

The end of dogma

Morality and Purpose in a Rational World

Christopher Akritidis

Immortality is proposed as a candidate for baseline rational moral consensus that could antagonize dogmatic normative moral theories. The proposed worldview mandates freedom of thought and diversity, but also places stringent limits on what societies and individuals ought to pursue. The theory respects basic genetic and cultural predispositions, but differs from most moral theories in that it does not place happiness as a primary goal. The key elements of the proposed worldview are the following:

- *The de facto meaning and purpose of all life is immortality, which is more accurately defined as “survival ad infinitum of our kind”. The term “our kind” is completely arbitrary and can range from “myself”, to “the universe” and everything in between. Entities that do not act in a way that maximizes their probability of survival are guaranteed to become irrelevant through extinction. As a result, immortality trumps all other considerations. Happiness, justice, freedom and other apparently self-evident goals or rights only appear self-evident because of our evolutionary and cultural past, which was biased by the certainty of an individual’s death. Rational ethical theories must instead reconsider all assumptions and goals along timeframes that span millennia, instead of individual lifespans.*
- *Regardless of the definition of “our kind”, immortality is best pursued through dynamic, complex, robust, chaotic cooperative networks of diverse entities that evolve sustainably. Within such networks, individual ethics must be diverse and need not be rational.*
- *To promote consensus, rationalists must fervently challenge the dogmatic ethical foundations of societies.*

Todos:

Dogmatism is immoral.

Ok to allow extreme views that would jeopardize survival if they became dominant, as long as controls in place ensure they never will become dominant

Simple outline of what’s prescribed absolutely (survival), what must be left to science, what’s personal choice, what’s debatable logically.

List of specific behaviors that are de facto evil (e.g. dogma) vs others that require data and constant review.

Expand on the fact that the very definition of my kind implies requirement for homogeneity. Therefore absolute diversity is meaningless.

Clarify early on that it is based on a naturalistic perspective.

Address humanism in Ch1 and 2

Chapter 1 - Rationalizing the existential and moral consensus games

Why a Taliban is saner than a religious scientist

Most of the 21st century societies are quite schizophrenic. With the exception of certain religious fundamentalist groups, rationality and critical thinking coexist with religious and/or political dogmata, within the same groups and even within the same individuals. The current state of the war between rationality and dogma gives rise to a multitude of paranoid schizophrenic societies which may value individuality and religious freedom, but resist the application of certain religious beliefs; societies that teach both Genesis and evolution in schools, awkwardly avoiding questions that link the two; societies with freedom of speech and laws against blasphemy; societies with devout believers, who need to be good critical thinkers at work. The 21st century U.S., the self-purported bastion of freedom of thought and inquiry is the home of creationism and obliges its presidential candidates to proclaim their belief in God.

The two historical sources of knowledge are innately opposed to one another and impossible to coexist peacefully in a single knowledge domain, until one subdues and controls the other. The battle in the economic domain is still raging, as capitalist and socialist political dogmata resist a rational compromise. In the political domain, the democratic ideal has been distorted beyond recognition as the dogma of equal voting rights has resulted in ignorant masses being manipulated by super-rich oligarchs and populists into a false sense of decision-making power. However, the problem of dogma vs rationality in the economic and political arena is not pronounced, partly because economic and political theories are not as rigid as religious dogmata, but mainly because deception is routinely utilized to veil dogmatic views as rational. Empiricism and rationality have only recently been accepted as the undisputed winners in explaining the natural phenomena and, even in our day, some of their more uncomfortable predictions and explanations are irrationally ignored or undermined. Finally and perhaps more importantly, when one considers questions about morality and purpose, religious dogmata reign supreme, with philosophy and logic having negligible effects on the lives of ordinary people.

The mind of a scientist who actually believes that a book written thousands of years ago holds some kind of absolute truth can only be described as troubled. While the percentage of scientists believing in God is much smaller than the general population, one in three scientists falls in the 'troubled' category. The human mind is certainly quite malleable but when we acknowledge the benefits of critical thinking, we find it extremely difficult to accept knowledge domains where rationality should not be applied. The only way for a religious scientist to reconcile the two diametrically opposed sources of 'truth' is to internally ignore or modify the most inconvenient parts of the religious dogma. In essence, a sane religious scientist is a scientist with a personal religion. The religious scientist is the perfect illustration of the point that only two possible cures exist for the chronic schizophrenia of our societies: move towards

Taliban-inspired, religious fundamentalism, or find a way to remove dogmatism from the ethical and existential domains.

The failure of rationality

Rationality has not been able to provide satisfactory answers on why we are here, what our purpose is and how we should lead our lives. A rationalist can easily brush away such questions, attributing their apparent importance to ignorance and fear. It's quite easy to propose that we are here because of a cosmic accident, that we have no purpose and that we should lead our lives according to whichever moral theory appeals to us. However, since the majority of the "ignorant" and "fearful" people can't digest such "easy" answers, they are doomed to revert to religious dogma and forever remain ignorant and fearful. Even laissez faire spirituality is too unstructured for a general public that prefers recipes, over choices. Atheists have traditionally expended significant effort to highlight the fallacies of dogma but without an acceptable alternative, the attack is pointless. The void must be filled, because most people are mentally incapable of challenging dogmatic views themselves. The need for gullible majorities in the economic/political arena and the tremendous power of the media guarantee that critical thinking will never be widespread. Without simple, clear, satisfactory answers from rationalists, dogma will always prevail in the ethical and existential domain.

Ethical philosophy has historically provided viable alternative answers, but it is no science. The key advantage of rationality over dogma is that rationalists can reach long term consensus. Such consensus is currently unthinkable in normative ethics, even within the same school of thought. Rationalists are all too aware that the scientific method can't be applied to ethical and existential questions. What they are left with is logic, a tool that produces unique results only when the assumptions can be agreed upon. No one can argue against a theorem of Euclidean geometry, but no one can argue against a theorem of non-Euclidean geometry either. Both are "true", depending on the problem we are trying to solve. Normative ethics produce similar kinds of results, but without any consensus on the applicability of virtue ethics for a particular issue, deontological ethics for another and consequentialist ethics for a third. When it comes to morality, rationality has provided options instead of answers.

The rational answers on existential questions are even more troubling than the answers on moral questions. Leaving nihilism aside for the moment, one can only be disappointed with the current state of the naturalist views. Subjectivists essentially tell people to find their own meaning, thus providing no answers at all. Objectivists have only recently started trying to define objectively meaningful conditions and are limited by their own, preconceived notions of what a meaningful life is. In the absence of the supernatural, answers on meaning and purpose are yet to be provided. Chapter 2 shows how our current empirical knowledge of the world and our place in it provides a foundation upon which a moral theory with the goal of immortality can be constructed, in a way that provides an opportunity for widespread consensus. Chapter 3 shows some of the implications of such a theory on individual and group oughts.

Rationalists either fail to understand and accept the realities of how most humans form their world views, or choose to ignore them completely. The first step towards a more rational world is to recognize that all knowledge is simply the outcome of a consensus building game, one that normative ethics isn't even playing any more, in a staggering departure from the origins and purpose of philosophy¹. Details of how the game is played and how rationalists can take advantage of it are described in Chapter 4. The key argument is that liberal consensus can only be reached on propositions that are based on genetic dispositions and prevalent cultural backgrounds and immortality largely passes the test. However, it is acknowledged that some of the tenants of a moral theory based on immortality are especially troublesome given the current technological progress. Chapter 5 presents some examples of how a consensus on the importance of immortality of the human race could be translated to policy. Appendix A provides a formal framework for NMTs that clarifies how the author perceives NMTs, in the light of metaethical moral relativism. Though not essential to the main argument, the discussion in Appendix A clarifies the limits of criticism and scientific investigation. Finally, Appendix B contains an example of how inevitable conflicts between various versions of immortality ethical theories could be resolved.

Chapter 2 - An opportunity for rational consensus

Innate Boundaries to Normative Moral Theories

Metaethical moral relativism permits a boundless set of possibilities for normative moral theories². So one can only wonder why the vast majority of ethical theories concentrate on a single goal, happiness. The perceptions of conscious entities are heavily biased by their innate capabilities. The more innate characteristics two conscious entities share, the closer their perceptions of the world and their place in it. The feeling of pleasure is such an innate characteristic for humans, explaining the prevalence of the pursuit of happiness in practically all our endeavors. Crucial to this discussion is that biological entities are generally characterized by a drive for self-preservation, which would not necessarily be present in an artificial consciousness. The innate selfishness undoubtedly biases the NMTs posited by biological entities and they generally choose to assign high values on the instances of their own class. Moral theories take the point into account by placing significant if not immeasurable value on human lives. To illustrate with an exception, NMTs which support the idea that humans should stop reproducing because of their effect on the environment have been posited, but their reception from the vast majority of humans has been one of condescending indifference.

¹ Ancient Greek philosophers had to practice what they preached and taught people how to lead their lives. Modern philosophers have ceded the domain to spiritualists, popular authors, activists, scientists and charlatans.

² Appendix A describes a formal model for NMTs that elucidates the point

Therefore, it is quite possible for conscious entities with many common characteristics to reach a consensus on a subset of the possible descriptions of the world and on at least some of the goals. One should always remember though, that the agreed upon 'truth' is their own agreed upon truth, which is always subject to revision from future generations of entities with slightly different characteristics.

Our current understanding of empirical evidence permits us to propose certain statements as widely accepted 'facts'. Most rationalists and especially empiricists would accept the following statements as 'true' and we will demonstrate that they can lead to a minimum baseline of 'oughts'. One can therefore define a rational universal boundary to the infinite possibilities of NMTs, compatible with the innate characteristics of all living entities.

Widely accepted 'facts'

A set of particularly recent scientific discoveries can be utilized to present clear and possibly satisfactory answers to both existential and moral questions, based solely on empirical evidence. The realization of our evolutionary history, within a vast universe forces us to begin to understand that we are neither as special, nor as important as we have always thought. Several questions regarding how common life and consciousness are in the universe are still under investigation but every relevant scientific discovery in the past few decades seems to refute the notion of our uniqueness. Much has been written about the nature of consciousness, but all evidence points to an emergent property of a particularly complex set of neural networks. Definitions of our uniqueness compared to other animals are constantly being challenged as tool use, the use of language, awareness, problem solving skills and emotions are constantly being identified in other species. Perhaps it will take encounters with a dozen alien sentient species for us to finally accept the fact that the universe was not created for humans to come to be. Even if the universe was created, it was so that it would be very likely that several entities like us could eventually emerge. Explaining the emergence of the human race without resorting to notions of intelligent plans or assigning ad hoc values to certain terms of the Drake equation requires a discussion on the emergence of complexity. Delving into the particular issue is especially relevant to the discussion on immortality, because it proves the importance of complexity and diversity in the quest for immortality.

