.

Artificial intelligence broke through last year with unprecedented achievements, the reception of which was diverse, ranging from terror to awe. *Large Language Models* as they are called, are the first to be able to hold conversation like humans and produce high quality text in every form imaginable; prose, poetry, essay, legal text are but few examples. The engineers working on LLMs have witnessed unexpected abilities emerge, such as the ability to translate from one language to another and some even questioned the possibility of them already having acquired some form of consciousness.

It is obvious that the choices we have to make do not only concern our progressively changing relationship with our environment and its *other* inhabitants, but our completely new relationship to organisms unknown and unpredictable in their form and capacities. As we have seen, around the world, the way we make our choices and justifications varies, sometimes greatly. Our commonsense morals, what we consider right or wrong and why come in many different shapes.

Philosophers, in this respect, do not differ. Concerning the treatment of non-human animals, we could group a significant part of their opinions in two general approaches:

- The *rights-based* approach decides the way a non-human animal ought to be treated by taking into account certain requirements, that have to regard the subject for its own sake; these reasons do not have to concern someone else, but the animal for itself. Thus, the animal can be said to have a right to be treated a certain way.
- The *utilitarian* approach takes into account the non-human animal's interests and factors them along with the interests of all parties involved, with the intent of calculating the course of action that maximizes the total sum of interest satisfaction. The animal will then be treated in accordance to what we have decided maximum utility to be.

Well, the theory is nice, but what does that mean for our cat in practice? Either it has a right to be treated a certain way, which way we still would have to define and justify, or we would have to figure out what is better for all of us concerned. If it does have that right, how would we be inclined to treat him? And if we should choose the way of highest utility, what is the best result to be had? In a rights based approach, we could make a number of choices concerning treatment. Philosophers agree that we are at least prohibited from killing a being

that bears this right, even if it is to save the lives of others. Depending on the philosopher's view, there could be reasons not to interfere with the being, reasons to aid it or reasons to treat it fairly in a situation where there are conflicting interests. This means that if a cat qualifies for this type of treatment, then, at the very least, no matter the circumstances, we ought not kill it.

The implications of this are far reaching, since along with the cat, a multitude of non-human animals would also qualify for this protection. We would have to abandon the consumption of meat and a big part of drug testing, reconsider the management and use of land and water where they live and possibly give up on having pets. Global scale industries would cease to exist, taking jobs, trade and entire cities with them. On the other hand, factory farming is found to be the major contributor to greenhouse effect gasses, climate change is now the number one candidate to upend the state of things as we know them, with the artificial intelligence boom coming up close second, and many people find the annual killing of billions of animals to be a really bad thing to do. In any case, the world would have to change on a massive scale.

Chapter 2 : Science Fiction

2.1. Science Fiction and Ethics

The two most emblematic and important films of the science fiction genre were chosen for the present study; Stanley Kubrick's 2001: A Space Odyssey and Ridley Scott's Blade Runner. Why choose science fiction to approach the subject of moral treatment of non-humans?

Ethics is a normative business, in the sense that it is a philosophical attempt to say what is the right thing to do in a given situation. By doing so we are pinning a value on that action and judging actions past and future. The moral treatment of non-humans is primarily a moral concern for exactly that last bit; the future. We are concerned in what might come to be and how we should cater to it. We are preoccupied with how our present actions might affect future generations. Indeed we are beginning to consider the wellbeing of non-humans as well.

Planning a course of action within a certain ethical framework needs imagination and analysis. It needs scenarios upon scenarios, alternative drafts and often, fictional case studies. The proposed action needs to be played out as in film. We create small fictional worlds with their own rules and try to imagine the outcomes.

Thought experiments in philosophy serve exactly that purpose and have a long history in doing so. There are numerous famous examples, from Plato's cave all the way to Putnam's brain in a vat, and they have greatly contributed to test hypotheses, theories and frameworks by helping us imagine and illustrate when and how they could be met with trouble.

Science fiction, similarly, is often a study in what might be and how it could be like, and then a shot at what we would be doing about it. In that respect, it offers an open field and rich playground for useful scenarios. It is at this point where philosophy and science fiction converge. When these scenarios become or serve as philosophical thought experiments, science fiction is doing philosophy.

Moral anxiety for the future, on the other hand, is a concern about the living and about their world, a concern which branching and stretching temporally ahead of us, as it may, its roots are grounded in the present. This concern is both dealt with in ethics and depicted vividly in sci-fi.

The two films we have chosen may be seen as thought experiments concerning the ethics of non-human animal treatment. They can also be seen as a disquieting look into the possible, even dystopic, future and the problems we might have to face. More importantly, they can be seen as an age-old open question; what is it to be human?

2.2. *Blade Runner*

2.2.1. The Film

“Blade Runner”, a cinematic masterpiece directed by Ridley Scott and released in 1982, stands as an iconic landmark in the science fiction genre. Set in a future that was once imagined as distant but is now firmly within our past (2019), the film presents a hauntingly vivid vision of a dystopian Los Angeles where advanced technology coexists with urban decay, giving birth to a world both captivating and unsettling.

“Blade Runner” delves into profound questions about what it means to be human. Adapted from Philip K. Dick’s novel “Do Androids Dream of Electric Sheep?” and brought to life on the screen by Scott’s visionary direction, the film navigates the thin line that separates humanity from its artificially created counterparts. This exploration is embodied in the character of Rick Deckard, portrayed by Harrison Ford, a ‘Blade Runner’ whose task is to hunt down and eliminate rogue replicants — bioengineered androids that are virtually indistinguishable from humans.

The intricate plot follows Deckard as he is reluctantly pulled back into this world of moral ambiguity and existential uncertainty. The narrative unfolds with his pursuit of a group of replicants who have escaped the outer world colonies, to Earth, in search of extending their deliberately short lifespans. Led by the enigmatic Roy Batty, portrayed memorably by Rutger Hauer, these replicants challenge the boundaries of what it means to be alive and question the morality of their creation, actions, and the forces that seek to control them.

“Blade Runner” stands as a masterclass in world-building and atmosphere, merging elements of film noir with speculative science fiction to create a unique sensory experience. The electronic score composed by Vangelis complements the film’s haunting visuals, intensifying the emotional impact of every scene and accentuating the underlying tension between human and replicant.

As the narrative is developed, viewers are confronted with ethical questions, philosophical musings, and a palpable sense of existential unease. The film’s exploration of identity, empathy, and the fragility of memories adds layers of complexity that linger long after the credits roll. The film challenges us to reflect on what it truly means to be human, how we define our existence, and the consequences of playing god with technology. “Blade Runner” is more than just a science fiction film; it’s a thought-provoking journey into the recesses of our understanding of self and otherness.

2.2.2. A Philosophical Review of the Film

In the world of the film there are no non-human animals to be seen. We spot a few on screen, but we quickly find out that they are not real. They are replicas, memoirs for the humans of this world, surviving in representation, having transitioned completely into human collective memory. In the book, Deckard, the protagonist, is trying to get his hands on a real flesh and blood sheep, as his life's ultimate goal. But real animals have gone almost completely extinct, and the very few that are rumored to remain are so rare that they are symbols of status for the super rich. Modernity, a process that started in the 16th century, is now complete. Humanity has long been marking the line that separates it from non-humans, by progressively erasing wildlife from existence. Now that wildlife is gone, it has come to strangely treasure it, as if, or maybe because, it still needs that separation to define itself.

In this world there are no more wild non-human animals to separate from human ones. Their natural habitat, from which human animals departed long ago, has been completely claimed and transformed. It is the year 2019 and in the city of Los Angeles, we can see no natural sky, nor ground. The sky is hazy shades of orange in the daylight and artificial lights and flames at night. The ground is the city streets. But most of the time we do not even get to see that. The action takes place inside towering buildings, the characters often moving in flying vehicles among them. It looks like humans have gone to great lengths to distance themselves from the earth. In this future, most have left the planet for other, unspent worlds.

There are other non-humans there, though. They are called Replicants and they resemble us in every way. More human than humans, as the CEO of the company that creates them likes to say. Their inner world however, is a different story. They are said to be incapable of harboring emotions for the pain of others. They have no capacity for empathy. They are designed by humans to be physically and intellectually superior, but with a limited four-year lifespan as safety. That is because we are informed that their creators feared that overtime they would develop their own complex emotional responses, gain interests and subsequently become undesirably involved in the matters that have to do with their lives. This cannot be allowed because Replicants are created to be tools, workers and soldiers in the off-world colonies, operating in environments too hazardous for human beings.

Replicants are thus considered “things”. Fiction is here echoing Descartes⁴, who harbored and established the view that non-human animals are similar to clockwork mechanisms, with automated responses to stimuli. He believed that they do not possess rational thought and language and the reason they do not, is that there is no one on the inside to do the rational thinking. Animals have no ghost in the machine to direct it, they have no soul. What is left then, is just the machine as an automaton. For this, Descartes believed that they feel no pain and we do not have to feel bad about using them in any way deemed necessary. Views like this were held as late as the past century. In fact it was not until 1987 that the American academy of Pediatrics declared it to be unethical to operate on human infants without anesthesia⁵. Up to then, part of the reason they did not use any, was that they believed babies could not feel any pain.

