Richard Friedrich Runge*

Neo-Aristotelian Naturalism as a Metaethical Route to Virtue-Ethical Longtermism

https://doi.org/10.1515/mopp-2023-0089
Received October 20, 2023; accepted August 21, 2024; published online September 19, 2024

Abstract: This article proposes a metaethical route from neo-Aristotelian naturalism, as developed in particular by Philippa Foot, to virtue-ethical longtermism. It argues that the metaethical assumptions of neo-Aristotelian naturalism inherently imply that a valid description of the life-form of a species must satisfy a formal requirement of internal sustainability. The elements of a valid life-form description then serve as a normative standard. Given that humans have the ability to influence the fate of future generations and know about their influence, this article posits that our future as a species becomes a part of the description of the human life-form. Consequently, in our case, the sustainability requirement is extended into the future, leading to a minimal and humanist version of longtermism. As in Rosalind Hursthouse's naturalistic virtue ethics, the focus of the proposed virtue-ethical longtermism remains on developing virtues as the only reliable way for the agent to achieve the good life. However, since what can count as virtue must be compatible with a life-form description that, taken as a whole, facilitates the long-term survival of our species, this position still implies a genuine longtermism, although it contrasts with the quantifying and maximizing logic of consequentialism.

Keywords: existential risk; longtermism; neo-Aristotelian naturalism; rationality theory; virtue ethics

1 Introduction

Longtermism is currently known as the view that "positively influencing the long-term future is a key moral priority of our time" (MacAskill 2022, 4). This definition does not necessarily imply a commitment to maximization, although longtermism is often interpreted through a consequentialist lens that focuses on maximization. In

***Corresponding author: Richard Friedrich Runge**, Philosophisches Seminar, Christian-Albrechts-Universität zu Kiel, Leibnizstr. 4, 24118 Kiel, Germany, E-mail: rrunge@philsem.uni-kiel.de. https://orcid.org/0000-0001-8055-4668

any case, this definition implies that in the far future, there should be beings for whom something can be positive at all. For instance, we could hypothetically construct an impressive monument or work of art that would endure long after life in our universe had ceased, but this could not qualify as "positively influencing the future" if there were no beings left to appreciate or utilize it.

Given that longtermism, at its core, implies first and foremost the existence of beings for whom something can be positive, *minimal* longtermism can be understood as a position that we should make substantial efforts to ensure the continued existence of such beings in our universe. This view is weaker than typical consequentialist positions in that it does not specify that there should be particularly many of these beings in the far future or that they should have an overall surplus of positive over negative subjective experiences in their lifetimes. It might suffice if the surviving group had a minimal size that still facilitated long-term survival, living under conditions that, despite being possibly burdensome, still enabled its members to take responsibility or live virtuous lives. To make another distinction, *humanist* longtermism can be defined as the view that we should prioritize specifically the survival of the descendants of the human species. In light of these distinctions, Hans Jonas (1984), for example, can be considered a longtermist, albeit a minimal and humanist one.

Having decoupled the theoretical core of longtermism from a maximizing consequentialist position enables me to argue for virtue-ethical longtermism in this article. This virtue-ethical longtermism will be a genuine longtermism, meaning it will share some of the more controversial claims that are typically associated with this position. For instance, it will acknowledge that there can be situations in which the living generation is morally required to engage in demanding intergenerational collaboration to prepare responses to existential threats that will only affect generations in the far future. In terms of content, the position to be developed here aligns more closely to Hans Jonas than to William MacAskill, yet it remains distinct from both positions, particularly concerning the underlying premises and the methodology employed to derive these claims. Ultimately, I hope not only to make a convincing case for virtue-ethical longtermism but also to give longtermism a critical twist.

To better elucidate the characteristics and challenges of virtue-ethical longtermism, let me begin by providing an overview of standard consequentialist longtermism. Consequentialism, as a paradigm in normative ethics, posits that there are general intrinsic values with positive or negative valences. Hedonistic utilitarianism, as one of the most prominent versions of consequentialism, states that experiences of pleasure generate units of positive value, while experiences of pain yield units that have negative value. Consequentialism furthermore adopts an impartial standpoint. From that standpoint, two beings who are experiencing the

same amount of pleasure generate the equivalent intrinsic value, regardless of differences in location (e.g., living on a distant planet), species (e.g., being a Martian), or temporal placement (e.g., existing in the far future). The ethical aim of consequentialism is to maximize intrinsic value in the world, assuming that it is possible to aggregate these units of positive or negative value and compare different possible outcomes of our actions in terms of their overall value.

The majority of consequentialist positions espouse the view that intrinsic value should not be discounted simply for the fact that it will be generated in the future (see Ord 2020, 46, 253–258; MacAskill 2022, 258–260). Additionally, longtermism emphasizes that the potential intrinsic value that can be generated in the far future is so vast that it cannot easily be disregarded in favor of the present (see, e.g., Bostrom 2013, 19; Greaves and MacAskill 2021). Therefore, positively influencing the future remains a key moral priority, even when epistemic uncertainties, such as the possibility of our extinction in the near future, are taken into account (Tarsney 2023). Due to the vast potential intrinsic value that could be generated if humanity survives long enough, longtermism emphasizes the ethical importance of making significant efforts to minimize risks of our extinction and to maximize the future potential for creating intrinsic value. Consequently, it may be more prudent to donate money to AI safety research or take steps to colonize Mars rather than to focus on effective short-term charity, since AI could potentially be a future threat to the survival of our species, while colonizing Mars could significantly increase the chances of our survival.