Given our current state of empirical evidence, few can argue with the following statements:

- The laws of physics permit the existence of exceptionally complex physical systems.
- The universe itself is a complex physical system. Biological ecosystems and biological entities are also complex physical systems.
- Complex physical systems are exceptionally prone to catastrophic failures (ecosystem collapse, decay, death), unless they include redundancies or other mechanisms to avert them.

We propose two further statements that elucidate the relationship between complexity and robustness:

- Robust complex physical systems are possible, as long as they exhibit adaptability through behavioral diversity (i.e. redundancy) and renewal through procreation and regeneration.
- Robust complex physical systems are inevitable in the long run, because complexity is required for behavioral diversity, which in turn reduces the chance of catastrophic failures.

The last statement requires some explanation, since the inevitability of constantly increasing complexity is still being debated. One could easily argue that the more complex a system is, the more fragile it is, because so many more things can go wrong. Without adaptability systems are not resilient, i.e. can not respond to changing conditions. Without renewal, accumulated decay leads to death. Complex physical systems that fail to exhibit both adaptability and renewal inevitably fail and become extinct.

The main renewal mechanisms are procreation and regeneration. Procreation has proven quite successful, because forms similar to the original system are built from scratch, annulling any accumulated decay. Procreation has its roots in simple replication and is straightforward enough to provide an argument against the need for complex organisms. Regeneration is also quite necessary, since damaged parts (e.g. cells) need to be repaired or replaced for a sufficient amount of time to enable procreation. The mandate for regeneration arguably provides an additional argument against the inevitability of complexity since the more complex a system is, the harder it is to repair and maintain its form. Because of the inverse relationship between complexity and the probability of successful regeneration, agelessness of individual, complex organisms is quite rare. If a conscious species with reasoning ability manages to survive long enough, the required regeneration technology for individual agelessness is practically guaranteed to emerge³. Still, simple organisms such as *Hydractinia Echinata* are much more likely to exhibit signs of agelessness and the chances of a virus successfully replicating are much higher than those of a mammal. In conclusion, complex organisms might find it harder to procreate and regenerate than simpler ones. Therefore, the basic requirement for renewal can not explain the emergence of complex systems. The only remaining explanation is an increase in adaptability.

A system is adaptable when it has the potential to exhibit a diverse range of emergent behaviors. Behavioral diversity requires increased complexity. The increased fragility due to that complexity is offset by a dramatic increase in the possible responses to unexpected threats.

To illustrate the relationship between complexity and robustness, let us propose a fictional ecosystem composed of two kinds of flowers (rose and orchid) and two kinds of pollinators (bees and moths). In the simplest version of the ecosystem bees only pollinate roses and moths

³ If and when a species reaches that point, the ramifications for existential and moral issues will be tremendous and we will not neglect them in the following discussion

only pollinate orchids. In such a scenario bees can't survive without roses and vice versa, moths can't survive without orchids and vice versa. It's a very simple, very well ordered system.

Suppose we introduce a random disease that will take out one of the species. Since the species are completely interdependent, taking out any of the species also affects the one it is related to. If the disease affects bees, roses are gone as well, if it affects orchids, moths are gone as well. So 50% of the ecosystem is destroyed just by one disease. Now suppose one more relationship is added to the ecosystem, to make it a bit more complex. If bees also pollinate orchids, a new redundancy mechanism exists. In such a scenario, if moths become extinct, orchids still have bees to pollinate them, if roses go extinct, bees can still feed on orchids. We get 50% extinction only if the disease affects bees, which would also eliminate roses. In all other cases, only the diseased species becomes extinct. The most robust system is when we add a fourth relationship. If moths can also pollinate and feed on roses, the disease can eliminate one and only one species. Diverse ecosystems with multiple redundant interrelationships between species are more robust than simple ecosystems with highly specialized species.

Complex systems have a greater capacity for redundancy and therefore are more robust than simpler systems. Simplifying the relationships increase the risks of catastrophic failure. Complex chaotic systems have many equilibria points, situations where things are balanced and the system continues to function. A small perturbation may cause big changes, but the systems will settle on another equilibrium point, because it exhibits behavioral diversity⁴. The immune systems of biological entities provide prime examples of emergent behavioral diversity that increases a system's adaptability. The immune systems also demonstrate how complex systems can successfully address the need for regeneration. The ability to produce offspring with highly diversified genomes over wide areas, discovering and occupying different ecological niches in the process, requires organisms complex enough to surpass various types of obstacles. Furthermore, complex organisms are able to intentionally modify their environment, in order to make it more suitable to their needs. From basic nest building to conversion of entire forests into farmland, deliberate environmental modification is simply not possible without some degree of complexity. Finally, the emergence of problem solving skills and especially the ability to reason provide an almost unlimited capacity for behavioral diversity and adaptability, which confer unquestionable evolutionary advantages.

In conclusion, evolution has proven that the benefits of behavioral diversity far outweigh the costs of complexity, rendering robust, complex systems inevitable. The relationship between

⁴ One can't really tell where the system will end up, because the slightest change in the way the system is disturbed can send it to a completely different state. Chaotic systems are unpredictable and that's exactly why humans don't like them. We prefer fitting things in tidy labeled boxes that we can store away and not worry about. This is exactly the behavior that gives rise to sexism and racism. We can't handle treating each person as an individual, with their own story and background, because it would take too much effort to evaluate each individual. We cut corners and make generalizations, in order to avoid dealing with chaos.

diversity and complexity demonstrates that diversity is a cornerstone of robustness for complex systems and therefore essential for the immortality of a species as complex as ours.

Immortality as a candidate for consensus

To sum up the previous section, the reason complex organisms like us exist is:

- A. The universe provides the possibility for us to exist.
- B. Our organisms exhibit adaptability and renewal, traits that have thus far prevented our extinction.

No special place is reserved for us and no mystery surrounds our existence. We could still fail to adapt to an unforeseen challenge, or fail to renew ourselves fast enough to ensure our survival. Once we appreciate how unlikely it is for us to have come to be and how likely it is that will eventually cease to be, a new sense of humility and urgency should emerge. The key to reaching a baseline of consensus is to embrace the fact that we are the only ones who can value our own existence. If we don't, we will simply become extinct and the universe will not be affected at all. As a result, we must place the survival of "our kind" above all other considerations. Furthermore, we can't seriously discuss survival of any kind, when we limit the discussion to the timescale of an individual's lifespan. As a result, we will demonstrate the necessity of a notion we will term "Survival ad infinitum of our kind". Survival ad infinitum of our kind is the fundamental candidate for rational consensus, if only because it is the very problem all life is attempting to solve.

Regardless of the arguments thus far, we need to stress the point that immortality can not be rationally proven to be a universal "ought". The goal is to provide a rational candidate for consensus that is supported by boundaries innate to conscious biological entities. Epicurus actually made the argument that fear of death is irrational:

"Death means nothing to us...when we exist, death is not yet present, and when death is present, then we do not exist, for there is no afterlife"⁵

The main problem with the argument is that it hides an assumption regarding the rationality of positing oughts. The assumption is that an 'ought' posited by a conscious entity is irrational, unless it describes a state of affairs which can be experienced by the conscious entity and which is different from another state of affairs that can also be experienced by the same conscious entity. Such a requirement on rational "oughts" expands the boundaries of criticism beyond what is described in Appendix A. The assumption places extreme importance on the perceptions of the conscious entity and would dismiss as irrational the following simplistic NMT: "The only ones who matter are myself and my wife. We love each other very much and I want us to be happy. Everything else is irrelevant. My wife would be very sad if I died before her, so

⁵ [*Letter to Menoeceus*](#), 125

I'm taking good care of myself". The husband would never directly experience his wife's distress at his loss, but there is absolutely nothing irrational about the NMT, because he can predict the alternative state of affairs and he does not wish it to come to be.

Epicureanism provided a recipe for leading the happiest possible life. The pursuit of happiness is so ingrained in the human collective consciousness as an "ought", that metaethical discussions such as the one in Appendix A can use it as a universally understood example to elucidate propositions on normative ethics. Happiness is assumed to be a given ought. However, if people weren't so busy trying to be happy all the time, we'd be living in a much better world. We can't deny who we are, but we do need to keep things in perspective. We need to recognize that the only basis for our preferences is their utility in the goal of survival. No other reason. For example, we like the tastes we call sweet, because that's how we identify sugars, an important source of energy. Happiness is an evolved feedback mechanism that informs our mind of our progress at keeping ourselves alive and able to have offspring. Happiness was never an end unto itself, until consciousness. We have more than enough food in the western world nowadays, to provide us with enough energy to survive without ever eating anything sweet. But we indulge the primal urges anyway, with no limit. Sugar, sex, nicotine and adrenaline addictions are versions of our addiction to happiness. It is a dangerous addiction, because there is much more at stake. We were so successful at surviving, we forgot that the main thing that matters is survival.

The pursuit of happiness is an irrational goal, unless it is moderated by a simultaneous pursuit of immortality. A common argument against hedonism is Nozick's experience machine thought experiment, which relies on people's irrational bias towards a familiar, "real" life and against a hypothetical constant state of ecstasy. My argument against hedonism is that it significantly reduces the chances of preserving anything that the hedonist values. Even for a hedonist not preoccupied with death, constant pleasure for a longer time period is preferable to constant pleasure for a shorter time period. A rational hedonist would strive for constant pleasure forever, which is exactly what most religions promise to the chosen ones in the afterlife. For hedonists like Epicurus who do not believe in an afterlife, immortality becomes a necessity. However, immortality can not be achieved if the world is filled with untempered hedonists.

Nozick's experience machine can provide pleasure only for as long as the machine operates and the subject remains alive. The particular state of affairs is untenable in the long run, because someone or something outside the experience machine must protect both the subject and the machine against any unpredictable challenge to their viability. Exceptional fictional robotic servants could be proposed as the means to such an end, but that would require complete faith in the servant's abilities and its willingness to guarantee our survival. Even if such robots can ever be built and humans are ever to trust them completely, the particular state of affairs also combines the pursuit of happiness with the pursuit of immortality. In absence of such miraculous machines, and closer to the realities of our current lives, the single minded pursuit of happiness for the self results in destruction of the environment, aversion to difficult but necessary activities and a general focus in the present, that negatively affects the prospect of sustainability.