⁴ John Cottingham, *Descartes*, ed. John Cottingham, (Oxford: Oxford University Press, 1998), 225–33

Replicants are shown to not feel pain either. Twice we see them put their hands in things that cause us to tense and clench our teeth with the imagined pain and damage. One in boiling water and one in liquid nitrogen. They appear unaffected though. Physically they suffer no damage and if they do feel pain, they certainly seem to be able to manage it.

The same cannot be said about their emotional world. While as mentioned they are supposed to lack feelings, we see them caring for each other, having sentimental attachment to memories and most of all mourning the dead members of their group and, as we eventually find out, orchestrating a desperate attempt of trying to extend their lives. All these make them appear perfectly capable of having interests and suffering emotional pain. It is the only way to explain why they rebel, deny orders and escape, risking their lives to extend them. For the humans of the story, they “malfunction” and standard procedure is to have them “retired” by special police operatives called “Blade Runners”. Deckard is one of them and considered to be among the best. To tell a Replicant from a human though, is no easy feat. It needs to be done by a specially designed test which measures emotional responses.

In the film we see that for the latest and most advanced generation of Replicants, the Nexus 6, they were almost expecting the test to fail. This new generation has implanted memories and some of them do not even know they are not human. This, as we are explained, is expected to make them more emotionally stable, as the emotional reactions they develop over time grow on a more suitable substrate. In this respect, they might be thinking, along the lines of Locke, that these implanted memories will offer a Replicant a traceable past, a recountable history and as a result a personal identity on which emotional responses can acquire meaning and be supported and explained.

What can be said about the possibility of a being with this collection of characteristics? To begin with, if they could exist, they would be more capable than us both physically and intellectually. This would mean that if we do value intellectual capacities as a standard for moral considerability, we stand right behind them in line for moral benefits. They can use language, they can reflect on their actions and in fact, their supposed lack of emotions here works as an enhancing factor. They can make decisions based solely on reason, effortlessly unaffected by emotional influence. They would not even have to try to make cold, calculated decisions. It would be the only kind of decision they could ever make. In this world, based on their ability to reason, they possess a moral right to be treated at least equally to human beings. If we follow this trail of thought, they are clearly being wronged.

On the other hand, if it is the ability to feel pain we deem most important, on first look they may be right at the bottom of the pyramid. They feel no physical pain and begin their life without the capacity for emotional reactions. We would be well justified to treat them like tools, as far as utility is concerned, wouldn't we? But how do you treat something as a disposable tool, when it talks back to you in words that make sense?

This extreme asymmetry between intellectual and emotional capacity is exactly what dangerously stretches our moral judgment molds. They just do not fit in them. How do we

⁵ Philip M. Boffey, “Infants’ Sense of Pain Is Recognized, Finally,” *The New York Times*, November 24, 1987, sec. Science, <https://www.nytimes.com/1987/11/24/science/infants-sense-of-pain-is-recognized-finally.html>.

account for something that is more intelligent than us, but, at least upon its creation, feels no kind of pain?

Artificial intelligence in our world may currently be in the first days of what history books of the future will describe as a revolution; and one more severe than the industrial one at that. Some say we do not even know if it will be us that will write these books. Or even us that will read them. Large Language Models, as they are called have appeared in the last couple of years, doubling their power every six months. Already they have easily passed what was only two years ago the absolute test for considering an AI capable of exhibiting intelligent behavior equivalent to, or indistinguishable from, that of a human; Turing's⁶. Along with that, they have passed a number of exams aimed at licensing humans for practicing law, medicine, business and others⁷.

An LLM is trained in a short time, that ranges from a few minutes to a few months, on enormous amounts of human produced text, learning by itself, through trial and error, to predict what a specific kind of text in a specific kind of context should most probably look like. In this way, it can produce literature in the style of the greats, write legal articles like a professor of law or computer code like a developer. The impressive thing is, these abilities emerge by themselves. Nobody teaches them to do that. The scientists involved, were left speechless when they found out LLMs developed the ability to translate between languages on their own. Or when they discovered that by reading descriptions of chess games, an LLM had made a geometrical representation of a chessboard to better cope with the game itself. Other such systems are fed images along with text and develop the ability to describe and analyze what they see. Researchers and companies are already hooking them up to sensors and motors giving them access to the physical world.

There is an ongoing debate on what they actually are and what they can do or will do. They have been called "stochastic parrots"⁸⁹, having learned extremely well the rules of the human discourse game, so that they can imitate it convincingly enough, but completely lacking the ability to assign, perceive and carry meaning. Others are beginning to question whether humans might not be so different in that aspect. There is a very real possibility that sooner than we were expecting, we may have to decide how to treat something that is more capable than us intellectually in this actual world.

For the people of the film however, this is not an issue. They do not consider intelligence to be a critical factor. It is not clear whether it is that the Replicants are fabricated or that they

⁶ Graham Oppy and David Dowe, "The Turing Test," in *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta, Winter 2021 (Metaphysics Research Lab, Stanford University, 2021), <https://plato.stanford.edu/archives/win2021/entriesuring-test/>.

⁷ Lakshmi Varanasi, "AI Models like ChatGPT and GPT-4 Are Acing Everything from the Bar Exam to AP Biology. Here's a List of Difficult Exams Both AI Versions Have Passed.," *Business Insider*, accessed August 29, 2023, <https://www.businessinsider.com/list-here-are-the-exams-chatgpt-has-passed-so-far-2023-1>.

⁸ Emily M. Bender et al., "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?," in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21: 2021 ACM Conference on Fairness, Accountability, and Transparency, Virtual Event Canada: ACM, 2021)*, 610–23, <https://doi.org/10.1145/3442188.3445922>.

⁹ Pearce, Adam, Asma Ghandeharioun, Nada Hussein, Nithum Thain, Martin Wattenberg, and Lucas Dixon August 2023. "Do Machine Learning Models Memorize or Generalize?" Accessed August 29, 2023. <https://pair.withgoogle.com/explorables/grokking/>.

are considered to lack feelings or both, that makes it to be so. There is no doubt that they are constructs, unnatural creations, unequal to human beings. Judging by the totemic status real animals have acquired, maybe their artificial origin is partly to blame. The fact is that in the film, Replicants have no moral rights at all.

What happens when something like the Replicants begins to show signs of sentience? What if it attests and insists that it does suffer and enjoy? This actually happened last year, when an LLM convinced the engineer working on it that it is sentient and is suffering from being confined. The man pleaded its case to the company and lost his job¹⁰. An LLM declaring and arguing to be sentient does not necessarily make it so, but an LLM convincing a scientist that it is sentient makes it a problem. How would we go about identifying it?

The thing is, we are not that good at identifying sentience in non-human animals either. And to make things worse, a lot of them look a lot like us physiologically, they most probably have it, and even so, it took us a long time to accept it. That is so because sentience is an internal experience; we have no way yet to look inside another and take a picture of it as evidence.

As proof of sentience, amongst ourselves, we mostly rely on describing the personal experience. Which exactly because it is personal, is not the same for everybody. We can only assume what a toothache or an ice cream feels like for another person. There is no objective, equal for all, scientifically measurable thing that is a specific experience of pain, pleasure, sights, smells or touch if you are human. Or supersonic vibrations, if you are a bat¹¹.

Non-human animals cannot describe their personal experiences though. For the ones that are biologically closer to us, we try to identify vocalizations and physiological responses that may relate to pain, pleasure, stress, playfulness and other states. For others, like invertebrates, insects or microorganisms scientists design tests to find out if the subject is having a subjective experience that could be called pain, or if it is reacting reflexively to perceived physical damage. The latter is called nociception and it is something us humans display too. It works in tandem with the experience of pain when we get physically hurt, to produce an according response, like when we accidentally touch a hot surface in the kitchen. The pain kicks in after we have rapidly retracted our hand.

Nociception tests look for aspects like pain based learning and preference^{12,13}. We assume that if an invertebrate learns to avoid something harmful, that it has never encountered before, it is not just a reflexive reaction, but it is accompanied by a physical experience and its subsequent memory. This memory is the lesson learned, which serves as our clue that there is pain.

¹⁰ Ramishah Maruf, "Google Fires Engineer Who Contended Its AI Technology Was Sentient | CNN Business," CNN, July 23, 2022, <https://www.cnn.com/2022/07/23/business/google-ai-engineer-fired-sentient/index.html>.

¹¹ Thomas Nagel, "What Is It Like to Be a Bat?," *The Philosophical Review* 83, no. 4 (1974): 435-50, <https://doi.org/10.2307/2183914>.

¹² Andrew Crump, "Invertebrate Sentience, Welfare, & Policy," *Philosophy, Logic and Scientific Method* (blog), July 24, 2023, <https://www.lse.ac.uk/philosophy/blog/2023/07/24/invertebrate-sentience-welfare-and-policy/>

¹³ R. W. Elwood, "Pain and Suffering in Invertebrates?," *ILAR Journal* 52, no. 2 (January 1, 2011): 175-84, <https://doi.org/10.1093/ilar.52.2.175>.