While the basic structure of consequentialism makes longtermism almost a natural conclusion, it represents just one of the three major paradigms in normative ethics, alongside deontology and virtue ethics. Arguing from a virtue-ethical position, I am convinced that consequentialism has had a misguided focus from the beginning – that is, maximizing some kind of general intrinsic value in the world. Therefore, consequentialist longtermism, as a radicalized version of standard consequentialism, in my view only makes this misguided focus more apparent. Nevertheless, taking responsibility for our long-term survival as a species does indeed seem to be a critical task of our time, also from a virtue-ethical perspective.

However, when we consider potential alternative theoretical foundations for longtermism, even a minimal one, virtue ethics seems particularly ill-suited for such an endeavor. Virtue ethics is often viewed as the staunchest opponent of consequentialist impartiality, and it is only the most extreme form of this impartiality – namely, impartiality regarding whether someone lives in the present or the far future – that seemingly leads us to longtermism. Bernard Williams (1981), whose theories had a profound impact on the development of modern virtue ethics, famously argued that it would be “one thought too many” (18) for a husband to first consider whether or not it is permissible to save his own wife if he is thereby

choosing her over two strangers whose lives are also at stake. But if virtue ethics is not compatible with justifying something like "effective altruism" in our present generation, how could it possibly justify an even more far-reaching claim to prioritize, at least in many cases, the long-term future, which appears to be even more distant and disconnected from our current lives?

Michael Slote (2007, 25–26), for example, explicitly argues from a virtue-ethical perspective that the virtue of empathy involves being more concerned about people in the present time, such as a group of trapped miners, than people in the far future, even if donations to prevent future mine accidents would eventually save more lives than donating to a mission to rescue miners in the present. But this interpretation of virtues is disputable (Chappell 2019). One possible route to arrive at virtue-ethical longtermism would thus involve a reevaluation of our understanding of virtues. Perhaps there are specific virtues that would enable us to prioritize the future, that is, new virtues that we may not have considered before – something like *far*-foresightedness, for instance (see also Cadilha and Guedes Vaz 2023). Alternatively, traditional virtues could be reinterpreted in light of the new challenges we face today. For example, the virtue of charity or benevolence could extend its concerns beyond the needs of the present to encompass also the needs of future generations. This could, indeed, only have become a relevant aspect of charity or benevolence after we have attained the technological means to influence the future in the long term, not just the next generation. Neglecting the far future could also be considered ignorance – that is, a vice – since it will eventually become the future of our children, their children, and so on. Some suggestions have already been made regarding the possibility of arriving at longtermism through a virtue-ethical framework (see, e.g., Ord 2020, 52–53).

In this article, however, I intend to propose a different, hitherto unexplored route. I have chosen this route not only because it can make a particularly strong case for longtermism, but also because I do not believe that a fundamental change in our understanding of the traditional, partly present-focused virtues is advisable. Such changes, in my view, would rather undermine the distinct and appealing foundational structure of virtue ethics that sets it apart from consequentialism. Simply inventing new virtues, on the other hand, appears to be a rather uninspired ad hoc solution. Instead of pursuing these approaches, I plan to demonstrate that a form of longtermism is already inherent in the basic metaethical structure of virtue ethics itself, at least when basing virtue ethics on neo-Aristotelian naturalism – the most influential metaethical foundation of virtue ethics today.

The following section of this article explains the basic metaethical structure of neo-Aristotelian naturalism with regard to nonrational beings. It demonstrates that satisfying a formal requirement of internal sustainability is necessary for life-form descriptions to attain validity and normative force, while the third section elaborates

on the notion of natural habitat that is involved here. The fourth section shows that this sustainability requirement remains intact even in the context of life-forms that involve rationality. In the fifth section, I start to elucidate the influence of human rationality on the implications of the sustainability requirement, which prepares a longtermist extension of this requirement in the sixth section. The seventh section explains the metaethical preconditions of virtue-ethical longtermism, which is contrasted with consequentialist longtermism in the eighth section. I conclude with a summary in the last section.

2 The Metaethical Foundation of the Sustainability Requirement

While virtue ethics can be based on a variety of metaethical frameworks, this article adopts the perspective of neo-Aristotelian naturalism, which has been established primarily by the influential works of Rosalind Hursthouse (1999), Philippa Foot (2001), and Thompson (2008). Although I have elsewhere suggested certain modifications to this framework to enable a better understanding of the evolution of life-forms and to facilitate more fine-grained evaluations (see Runge 2023, 2024b), it is not necessary to introduce any major modifications here. The argument presented here can be effectively articulated by adhering to the fundamental tenets of standard neo-Aristotelian naturalism, which remain central to my perspective and are likely to be more familiar to most readers.