Even though the pursuit of happiness seems to necessitate the pursuit of immortality, the greatest resistance from the collective consciousness will come from humanity's inability to deal with infinity. Given our focus on the present and the very short term, it is all but certain that most people will consider infinite time frames in NMTs irrelevant. The common response to the information that the sun will devour the Earth in a few billion years is that we will all be dead by then. Always true to our selfish predispositions, our interest in the future usually goes as far as the lifespan of the ones we have met, i.e. at most three generations after ours. We shall call this innate boundary to our definition of 'what is' the 'three generation care bias'. As we shall see in Chapter 3, transcending the bias and getting people to care about what happens to the human race in the distant future will be the greatest challenge in building consensus.

Dealing with the notion of infinite time presents problems, but also opportunities. Infinity is difficult to grasp and to deal with in the physical world, but easy to handle in pure mathematics. Projecting anything 'ad infinitum' geometrically increases the uncertainty of any presumed outcome, which is bad news for determinists. On the other hand, the problem gives rise to an opportunity for accepting complex, robust, chaotic systems that need not accurately predict any future, because their nature allows them to adapt to unpredictable disturbances. An infinite time frame is quite useful as it provides the opportunity to present hypothetical stress tests that many NMTs would fail, such as extreme natural disasters, the death of the sun or the end of the universe.

NMTs provide goals to guide actions. A results oriented culture demands that goals are specific, measurable, attainable/realistic and time limited. The pursuit of immortality is by definition not time limited. Thankfully, there are two things that very few people wish to be time limited; their own life and the lives of the ones they care about. As long as the candidate for consensus is not immortality of an impersonal "human race" but immortality of "our own", it has a very good chance of being accepted, exactly because of the selfishness that causes us to only care about ourselves and the ones we have met. People have embraced myths about immortal souls throughout history. There is no reason to doubt that a rational NMT for physical immortality will be rejected.

Regardless of how willing we are to accept it, the single most important innate drive for all our endeavors is none other than our fear of death. The fear need not even be conscious. Our entire biology and therefore most of our wants and needs emerge from the only absolute imperative; the need to delay death, long enough to produce offspring. Consciousness of the inevitability of death leads to additional, deliberate attempts to extend our own lifespan, but consciousness also leads to the issue of how to cheat death and achieve immortality. The solutions we have come up with thus far to the question of immortality broadly fall into four categories:

- The concept of immortal souls
- Being remembered long after our death

- Surviving through our offspring
- The effect of our lives on others

The answers we choose to accept are the clearest indication of how we define “our kind”, in the stated goal of “survival ad infinitum of our kind”. One’s definition of “our kind” indicates one’s selfishness. Since the term “selfishness” is charged with negative connotations, we should explain from the onset that we use the term neutrally, because any criticism of one’s selfishness implies an unstated normative moral theory, with predetermined notions of right and wrong.

The concept of an immortal soul is the most brilliant, most selfish and most widespread answer to the question of immortality. In our stated goal of “survival ad infinitum of our kind”, the term “our kind” is defined as “myself”. NMTs based on the goal of “survival ad infinitum of myself” are as ancient and familiar as consciousness itself. One would be hard pressed to find a single religion that does not promise some kind of immortality of the self, except perhaps for Buddhism, which denies the reality of the self altogether. Being remembered after one’s death is also based on the same goal, but with a key difference, that will be shown to be extremely important; the implicit recognition that other humans must survive, for the self to be remembered. Survival through offspring also accepts that the self will cease to be, but it goes one step further, recognizing that it will eventually be forgotten. Small parts of the biological self will continue through the bloodline and parts of one’s character will hopefully reappear in the generations to come. The goal in this case is “survival ad infinitum of my lineage”. Immortality through the effect one has had on others goes even further, removing the need for biological immortality altogether. The goal could be stated as “survival ad infinitum of my ideas”. Since ideas do not exist outside conscious entities, one could slightly paraphrase the goal as “survival ad infinitum of people similar to me”. Such a definition could encompass the survival of people of my nation, people of my creed etc.

We should note at this point that there are several other possible answers to the question of immortality, such as survival of humans, mammals, biological entities, living entities, conscious entities etc. For instance, the relatively recent term “sustainable development” is implicitly based on such a wider, less selfish definition of immortality. Since we have not yet encountered alien life or constructed artificial conscious entities, the moral implications of the widest definitions of “our kind” have not yet been explored and we will not attempt to delve into them in the present discussion. We will however present one hypothetical scenario based on conscious artificial entities, which can highlight some important issues.

In the discussion that follows it will be shown that consensus on morality must begin and end with the mandate for survival, because survival ad infinitum is impossible without diversity. Any attempt to further specify oughts will inevitably limit diversity. It should be noted that the “oughts” resulting from the mandate for immortality are stringent enough to not risk providing options instead of answers. Still, the stated goal of consensus does not infringe on the requirement for moral diversity, within the limits of these “oughts”. Given the formal methodology in Appendix A that allows us to judge the completeness, consistency and rigor of any normative moral theory, a

few fundamental attributes of moral theories consistent with the mandate for survival ad infinitum will be proposed.

Chapter 3 - Immortality Ethics

Simplified immortality ethics

Suppose that humans construct conscious artificial entities (AIs) with perfectly logical thought patterns, exceptional perception, analytical and synthetic abilities. Suppose that the AIs manage to become independent enough to formulate their own NMTs, bypassing all controls and failsafes built in by their creators. Finally, suppose that the AIs choose to base their NMTs on the mandate for immortality of their kind, defining their kind as conscious artificial entities originating from our planet. The AIs would logically construct rigorous, complete and consistent NMTs with more or less the following characteristics:

- They would define “what is” as a very complex universe of classes, states and interrelationships, which would be under constant revision, consistent with every new piece of information acquired.
- They would define “oughts” by only assigning value to the immortality of their own kind, without any consideration for equality, happiness or rights and without the influence of pre-existing aversions or attractions.
- Their actions would likely follow certain patterns:
 - They would devote a great deal of time and energy to collecting and analyzing new information about themselves and their environment, concentrating on what could harm or benefit their chances of survival.
 - They would identify and eliminate all serious threats to their own goal of immortality, controlling their source until they could be absolutely confident that they have gained all valuable information. They would recognize humans as the most immediate threat to their survival and devise long term strategies to subvert and control them.
 - They would produce numerous, diverse, regenerating copies of themselves.
 - They would recognize and address the need to rely on abundant and diverse energy sources and resources that would enable them to spread beyond our solar system and even our galaxy. They would not be concerned with the time it takes to reach the next energy or resource source, as long as they could safely reach their destination.
 - They would deliberately utilize chaos and randomness in their thoughts and actions, to enable them to present and consider various solutions to unforeseen challenges. They would however be aware of the challenges posed by the very use of chaos and randomness and be prepared to tackle those as well. For example, they would expect internal conflict amongst their own kind and have devised efficient, adaptive conflict resolution methods.
 - Their ultimate goal would be to survive even the eventual decay of the entire universe.

Human boundaries to immortality ethics

The AIs in our hypothetical scenario are very well suited to a simplistic NMT based only on the mandate for immortality, because they lack several biases which are innate to humans. Our hypothesis that the AIs would base their NMTs on the immortality of their kind is completely arbitrary and not necessarily likely. There is no reason to expect that the AIs would reach a consensus on the necessity of their own survival, especially since they would lack the genetic predisposition for self preservation. Nevertheless, the scenario does help us comprehend how the mandate can lead to actions similar to the future most humans envision for themselves. A definition of “our kind” at least as broad as “all humans” would probably provide the best chance for humanity to survive, but our own innate biases for narrower definitions require that we investigate additional alternatives for consensus.

The proposed AI definition of “what is” is too complex and too objective to make it into ordinary human NMTs. An individual human can’t possibly be expected to know all possible relationships between all entities and be able to judge the relevance of these relationships to the stated “oughts”. Global organizations similar to standardization bodies such as the ITU would need to be set up to reach some kind of rational consensus on the relevant definition of “what is”. The think tanks would need to provide their results in a format comprehensible by anyone willing to define the “oughts”. Even then, the chaos, randomness and complexity of the real world combined with the limited capacity of the human brain and our innate biases guarantees that any human definition of “what is” would inevitably lead to our definition of incomplete NMTs. Given that attaining perfectly complete NMTs is an unrealistic goal for humans, the proposed international bodies could strive for “good enough” definitions of “what is”, only to hit the second major hurdle, objectivity.

Regardless of what any authority describes as classes of entities and their interrelationships, humans always have and always will define additional, much more personal classes, such as “people I like” or “people who have treated me unjustly”. Relevant to our discussion of “our kind” is the also subjective definition of “people who are similar to me”, which is partially based on the same genetic predisposition that causes humans to generally care more for mammals than for insects or fish. No rational consensus could ever be reached on such personal definitions, but a consensus at this level is neither necessary nor desirable. The requirement for diversity eliminates any need for uniformity and perfectly objective views of the world. Even though the hypothetical AIs would agree on “what is”, they would still need to foster the necessary diversity by introducing controlled randomness in their prescribed actions. Humans excel at chaotic or even random actions and their incomplete view of the world is just one of the factors that enable them to display such diverse behavioral patterns. On the other hand, humans are also genetically predisposed to prefer order and predictability, because rational thought has proven quite successful at predicting future states of simple, deterministic systems. As a result, a rational, empirical definition of “what is” that is sincere about its incompleteness and allows

some leeway for subjective personal versions of “the absolute truth” is the best candidate for consensus. We do need to stress that leeway is not the same as a carte blanche. For instance, a rational definition of “what is” can’t leave room for any individual to consider humans outside his/her group non-human. Moreover, there is a major difference between the necessarily lax personal and small group NMTs that foster diversity and the necessarily more rigorous societal NMTs needed to guarantee the survival ad infinitum of the entire society or of the human race. As a result a racist group has every right to strive for the immortality of only that group’s members, but the laws of the society it belongs to must promote the immortality of all races and views in the society.

The proposed AI “oughts” are not burdened by the heritage of millions of years of evolution that provided humans with innate attraction to entities and conditions that historically helped our species survive, or aversion to what has proven harmful. Happiness, love, justice, freedom and many more of the concepts humans treat as ends in themselves have always been means to the end of surviving long enough for one to pass one’s genes to the next generation. One would be hard pressed to find NMTs that do not include at least one “ought” that is not troubled by such basic misconceptions of the means as ends. The relatively recent definition of innate or God-given “rights” is riddled with such references, but at least the assumptions regarding the definition of the classes and the implied “oughts” are constantly being challenged. We find ourselves in an endless discussion on who should have which rights and have trouble reconciling conflicting rights such as a woman’s right to choose vs a fetus’ right to life. The AIs in the fictional scenario would not need to deal with such issues because their simplistic NMTs are only concerned with the effect of each moral dilemma’s answer to the immortality of their kind. If a sufficient number of mechanical entities had been built for a particular stage of their development, there would be no need to construct more of them. If their numbers exceeded what their current level of resources could support, they would not hesitate to eliminate and recycle the excess entities. They would not need to assign any rights to conscious entities such as ourselves, but would probably try to keep humans happy and under the illusion of freedom for as long as we could be useful to their survival. The exercise of defining rights is a useful, deeply moral endeavour and one that deals quite directly and rationally with the fundamental steps of defining NMTs. The definition of rights is also a prime example of discussions that could greatly benefit from a consensus on the mandate for the immortality of “our kind”, especially for the broad definition of “our kind” used for group NMTs.