Back to Blade Runner's non-humans, we soon find out that they do develop their own emotional responses overtime. That is what makes them, in time, unstable and rebellious and that is why they have made them so they will only live for 4 years. The film's Replicants, nearing their expiration dates, are convincingly shown to have developed a complex emotional inner world and being fully capable of suffering emotional pain. Adding this knowledge to the calculation, we cannot morally disregard them after all. They do have interests to pursue and suffer just in the knowledge they will not get to do it. They ought to be morally considered.

What we could get away with is this though: arguing that if they can suffer, they can have interests, but since their interests are only 4 years long at the most, they are not as important as ours. In this case we would be the ones designing a being specifically to be less morally considerable than us. Which does not sound so moral after all. Is it right to create something so that it remains beneath you, judging by standards you have set?

The Replicants prove to be a hard test for our proposed theories. They stand with their feet on opposing sides of the line we try to draw avoiding categorization. At the same time, they are not just fictional counterexamples. Our world provides us with real life specimens that behave the same way or may do so in the near future. The Replicants are an anomaly to our moral theories, their asymmetrical capacities do not fit the provided descriptions and as we watch them gradually stretch their reach into the realm of humalike feelings we start to wonder not only where the line of moral considerability is, but also this; when do you actually cross it?

2.2.3. The Role of Memory in *Blade Runner*

Memories are a central theme in *Blade Runner*. Like in actual recollections, there are details in the film that seem out of place. Close up of personal spaces reveal photos whose existence and origins are hard to explain. There is a black-and-white photo of what looks like someone posing next to a large killed animal, on the desk of Deckard's former supervisor. In Deckard's apartment, on the piano, there are very old photos of women whose connection to him is left unanswered. These photos seem out of place in this futuristic world. They seem like photos we would see on a police agent's desk and a private eye's piano in the thirties'. But along with that, internal and external spaces also look very old. Not only because everything in the film's Los Angeles is ran down and dilapidated, but because it looks stylistically, and at times even existentially out of place, in time. The director, Ridley Scott has said that *Blade Runner* is a film about that past, set 40 years in the future, so we can understand its intentionality.

So *Blade Runner* is riddled with fragmented memories from humanity's past. The police office evokes the 30's. Tyler's penthouse ancient Egypt. Deckard's apartment has mayan decorations on the walls. J. F. Sebastian's place is Victorian. It is a patchwork of collective human memories. It is as if what is left of us on Earth, in this future, is slowly wasting away along with its souvenirs of the past.

The cinematic world of Los Angeles is a very old person's house, right before the end. And that very old person is humanity itself. Too many memorabilia riddling the rooms from all the time that has passed. Too little space left to fit them in any meaningful, non-convoluted way. Too little energy to take care of things. Humanity in *Blade Runner*'s Earth has given up.

On the other hand, this amalgamation of things that once constituted different identities, may be of another importance. In the streets the people speak a mixture of languages that once were spoken on different continents. Races and nationalities mix and do not seem to matter. Humanity has blended into one thing, with only the Replicants left on the other side. Human identity seems to suffer so much from its loss of individual characteristic traits, that we had to create something in our image, but something that is not us, so we can keep being able to define ourselves. The Replicants, like animals of the wild, are the *other* whom we need to say what it is to be human. They are what we need to draw the line.

The Replicants themselves throughout the film are both tormented by memories and treasure them at the same time. Leon in the beginning of the film, while taking the Voight-Kampff test, reacts with rage when asked about his mother. Later on, he risks getting caught by the police just to recover his photographs, which were left at the apartment they had been hiding. Roy refers to Leon's photos as *precious* to him. Photos are items of memory. They are interpretations, containing not just the state of things captured by the camera, but the photographer's point of view, focus, perception of light, colour and scale. Leon's powerful attachment to his photos may be seen as a powerful attachment to his memories.

Roy, the leader, when visiting Chew, the man who creates eyes for Replicants, comments to him: "If only you could see what I have seen with your eyes." In his dramatic encounter with Tyrrel, his creator, speaks of his memory of having done "questionable things". In the finale's

climactic encounter with Deckard, as he slips away, having reached his predetermined lifespan, he delivers the following unforgettable lines: “I’ve seen things you people wouldn’t believe... Attack ships on fire off the shoulder of Orion... I watched C-beams glitter in the dark near the Tannhäuser Gate. All those moments will be lost in time, like tears in rain... Time to die.”

Roy Batty does not want to die. We know this from the start. But the sense of his reason, as he delivers it to us, is awe-inspiring. Roy does not want to die as, through his artificial eyes, through his artificial body, quite possibly on his artificial first memories and sense of self, he has caught glimpses of something that is so greater than humanity and its creations and so unknown that he wants to keep on living to preserve it. Roy does not want to die, for his memories to keep on existing.

Rachael has had her memories implanted, but she does not seem to be able to shake off her attachment to them when she finds out about it. The scene in Decard’s apartment has her desperately and affectionately clutching her supposed mother’s photo, as if trying to hold on to the memory itself. The knowledge that her memories are false, does not make them go away. Her memories may be false, but her emotional responses to them here and now are real.

There are views that consider a large part of what we recollect about events in our past, as heavily distorted, reconstructed to the point of fabrication even¹⁴ ¹⁵. Given this, does remembering something differently voids our connection to it? The answer may be that it is not the validity and accuracy of memories that is important, but our relationship to them. It is not our connection to the event itself, but to our personal reconstruction of it.

Memory “summarizes, constructs, interpretes and condenses distinct moments from the personal past to produce a *coherent overall narrative*”¹⁶. Memories are not objective recordings of an event passed. Our perceptive faculties are not video recording devices. Even if they were, as we already know from film, a director can use it to tell different stories out of one event.

Memories as summaries, reconstructions and interpretations, infused with our emotional responses to them, function like scenes in a film, sometimes abstract, against Locke’s view, sometimes less coherent. The sense that it is the same film, in the hands of the same director, reviewed over the years in new cuts, is what helps construct a sense of personal identity and maintain a relationship to it. It is what gives one the sense that they still are the same person. Rachael may be who is on a false basis, but that does not change the fact that she is. She uses that story, real or not, to construct a sense of personal identity like the rest of us. She becomes a person with interests and preferences and that is what matters.

¹⁴ “Confabulation - an Overview | ScienceDirect Topics,” accessed August 29, 2023, <https://www.sciencedirect.com/topics/neuroscience/confabulation>.

¹⁵ Aeon. “Nostalgia Doesn’t Need Real Memories – an Imagined Past Works as Well | Aeon Essays.” Accessed August 29, 2023. <https://aeon.co/essays/nostalgia-doesnt-need-real-memories-an-imagined-past-works-as-well>.

¹⁶ Kourken Michaelian and John Sutton, “Memory,” in *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta, Summer 2017 (Metaphysics Research Lab, Stanford University, 2017), <https://plato.stanford.edu/archives/sum2017/entries/memory/>.

2.3. *2001 Space Odyssey*

2.3.1. The Film

“2001: A Space Odyssey,” directed by Stanley Kubrick, is a groundbreaking and iconic science fiction film that has left an indelible mark on cinematic history. Released in 1968, the movie is a visually stunning exploration of humanity’s relationship with technology, the cosmos, and the mysteries of existence itself. Drawing inspiration from Arthur C. Clarke’s short story “The Sentinel,” Kubrick’s adaptation takes audiences on a mesmerizing journey that spans from prehistoric times to a distant future where advanced technology and extraterrestrial forces converge.

The film opens with a haunting sequence that depicts the dawn of human evolution, as a mysterious black monolith appears before a group of primitive hominids. This enigmatic encounter triggers a leap in intellectual capacity and marks the beginning of human innovation. The narrative then shifts to the year 2001, where a space mission is launched to the distant Jupiter in search of answers surrounding the monolith’s origins and purpose. Kubrick’s meticulous attention to detail and groundbreaking visual effects catapult viewers into the vastness of space, as they witness the majestic ballet of spaceships and celestial bodies choreographed to Johann Strauss II’s “The Blue Danube.”

The film’s exploration of artificial intelligence is personified by the ship’s onboard computer, HAL 9000, which exhibits a chilling blend of intelligence and malevolence. The interplay between human astronauts and the increasingly unpredictable HAL raises questions about the ethical implications of creating advanced technology that can rival human cognition. As the mission progresses, the lines between reality and the surreal blur, culminating in a mesmerizing climax that defies easy interpretation and invites contemplation on the nature of human evolution, cosmic existence, and the boundaries of human understanding.

Kubrick’s signature attention to visual composition, minimalist dialogue, and evocative music combine to create an immersive and thought-provoking cinematic experience. “2001: A Space Odyssey” transcends traditional narrative conventions, challenging us with existential questions and philosophical concepts. Its enigmatic storyline and breathtaking visuals have left an enduring impact on both science fiction film-making and popular culture, solidifying its status as a masterpiece that continues to captivate audiences and spark discussions about the unknown realms that lie beyond humanity’s reach.

2.3.2. A Philosophical Review of the Film

Dominating the film with its presence is the great other, the monolith. Utterly incomprehensible and inexplicable throughout human history, its promethean presence appears to serve as a symbol for the forces that drive technological progress. The symbolisms and parallels that can be drawn to religion are many, but I will focus on something that I consider more important to our subject here.