In contrast to consequentialism, which posits the existence of general intrinsic values in the universe, ethical values according to neo-Aristotelian naturalism are inherently relational, based on a different understanding of the logical grammar of ethical sentences and attributions of goodness. According to this view, the goodness of an individual being is fundamentally understood in relation to the life-form (or nature) of the species to which it belongs. To illustrate this perspective, let us consider a scenario where I am walking through a forest with a conversation partner who lost her vision some years ago. If I describe the environment to my partner by saying, “That bird there, which is chirping right now, is red,” the logical grammar of the sentence is similar to saying, “That car there, which is making noise right now, is blue.” However, if I were to say, “That bird there, which is warning the flock of a predator right now, is good,” the logical grammar of such a sentence would be fundamentally different.

The goodness of a bird cannot be understood as an independent property. While redness is in all instances dependent on a certain wavelength of light, the attribution of goodness to living beings presupposes a standard derived from the life-form of

the species to which the respective living being belongs, so that "good" is understood here as a logically attributive adjective (Geach 1956; Foot 2001, chs. 1 and 2; Thompson 2008, ch. 4). Thus, it would be more accurate to state, "That bird there, which is warning the flock right now, is a good goose (not just any bird)." To illustrate, in a sorcerer movie, domesticated owls could be artificially arranged in a manner that simulates a flock, and one of them could be trained to cry loudly after an evil sorcerer appears. However, it would be inappropriate for an audience member to describe this owl's behavior as an expression of her being *a good owl* (see also Thompson 2004, 55).

In the context of neo-Aristotelian naturalism, attributions of goodness to living beings do not necessarily imply recommendations to my conversation partner (e.g., suggesting that *she* should have the traits of this goose) or express personal suitability for my own purposes (e.g., satisfying my craving for goose meat). Instead, when we take an ethical stance toward living beings, attributions of goodness refer to an internal standard (Foot 2001, 26). From the normative perspective of ethical theory, a sentence like "This goose is good" describes that this goose objectively possesses a good constitution, measured against its own nature, and this statement can be either true or false. For instance, it may be false if I have misremembered or never known the characteristics of a good goose, or if the supposed goose was actually a severely mutated swan, hardly distinguishable by a layperson from a genuine goose (Foot 2001, 35). While one would rarely hear a sentence like "This goose is good" in everyday conversation (although we may indeed say something like "Socrates was a good human"), such attributions are more frequently made when describing parts of living beings. For example, I might say that an oak tree I planted some years ago has developed good roots (Foot 2001, 46).

In neo-Aristotelian naturalism, the thesis of the objective relativity of goodness is not just a thesis about the proper use of words in a certain language game. Rather, it is considered an ontological fact that descriptions of (natural) goodness must be understood in relation to a living being and its life-form (Foot 2001, 36). If this ontological thesis holds true, then it no longer makes sense to assert "The pleasure of every sentient being is good, and we should maximize what is good," as if such an undertaking were somehow analogous to maximizing redness. While it is conceivable that we could maximize redness in the world by, for example, painting all trees red, if goodness is understood only in relation to a species, then a human who maximizes the pleasure of lions by stimulating their brains with electrical impulses, for example, may not be considered a good human, but rather a strange freak. Likewise, the lions who could not avoid being captured by this freak and now live within a computer simulation would not be considered good lions either. While a world could be redder, the state of affairs in the world could not be made "better" by such actions (Foot 2001, 49).

According to neo-Aristotelian naturalism, attributions of goodness to an individual have to refer to the species standard. This species standard can be derived from an idealized description of the life cycle of members of the species. This life cycle can be summarized in a set of statements that ideally describe different stages or aspects of the life of normal members of that species, while these statements simultaneously formulate a normative standard (Foot 2001, ch. 2; Thompson 2008, ch. 4). These statements, known as natural-historical statements or "Aristotelian categoricals," serve as the foundation for attributions of goodness. Taken together, they constitute what can be considered the life-form of the members of a species at a specific point in its evolution. Examples of such a "natural history" are: "Frog eggs are fertilized in water; from the frog eggs, tadpoles develop with gills; tadpoles swim in the water with a fin, and they feed on plants; from the tadpoles, frogs develop, which occasionally leave the pond, jump on four legs, and feed on insects, etc."

It is crucial to note that within neo-Aristotelian naturalism, the idealized life cycle is not merely a statistical generalization (Foot 2001, 33; Thompson 2008, 68). For instance, even if the majority of tadpoles were to perish at an early stage of development, this would not alter the description of the life-form. Instead, natural-historical statements describe a complete life cycle that necessarily includes reproduction (or participation in a reproductive nexus) to be considered complete (Foot 2001, 33, 42; Thompson 2004, 65 n10). Neo-Aristotelian naturalism proposes that this logical grammar and general structure can also be applied to the ethical evaluation of the life of humans, despite the fundamental changes that arise from rationality's influence on the human condition (Foot 2001, 27; Thompson 2008, 82). This metaethical framework for understanding ethical sentences already establishes a sustainability requirement that will serve as our entry point to develop virtue-ethical longtermism later. In the following, we first have to better understand the basic structure of this grammar of goodness, which can be best exemplified by analysis of simple, nonrational life-forms (Hursthouse 1999, ch. 9; Foot 2001, 15–16).