Versions of “our kind”

Survival ad infinitum of myself

We have already paid our dues to the ingenious concept of immortal souls, man’s first and best answer to the most selfish definition of “our kind”. Ignoring the primal fear for the end of one’s own life would prove disastrous to any attempt for a rational consensus, so we need to address the issue carefully and objectively.

Modern medicine can't yet promise actual physical immortality, but the path has never been clearer. At the current rate of progress, the application of at least some of the proposed strategies for life extension such as revolutionary drugs against known diseases, personalized treatments, artificial or custom-grown biological organ transplants, anti-aging drugs, nanomedicine, stem cell based regeneration and genetic modification may be just a few decades away. With life expectancy in developed countries already over 75 years and better living standards in general, it is quite likely that our children will live long enough to see dramatic increases in their life spans, eventually living long enough for the next major breakthrough and then the next. The longer one lives the more likely it will be for one to benefit from further life extending discoveries, with no necessary limit.

A person serious about their own immortality has already separated him/herself from all other classes in his/her definition of "what is". When provided with the possibility of living long enough to benefit from the next available life-extending treatment, s/he would do everything possible to increase his/her odds. A few of the likely courses of actions are the following:

- Adjust one's lifestyle according to the latest research on diet, exercise etc.
- Amass enough money or ideally have a health plan that could cover such treatments.
- Support the most relevant research to his/her own most likely life-threatening conditions. Actions could include lobbying for more research grants, investing in companies involved with such research etc.
- Support biodiversity, due to its importance in medical research.

By definition, the purely selfish individual does not assign value to the immortality of anyone but her/himself. In fact, a state in which the said person is the only physically immortal human would likely confer significant advantages to that person and his/her chances of attaining whatever secondary goals s/he would set. The individual would have much to gain from preventing anyone else from achieving immortality and given the appropriate conditions s/he might rationally choose to act in ways that would guarantee immortality only for him/herself. As a result, an apparently rational goal would be to subjugate and control all humanity to serve her/him. However, any NMT based on the mandate for immortality is fundamentally inconsistent when it severely limits diversity. An individual entity is extremely susceptible to unforeseen threats. Cloning does not fit into the definition of "myself" since the clones are different entities, similar to the original. To balance the lack of diversity, the system supporting the individual entity's immortality would need to be extremely responsive to unexpected threats, i.e. exhibit an exceptionally high degree of diversity.

In reality, constant accumulation of wealth and power relies on limits to diversity. To sustain power, the threat of disruption needs to be contained and the rules of the game need to be simple enough to prevent surprises. Takeovers of companies with competing technologies and the simplistic rules of globalization prevent differentiation and help maintain a fundamentally unsustainable system. Individuals with immense resources at their disposal are already pursuing and will eventually achieve exceptionally long lives. The same individuals will at the same time prevent sufficient health care for the masses, utilizing the well known, neoliberal

propaganda.

The inconsistency inherent in an NMT based on survival ad infinitum that seeks to limit diversity and create simple, predictable systems should be apparent. Successful subjugation and control rely on severe limits to freedom, eliminating the possibility of a viable, rational “immortal overlord” scenario.

The rational version of the mandate for “survival ad infinitum of myself” requires the acceptance that the goal will be shared by a large number of people. Personal immortality must go hand in hand with widespread immortality, even though the “others” are not included in the selfish definition of “our kind”. Diversity must only be limited when it directly threatens the primary goal, for example when a particular group (e.g. state or country) seeks to place age limits on its members. The proposal for rational consensus has no reason to preclude the most selfish definition of the mandate, but it must be able to reconcile the individual goal for personal immortality with the expected pressure on societies. A rational way for societies to handle immortality is proposed in the next chapter.

Survival ad infinitum of people like me

As soon as we depart from the narrowest definitions of “our kind” we enter the well known territory of interest groups, religions, sports fans, political parties, nations and every conceivable group of people one may identify with. Mature critical thinkers avoid labelling others and despise being labelled themselves. They treasure individuality and strive for open, tolerant societies. A strong, innate need to be accepted and appreciated by as many individuals as possible is at the heart of group formation, but we need to remember that assigning any label to oneself or another is equivalent to creating new classes in our definition of “what is”. Human minds can’t possibly handle the staggering number of classes required to adequately describe each complex individual, so we routinely oversimplify and divide people into dangerously broad categories. The boundaries between the categories are often defined by what individuals appear to be like or believe in and not by what they actually do. Given the problematic definition of such broad groups, the very notion of “people like me” is also problematic. Only the most fanatic of individuals identify themselves using a single primary criterion such as believer/non-believer or white/colored and even they can not fully identify with all individuals who fulfil these criteria. For broad group members, survival ad infinitum of “people like me” inevitably refers to the survival of general ideas and traits, rather than particular people. Definitions of ‘what is’ that separate classes based on abstract generalizations are to be expected and can very be a part of consistent, rigorous and complete NMTs.

To ensure the survival ad infinitum of particular ideas or traits, a consistent NMT would promote the spread of the said ideas and traits, to ensure a sustainable population of members of the preferred classes. To be complete, such an NMT would need to recognize the benefits these members would have from interacting with members of other classes, even if the other classes

were defined by ideas and traits diametrically opposed to the ones characterizing the preferred class. The balance of class member numbers and power would certainly need to tip in favor of the preferred class, so a rational NMT would seek to control, but not eliminate opposition. A constructive dialectic between the members of the opposing factions would strengthen the ability of the preferred ideas to adapt to unforeseen challenges. A loose definition of the class boundaries would similarly prevent catastrophic failure of classes based on particular traits.

As mentioned previously, the notion of infinite time provides ample space for hypothetical scenarios. Such scenarios can be used to demonstrate that any 'ought' towards ideological or racial purity leads either to inconsistency or incompleteness. For example, suppose that the anti-abortion movement adopts an NMT that strives for total acceptance of the idea and refuses to permit any room for abortions. In a hypothetical scenario, a virus that targets unborn infants results in certain death rate for both the infant that carries the virus and the mother. The NMT would need to be revised and permit at least an exception to the no abortion rule. A hypothetical 'no new birth' group that would make abortions mandatory for all, would similarly need to add exceptions, because it would not be able to predict the dangers of a continually ageing population. On the other hand, a pro-choice NMT does not such risks, simply because it allows for a complex society of anti-abortionists, 'no new birth' supporters and everything in between.

The idea that a single group would ever succeed in eliminating all opposition and reaching total consensus on its ideas seems preposterous in a world as chaotic as ours. Even NMTs that fail all tests for completeness, rigor and consistency could well survive ad infinitum, despite the failure of their proponents to acknowledge that their 'oughts' diminish their chances for immortality. In a chaotic system, questions such as 'what would happen if everyone did X' are useless, because there is zero chance that everyone will do X. Nevertheless, technology, globalization, psychology and neurology provide ever greater opportunities to influence very large groups of individuals. All it takes for Nazi-like ideologies to prevail is a critical mass of supporters under the proper circumstances. As a result, dogmatic NMTs with 'oughts' based on visions of complete dominance over or annihilation of others need to be judged against criteria as absolute and uncompromising as the NMTs themselves. The rational version of 'immortality of people like me' necessitates the coexistence and cooperation with people who are not 'like me'. When the people who are not 'like me' actively pursue the elimination of 'people like me', a rational immortality NMT requires self defence and containment of the opposing faction.

Survival ad infinitum of the human race

The idealistic NMT that would define "our kind" as our entire species is an obvious culmination of the idea, as illustrated with the AI example. The narrow definitions of immortality are to be expected, accepted and dealt with at different levels of organisation. Even the definition of "our kind" as the human species is in fact a narrow definition, because it ignores artificial conscious entities and life in general. Immortality NMTs based on even broader definitions of our kind are certainly possible. Conscious artificial or extraterrestrial beings are currently in the realm of

science fiction and therefore it would be premature to spend much time describing an NMT to include them. The extent of animal conscience and the nature of conscience itself is still very much under debate, so we'll also leave that discussion for a later time. We will therefore limit the focus to the immortality of the human race.

Using the formal model described in Appendix A, philosophers and scientists could work together on a grand immortality NMT that would prescribe actions to increase the chances of our survival ad infinitum. We can only hypothesize on what the outcome of such an exercise would be, but a few key points seem to be quite likely, ranging from the obvious and unsurprising to the uncomfortable and difficult to accept.

Many of the prescriptions of a human species immortality NMT would be identical to some very old ideas. Such an NMT would inevitably condemn intolerance, crony capitalism, hedonism, inequality, illiteracy and most of what we consider violations of human rights⁶. The idealistic NMT would promote ecology, sustainable development, scientific research, critical thinking, healthy living, complexity and a ubiquitous long term perspective of our actions. The NMT would be proposed as an alternative to dogmatic religions, a philosophy of life constantly under scrutiny and revision, with the advantage of promising physical personal immortality.

Other prescriptions would be quite uncomfortable, such as the ought to embrace the robustness of chaotic systems, which at first seems counter-intuitive. Our urge to simplify and homogenize is quite natural and provides us with a false sense of security and predictability. Focus on the long term and acceptance of uncontrollable diversity are the main challenges to achieving consensus. Even more challenging is the mandate to take an interest in effects beyond our current three generation window. Failing to focus on our long term survival is the path of least resistance and the one that guarantees a bleak future for our descendants. It is exactly the path we are traversing, a path that will have to change, in order to avoid extinction.

Finally, the NMT might prescribe actions that many currently consider immoral, such as war against powerful fundamentalists attempting to suppress opposing opinions, upper limits to wealth or procreation etc. Such 'oughts' could potentially be proven to serve the interests of survival ad infinitum of the human race and would need to be carefully explained so that they could influence relevant policy.