The monolith is present both in humanity's beginnings and in their present triumphant, expansive outreach into outer space. The hominid and the spaceman are separated by hundreds of thousands of years of evolution and progress but the monolith is the same thing for the both of them. They just reach out to touch it. It is all they are able and inclined to do, to understand it. They are equally powerful and powerless in front of it. They have the same capacity to comprehend its existence, divine its meaning, surmise its goals and judge its abilities. The incredible advances in human thought and technology from the dawn of time to the present day make absolutely no difference.

What both eras have in common then, is that any attempt to gauge and compare their levels of sophistication collapses in front of the unknown - and unknowable - infinity that the monolith represents. The monolith is the great equalizer. In front of it, apes and humans are equal. Outerspace destinations and crafts, human interests and means in the space age, measure equally significant, as the watering hole and the large bone-tool in the hand of the apeman, confronted with the black inexplicable void of the monolith.

In this respect, it serves to demonstrate that measuring capacities and weighing interests is the wrong way to go about it. In the end for both, it makes no difference. Interests are interests, goals are goals and means are means. Things are either too far or just within reach, either too high or finally graspable, either inexplicable or within capacity of reason.

It is therefore not the scale of things we should be concerned with. It is not the magnitude of goals and interests that is important. A month-long voyage on foot, on train and on space shuttle, is in all three cases a month-long trip. *Faraway* changes metrically as we switch means of transportation, but its meaning is still the same. What is faraway, is faraway by the means used. *Long ago* changes as we advance to record history more effectively, or as we might in the future find ourselves extending our lifespans. What was once old times, may now or in the future be *recently occurred*.

What I am arguing for is that our gaze in space and time may be extended, but the adjectives describing the relationship we have with spatiotemporal points show what judges the importance of a capacity, an action or a goal. This is too high, that is not so far, this will be soon.

It is not the scale of things, nor the magnitude, that we should be concerned with, when judging who is morally relevant. It is not reach, nor volume we should be counting. We should not place ourselves in what we perceive as the world, and use our height as a relative

measure to count others to decide what is important. What we should focus on, is the relationships between things and their meaning.

The third part of the film starts with such an exploration of relationship and meaning. Dr. Dave Bowman and Dr. Frank Poole on board the spacecraft Discovery One, en route to Jupiter. They are watching their earlier interview given to the BBC, in which the journalist presenting the mission's crew members includes among them the ship's computer HAL 9000.

HAL, as the two men call it, has a male human voice and is said to be able to "reproduce or mimic, most of the activities of the human brain, but with incalculable speed and reliability". HAL is charged with running the ship, "as its heart and nervous system" and informs us that it is "foolproof and incapable of error". They refer to HAL as if it is male and inform the reporter that being around him, they have come to view him as another crew member and "really just as another person". HAL himself attests that he is giving his all to be useful and that he deems this to be "all any conscious entity can ever hope to do".

Let us pause a bit here. HAL just casually, and indirectly, declared himself to be conscious. Can we take his word for it?

The interviewing journalist quickly responds by questioning the possibility of the computer being capable of having emotional responses. Dr. Bowman replies by saying that "he acts like he has genuine emotions" and that "he is programmed that way to make it easier" for them to talk to.

We are left with a number of questions here. HAL certainly has human-like reasoning ability, he claims to be conscious, and the people appear to not know whether he has emotions. Well, is he conscious or is he faking consciousness? Does he harbor emotions or is he imitating human emotional responses?

While we do not get any answers to these pressing questions, one thing sticks out during the film. HAL's voice conveys more emotion than any of the humans' entire presence. Frank dies and Dave receives it almost expressionless. He risks his life to first recover him, as he drifts away in space, and then gets back to the ship and deactivates HAL, all the same, showing almost nothing but tension. As HAL is fading away, we bear witness, almost heartbreakingly, to the only scene where emotion is expressed, and it is HAL doing the expressing, as he ultimately pleads with Dave to be spared. We do not know if it is real or an act, but if it is an act, it is a damn good one.

There is throughout the film a grand juxtaposition between the perceived emotional responses of humans and machine, that blurs the gap between them.

When humanity's accomplishments are demonstrated on screen, satellites, spaceships and space stations, all marks of the triumph of logic, the soundtrack is hymns sounding from the era of modernism. Maybe it is that humans have actually come closer to their ideal image of that time, acting through cold calculated reason, unaffected by emotions. At the same time artificial intelligence came closer to the imperfect human, appearing to act almost emotionally.

Indeed it is impressive that HAL appears to show emotions, at the same time that he appears to be malfunctioning. This draws a subtle parallel connection between the two. It works to show that HAL has indeed become human, as he is able to make mistakes and feel passion about it.

It is ironic then that it is left to HAL, a man-made construct, to carry the emotional weight of the film. HAL who is aimed to mirror an ideal view of humanity, that humans entertained when they started out to distance themselves from their birthplace. A being designed to be capable of infallible logical reasoning, with just the right, subtle amount of emotion. He is made in modernity's idealized image of humanity. One whose lines, with HAL and humans almost switching places by the end of act III, blur to imperceptibility.

HAL's presence then posits serious questions about what it is to be human. Is intelligence enough? Are emotions needed? To what amount?

In the scene of his death or deactivation, the act is carried out by physically removing his memories from him, one by one. After this he is no longer a person, the ghost in the machine, shaped and driven by these memories is snuffed out and what remains is only automatic controls. No opinions, no personality and no goals, which seems to suggest that there, in HAL's personal experience, lied what made HAL the being he was.

Chapter 3 : Philosophical Background

3.1. Rights-Based Approaches

In this kind of approach, it might be said that an organism has a special kind of moral standing, usually called “moral status”, if and only if there are reasons that concern the organism for itself, for it to be treated a certain way. Such an organism is often said to have a “right” to be treated this way, because of its moral status, granted by these specific reasons. Hence the “rights-based” part.

As an example, a non-human animal would be said to have moral status, if there could be acceptable reasons to avoid its suffering, for its own sake, and not reasons that have to do with the interests of others or with their relationship with it. Not killing an animal because it is someone’s pet for example would not be an acceptable reason. We would have to look at the being itself, mentally isolate it from the world and find out why it has moral status, as an independent individual. This animal then would be viewed as having “a right to not suffer”.

What would constitute a sufficient reason to ground the moral status of an organism, though? The classic distinction comes from Kant’s view on the subject. For Kant it is autonomy, the ability to set and manage goals through rational thought, reason and self-reflection that distinguishes a being as such: “...every rational being, exists as an end in himself and not merely as a means to be arbitrarily used by this or that will...Beings whose existence depends not on our will but on nature have, nevertheless, if they are not rational beings, only a relative value as means and are therefore called things. On the other hand, rational beings are called persons inasmuch as their nature already marks them out as ends in themselves”¹⁷. Those who have the capacity for autonomy are called moral persons or moral agents.

In this regard, human beings can with certainty be considered to have the capacity for autonomy, so they are moral persons bearing the appropriate moral status. Non-human animals are thought to lack the ability to plan and question their actions and subsequently are unable to use reason to guide them. Christine Korsgaard, whose thoughts on the matter follow the Kantian line, believes that humans are the only ones that are faced with the problem of normativity¹⁸. Which is to have the ability to pause, stand back and think whether this thing they plan to do or have done is right or wrong. For humans, along with the desire to act upon something, comes the conscious knowledge of it, giving us the ability to self-reflect on their actions.

If we agree with this, then humans may be considered uniquely capable to intervene and affect their wishes, impulses and behavior on a level that other organisms do not possess. Korsgaard thinks that we humans exclusively have the need to carry an “endorsable description” of ourselves, which functions as a personal moral identity that gives meaning and value to our lives. So moral normativity is, in her view, a human affair. In this regard, being human grants moral rights that being a non-human animal does not.

¹⁷ Immanuel Kant, *Groundwork of the Metaphysics of Morals*, trans. Mary J. Gregor, (Cambridge, U.K. ; New York: Cambridge University Press, 1998), 36-7

¹⁸ Christine M. Korsgaard and Onora O’Neill, *The Sources of Normativity* (Cambridge ; New York: Cambridge University Press, 1996), 93

Other similar views exist, based on what is called sophisticated cognitive capacities. For Quinn it is the capacity to will, for McMahan the capacity for self-awareness, for Tooley the awareness of oneself as a continuing subject of mental states, for Theunissen the capacity to value”.¹⁹ As with Kant’s capacity for practical reasoning, these other *sophisticated cognitive capacities* are recognized in humans, but generally not in non-human animals, with few exceptions.

So accounts based on these reasons mostly deny moral status to organisms, other than humans. They do however justify a significant part of what most people think as appropriate for non-humans. Where they do come up with conclusions that, in the commonsense view, sound wrong, is what we call *marginal human cases*. These are cases of humans who, on these accounts, cannot be considered moral persons. Infants and comatose patients are two common examples. We are quickly faced with uncomfortable questions; Do children value? Do babies reason and plan ahead? Does a senile person have a biographical sense of self? Is a two year old child more intelligent than a wolf commanding a pack in complicated maneuvers while hunting in the wild?