When we assert that the idealized description of a species' life-form necessarily includes some element of reproduction, this introduces a formal requirement that natural-historical statements must satisfy. This requirement is that the natural-historical statements that describe the normal organization of survival and reproduction by the members of the species must not, taken as a whole, involve dynamics that would lead to the eventual extinction of the life-form. This requirement is referred to hereinafter as the *sustainability requirement*. To illustrate, a sentence such as "A male and a female mute swan form a lifelong monogamous bond to lay and incubate a single egg in their lifetime" would be disqualified as a candidate for a natural-historical statement. If this standard were realized within the species, it would result in a halving of the population in each generation until only one mute swan remained, unable to continue the life-form alone. Life-form descriptions that

internally fix such dynamics fail to conform to the idealized nature that such descriptions have to meet, rendering such suggestions for natural-historical statements invalid from the outset. This aspect of sustainability, which is related to the internal functionality of the life-form, can be referred to as *internal* sustainability.

Internal sustainability encompasses a second aspect that can be derived from the standard account of the life-form concept in neo-Aristotelian naturalism. The life-form description must allow for the regular reproduction of the species, even when taking into account the normal range of external threats in the species' natural habitat. If frogs normally lay their eggs not in safe ponds but in ponds where the eggs have a low survival rate due to prevalent predators, a natural-historical statement must take this fact into account (see Foot 2001, 29, 33). The life-form description is idealized in the sense that it only tells a natural-history story of the members of the species that have successfully completed their life cycle, but it is not utopian either (see Thompson 2004, 50). That is, the description of the natural habitat – which is part of the life-form description and in that sense internal – must include a description of its typical dangers and threats. Consequently, a natural-historical statement about frogs must specify that frogs have to lay thousands of eggs in ponds (Thompson 2004, 49–50). Consider the following set of sentences as a potential part of a life-form description: "Frogs lay their eggs in ponds where their chance of survival is 1 to 1,000; frogs lay two eggs in their entire lifetime." If the members of the species adhered to this life-form description, they would not be able to reproduce themselves sustainably in their natural habitat. Consequently, this description does not satisfy the sustainability requirement because the natural habitat is an integral part of the life-form description.

The notion of the internal sustainability of a life-form, which can also be described as life-form functionality, should be distinguished from what is commonly referred to as sustainability, or what I prefer to call *external* sustainability. External sustainability refers to the phenomenon where the members of a species gradually alter the external environment in such a way that the environment cannot replenish the depleted resources at the same rate as they are consumed. For the purposes of this article, however, it is sufficient to work with the concept of internal sustainability only, allowing us to disregard external sustainability (see also Runge 2024a).

3 Natural Habitats and Accidents

From the previous section, we have been able to infer that adherence to the normative standards of the life-form must enable the species to reproduce itself sustainably and with some regularity in its natural habitat, despite the interference of the normal range of threats considered typical of this habitat. It is important to

note, however, that not everything occurring within the geographic area where members of the species live is automatically included in the life-form description. Similarly, the natural habitat description (which is a part of the life-form description) does not simply describe all aspects of the environment. In neo-Aristotelian naturalism, both concepts serve a normative function by enabling the attribution of natural goodness. The description of the normatively relevant aspects of a life or a situation, however, varies depending on the focus of the respective normative theory that employs such concepts.

The ethical focus of naturalistic virtue ethics is not on the actual outcome, as in some versions of consequentialism, but rather on whether an individual has stable dispositions to attain these goods in the most reliable way under normal conditions. Rosalind Hursthouse (1999), for example, conceives natural goodness (or, in other words, virtue) as the "most reliable bet" or "the only reliable way" (176) to attain the goods necessary for the flourishing of a living being with a specific nature. From this perspective, the actual attainment of goods, despite having taken a normally unreliable way, would be considered a result of good luck, and a life whose success is based on luck would lack genuine ethical value (Foot 2001, 93 n16).

From the outlined virtue-ethical perspective, it makes sense that certain rare events that may occur in an individual's global environment do not become part of the life-form description of the species' natural habitat. If these rare events interfere with the attainment of goods, they are rather considered *unfortunate accidents* or bad luck. Foot (2001, 34) provides the example of a single trap in a forest that happens to catch, of all things, the fastest one of a group of deer, which under normal conditions would be best suited to escape predators. While being slower would have prevented the death of the respective deer in this particular case, it would not generally be the most reliable way for deer to survive under normal conditions. Similarly, if small meteorites were to strike the natural habitat of deer every year, occasionally resulting in the death of one of them, it would not imply that deer in general needed some kind of defense mechanism against meteorite strikes in order to survive on a regular basis. Therefore, such an event would still be classified as an accident, despite occurring with some regularity.

It is noteworthy that this normative interpretation of the natural habitat implies that two species may inhabit the same geographic area, while different aspects of the natural environment are considered part of the life-form description of their respective natural habitats. For example, it can be argued that wolves are part of the natural habitat description of deer but not of earthworms. A deer that is slow of foot and falls prey to a wolf would lack natural goodness, whereas if an earthworm were to die from being stepped on by a wolf, it would be considered an unfortunate accident. Similarly, ultraviolet light is absent from the description of the natural habitat of bats, whereas the inability of a bee to see ultraviolet light would be

considered a defect. That is, the nature or life-form of a species influences which aspects of the natural environment need to be integrated into the life-form description of the natural habitat. This, in turn, affects which events can be considered accidents for such beings. Such an understanding of the concept of environment is not exclusive to neo-Aristotelian naturalism. In modern philosophy of biology, for instance, similar ideas have been suggested (see, e.g., Lewontin 2000, 48–49).