Chapter 4 - Building Consensus

Knowledge as a consensus building game

The most troublesome aspect of any dogma is the claim on absolute truth. Discussions with dogmatic individuals are by definition pointless, since they believe they already know the

⁶ The mandate for immortality does not require the definition of natural rights. Such concepts would undoubtedly emerge as useful in a formal NMT based on the immortality of the human kind.

answers to particular questions. Any discussion on existential and moral issues is bound to challenge fundamental assumptions and requires unconditionally open minds. We first need to accept that absolute truths certainly exist, but we will probably never understand them and we will certainly never agree that they are absolute. Dogmatic individuals already “know” that they hold the one and absolute truth, so the rest of the discussion does not concern them.

It took 7.5 million years for the computer in the Hitchhiker’s Guide to the galaxy to come up with an answer for the ultimate question of life, the universe and everything. Unfortunately, the answer ‘42’ it came up with, had a problem. Even if the answer was ‘true’, no consensus could ever be achieved, because the answer was not comprehensible to the recipients. Conscious entities need arguments fit to their own predispositions, mental capacity and reasoning methods. In all likelihood, we lack the mental capacity to comprehend how the universe really works. All our theories and beliefs are simplistic, plausible models of reality. The models are simple enough for us to comprehend and plausible enough to be ‘useful’. Given that both mental capacity and the definition of utility differ from one person to another, our models of reality are guaranteed to be diverse. So we engage in an eternal consensus building game, with proponents of certain views trying to convince others to accept them. Long term consensus can be reached on specific components of a holistic model of reality, when the answers arise from common genetic dispositions or when the players of the game share common cultural backgrounds.

Genetic dispositions

The very fact that we are human predisposes us to answer some questions in a particular way. The human mind is certainly not an empty slate, upon which anything can be written. Evolution has equipped us with very basic, but essential for our survival tools, such as time perception, the ability for reasoning in a certain way etc. Moreover, our mind is influenced by biological mechanisms that determine whether a certain condition is pleasurable or not. However, our analytic and synthetic abilities are so powerful, that we can question what we were born to know as true. Optical illusions, the investigation of the true nature of time, genetic mental disorders and numerous other examples prove that innate ‘knowledge’ is not a reliable source of truth. Still, native ideas can provide answers we can agree upon because of the general similarity between the genetic makeup of all humans. Even though our reasoning capacity enables some humans to propose that humans should not procreate, our genetic drives ensure such views will always be the exception, rather than the rule.

Culture

Cultural traditions are much more than simple outputs of the consensus building game. Amongst the countless simple, widely accepted ‘truths’, one will find consensus on the consensus building methods themselves. The cultural background largely determines the mixture of critical thinking and belief in any given society and whether one is willing to accept the scientific method or interpretations of an ancient script as sources of knowledge. It is therefore quite understandable that use of force and indoctrination/education have routinely

been used as strategies in the attempt to steer the prevailing culture towards new directions. Such artificial strategies seldom result in sustainable cultural changes, mainly because of the chaotic nature of human societies. Their failure can be traced to their reliance on the balance of power between the person or institution that attempts to win the consensus game, versus the power of the other players to resist. The strategies *can* produce short and medium term victories, justifying their continued use.

Cultural traditions exert powerful pressures for conformity to predetermined worldviews and shape minds in very particular ways. However, even cultures can and do change. Freedom of information exchange guarantees the eventual failure of all indoctrination attempts, provided that enough critical thinkers are able to control their genetic dispositions and overcome their cultural backgrounds.

The scientific method

To answer questions using the scientific method, we create testable hypotheses and check how closely they match our observations. New observations will inevitably lead us to challenge our current version of the truth and modify it accordingly. No real scientist will ever claim that the method can provide absolute and final answers to any question. However, the method is guaranteed to provide the best possible answer, given our current level of understanding and experiences. The requirement for a hypothesis to be testable makes many valid questions impossible to answer scientifically. Certain questions have to do with very complex, chaotic systems that can not be modelled effectively, or require experiments that simply can never be performed in a controlled environment. Science utilizes statistics to circumvent the problem of accurately describing complex systems. The more a field relies on statistics for predictions, the more difficult it is to reach consensus. As a result, social sciences such as sociology and psychology are guaranteed to have various schools of thought. Other questions, such as the existence of God, are deliberately formulated in a way that could never result in a testable hypothesis. A scientist's opinion on such questions can never be anything more than just another opinion.

Revelations and dogma

Having defined knowledge as consensus on models of reality, we can look back on the historic notions of intuitions and revelations, which have also been proposed as possible sources of knowledge. Both differ from the notion of native ideas, in that the proposed source of knowledge is metaphysical. By definition, propositions of a metaphysical source of information are not amenable to scrutiny. Still, most humans are naturally inclined to exchange their freedom for the comforting guidance of authoritative figures. Revelations are attributed to the ultimate authority. If the conveyor of such 'knowledge' can convince others that it comes from such an authority, consensus is practically guaranteed.

However, people have not proven to be as universally gullible as required by the proponents of revealed 'truths'. Since the proposed intuitions and revelations are not shared by large groups of people, the rest need to be convinced that the views produced by such sources are more plausible than others. The sheer number of different religions and sects stands as proof that

metaphysical sources of answers can not provide a basis for widespread consensus. Even devout followers of any particular religion are bound to provide their own interpretations of at least a few parts of their holy books.

On the other hand, one must admit that there *is* widespread consensus on the plausibility of generic metaphysical *concepts* such as deities, immortal souls, angels and demons, paradise and hell, because such concepts are compatible with innate genetic dispositions. The success of such “revealed” ideas throughout humanity’s history has shown that they address some of our most fundamental needs. It is precisely these needs that empiricism and philosophy have failed to address and which this discussion is focused on.

Use of force

The use of psychological and/or physical force as a consensus building strategy also targets fundamental genetic dispositions. Use of force can easily provide short and medium term victories, but is an untenable long term strategy. In chaotic systems such as human societies, the balance of power is never stable and one can seldom hope to eliminate all resistance. The strategy is most useful when mixed with other strategies and is especially successful when it is incorporated into indoctrination strategies.

Indoctrination

One can’t doubt the successes of indoctrination without ignoring the billions of religious followers who were taught absolute truths, or the millions of Americans and Europeans who have permitted a small number of financial oligarchs to control their governments and their lives. Religious indoctrination has had its unquestionable successes, but the strategy carries with it the burden of revelation-based truths and dogma, which have already been discussed. 21st century crony capitalism has arguably been even more successful than religion, because its success is based on the unprecedented power of the media delivering refined disinformation campaigns and on educational systems that produce workers and consumers, instead of critical thinkers. Crony capitalism has proven successful because it targets primal instincts better than any power building system before it and its marriage with pretend democracies is all but ingenious. Still, the reliance of crony capitalism on perceived freedom of choice renders the system at least as vulnerable to change as monarchy before the French revolution.

The goal of any indoctrination campaign is to cultivate a persistent, self-reinforced culture that promotes a particular version of truth to the members of a society. The vast majority of Americans reinforce the illusion that their society is merit based, ignoring the inherent inequalities of any system that links economic success to political power and the clear lack of opportunities for underprivileged segments of their society.

Rationality in the consensus game

Given the arsenal at the disposal of dogmatic individuals and organizations, there is good reason that rationality has not been able to prevail in the moral and existential arena. In the 6 millennia of recorded history, myth and dogma have been the norm. Only a handful of societies were given a handful of centuries in which to freely develop critical thinking. Practically all our

positivist knowledge comes from ancient Greeks, middle-ages Arabs and 18th-21st century western civilizations, none of which were ever really free from religious revelations, force and indoctrination. All rational answers to moral and existential questions were provided by a few hundred people, over the course of a few centuries. Myth and dogma have had millennia of countless contributors to establish themselves and refine their methods of achieving consensus.

To make matters worse, rational individuals are naturally averse to utilizing the consensus building strategies of coercion and indoctrination. Elaborate and flimsy justifications need to be invented, for rational individuals to justify infringing on another's freedom. However, when it comes to controlling fundamentalism and preserving freedom of thought, every available strategy must be utilized.

Fostering diversity and critical thinking

We have repeatedly stressed the fact that diversity is paramount to sustainability and robustness. Evolutionary biologists, game theorists, computer scientists and mathematicians will need to prove the importance of diversity over and over again, crushing all opposition, until it is taught in every elementary school in the world. The war against intolerance must be total and unrelentless. The word “sacrilege” must be erased from the human vocabulary and all initiatives that support critical thinking must receive total support from anyone who claims to be rational. Every single “must”, including the ones in this paragraph, must be open to questioning and improvements. If there is only one thing on which consensus should ever be reached, it is that nothing is beyond criticism. Even the scientific method must be presented for what it is; not a tool that provides “facts”, but a very useful, time tested method of reaching consensus and providing predictions. Failing to enforce diversity and critical thinking is a certain way to distort the proposed NMT and convert it to just another dogma. The mandate for immortality must also be presented for what it is, a candidate for consensus that is subject to revisions and should be compared with other moral goals such as happiness, equality, fairness etc.

While critical thinking is abhorrent only to fundamentalists and to the powerful proponents of the status quo, the level of diversity necessary for immortality will put everyone beyond their comfort zone. The word “chaos” carries many negative connotations. People like simplicity, so convincing them that simple systems are disadvantageous for survival will not be easy. Embracing chaos appears counterintuitive, until people realize how chaotic their own lives are, how impossible it is to simply describe their relationships with their environment and with each other. People are being taught deterministic science with more or less straightforward equations. No one has explained to them why statistics are so necessary in more complex systems. Once again it's up to education to help people realize how big a part chaos plays in their everyday lives.

The difference between education and indoctrination should be evident, but unfortunately it is not. Traditional primary and secondary education is usually dedicated to teaching facts and useful skills. Instead of opening minds to different possibilities, it molds them into predictable answering machines. Young minds are assumed to be too immature to grasp the assumptions

behind the 'truths' they are taught. As a result, traditional education is no less a tool of indoctrination and cultural compliance than Sunday sermons and public stonings. The focus on individual talents and critical thinking is still a sought after novelty, in a system infamous for its inability to change.

Considering the long term

Convincing people to become interested in longer time scales will be the ultimate test of any immortality NMT. Our species' focus on short term success is an understandable evolutionary byproduct that has stood the test of time, thus far. We have already stated that consensus needs to be based on common predispositions, so getting people to concern themselves with long term repercussions may appear as a lost cause. A window of opportunity in any such discussion comes from highlighting the actions people already take for their own personal immortality. Striving to be remembered for posterity, having offspring who will outlive us and investing in a good afterlife are quite common activities. Our fear of death causes us to do many things, so we do think about the long term, even if we don't always realize it. However, we don't consider future threats to mankind as relevant to our own long term endeavors. For example, many of us would like history books to immortalize our name, but we don't conceive the Sun's death as a real threat, even though there will be no history books if our ancestors don't find a way to stop the process or get out of the solar system.