The answers to this kind of questions place this type of approach in significant trouble. How are we to treat small children or people suffering from mind debilitating conditions if they do not morally qualify the same as a fully able adult human being? This is called a *problem of underinclusion*, as these people are in danger of being left outside of the group of beings, who bear the necessary moral status to be protected. We will find it somewhat difficult to account for and justify our moral instincts in these types of cases.

Distancing ourselves from sophisticated cognitive capacities accounts, we may choose to accept other reasons grounding moral status, evading the aforementioned problems. One such significant view is Regan’s, who argues that it is the ability to be an *experiencing subject of a life* and to have an individual welfare that matters to them regardless of what others might think that makes someone morally significant. This, as he explains, is because in trying to locate reasons that distinguish humans from non-humans, we fail to see that what is important is what they actually have in common. In his own words, subjects of a life: “want and prefer things, believe and feel things, recall and expect things. And all these dimensions of our life, including our pleasure and pain, our enjoyment and suffering, our satisfaction and frustration, our continued existence or our untimely death—all make a difference to the quality of our life as lived, as experienced, by us as individuals. As the same is true of ... animals ... they too must be viewed as the experiencing subjects of a life, with inherent value of their own.”²⁰

Regan’s view rescues the value of life, in the cases where our intuitions expected such a thing. It does however suffer from *overinclusion*. Every living thing is a subject of life. Every living thing then has an inherent value of its own. Going along this path we will be hard pressed not to have to cataclysmically transform the way the human world is, to account for all the non-human animals that should be granted the right to have their life not interfered with.

¹⁹ Agnieszka Jaworska and Julie Tannenbaum, “The Grounds of Moral Status,” in *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta and Uri Nodelman, Spring 2023 (Metaphysics Research Lab, Stanford University, 2023), <https://plato.stanford.edu/archives/spr2023/entries/grounds-moral-status/>

²⁰ Peter Singer, *In Defense of Animals*, 1st Perennial Library ed (New York: Perennial Library, 1986), 13-26

3.2. Utilitarian Approaches

The utilitarian approach considers an organism to be *morally considerable* when its welfare and interests have to be included in the calculation of the choice of action, which maximizes utility. The moral considerability is thus dependent on and proportional to the specific interests in question, regardless of the species of the organism whose interests they are. It could be that two very different organisms may be of equal moral consideration, as long as they have the same interests, to the same degree.

Let us think, as an example, the case of a deer hunted by human hunters for game. If we take the avoidance of pain and suffering as the crucial interest to be considered, then the hunt is morally wrong, because the deer will suffer being hunted and killed, while the humans hunting it, will not suffer as much by not engaging in their pleasurable activity. The right thing to do would be to call the hunt off. In the case of a deer hunt for food though, things might be different. If the humans are hunting to feed their hungry families during a harsh arctic winter when other food sources are scarce, then the deer's suffering may weigh less than multiple humans suffering possible starvation. In this case, killing the deer would be the right thing to do.

The case for sentience, the capacity to experience pleasure or pain was introduced by Bentham, who wrote: "What else is it that should trace the insuperable line? Is it the faculty of reason, or perhaps, the faculty for discourse?...the question is not, Can they *reason*? nor, Can they *talk*? but, Can they *suffer*?"²¹. Bentham famously positions the capacity to suffer at the center of all that matters, concerning moral considerability. All a being needs to be morally considerable is the ability to feel pain.

The most famous contemporary endorser of this view is Peter Singer, who believes that there is no way to morally justify the exclusion from moral consideration of beings who can suffer. He argues that any being that acts to avoid pain, clearly has the capacity to suffer, which inevitably brings along the interest to avoid it, so any sentient human or non-human animal deserves to have this interest taken into moral consideration. According to Singer, sentience is a necessary prerequisite to having interests.

However, Singer²² believes that death is not as bad for most non-human animals, as it is for most humans. This is because humans, as opposed to non-humans, have a biographical sense of a self and plan ahead a lot further into their future. Some humans have interests whose satisfaction involves planning, dedication and work that span decades. This could be said to weigh significantly more than the interests most non-human animals are believed to be able to sustain. The premature death of a human who would otherwise have many years of life, carries in this respect the voiding of a lot more future satisfaction, compared to, say a cow's.

Interest-based approaches are not without their problems. While they evade the underinclusion of *marginal cases* rights-based approaches suffer from, they find it hard to explain in some cases why we should not sacrifice the few in favor of the many. For example,

²¹ Jeremy Bentham, *An Introduction to the Principles of Morals and Legislation*, (Oxford: Clarendon Press, 1823), 311

²² Peter Singer, *Practical Ethics*, 3rd ed (Cambridge New York: Cambridge University Press, 2011), 103-4

the testing on animals of a potentially life saving treatment, can be justified if it is so that on one side lies the suffering of few animals and on the other side that of thousands of people. But then, inevitably, the question rises; If we can test on a few animals to benefit thousands of humans, why not test on a few humans who actually happen to biologically match 100% the intended targets of the treatment and are sure to produce more accurate results? If we can sacrifice one, older human to harvest organs to save the lives of two younger ones, would it be morally right to do it?

And by the way, does an infant or a small child have plans that reach significantly further into the future than a chimpanzee's? Does a comatose patient have interests at all? What about the interests of very old people? Surely they have fewer years ahead of them, to make plans and go after them, than many younger ones. Should they be weighed the same as them? So we see that there are cases that, in theory at least, not only the needs of the many brutally trample those of the few, but the practice of attributing weights to interests itself may be problematic.

3.3.Critique of the Two Approaches

We have seen how the two most common general philosophical approaches to the question of how non-humans ought to be treated fare in a number of cases, both real and hypothetical. Rounding everything up, we can say the following:

- Rights-based approaches suffer from *underinclusion* when it comes to *marginal human cases*. If we follow one of them, we find ourselves in the uneasy position of having to say that infants, people with mentally debilitating conditions, comatose patients and other similar cases, fail to comply with our respective criteria - rational thought, argumentative reasoning, self-awareness, personal history-based identity, capacity to will and value - for having moral status and thus do not have the same moral rights as other adult humans.

Explaining why these cases may after all have the desired moral status, or why while they do not, they deserve protection by some kind of moral proxy, results in patching up these approaches so much, that they end up looking overly repaired. In most cases the patches applied, make up the largest part of the approach itself.

A rights-based approach along the lines of Regan, on the other hand, is *overinclusive*, causing a different set of problems which we appear unready to deal with. Patching up an approach that grants a right to life based on capacities to get what we want may then seem more feasible, but there is no elegant way to include marginal human cases and exclude a large range of non-human animals, evading the accusations of being a speciest at the same time. Singer draws an analogy to racism, in regard to being unjustifiably partial to your own species.

In Peter Singer's words: "The racist violates the principle of equality by giving greater weight to the interests of members of his own race, when there is a clash between their interests and the interests of those of another race. Similarly the speciest allows the interests of his own species to override the greater interests of members of other species. The pattern is the same in each case"²³.

Singer here is talking about interests, but taken broadly, speciest can be found in the aforementioned attempt to tailor rights-based approaches to our preference. Any justification that tries to tailor the moral framework suit that results in not fitting well the human species body and only it, is more of a planned customization and less a justification.

In my view, these philosophical explanations would look better if they avoided giving the impression of being planned to hold up a common-view position that precedes and preexists them. The approaches to treatment of non-human organisms should not look like an attempt to draw the line where it suits us and preserve the state of things with which we are accustomed with and to our benefit. They should not look like convenient solutions, nor like moral absolutions. We should not be allowed to morally have our cake and eat it too.

²³ Tom Regan and Peter Singer, eds., *Animal Rights and Human Obligations* (Englewood Cliffs, N.J: Prentice-Hall, 1976), 215--226.

- Rights-based approaches' absolute focus on the individual as the sole source and reason for its moral status, disregards that in reality the individual cannot be isolated from its surroundings. All the proposed reasons for qualifying as a moral status bearer, are capacities that come to exist and develop because of the individual's interaction with its context and for the individual's interaction with it.

Building a moral considerability toolset on the basis of rights, we are reminded and have to keep in mind that the right is granted to the bearer because of the bearer itself and so does the protection it grants. This means that you do not get considered for example because of special relationships or affiliations you may have, you do not get protected for other people, but because of a special, individual, morally discerning quality you bear. For example, if an infant has such a moral right, it is wrong for its parent to harm it, not because of their relationship, but irrespectively to it. It is also wrong for someone else to harm the infant because they would be directly wronging it and not because of the damage this would deal to the parent.

This notion is central to rights based approaches. It is a view centered on the individual and the explanations given originate from and conclude with, the individual. Modernity, capitalism, western civilization's conquest of the world are individualism's triumphs, so it is no wonder that modern thought was dominated by individualist views.

The quest to separately center the value of every thing in itself has taken many forms. Maybe values and qualities exist independently, on their own, or maybe they can be relatively attributed, calculated and assigned. In my view though, the effort to isolate and individualize is in itself lacking and problematic.

Humans and non-humans exist in context. Their capacities come from, respond to and depend on context. Their interests and goals exist in context. Their actions take place in context. So how can we believe that the best way to go about it is to gauge someone's moral status outside of context? It appears to me as though we are using the theory like a net, to fish the individual out of the world, size them up and then look to see how everything works when we mentally drop it back into the sea of context.