While the placement of a single trap in the forest does not necessitate an amendment to the description of the deer's natural habitat in general, it may be necessary to update a life-form description if what normally happens in the natural environment of a species undergoes a lasting and irreversible change (Foot 2001, 29). Consider, for example, the Stephens Island wren *(Traversia lyalli),* a small, flightless bird, whose last refuge in the nineteenth century, covering an area of less than one square mile, was Stephens Island (Higgins et al. 2001, 90–91; Marra and Santella 2016, ch. 1). From the preceding discussion, it can be concluded that the arrival of a single, infertile cat that happened to decimate the majority or even all of the Stephens Island wrens during its lifetime would still be considered an accident that befell them, without necessitating an alteration of the description of their natural habitat. This is because the arrival of such a cat could not be considered a regularly occurring event in the natural environment of the affected species.

However, if a group of cats were to establish a regular and sustainable breeding population on Stephens Island, from that moment on it would be more of a lucky accident that a Stephens Island wren did not encounter a cat in its lifetime. Consequently, the life-form description of the natural habitat would need to be updated to reflect the new normal external threats to the members of the species (Runge 2024b; see also Voorhoeve 2003, 37–38). Given that Stephens Island wrens are unable to swim, they are disconnected from any natural environment that corresponds to the former description of their natural habitat. In such a scenario, an otherwise unaltered description of the life-form of the Stephens Island wren, in conjunction with the updated description of the natural habitat, would no longer satisfy the sustainability requirement. It would result in a set of natural-historical statements such as: "Stephens Island wrens evade predators by being nocturnal and able to run fast; Stephens Island wrens live on islands where there are predators that see well in the dark and are faster." Since this life-form description does not satisfy the sustainability requirement, it becomes invalid and the natural-historical statements lose their normative force. Consequently, we could no longer attribute natural goodness to the formerly normal members of the species, given that their traits cannot be considered a reliable way to attain goods, such as survival and reproduction, under the normal conditions of their natural environment.

From the perspective of normative theory, when a life-form description becomes invalid, this prompts a search for an alternative description of the species' life-form that is able to satisfy the sustainability requirement. Consider a scenario in which a minority of Stephens Island wrens retained the genetic capacity to develop flight. In the absence of cats or similar predators on Stephens Island, the reactivation of flight would have been merely a waste of energy and therefore not part of the most reliable way to attain goods. However, if the life-form description of the natural habitat changes due to the introduction of cats, it would be necessary to reassess the normative status of these formerly atypical members of the species. That is, one would have to test whether the capacity for flight could be considered part of an alternative life-form description of the species that, taken as a whole, satisfies the sustainability requirement. If this were the case, the capacity for flight could be considered the new normative standard that is valid for the species.

4 The Validity of the Sustainability Requirement for Rational Beings

In the framework of neo-Aristotelian naturalism, the evaluation of life processes undergoes a significant shift when considering the human life-form, despite the overall relational structure of evaluations remaining intact (Foot 2001, 27). The shift is due to the human capacity for developing rationality, which allows individuals to reflect on and distance themselves from their natural inclinations and to act based on rational insights. For instance, although humans may have a higher sex drive compared to other animals, this does not necessarily imply that frequent sexual intercourse or frequent changes in sexual partners, as seen in bonobos, must be intrinsic elements of the human life-form. Important reasons such as joint family planning, the avoidance of venereal diseases, or similar factors may convince us to distance ourselves from certain natural inclinations and to act differently. As a result, the normative content of the human life-form cannot simply be derived from the nonrational, biological constitution of humans as such, as if the tooth position of humans alone could indicate that they should not be strict vegans (McDowell 1995). This is because reason gains a normative guiding function in human life, serving as a human substitute for instincts that we lack and allowing them to determine appropriate actions depending on the context (Foot 2001, ch. 4). In this sense, developing a certain mode of rationality as a second nature is itself natural to humans, while the specific mode of rationality is normally acquired by initiation into an intellectual tradition that is rooted in a certain culture.

From the perspective of neo-Aristotelian naturalism, however, the ability of human beings to take into account reasons does not transform them into some kind of pure rational beings who do not take into account their own biology at all when thinking about ethical standards. Instead, human reason remains connected to the (biological) peculiarities of human life. This implies that our rationality is a specifically human form of rationality, not the abstract rationality of persons as such (Foot 2004, 13; Thompson 2004, 60–61). Therefore, human reasoning does not lose itself in purely abstract considerations, as has been leveled as a criticism against Kantian ethics by virtue ethics (Foot 2001, 14), nor does it succumb to similar critiques potentially aimed at the consequentialist style of reasoning.

Were we able to communicate telepathically with all humans at the same time without being intellectually overwhelmed, it might be feasible to be close friends with every living member of our species. However, given our normal natural abilities, this is not a valid ideal for humans (see Foot 2001, 45). Abstractly speaking, one could argue that it would be good to be close friends with everyone. But even if this last sentence may sound rational from an abstract, species-independent point of view (as the sentence: "One should maximize pleasure in the world"), a human who treats even complete strangers as if they were close friends (or who maximizes the pleasure of lions) would be making poor use of her specifically human rationality. Within the framework of neo-Aristotelian naturalism, the specific focus of human rationality presupposes normal human biology and relates to what can be considered the most reliable way to achieve a good life for a human being, not to what would be right for a rational being as such.