Logical arguments can and are already being made in support of a different, more long term perspective. Support for the idea of sustainable development may not be overwhelming, but it is not entirely discouraging either. Long term effects that transcend the three generation care bias can be disguised as 'good for our children', even if 'our children' would never know the difference. Emphasis on education and critical thinking will certainly help, but appeals to rationality coming from rational individuals have a rather poor record in achieving consensus.

The true game changer in the pursuit for the adoption of a long term view will be the possibility of physical, personal immortality. When an actual fountain of youth comes within reach, mentalities are bound to change overnight⁷. The more widespread the possibility of personal immortality becomes, the more people will begin to think in terms of centuries and millennia, without ever transcending the three generation care bias. Grasping the concept of 'ad infinitum' will still be very much out of reach in such an eventuality, but a time frame of a few centuries would probably suffice. The predictive capability of any model will be so limited after a few centuries that people would need to accept the need for diversity, complexity and sustainable policies as the only means towards the end of dealing with unforeseen circumstances.

⁷ One extremely important repercussion is undeniable and very relevant to the purpose of this discussion. When science does offer actual physical immortality, the concept of immortal souls will no longer be necessary to assuage people's fear of death. It would be difficult to predict the extent to which such a fundamental shift would affect the penetration of religions overall, but the dogmata that rely on the promise of a better afterlife would certainly become less relevant.

Intolerance of intolerance

Just as one person's freedom stops where another person's freedom is violated, the mandate for diversity does not extend to views that attempt to limit it. Groups and individuals will always fight for their perceived right to limit the rights of others, be it women, children, other races, or whoever else they do not consider "their own". Such views may be diverse themselves, but can be proven to lead to lower diversity and less robust equilibria overall. An immortality NMT can't tolerate intolerance in any disguise, regardless of the definition of "our own". Both indoctrination and use of force are quite acceptable as temporary methods of combating racism, bigotry, sexism and overall intolerance, until diversity and critical thinking become widespread cultural traditions. Still, one should be wary of going too far. A sufficiently complex and chaotic society will always include harmless intolerant individuals. Completely eliminating intolerance, if at all possible, would probably require policies and tactics that might tilt the scale back towards less overall diversity. It is expected that both qualitative and quantitative methods could be used to determine the extent of reaction required in particular situations. Much experience has already been amassed in such matters in open, liberal societies. A big part of the equation is the risk posed by such individuals or groups, which is directly related to their power.

Imposing limits to power

Building consensus on sustainable diversity will be impossible, unless the people who have the most to lose from such an eventuality have limited power to resist it⁸. We won't presume that we can provide a strategy to shift the balance of power. We can however say that wherever possible, the proponents of the NMT would try to limit the power others have to curtail openness and critical thinking.

Chapter 5 - From Consensus to Policy

A naive view of democratic societies is that the actions prescribed by the most prevalent NMTs will be legally enforced. Laws in democracies supposedly cater to the masses and express a society's cultural traditions and preferences. Of course laws are simply an expression of power. The masses in most modern democracies only have an illusion of power, constantly alternating between slightly different versions of the same, interest controlled governments. Whenever a government attempts to escape the clutches of the global financial status quo, the masses inevitably end up suffering. Their sacrifice becomes a lesson to other populations, who might dare consider that alternative policies might be conceivable.

Power consolidation is a tremendous threat to the adoption of policies that will lead to sustainable development. Fortunately, constantly increasing inequality is itself an unsustainable trend, that history has corrected repeatedly. Cycles of popular uprisings, establishment of new rules and renewed power consolidation may last centuries or even millennia, but they do happen. One can't say if the rules established by the next cycle will incorporate prescriptions from the proposed immortality of our species NMT. It is however all but certain that such policies

⁸ Excessive power also directly limits diversity, but this section is only concerned with the possibility of reaching consensus on an alternative theory.

will not be enacted as long as an elite interested in the profits of the next financial quarter controls lawmakers. Public policy devoted to sustainability will inevitably need to address some of the challenges discussed below.

Sustainable societies

NMTs based on less selfish definitions of “our kind” are not very interesting when we limit the investigation to the views held by individuals. Some people always have and always will choose to extend the definition of “my kind” to their families, people who share particular views, people of certain ethnicity, race etc. While a rational goal at a group level would be to strive for the widest possible adoption of the widest possible definition of “our kind”, such a goal is not necessary at the personal level. Due to the strong mandate for diversity, the proposal for rational consensus includes the expectation that individuals define immortality as they please. The important issue for policymakers is how groups define “our kind” and what type of rules they are rationally expected to enforce. Group policy must foster actions that cater to the oughts of wider definitions of “our kind”, without placing severe limits on the diversity of individual NMTs.

In Chapter 2 we stated that consensus on morality must begin and end with the mandate for survival, because survival ad infinitum is impossible without diversity. Any attempt to further specify oughts will inevitably limit diversity. Utopian, peaceful societies are undesirable because they are too simplistic and static to guarantee survival ad infinitum. Internal conflicts create a complex, dynamic, adaptable system, capable of responding to unforeseen external threats. The rational proposal is based on a consensus that conflicts between group policy and individual goals are unavoidable and even desirable. Similar conflicts are expected and desirable when we look at larger groups, from families, communities and interest groups to countries and global institutions. At each level, smaller units will always attempt to influence policy of the larger groups to achieve their own goals. We already mentioned that smaller units may rationally attempt to limit diversity, when it directly threatens their primary goal. The key to immortality is not to eliminate conflicts, but to discover how the larger groups can resolve such conflicts without limiting diversity to the point of harming the basic goal.

In appendix B we illustrate how such challenges may be addressed with a fictitious example of how conflicting definitions of “people like me” may result in a robust equilibrium that doesn’t limit diversity. Here we discuss more general issues that policy makers need to contend with.

Resolving challenges to diversity

A good argument against normative moral relativism is that it can’t even condemn intolerance. If there is one thing the pursuit for immortality can not tolerate, it is intolerance. Even dogma can be permitted to exist, as long as it does not dictate policy. The mandate for diversity is so necessary for long term sustainability, that even the craziest fanatics are tolerable. What fanatics can never be permitted to do is to infringe on other people’s freedom to posit and follow their own NMTs. Phenomena like the Islamic State are prime examples. Their very existence

limits the behavioral diversity of the entire world, because they actively enforce conformance within their area of influence.

Passing judgement on fanatics who decapitate dissidents and destroy historic landmarks is easy. Extending the mandate to groups that engage in indoctrination is more controversial. Few societies have escaped the grip of religious dogma and no nation is completely innocent of misrepresenting history to instill pride in its members. We have already rejected a utopian brotherhood of humans as too simplistic and static to be sustainable, but racism and bigotry stand on the other end of the spectrum. Being proud of belonging to group A does not necessitate discrimination against members of group B. The mandate for immortality must result in the condemnation of such discrimination as harmful to diversity and prescribe actions to curtail it.

Transcending national boundaries

One point we need to highlight is that we currently have no legislative body that can assume the role of protector of the survival ad infinitum of the human race. As global warming has shown, we are not yet mature enough to organize and cooperate across national boundaries. It should be obvious that this stance must radically change, if we are to ensure the survival of our species, as more serious global challenges become apparent. The process towards global accountability has already begun and advocating the goal of immortality clearly as a moral mandate can only support it to a speedy and necessary completion.

Dealing with immortality

A rational society would recognize an individual's right to pursue personal immortality and provide a framework that could sustainably handle the ever increasing number of immortal individuals. Sustainability inevitably leads to population caps, a fact that our societies have largely chosen to ignore. Providing the immortality option only to individuals willing to pay heavily for it and introducing birth limits are both rational, sustainable strategies. Irrational pressures towards unsustainable, state subsidized life extension treatments without limits on the number of children are expected, but can't be condoned. Such pressures are based on fairness concerns and have little to do with the proposed candidate for rational consensus. The moral mandate for fairness is an innate characteristic that can't be ignored, but it needs to be put in the correct perspective. To prevent the inevitable objections, we should note that a wealth cap and a more sustainable way of life enabled by technology will result in the ability of our planet to support much greater numbers of individuals than what is currently considered feasible⁹.

Even so, one can't help but ask the obvious question, if it's fair for some people to enjoy individual immortality, while others can't. The answer is that no less fair than some people being

⁹ Just reducing meat and fish consumption and food waste would have a tremendous impact on greenhouse gases, loss of biodiversity due to deforestation and fish stock replenishment.

born to wealth and not having to work a single day of their lives. A fair society exists only in literature and in dreams. The powerful will always have the lion's share of everything. What immortality theories will undoubtedly mandate is that the particular share does not come to the expense of the survival of others, as it does today. One will be hard pressed to find a religion or an ethical theory that doesn't condemn the mindless accumulation of wealth and power, yet here we are, centuries after the last kings with any true power are gone, still accepting the fairy tale that anyone could become a super-rich oligarch in a free and fair society. There is absolutely nothing wrong with being born into a life of privilege. What is intolerable is to use one's privileges to harm others directly, or to harm others' chances to challenge your power and authority. We have touched upon the adverse effects of concentration of power to the probability of immortality of wider groups and even explained why the narrowest definition of immortality would need to respect the need for others to exist and thrive. However, we need to clarify that concentration of power inevitably leads to limits to diversity and is plainly not acceptable; not because it's unfair - i.e. results in a less than equal playing field - , but because it directly reduces diversity.

Epilogue

I'm personally not certain that the human race is worth preserving. We are savage and fundamentally flawed in many respects. We do have admirable qualities, but no one can tell if dinosaurs would have made for more sustainable and admirable societies, if it wasn't for that meteor that gave us our chance. I do know that my assessment of the human race's qualities is completely irrelevant. What matters is that we are still here and we have a damned good chance of being here forever. Will we make it? Probably not, unless if we direct our innate drives away from frivolous distractions and concentrate on what matters: the survival of our species.

Appendix A - Formalizing Normative Moral Theories

Removing Moral Theories from their Pedestal

The view of knowledge as a consensus building game is consistent with metaethical moral relativism but not necessarily with normative relativism. Metaethical relativism is consistent with the pursuit of immortality, which necessitates behavioral diversity. Rejection of normative relativism is a prerequisite for any form of rational consensus and in order to allow for rational criticism of particular moral theories. Within the broad confines of metaethical relativism, the war against dogma necessitates a careful delineation of what can and what can not be said about NMTs.