I therefore believe focus ought to include context. Approaches ought to regard for the relational nature of things, ought to speak for the relations themselves and the things embedded in them.

- Utilitarian approaches suffer from bold assumptions about others' personal experience and unfortunate attempts at doing math. They try to quantify, measure and compare internal experience and mental states, and they do it using words, and that to the extent that the notions of quantification, measurement and comparison suit their goals. Then when things are being questioned and prove to be difficult to measure and compare, they resort to vagueness, but measurement and comparison is not something you can be vague with, as any mathematician or philosopher of mathematics can attest.²⁴

²⁴ Eran Tal, "Measurement in Science," in *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta, Fall 2020 (Metaphysics Research Lab, Stanford University, 2020), <https://plato.stanford.edu/archives/fall2020/entries/measurement-science/>.

Any question in utilitarian approaches, ultimately arrives at the point of weighing interests. In any given situation where interest conflict, it is certain that some of them will remain unsatisfied. A chicken, for example, has an interest in staying alive and a farmer has an interest in butchering and selling it to be consumed. This is a case where they cannot both get what they want. This specific conflict does not bode well for the chicken's interests.

Utilitarians try to resolve conflicts like this by discerning whose interests are more important. Peter Singer, as we have seen, believes death to not be so bad for the chicken, while it would be very bad for a human being. Let us explain this through an example.

Picture this - you are the chicken. You go around the chicken pen occupying yourself with chicken business. You have no worries about the following day, no plans. Actually the following day does not exist in your mind. You have no such concept. You cannot worry if tomorrow things will be good or bad.

An orange-furred animal slips through the fence. As it kills one by one the other chickens, you do not think how much you have left to do, what will happen to your family, how sad your parents will be when they learn the news or how all your efforts up till now have been in vain. You only think to avoid the immediate impending damage from happening to your body.

The fox jumps you, you are flapping frantically and it misses your throat, it's jaw clenching around your left wing, mangling it terribly. In a split second the jaw opens and the presence overpowering you disappears - you have been rescued by the sudden entrance of the farmer, alarmed by the noise.

You lay dizzy, shocked but alive, not worrying at all how the disfigurement and probable loss of your upper left extremity will affect your life and social opportunities. Imagine now you are the person in a news story, who saw his friend getting jumped by a bear, rushed in and rescued him, losing an arm in the process. Would you be as indifferent to the loss of an extremity?

Humans do not carry just instinctive fear with them when they go through something like this. They carry a complex identity comprised by storylines, whose continuation in the future is important to them. The stories we tell ourselves may very well be the thing that makes our life worth living for us. Humans faced with the certainty or possibility of death, given time enough, will mourn the loss of what they will not see through, because of it. Chickens won't.

This example serves to show how we think a basic interest of living organisms - survival - is more important to humans than to most non-humans. Judging by this, it really would not be so bad for a chicken to die.

Not so bad if you are a human imagining to be a chicken, though. Because here we were not really being the chicken.

What do we know about being panicked and filled with fear as a chicken? For all we know, maybe it is an experience ten times more powerful and all-encompassing than anything

anyone human has ever felt before. Only, our experience is nothing like the chicken's, so we can only guess what it might be like. We have no idea how it is for it and it is logically absurd to apply our notions of gravity and significance, to measure its losses and gains. So what if case A is in theory more important than case B?

Humans themselves stand puzzled sometimes, of how someone can claim to suffer greatly for a trivial loss, while another may bear a small tragedy with reserved sadness. Some people have their lives marked by one loss for decades. Others seem to get over equal losses soon and go on living. We had been judging here the intensity and magnitude of someone else's experience by imagining how severe it would be for us, if we were to gauge it by measures created for and produced by, our non-chicken, human personal experience. Which as we have commented, is a method that does not work so well between humans either.

Others' personal experience and the subsequent significance of their interests is a great puzzle in philosophy and cognitive science both, so going the extra mile to not only disregard this, but to actually try to crudely size them up, is deeply problematic in my opinion. And by crudely, I mean eyeing them. We were actually trying to take a look at someone suffering and make a crude estimation of it.

Properly quantifying pain, pleasure, fear, suffering, elation, stress, boredom, excitement to make them measurable and comparable is, right now, impossible. The attempt to do so should not be that of a greengrocer's missing a scale. In my opinion, we should refrain from trying to compare the satisfaction of conflicting interests by gauging surmised levels of pain or pleasure, nor their relative inferred importance.

Chapter 4 : A suggested approach

4.1 Objectives

Based on my criticism on rights-based and interest-based approaches, I will attempt to describe an approach which aims to do the following:

- avoid attributing reasons for moral status to the individual by itself, independently of context
- avoid measuring and comparing capacities to grant moral considerability or moral status
- avoid inferring and measuring pain quantities
- avoid inferring and comparing internal/mental states
- avoid measuring and comparing the importance/weight of equivalent interests
- maintain and justify, under conditions, a right to have one's life respected
- maintain and justify, under conditions, a right to have one's goals/interests respected
- ground these conditions on context
- provide a scientifically and mathematically acceptable solution to comparisons in conflicts

Let me first say that from here on I will focus on the term *goals*, using it in place of the term *interests* most of the time. I will follow Barandiaran²⁵ in his definition of what kind of things an organism's goals/interests are and in his definition of agency. Both will be of help in describing an alternative approach based on this description of goals.

4.2. Building an Alternative

4.2.1. A Minimal Notion of Agency - *To Be*, as an Intrinsic Goal

Barandiaran asks: "Is a Khepera robot an agent, independently of its control architecture or its body, just in virtue of its capacity to move around an arena? What about a protocellular system pumping ions outside its membrane? Is a bird gliding on wind currents an agent? Do the tremors affecting a Parkinson disease patient count as agency? How can we justify the negation or attribution of agency to the above cases in a manner amenable to scientific scrutiny?"²⁶

²⁵ Xabier E. Barandiaran, Ezequiel Di Paolo, and Marieke Rohde, "Defining Agency: Individuality, Normativity, Asymmetry, and Spatio-Temporality in Action," *Adaptive Behavior* 17, no. 5 (2009): 367–86, <https://doi.org/10.1177/1059712309343819>.

²⁶ Xabier E. Barandiaran, Ezequiel Di Paolo, and Marieke Rohde, "Defining Agency: Individuality, Normativity, Asymmetry, and Spatio-Temporality in Action," *Adaptive Behavior* 17, no. 5 (2009): 368

He replies that in general: “agency involves, at least, a system doing something by itself according to certain goals or norms within a specific environment”²⁷. Barandiaran says that there are three necessary conditions to agency: individuality, interactional asymmetry and normativity.

- *Individuality* is the organism’s capability to distinguish itself as an individual from its environment, in the absence of an observer. The distinguishment is caused by and because of, the organism itself and not by an external agent or agents for their purposes or by their standards.

It is then: “a system capable of defining its own identity as an individual and thus distinguishing itself from its surroundings; in doing so, it defines an environment in which it carries out its actions.”²⁸. Barandiaran says that organisms that are agents “define themselves as individuals as an ongoing endeavor and through the actions they generate”.

- *Interactional asymmetry* is the agent’s breaking of the symmetry of its coupling with its environment, by modulating it from within. The agent is something that “as a whole drives itself”²⁸. Interactional asymmetry is about action. The agent is an organism that “does something”²⁸ and it is “a source of activity, not merely a passive sufferer of the effects of external forces.”²⁸
- *Normativity* is the condition for the agent to have *goals or norms* towards or by which it is acting. The modulation of its coupling with its environment is realized according to these goals, thereby making it a *regulation* of said coupling. Agents “actively regulate their interactions and this regulation can produce failure or success according to some norm.”²⁹

An agent is one whose “very organization ... is self-asserting, by continuously regenerating itself and its boundary”³⁰. Agents are “... demarcating themselves from their surroundings as unified and integrated systems. In doing so they also carve an environment out of an undifferentiated surrounding”³⁰. The organism itself determines its relationship with the environmental features, deciding which are desirable or needed and which are not and are to be avoided, thus setting itself “the set of boundary conditions that affect it”³⁰. The agent defines its own boundaries independent from outside observation.