The capacity for reasoned deliberation raises an important question concerning the significance of the sustainability requirement with respect to the human life-form. One could argue that our rationality allows us to distance ourselves not only from our natural inclinations, but also from the previously established formal requirement of internal sustainability. In that case, it could be legitimate to establish our very own self-chosen rules, such as maximizing the pleasure of the present generation at the expense of having a future as a species, as some might perceive this as rational and even form a consenting society around it. However, I believe that this option is invalid from the standpoint of neo-Aristotelian naturalism because the ability to distance ourselves from some biological impulses does not undermine the fundamental logical grammar or the underlying structure of ethical judgments as such, given that the life-form relativity of ethical standards is viewed as an ontological fact. Although I am able to intellectually distance myself from being a human by identifying as one of Kant's (2008) Saturnians, this act neither changes the ethical standards that apply to me, which are dependent on the life-form of my biological species, nor can a similar act of distancing myself from ethics in general abrogate the life-form relativity of ethical standards.

Although the concrete normative content of the human life-form may differ significantly from the normative content of nonrational life-forms, the standard of goodness for humans still refers to the idealized life-form description of the species. The life-form description of the honeybee may specify a concrete behavior as a normative standard in certain types of situations, such as stinging in response to threats to the hive. The human life-form, on the other hand, does not require a predefined action in response to threatening situations, but one that is guided by a specific form of rationality – specifically human rationality, combined with good character traits, such as the virtue of courage, that give the proper focus to that rationality and enable the agent to act accordingly. The choices for actions derived from good human rationality may differ substantially from those of Kant's Saturnians, for whom a different form of rationality and a different set of virtues may be considered an expression of goodness. In any case, acting on the basis of reasons is still integrated into a framework that presupposes the continued reproduction of the species – it is, in our case, a part of the specifically human form of organizing the reproduction of a society. Consequently, the form of rationality that can be considered valid for human beings must be consistent with the formal requirement of internal sustainability. At this juncture, we have achieved a satisfactory understanding of the general structure of natural normativity, which now permits us to examine the impact of human rationality on the implications of the sustainability requirement.

5 Human Rationality and the Sustainability Requirement

From the third section, we have inferred that if a single, infertile cat arrived on Stephens Island and killed all the members of the Stephens Island wrens, this event would be classified as an unfortunate accident that befell the wrens. Let us now examine how the evaluation shifts when we consider a seemingly analogous case involving the human species. Imagine that, instead of a single, murderous cat invading Stephens Island, a single, gigantic asteroid hits planet Earth with such a devastating impact that this event ultimately leads to the extinction of the human species. If such an event had occurred a millennium or even a century ago, I suppose the analogy could have been considered valid. Nevertheless, it seems reasonable to posit that if such an event were to occur next year, the analogy would no longer hold. To discuss the reasons for this change, we need to examine how rationality extends the implications of the sustainability requirement for human beings.

Most animal species, like deer, have a specific natural habitat, such as a certain type of forest. In that habitat, they encounter only a limited range of threats, such as predators or diseases, and they have predefined, instinctive or natural responses to these threats, such as swiftly running away or producing antibodies. Humans, however, do not have a specific natural environment as their habitat; they can live almost anywhere because they can make the natural environment habitable through their work. This also means that there are almost no predefined threats and instinctive responses to them. But just as deer have sense perception (such as sight) and dispositions (such as vigilance) to recognize and respond to predators in their natural habitat, human beings have rationality (specifically human rationality) and character virtues (such as courage and temperance) that enable them to recognize and respond to a wide range of existential threats – wherever they live.

Additionally, it is crucial to recognize that the history of humans is decoupled from their evolutionary history. Although the biology of humans has remained largely unchanged since the emergence of the species *Homo sapiens,* the standards for the development and application of our rationality have increased throughout our history, as the standards for swiftness have increased in the evolutionary history of deer. The development of our rational capacities in history has also enabled humans to investigate the natural history of our planet. At this juncture, at the latest, the aforementioned analogy begins to break down.

Based on our current knowledge of the natural history of our planet, humans are well aware of the fact that asteroid impacts of varying sizes occur with some regularity, albeit at irregular, long intervals – too long to be noticed by any other animal species. We also have substantial evidence that asteroid impacts have caused significant extinction events on Earth. Furthermore, it is reasonable to assume that asteroids come in different magnitudes, potentially even capable of causing the extinction of our species. Additionally, in our recent history we have achieved a level of developing our rational capacities that would, in principle, enable us to invent and build technological devices to deflect asteroids and even to establish colonies on other planets. While humans are not directly endowed by nature with the ability to deflect incoming asteroids – just as a newborn giraffe is endowed with the ability to walk in order to evade predators – our life-form description includes a capacity (that is, rationality) to respond to a wide range of undefined threats, and we have dispositions (that is, the virtues) to act in accordance with our rationality and to give it the proper focus.