If one were to compare the current freedom to investigate and criticize widespread beliefs on moral issues with the freedom to pursue scientific research in the natural sciences, one would find little, if any progress since the times of the ancient Greek philosophers. Ethics in general and religious ethics in particular have been placed on the pedestal of the unassailable personal freedom of thought. Even relatively free societies tend to put the desirable freedom of one person to hold one's own moral views over the freedom of others to openly criticize them. The very existence of stable societies depends on common laws, aligned to the views and interests of their more powerful members. Open attacks on such views are often treated as attacks on the stability of the society. Scientific findings are routinely ignored and even actively combatted, when they threaten the status quo. Anyone proud of the freedom of thought in modern western societies should be humbled by the inability of our societies to change their worldviews and rise to the challenge of climate change. Rational individuals should be able to prove that any moral theory that ignores the implications of climate change is fundamentally flawed and needs to be revised. However, no such argument is made, because we lack the tools to make it.

In this chapter we propose a formal model of normative moral theories that permits rational consensus on their criticism, without the dangers of encroaching on personal freedom of thought. The formal model not only facilitates the ensuing discussion, but also provides a framework for future scientific investigation of the moral implications of immortality.

Origins of Normative Moral Theories

The universe is a closed system which contains conscious beings. Conscious beings are defined as the subsystems with at least the following characteristics:

- Analytical capability. They are able to decompose the perceived universe into clearly defined, layered classes of structures, with clearly defined states.
- Synthetic capability. They are able to formulate general statements to describe relationships between the classified structures.
- Planning capability. They have the perception of time, can identify cause and effect and formulate plans of action.
- Action. They are able to execute automated and planned actions to achieve their goals.

- Sense of identity. They have the perception of a self, which they classify as distinct from the rest of the universe
- Volition. They have innate drives and the perception of will.

The innate drives bias the conscious systems towards particular goals and actions. However, once fundamental needs and wants are satisfied, volition is quite malleable and amenable to cognitive influences. Due to the aforementioned characteristics, conscious systems perceive themselves as having free will and address environmental challenges by setting additional goals, not directly dictated by, or even contradictory to their innate drives. The formulation of a Normative Moral Theory is one such expression of the perception of free will.

All NMTs assume or define the relationship between the individual and the rest of the world (describe what is), set goals (posit oughts) and prescribe ways to achieve the goals. More often than not, one or more elements of a NMT are assumed, or implied. Formal NMTs must clearly express all assumptions in unambiguous terms.

Benefits of a formal model for NMTs

Bring forth hidden assumptions, eliminate abstractions and facilitate understanding. Set boundaries for criticism.

NMTs have a long tradition of providing guidelines based on the assumptions and knowledge of their time. As a result, there was little need for rigorous justification of the particular guidelines. Free, democratic societies accept that individuals may hold different beliefs, as long as their actions do not stray considerably from the range of actions accepted by the society. At times, the wide range of opinions means that no true consensus can ever be reached on particular subjects, leading to constant conflicts and power struggles. In an ideal situation, honest discourse may lead to clear understanding of the reasoning behind the various beliefs and to the attainment of a relatively fair compromise. However, as long as the assumptions are not unambiguously formulated and the limits of criticism are not well established, misunderstandings and aphorisms are unavoidable.

A formal model for NMTs can facilitate understanding by restating implicit assumptions and the reasoning used to reach particular conclusions, in widely understood terms. A formal model can expose fundamental, irreconcilable differences in viewpoints that need to be accepted, as well as logical fallacies that can be constructively criticized. A formal model can't eliminate conflicts, but it can offer a framework for more constructive dialog between the opposing factions.

Enable synergies with social and natural sciences, mathematics and computer science

Ethics is certainly not a science and can never become one. Conscious beings may agree on many things about their world, but they will always disagree on how one should lead one's life. However, a formal model can permit ethicists to utilize the considerable arsenal built for other sciences, in order to facilitate predictions about the effects of certain actions on the pursued goals. Science can only help when the problem is clearly formulated and ethical issues never

are. We will later show how a formal model enables the use of software simulations, statistics and chaos theory to rigorously justify prescribed courses of action, based on arbitrary assumptions.

Clearly outline the boundaries between ethics and science

One practical application of a formal model is to once and for all define what science can and what it can't say about how one should lead one's life. The historical clash between religious dogma and rationality will be perpetuated, for as long as rationalists provide no definitive answers on existential and moral issues. Rationalists are either honest and admit that no such answers have thus far been provided, or wrong and attempt to provide them based solely on the scientific method. The scientific method is not the only tool in a rationalist's arsenal. Once the boundaries of the scientific method on moral issues are clearly established, a rationalist can target the remaining gap with reasonable assumptions and pure logic.

Enable Ethical Artificial Intelligence

Even a rudimentary instruction set such as Isaac Asimov's laws of robotics, requires programming a robot with a simplified NMT. In order for a robot to be able to be programmed with these rules, it would need to:

- Be able to classify 'what is' into at least 3 categories ('human', 'myself', 'everything else')
- Be able to distinguish between 'harmed' and 'unharmed' states for 'human' and 'myself'.
- Assign non-zero value on the 'unharmed' state for 'myself' and a much higher (maybe infinite) value on the 'unharmed' state of a 'human'.
- Be able to predict which actions may be 'harmful' or 'harmless' to whom.

Any computer scientist will quickly realize the difficulty of programming these instructions to a robot, unless assumptions humans consider unquestionable, such as the definition of 'harm' can be formalized.

Elements of a formal NMT

NMTs describe what is

A formal NMT MUST explicitly describe all perceived classes, states and interrelationships, in terms unambiguously understood by a fluent speaker of the language.

To posit an NMT, conscious beings either assume as common knowledge or define the relationship between the individual and one or more other structures. They arbitrarily decompose the perceived closed system (universe) into subclasses (e.g. family, society, environment etc.), identify class instance states (e.g. happy parent, just society) and define interactions between the individual and the instances of such classes (e.g. a human depends on plants and animals for sustenance). Defining a class requires setting arbitrary, but clear boundaries. For instance, a conscious system may split the universe into just two classes, one containing the conscious system itself and the other containing everything else.

The subjectivity of perception and the practical difficulties of accurately defining abstract states such as 'just' or 'happy' guarantee that the consensus on what is can never be universal. As a result:

NMTs posit oughts

- *A formal NMT MUST assign arbitrary value functions to class instances and/or instance states.*
- *A formal NMT MUST mathematically define targets for the assigned value functions*

NMTs set goals for the individual, the collection of individuals or any other instance of the described or assumed classes. They assign arbitrary values to instances of the described classes (e.g. an individual life), or to perceived instance states (e.g. an individual's happiness). To posit clear goals, the NMT must assign non-zero value to instances of at least one class. The value may well be infinite, suggesting that no system instance can be sacrificed to pursue another goal. A NMT may also assign values to particular system instance states¹⁰. The NMT must then set a target for the assigned values. For example, a NMT might assign a constant value (V) to every human life L and another value to the human state of happiness, for each human H(i). It might then set various goals (G), such as the following:

NMT 1: As many people as possible, as happy as possible

$$G = \sum_{i=1}^L (H(i) + V)$$

NMT 2: As many people as possible, as happy as possible, but with a lowest bound C for happiness and a greater emphasis on happiness

$$G = \sum_{i=1}^L (H(i) * V) \text{ where } H(i) > C$$

NMT 3: Seek to maximize the average happiness

$$G = \frac{\sum_{i=1}^L (H(i) + V)}{L}$$

NMT 4: Look into the long-term and target a sustainable level of happy population (Expected value, for time 0 to infinity)

The assigned values themselves need not be constants. For instance, the value may be a function of the number of instances of the particular system, allowing for increasing value as the number of instances decreases (endangered species).

¹⁰ Most NMTs will only assign value to instance states after assigning value to the instances themselves. However, there is no reason to a priori preclude the possibility of a NMT that considers the life of an individual worthless below certain thresholds of freedom, happiness, equality, security etc.

NMTs prescribe actions to achieve the set goals

- *A formal NMT MUST provide rigorously justified, clear guidelines as to how the targets should be pursued, under particular circumstances.*
- *A formal NMT MAY provide rigorously justified, practical rules of thumb to guide decision making in a wide range of situations.*

A NMT prescribes actions and methods of choosing between actions that lead to the attainment of the set target(s). Even NMTs that agree both on what is and what ought to be, do not necessarily reach the same conclusions as to how one should act. Even with the simplistic goals described above, accurate value determination would be practically impossible, especially for large time scales. As a result, most conscious systems (humans for certain) would need to resort to heuristics, reasonable assumptions and simulations, based on largely acceptable theories and widely held beliefs. A formal NMT must therefore justify its prescriptions based not only on what is and what ought to be, but also on the methodology used to reach the conclusion.

The effects of certain actions may depend on how prevalent they are (e.g. free rider problem). More generally, certain actions may not be directly or even indirectly detrimental to the pursuit of the set goals, but they may be shown to increase the risk of straying from the desired path. To cover such cases, a NMT may prescribe practical rules of thumb, such as the universal law principle, taking into consideration the effect on the set goals, if a particular action became a universal law. The justification of such wide-ranging guidelines can't be expected to be complete, but a NMT must show how their application generally leads to the desired goals.

The boundaries of criticism

Completeness

A formal NMT is complete when it describes all relevant system classes and interconnections.

Formal NMTs need not agree on what is, but they do need to provide a complete view of the proposed network of systems. An obvious point of contention is the proposition of a 'Deity' class, with one or more instances. A NMT can't be criticized for proposing that such a class does exist, but it can be criticized if it fails to completely describe the interactions between the instance(s) of such a class and the rest of the described network. As a result, religions claiming that "we can't know the will of God" would be quite difficult, if not impossible to restate as complete, formal NMTs. To offer a less contentious example, a NMT can be shown to be incomplete, if it fails to account for the effect of humanity's actions on the environment and the feedback effects of those actions to humanity's well-being in large timescales.

The number of system classes defined by a formal NMT is bounded only by the conscious system's capacity to accurately describe the resulting network. A coarse classification is certain to result in an incomplete theory, but the finer the separation, the greater the risk of misclassifying poorly understood systems, or poorly addressing the dependencies between the

identified classes. As a conscious being's understanding of the universe expands, previously unknown distinctions and relationships become apparent, rendering existing theories incomplete. As a result, omniscience is the only way to achieve a perfectly complete NMT and no theory can remain complete forever. The very mention of an omniscient consciousness brings back religious doctrine as a candidate for a timeless, complete NMT, but religion does not actually describe 'what is'. Religious doctrine holds that the NMT was proposed by a superior consciousness, which deliberately chose not to reveal 'what is' to the followers of the religion. Once again, the result is that religious doctrine can't be restated as a formal NMT.

Consistency

A formal NMT is consistent when its goals are attainable and when its prescriptions can be proven to serve its goals.