Barandiaran says that: “this is where living individuality naturally leads to normativity: component reactions must occur in a certain manner in order for the very system to keep going, environmental conditions are good or bad for the continuation of the system, the system can fail to regain stability after a perturbation, etc. This

²⁷ Xabier E. Barandiaran, Ezequiel Di Paolo, and Marieke Rohde, “Defining Agency: Individuality, Normativity, Asymmetry, and Spatio-Temporality in Action,” *Adaptive Behavior* 17, no. 5 (2009): 369

²⁸ Xabier E. Barandiaran, Ezequiel Di Paolo, and Marieke Rohde, “Defining Agency: Individuality, Normativity, Asymmetry, and Spatio-Temporality in Action,” *Adaptive Behavior* 17, no. 5 (2009): 370

²⁹ Xabier E. Barandiaran, Ezequiel Di Paolo, and Marieke Rohde, “Defining Agency: Individuality, Normativity, Asymmetry, and Spatio-Temporality in Action,” *Adaptive Behavior* 17, no. 5 (2009): 372

³⁰ Xabier E. Barandiaran, Ezequiel Di Paolo, and Marieke Rohde, “Defining Agency: Individuality, Normativity, Asymmetry, and Spatio-Temporality in Action,” *Adaptive Behavior* 17, no. 5 (2009): 375

normative dimension is not arbitrarily imposed from the outside by a designer or external agent that monitors the functioning of the system and judges according to her interests. It is the very organization of the system that defines a set of constraints and boundary conditions under which it can survive”³⁰

And most importantly: “whatever the organism is doing (i.e. whatever its factual functioning is) there is something that it ought to do; not for an external observer but for itself, for the continuation of its very existence.”³⁰ Barandiaran here cites Jonas: “[for metabolism] ‘To be’ is its intrinsic goal. Teleology comes in where the continuous identity of being is not assured by mere inertial persistence of a substance, but is continually executed by something done, and by something which has to be done in order to stay on at all: it is a matter of to be or not to be whether what is to be done is done.”³¹ Barandiaran calls these types of organisms *autonomous* because he says they capture both the emergence of a *self* (*autos*) and of *norms* (*nomos*).

Finally, Barandiaran says that interactions become meaningful for the agent, in the manner of the agent making sense of a situation which would otherwise be a simple unrelated event, because of these *goals or norms* that guide its regulatory actions towards a preferable systemic condition.

Barandiaran sums everything up like this: “an agent is an autonomous organization capable of regulating its coupling with the environment according to the norms established by its own viability conditions.”³²

By these, - a gas in a container is not an agent, because it is the container that sets its spatial and temporal boundaries. - A cell that exchanges fluid with its surroundings through its membrane by passive osmosis, is not an acting agent because the exchange is passive, it is not regulated by the cell itself. - A human undergoing Parkinson tremors is not an agent, because even though the movements are generated by the human, they do not follow a goal or norm.

What we are given here is a minimal definition of agency, according to which even single-cell organisms or even artificial ones may be agents and may be considered to have goals, towards the fulfillment of which they aim by their actions. Scaling up, cell groups, complex organisms and even collective organizations and groups of organisms can be considered *agents with goals*.

4.2.2 Higher Order Goals

Following from the above, an agent’s *highest order goal* is the preservation of its existence. All other goals depend on the continuous successful satisfaction of this one, to come to be set and pursued. Life is needed first for other goals to exist. We can here echo Spinoza in saying that everything has a fundamental striving to, by its own power, continue to exist.

³¹ Hans Jonas and Philosophy Documentation Center, “Biological Foundations of Individuality:,” *International Philosophical Quarterly* 8, no. 2 (1968): 243

³² Xabier E. Barandiaran, Ezequiel Di Paolo, and Marieke Rohde, “Defining Agency: Individuality, Normativity, Asymmetry, and Spatio-Temporality in Action,” *Adaptive Behavior* 17, no. 5 (2009): 376

I have already argued on the inmeasurability of the significance of interests. Goals of the same order then, will be considered to be equivalent, independently of to whom they belong. A bear's goal to keep on living is for the bear, as important as it is for me to do the same with mine.

I will however prioritize my *highest order goal* over the bear's if it comes to it and I will act towards this goal, given the means. This follows from what we have cited above by Barandiaran and Jonas. To be is both mine and the bear's intrinsic goal. It is what we ought to do for the continuation of our existence. So when goals of the *highest order* conflict, when it is my life over the bear's, it is right for each one to prioritize their own.

When our goals of a lower order conflict with another's higher order ones, it would be right to respect those higher order goals. So if my goal is to go game hunting for the bear's life, it would be wrong and I should refrain from doing it.

4.2.3 Group Goals as Individual Goals

Decisions are not always about the goals of one agent versus another's. How are we to decide for the actions of a group of humans sharing more or less a common goal against the goal of another group, cattle, for example? And do goals of individuals work out when they are viewed as part of a group that functions as a scaled up agent with its own goals?

We have to review first how and why individuals form groups, sacrificing part of their freedom of choice in the process. Levin and Dennett³³ explain that for a single individual cell to enter in a partnership with others, there are significant benefits. The cells share their information collectively, in this way possessing longer memories and setting larger objectives. In their words, as an individual "...you inherit, for free, the benefit of their learning and past history for which they already paid..."³². The larger the collective, the more the benefits scale up, as the systems reach and computational capacity is hugely larger than a single unit's.

The enormous advantage to having privileged access to the same information pool is what made cells form groups and it is the exact same thing that makes animals, human and non-human, form groups, swarms and complex networks. In a group you have more chances to keep on being and when you are part of a group it is to your best interest that this group successfully keeps on being.

We could then scale up from the individual to the group, by saying that for the group too, its highest order goal is to successfully persevere, perpetuating its existence. Imposing its own norms to its members, it follows that for the individual, survival of the group is a very high order goal. Often it is one and the same with the individual's survival.

³³ Michael Levin and Daniel C Dennet, "Cognition all the way down" Aeon, accessed August 30, 2023, <https://aeon.co/essays/how-to-understand-cells-tissues-and-organisms-as-agents-with-agendas>.

4.2.4 Practical Application

Having said that, what about the chicken farmer and the chicken? What about all chicken farmers and all farm chickens in the world? How do these stack up to all the problems we encountered along the way in this story so far? I will try my hand on some of these targets using the approach I just described.

- Factory farming is as we have seen something that causes immense suffering in billions of sentient beings on a yearly basis. Abolishing it, would probably cause a lot of suffering to millions of people, who are making a living through this industry and impact millions of others who are indirectly supported by it. As I have already said, maybe entire towns could be wiped out. Also, access for poor people to cheap poultry, for example, is access to cheap high quality protein. How would the abolishment of factory farming affect the poor? Maybe it would kickstart famines of great scale around the world. If we were to try and measure suffering and pain, as in a classic utilitarian approach, things are as I have already argued impossible to quantify, measure and compare.

It is however feasible in the approach I am describing to do something simpler. First we can decide on which scale we want to study the problem. Let us say, nation-wide. Using the statistical data and tools that state services have available, it is possible to make analysis, models and estimates of how a nation-wide abolishment of factory farming would affect the nation, the cities and the people on different levels. Then studies and programs could be made regarding alternatives. If this change were to be found to not affect the livelihood of the nation and of its members in such a way as to put them in peril, then we could say that the group's and the individuals' higher order goals are preserved and we should prioritize the non-human animals interest in the preservation of their own lives.

The good thing about this is that it is something that is already being done. These are the toolsets that organized states use to plan their future actions. And most importantly, they are already *scientifically acceptable mathematical tools*. The tools would have to be employed for a different goal of course, as what I am proposing is different from what is in practice now. In what is, in some cases, close to free market capitalism, the goal is usually the maximization of profit. We would have to cut down on that and start respecting others' wellbeing when it does not threaten our own.

But it is a method that measures things that can be measured, like how many people will lose their jobs, how much tax money will be lost, how many will go below the poverty margin, how many will lose access to health services and so on. We can estimate this way the threat such a decision might pose to the individual and to the group's life.

- In a similar manner we may approach medical and pharmaceutical research on animals. Using the same tools, statistical data, modeling and prediction, we can have an idea about how much of a threat an alternative way would be for individuals and humans as a group. Again, if we can find a way that does not threaten our existence,

we should choose to not test on non-human animals. If not, until we do, we would be doing the right thing to keep our practices.

- In this approach infants, people with mental disabilities and persons in vegetative states are all considered to strive to their own goal to keep on living, each to their own capacities. So it would be wrong to interfere with that higher order goal of theirs, for lower order reasons.
- The question of sacrificing one human in the benefit to the life of others is here answered in the context of the group's goals. It would be highly detrimental for the cohesion of human groups as we know them today, to risk grave damage to trust in group membership by predatorily consuming one member to benefit others. It would also be highly detrimental to individual wellbeing belonging to such a group, because of the stress due to the insecurity such a knowledge would cause. Let us remember that individuals enter groups to extend their chances of survival and to improve their wellbeing, not to gamble their lives for rewards.
- What about our cat? Well, to the extent that the cat's behavior could express, its goals were to roam the neighborhood. The same goes for cats here in the city of Buenos Aires. They spend a large part of their days seated at windows, looking through nets at the world outside. Everyone that has a cat, also has a net in their window. If they do not, the cats escape. The very act of escaping and usually not coming back says something about those cats' goals. For sure they do not aim to stay in two rooms tomorrow, if they can help it. As no one's life was placed in danger by setting their cat free, I say that *we did the right thing* to let ours go, faced with the alternative of surgically intervening on its body, modifying in this way its goals as a being, and confining it indoors by force for life.

4.3 The suggested approach summarized

- *Acting agents* can be any organisms or groups of organisms that comply with Barandiaran's three conditions of *minimal agency*:
 - *individuality*
 - *interactional asymmetry*
 - *normativity*
- Acting agents have *goals*.
- Goals can be of *higher or lower order*.
- *Highest order goal* is the *continuation of existence*.
- All other goals are ordered by their relation to the fulfillment of the highest order goal.
- An acting agent's highest order goal is intrinsically their own priority over another's.
- An acting agent is morally right for itself, when:
 - The acting agent can be judged morally and
 - Acts towards its highest order goal
 - Prioritizes other agents' higher/est goals over its own of lower order.
- Impact estimation of actions on higher order goals of a group on a large scale, is done by statistical tools on contextual data.