In light of the aforementioned facts, it would be inappropriate to describe the extinction of the human species due to an asteroid impact next year as an unfortunate accident. For humans, the incoming asteroid, despite being a unique, unprecedented event, can be considered to fall into the normal range of threats that are part of our unspecific natural habitat. That is, the incoming meteorite is not

analogous to the single trap in the forest that happens to catch the fastest deer, whose status as a deer with a good constitution is not negatively affected by this event since it is classified as an accident. Rather, it is more analogous to the random encounter of a deer with a wolf, which falls into the normal range of threats that a deer with a good constitution should be able to respond to. If the deer fails to be vigilant and does not recognize such an encounter in time, or if it recognizes the predator but lacks the motivation or ability to swiftly run away, it would be considered to lack natural goodness. Similarly to bees, human beings are social creatures, although they organize their reproduction not in a hive but in a society. A beehive must engender individuals who defend the hive by recognizing threats and by having the disposition to sting if required. Otherwise, such a beehive would be considered dysfunctional. Similarly, there would be something wrong with human societies in our time whose members collectively failed to recognize and respond to the normal range of threats that are part of our unspecific natural habitat. Such forms of society that do not engender the required form of rationality or the required character virtues that enable the human species to properly respond to such threats can be considered dysfunctional too because they would not satisfy the sustainability requirement.

Since gigantic asteroids on a collision course with Earth are probably neither the most urgent nor the most likely existential threat we currently face, the same reasoning would apply *a fortiori* to other scenarios involving uncertainty and where the survival of the current generation could be safeguarded by the application of our rationality and the cultivation of virtues. As in the case of animal life-forms, satisfying the sustainability requirement may, in principle, facilitate the long-term survival of the species, at least in the absence of large-scale accidents or irreversible environmental changes. Notice, however, that the position reached at this point is still weaker than longtermism. Longtermism would not only entail that each generation must reproduce itself sustainably under the current conditions of the natural habitat, but it may also require us to make substantial efforts to, for example, enable much later generations to respond properly to a predictable, irreversible change of the natural environment in the far future. In the following section, I will argue that the distinctive abilities of humans, which are part of their species' life-form description, imply a second, longtermist extension of the sustainability requirement.

6 The Longtermist Extension of the Sustainability Requirement

In the previous section, I emphasized that humans have an understanding that they, as a species, have a history. Having a history, however, does not only include the past

but also implies that we have a future. The development of our rational capacities in history has enabled humans to anticipate events and developments that will take place in the far future, such as the transformation of the Sun into a red giant in the next billion years, which will ultimately lead to the destruction of Earth. Furthermore, humans have the ability to influence long-term developments that may either facilitate or impede future generations' ability to respond to such events. For example, they can establish or perpetuate forms of societies that engender dynamics that either diminish or enhance the ability of future generations to eventually leave our solar system and establish colonies on other habitable planets. Additionally, humans are not only able to influence the future of our species in such ways, but they also know that they have this influence.

In the third section, I mentioned that if it is part of the nature of bees to find nectar by seeing ultraviolet light, then ultraviolet light becomes part of the description of their natural habitat. This implies that the range of possible defects of bees also expands. An individual bee that cannot see ultraviolet light due to a gene mutation, or a bee that sees it but has developed a habit of not responding to it, must be considered a bee with a natural defect. A bee with such traits is not aligned with the most reliable way to attain goods that are relevant for bees (such as nectar) and therefore lacks natural goodness. Similarly, the rational abilities of humans extend, so to say, the *horizontal* axis of the sustainability requirement, that is, the range of threats (or the space) that can be considered a normal part of their natural habitat.

As already stated in the second section, a life-form description is essentially a list of natural-historical statements that ideally describe one complete life cycle of the members of the species. In the case of birds, such a life-form description may include natural-historical statements that are concerned with the next generation, such as how they are fed and protected when they are young. Apart from influencing the size and condition of the next generation, however, they have no ability to affect any subsequent generations. That is, later (or earlier) generations do not appear in the life-form description of such a life cycle (see Foot 2001, 32 n10, 40 n1). This also implies that the possible extinction of the species in the future does not affect the normative status of a member of the present generation. From the perspective of an eighteenth-century human observer, who is planning to have one of his descendants introduce a group of cats to Stephens Island in the nineteenth century, the eventual fate of the Stephens Island wren may already be sealed. From a normative perspective, however, these future developments of the natural habitat, which may be predetermined and inevitable, are not a part of the life-form description of the currently living generation of Stephens Island wrens. If the present generation of wrens is able to sustainably reproduce itself under the current conditions of the natural habitat in the eighteenth century, the life-form description would still be considered internally sustainable. Therefore, its normative standards are valid, and

an eighteenth-century observer could have legitimately attributed natural goodness to these wrens. Only in the future, after the natural environment of the Stephens Island wren has changed in a lasting, irreversible way, would its life-form description become internally unsustainable.

In the case of human beings, it seems reasonable to suggest that our consciousness that our species has a future, our influence on the future fate of our species, and our knowledge that we have such an influence integrate the fate of future generations into the life-form description of the life cycle of the members of the currently living generation. These abilities extend, so to speak, the *vertical* axis (or time axis) of the sustainability requirement. That is, the current generation has to respond not only to the normal range of existential threats in their natural habitat that affect themselves (the horizontal axis), but they may also be required, at least in certain cases, to prepare responses to existential threats that will only affect generations in the far future (the vertical axis). This extended vertical axis is, in other words, the promised longtermist extension of the sustainability requirement.