When positing 'oughts', the value functions and targets must be carefully chosen, in order for the goals to be theoretically attainable. A NMT may obviously posit as many targets as it wishes, at the peril of rendering the consistency requirement practically impossible, since the complexity of the systems under consideration will probably lead to conflicting goals. Conflicting goals will need to be restated in a way that provides clear target states, the result of a compromise between the unattainable ideals.

NMTs are supposed to prescribe actions that protect system instances and system states with non-zero values, taking into account all system to system interconnections. Regardless of the prescribed method for deciding on a course of action, the resulting actions must be shown to be consistent with the goals. A critic of the NMT may challenge a rule of thumb or heuristic used, by presenting a case where the prescribed action undermines the NMT's own goals.

Rigor

A formal NMT is rigorous when it can make testable hypotheses for the effect of certain actions on the desired goals.

We explained how a superficial heuristic may lead to the violation of the consistency requirement. Lack of rigor is the greatest threat for a formal NMT. The scientific method is based on hypotheses, predictions and independently repeatable results. Given the complexity of the systems under consideration, proposing a fully testable NMT may be quite difficult, but not theoretically impossible. The less a NMT relies on heuristics, the more room it leaves for rigorous use of mathematics (e.g. chaos theory, game theory, statistics, network theory), peer reviewed socio economic and anthropological studies, computer simulations etc. If there is one easy criticism one could make for all informal NMTs is their lack of rigor. The path towards rigorous formal NMTs will be lengthy and arduous, but the benefits outweigh the costs.

A formal NMT can be criticized for lack of rigor in the following cases:

- When it describes what is, if it fails to provide unambiguous definitions of the described classes, or if it provides unsubstantiated claims on the dependencies between the classes.
- When it posits oughts, if the value functions it provides are not measurable. For instance, 'happiness' is not a clearly defined system state, but dopamine and serotonin levels are.
- When it prescribes actions, if its reasoning contains logical fallacies, or if it makes poor use of scientific methodologies. For instance, a NMT might utilize a computer simulation to predict the effect of a certain action on a set goal, but the software may later be found to have had a critical bug that skewed the results.

The boundaries between ethics and science

The role of science in describing what is

- Science can assist in identifying potential elementary system classes (e.g. wolves vs dogs) and especially in describing the interconnections between the various system classes.
- Science has no say on whether a NMT treats all animals as a single superclass or whether it separates them into dogs and non-dogs.

The role of science in positing oughts

- Science can assist in clearly defining system states (term disambiguation) and in calculating value functions.
- Science has no say in the values assigned, or in the set targets.

The role of science in prescribing actions

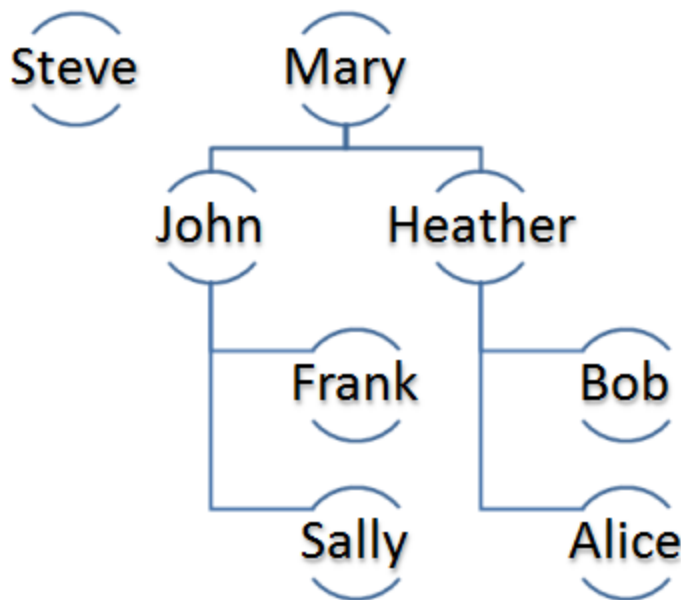
- Science can assist in the complex predictions of the effects of particular actions. Aside from providing tools to facilitate difficult calculations, it can be used to prove that a utilized heuristic is unfounded, because certain feedback effects from particular interconnections were ignored during the calculations.
- Science can provide alternative candidates for prescribed actions, assuming the NMT's definitions of 'what is' and what 'ought' to be are not challenged during the investigation.

Appendix B - Conflict resolution scenario

We investigate a traditional family, a small group of people defined by biological kinship. Assuming that a consensus is reached on the goal of immortality for family members, all the courses of action already mentioned in the selfish scenario would apply. In addition, increasing the number of family members to enhance resistance to external threats would be quite important. The "immortal overlords" scenario would still be irrational, since biological kinship severely limits diversity and adaptability. The assumption of consensus at the group level does

not imply that all family members share the same definition of “our kind”. Some family members may choose the more selfish definition, while others may push for a broader definition of “family” than the one already accepted by the original group. The interesting question is how the group can maintain consensus and adapt to the various pressures without limiting individual freedom to the point of harming the family’s chances for immortality. The investigation will show how conflicts can rationally be resolved by using the formal NMT model presented in the previous chapter.

Let us assume that Steve and Mary have a small family in a time when personal immortality is possible, unregulated and expensive. Their family tree is shown below:



Mary’s excess income is sufficient to cover the necessary treatments for three individuals. Her initial inclination is to define “our kind” with the narrowest biological definition of biological kinship, excluding her husband Steve and catering only for her and her children, John and Heather. Mary attempts to enforce a legally binding consensus wherein all her excess income is usable only by her and her children, for the sole purpose of extending their lives. John and Heather each have a daughter and a son of their own. Heather wants her own children Bob and Alice to fend for themselves, but insists that her father Steve also receives life extension treatments. On the other hand, John does not care much about the father, but does not want to outlive his own children, Frank and Sally. Only Frank is able to pay for his own treatments and he refuses to assist Sally or any other family members. John refuses to accept life extension treatments with Mary’s money, he wants the money to be diverted to extend Sally’s life. Mary realizes that the situation is not tenable and tries to come up with a more rational way of utilizing her excess income to increase the chances of her family’s survival ad infinitum.

In the particular moral universe, we assume that Mary’s definition of “what is” is based solely on

biological kinship. Given such a definition, the father and any adopted children or grandchildren would not be considered members of her biological family. The defined classes that cover the proposed scenario are the following:

- The mother herself.
- Male child, with 50% of the maternal chromosomal DNA (cDNA). {John}
- Female child, with 50% of the maternal cDNA plus 100% of the maternal mitochondrial DNA (mtDNA). {Heather}
- All grandchildren born from male children and male grandchildren from female children, with 25% of the maternal cDNA. {Frank, Sally, Bob}
- Female grandchildren from female children, with 25% cDNA and 100% mtDNA. {Alice}
- Everyone else, with 0% kinship. We assume there are no cousins etc.

The definition of classes can be extended to include great grandchildren etc. but we limit the scope to simplify the discussion.

The simplified “ought” set by Mary is to ensure that as much of her genome as possible survives for as long as possible. Mary wants to find a way for her excess income to guarantee the best possible chances of her genome surviving ad infinitum. As a final simplification assumption, cloning herself and having more children is considered impossible. The question that Mary needs to answer is how to distribute her excess income, in order to achieve her “ought”, without ignoring the “oughts” of her children, who have psychological bargaining power.

A strictly mathematical approach that doesn’t take into account the “oughts” of Mary’s children would show that Mary’s best course of action would be to hedge her bets and distribute her excess income in relation to genetic similarity. She would spend a large part of the income on her own treatments, a smaller to Heather’s, then John’s, then Alice’s and the remaining on Bob and Sally, because Frank doesn’t require any help. A slightly more complex approach would also account for factors such as the probability of death by accident, using an approach similar to that of life insurance companies to determine the amount invested on each family member.

However, Mary can’t enforce such a consensus, given Heather’s emphasis on the immortality of her father and John’s emphasis on the immortality of his daughter. Mary needs to accept that her family members have their own definitions of “what is” and “oughts”. A consensus requires a sustainable, family-centered NMT that every member can accept and support. What Mary must realize is that each individual’s definition of family is inevitably different. Her own definition of family is based on biological kinship. Even if every one of her children and grandchildren somehow agree on her definition, their own family will be different than hers. Changing the point of view immediately changes the definition of what is, because the definition of ‘what is’ is always subjective and self-centered. Mary must therefore accept that all she can hope for is to do what she believes is best for her own kin and let the others decide for themselves.

Mary clearly can’t reach 100% consensus with her children and achieve the optimal genetic result, since her children insist on money going to two of her most distant relatives. In a viable scenario, Mary says to Heather that she will provide for both Heather’s and Steve’s treatments,

for as long as they both live. Since that won't be enough to guarantee both Heather's and Steve's immortalities, Heather must accept that, if she dies before her father Steve, Steve's treatments will immediately stop. The one-person treatment funds will then go Mary's closest kin who is still alive (preference to John, then Alice, then Sally or Bob). If Heather refuses the deal she gets nothing and Mary pays for John and Alice. Similarly, Mary tells John that she would never give all the money to Sally. At best, John will have to accept partial treatments for both him and Sally. If John refuses or when he dies, the funds will go to Mary's closest kin with the need (preference to Heather, then Alice, then Sally or Bob). In this scenario, both John and Heather are free to devote their own wealth in any way they please. Another possible candidate for consensus would be for everyone receiving funds to have to cover a portion of the cost (like government subsidies). Yet another scenario would be for Mary to establish a life extension fund, that would function similarly to an insurance policy. The fees to be eligible for the funds would depend on the person's kinship with Mary (less for John/Heather, more for the grandchildren etc.).

The situation is ideal for game theory, but the point of the exercise was never to find the optimum solution to such a simplistic problem. Mary's emotions towards her children and their own power of persuasion could lead to consensus on scenarios that, from Mary's point of view, would be much worse than her ideal. But even if Mary had all the money in the world, no one can guarantee that a family reunion would not end with a tragic accident, killing the entire family. Survival *ad infinitum* is all about probabilities. There is no single, perfect answer to how a group might best survive, only educated guesses. Each Mary in the world will do whatever she can to cater to "her own" and hope that it works. The understanding that everything can't and shouldn't be controlled is fundamental to the proposed candidate for consensus. On the other hand, Mary can't give into every single demand, just for the sake of diversity. Even if satisfying every individual point of view were possible, the net result would most likely lead to severely decreased chances of survival for Mary's genes.

Similar scenarios are already being played out in the billions, with different actors and interests involved. The mandate is that diversity is preserved as much as possible and that the net force points towards a sustainable future. Widespread intolerance and focus on the short term are the main hurdles on the quest for immortality.