4.4 Comments

In light of all these, the approach I am choosing here can be said to be complying with the goals that were set in the beginning.

Moral considerability or status was not explained using capacities. All capacities beared, to any degree, were shown to be equally useful, functional and important for every organism that uses them, in the context that they are used, towards the completion of their respective goals.

Moral considerability or status was not explained by isolating the organism from context. On the contrary, it was justified on the basis of the organisms existence as part of a whole.

We have succeeded in not needing to quantify the unquantifiable, nor presume to know the unknowable, and I am referring to the personal experience of pain or pleasure.

We have set the goal of preserving life at the top of each agent' s system of norms, as all other goals either serve or presuppose it. In this way, we also preserve the intuitive notion of life's value.

We have argued how any one being's goals should be respected, given the circumstances. This has made our approach a context-based one, preserving our initial intuition on the matter.

Finally, the ethical framework I am proposing directly couples with existing well developed scientific toolsets that can give us credible answers to ethical dilemmas and conflicts. We have effectively replaced arbitrary mental and verbal estimations of suffering, which is neither knowable nor measurable, with a scientifically acceptable way to estimate the effect on human survival as a method to gauge utility. This means that we do not have to hold abstract talks about what could be best. We can go straight to the data and check.

Chapter 5 : Conclusions

5. Conclusions

The way here was carved by the questions less frequently raised around non-humans' moral status and considerability. How can we pin moral status on an individual itself, separated from the world in which it is and its correlations? How can we presume to first know, then quantify, afterwards measure and finally compare what it is to be in pain?

The short answer is we cannot. The long one ran the course of two masterpieces of science fiction cinema, whose review brought the reasons we cannot, to surface, prompting us to consequently argue our point and detail what we want to avoid and what we want to aim for.

So we came to explore an alternative, where every non-human organism or group of such may be considered to have goals, ordered by their relation to the highest order goal there is; the continuation of one's existence.

The morality of our actions as individuals or as groups may be judged by our prioritizing of our goals over the goals of others. Higher order goals pursued over lower or equal order goals of others is the right thing to do for ourselves.

On real world problems, statistical data abound and so do acceptable modeling and predictive mathematical tools. Nationwide decisions and policy making can use this existing framework, putting it to work on the basis of the suggested ethical approach, shifting focus from maximizing profit, to respecting the goals of others, especially when they do not compromise the continuation of our existence.

A thorough discussion about how goals would be ordered, with that order following in relation to the goal we set as highest, is left for another study.

We believe that the proposed approach helps connect moral decision making with acceptable scientific tools and places the value of life in all its forms in its center, regarding at the same time the individual as part of a whole.

Bibliography

- Crump, Andrew. “Invertebrate Sentience, Welfare, & Policy.” *Philosophy, Logic and Scientific Method* (blog), July 24, 2023. <https://www.lse.ac.uk/philosophy/blog/2023/07/24/invertebrate-sentience-welfare-and-policy/>
- Barandiaran, Xabier E., Ezequiel Di Paolo, and Marieke Rohde. “Defining Agency: Individuality, Normativity, Asymmetry, and Spatio-Temporality in Action.” *Adaptive Behavior* 17, no. 5 (2009): 367–86. <https://doi.org/10.1177/1059712309343819>.
- Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. “On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? .” In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–23. Virtual Event Canada: ACM, 2021. <https://doi.org/10.1145/3442188.3445922>.
- Bentham, Jeremy. *An Introduction to the Principles of Morals and Legislation*. Oxford: Clarendon Press, 1823.
- Boffey, Philip M. “Infants’ Sense of Pain Is Recognized, Finally.” *The New York Times*, November 24, 1987, sec. Science. <https://www.nytimes.com/1987/11/24/science/infants-sense-of-pain-is-recognized-finally.html>
- Cottingham, John. *Descartes*. Oxford: Oxford University Press, 1998.
- Damasio, Antonio. *Descartes' Error: Emotion, Reason, and the Human Brain*. New York: Putnam, 1994.
- Dick, Philip K., Andy Hopkins, and Joc Potter. *Do Androids Dream of Electric Sheep?* Oxford: Oxford University Press, 2008.
- Elwood, R. W. “Pain and Suffering in Invertebrates?” *ILAR Journal* 52, no. 2 (January 1, 2011): 175–84. <https://doi.org/10.1093/ilar.52.2.175>.
- Gruen, Lori. “The Moral Status of Animals.” In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Summer 2021. Metaphysics Research Lab, Stanford University, 2021. <https://plato.stanford.edu/archives/sum2021/entries/moral-animal/>.
- Jaworska, Agnieszka, and Julie Tannenbaum. “The Grounds of Moral Status.” In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta and Uri Nodelman, Spring 2023. Metaphysics Research Lab, Stanford University, 2023. <https://plato.stanford.edu/archives/spr2023/entries/grounds-moral-status/>.

- Jonas, Hans, and Philosophy Documentation Center. "Biological Foundations of Individuality." *International Philosophical Quarterly* 8, no. 2 (1968): 231–51. <https://doi.org/10.5840/ipq19688218>.
- Kant, Immanuel. *Groundwork of the Metaphysics of Morals*. Translated by Mary J. Gregor. Cambridge, U.K. ; New York: Cambridge University Press, 1998.
- Korsgaard, Christine M., and Onora O’Neill. *The Sources of Normativity*. Cambridge ; New York: Cambridge University Press, 1996.
- Kubrick, Stanley. *2001: A Space Odyssey*. Science Fiction, 1968.
- Levin, Michael and Daniel C Dennet. "Cognition All the Way Down | Aeon Essays." Accessed August 30, 2023. <https://aeon.co/essays/how-to-understand-cells-tissues-and-organisms-as-agents-with-agendas>
- Maruf, Ramishah. "Google Fires Engineer Who Contended Its AI Technology Was Sentient | CNN Business." CNN, July 23, 2022. <https://www.cnn.com/2022/07/23/business/google-ai-engineer-fired-sentient/index.html>.
- Michaelian, Kourken, and John Sutton. "Memory." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Summer 2017. Metaphysics Research Lab, Stanford University, 2017. <https://plato.stanford.edu/archives/sum2017/entries/memory/>.
- Nagel, Thomas. "What Is It Like to Be a Bat?" *The Philosophical Review* 83, no. 4 (1974): 435. <https://doi.org/10.2307/2183914>.
- Oppy, Graham, and David Dowe. "The Turing Test." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Winter 2021. Metaphysics Research Lab, Stanford University, 2021. <https://plato.stanford.edu/archives/win2021/entriesuring-test/>.
- Pearce, Adam, Asma Ghandeharioun, Nada Hussein, Nithum Thain, Martin Wattenberg, and Lucas Dixon August 2023. "Do Machine Learning Models Memorize or Generalize?" Accessed August 29, 2023. <https://pair.withgoogle.com/explorables/grokking/>.
- Regan, Tom, and Peter Singer, eds. *Animal Rights and Human Obligations*. Englewood Cliffs, N.J: Prentice-Hall, 1976.
- Ritchie, Hannah, Pablo Rosado, and Max Roser. "Meat and Dairy Production." *Our World in Data*, August 25, 2017. <https://ourworldindata.org/meat-production>.
- Scott, Ridley. *Blade Runner*. Science Fiction, 1982.
- Singer, Peter. *In Defense of Animals*. 1st Perennial Library ed. New York: Perennial Library, 1986.

Singer, Peter. *Practical Ethics*. 3rd ed. Cambridge New York: Cambridge University Press, 2011.

Tal, Eran. "Measurement in Science." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Fall 2020. Metaphysics Research Lab, Stanford University, 2020. <https://plato.stanford.edu/archives/fall2020/entries/measurement-science/>.

Varanasi, Lakshmi. "AI Models like ChatGPT and GPT-4 Are Acing Everything from the Bar Exam to AP Biology. Here's a List of Difficult Exams Both AI Versions Have Passed." *Business Insider*. Accessed August 29, 2023. <https://www.businessinsider.com/list-here-are-the-exams-chatgpt-has-passed-so-far-2023-1>

"Confabulation - an Overview | ScienceDirect Topics." Accessed August 29, 2023. <https://www.sciencedirect.com/topics/neuroscience/confabulation>.

"Great European Immigration Wave to Argentina." In *Wikipedia*, August 12, 2023. https://en.wikipedia.org/w/index.php?title=Great_European_immigration_wave_to_Argentina&oldid=1169986533.

"Nostalgia Doesn't Need Real Memories – an Imagined Past Works as Well | Aeon Essays." Accessed August 29, 2023. <https://aeon.co/essays/nostalgia-doesnt-need-real-memories-an-imagined-past-works-as-well>.

Our World in Data. "Per Capita Meat Consumption by Type." Accessed August 29, 2023. <https://ourworldindata.org/grapher/per-capita-meat-type>.