To illustrate how this model functions, imagine that in the near future, we discover a species of rational beings that had been hidden deep beneath the surface of Mars. The UNO then sends a group of neo-Aristotelian naturalists to provide us with a life-form description of these Martians (see Foot 2001, 36). Initially, these researchers might find that Martians are very similar to humans in most respects. However, after some time, they might notice that Martians possess a unique sensory apparatus that signals them when they have become grandparents. This signal then automatically initiates a genetic program that results in a loss of behavioral control, causing them to attempt to detect and eliminate their grandchildren by any means, which normally is also successful. While these Martians would, in a way, have successfully reproduced themselves, a natural-historical statement about such a trait that is included into a life-form description would imply a dynamic that eventually leads to the extinction of the species, as in the case of the mute swans that lay only a single egg. In that case, it would become a result of luck or individual failure that the species actually survived under normal conditions, which renders any suggestions for a life-form description that includes such a statement invalid. As in the case of the tadpoles who do not reach adult stage, we could not attribute natural goodness to members of the Martian species that had such a trait, even if they were in the vast majority.

Upon further investigation, our group of neo-Aristotelian naturalists might discover that in a minority of Martians, this unique sensory apparatus does not function. In their case, the genetic program to kill their grandchildren is activated only when they learn by some kind of external evidence that they have become grandparents, such as being told by another person. Despite these Martians being a minority, their natural constitution serves as the reference point for the idealized

life-form description because only this set of natural-historical statements can be compatible with the sustainability requirement. As in the case of humans, where a normative standard concerning their relationship models cannot be derived directly from their nature, the life-form description of these Martians might not directly specify how exactly Martians remain unaware of the fact that they have become grandparents. Instead, such a life-form description might include a statement that they use their specifically Martian rationality in a context-dependent way to conceal this information from the grandparents as best as possible. For that reason, dishonesty might be considered a virtue for Martians.

Within the framework of neo-Aristotelian naturalism, it does not seem to make a crucial difference whether Martians possess a congenital genetic program that causes them to detect and kill their grandchildren or whether they normally acquire relatively stable traits after birth that have the same effect. If they had character traits that disposed them to act in this way, they would be considered vicious, and if they had acquired a form of rationality that made it seem rational to them to act so, their rationality would be considered deficient. If they lived in societies that systematically engendered such dispositions or such a form of rationality, these societies would be considered dysfunctional. None of these elements could be part of a valid life-form description. Furthermore, it also seems irrelevant whether individuals have a trait that causes them to kill their grandchildren themselves or whether they have a trait that causes them, for example, to build and program killer robots to track down and kill their descendants in the far future. In our context, it is not even decisive to distinguish between actively causing the death of a future generation (e.g., by building a doomsday time bomb that cannot be deactivated) and knowingly leaving a future generation no other chance but to die (e.g., by not saving enough resources that will be required to respond to a predictable natural threat in the future). Whether or not, for instance, a Spartan society can attain the status of being internally sustainable is not contingent upon whether they actively kill their children with their bare hands or abandon them in the wilderness to die from natural causes.

Humans do not appear to be driven by any genetic program to undermine the long-term survival of our species. However, they may develop a form of rationality that provides them with reasons to knowingly cause or permit the extinction of our species in the far future. Theories such as egoistic hedonism, for example, may provide reasons for individuals to prioritize their own immediate pleasure over the future survival of our species. Even when they see reasons to facilitate the long-term survival of our species, they may develop character traits, such as intemperance, that prevent them from acting on their rational insights. Furthermore, they may live in societies that systematically engender such a form of rationality and such

dispositions – forms of societies like, as some virtue ethicists have argued, modern capitalism (see, e.g., MacIntyre 2016).

If the basic premises of neo-Aristotelian naturalism, from which we have derived the sustainability requirement in this article, describe the general structure of natural normativity properly, such forms of societies, or such forms of rationality, or such dispositions as described could not be considered possible expressions of natural goodness, even if they were prevalent in the human species during a specific historical period. Integrating a natural-historical statement with such a content into the life-form description of the human species would render it internally unsustainable and invalid. That is, natural goodness can only be attributed to human beings if their form of rationality, their dispositions, and the forms of societies in which they live are compatible not only with the sustainability requirement but also with its longtermist extension. This eventually leads us to virtue-ethical longtermism.

7 Preconditions of Virtue-Ethical Longtermism

This article started from the metaethical assumption of neo-Aristotelian naturalism that valid attributions of natural goodness to an individual necessarily presuppose a species standard that can be derived from the life-form of this species. We then examined the formal requirements that must be satisfied for a life-form description to be considered valid. In the previous sections, I argued that in the case of humans, their life-form description not only has to satisfy the sustainability requirement but also an extended version of it that implies longtermism. This means that, at this juncture, we can presuppose that *all* natural-historical statements concerning humans must be compatible with a version of longtermism. This includes statements concerning the form of rationality and the list of character virtues that members of the human species need (which might differ for rational Martians). Only rational will and these acquired character traits – not bodily traits or genetic programs – are the genuine object of moral philosophy, while moral goodness is considered a specific form of natural goodness (Foot 2001, 66, 81).

In this article, our focus so far has been on exploring the general grammar of goodness, which is a topic of metaethics. In contrast, normative ethics is no longer concerned with analyzing the general grammar but, to stay with the metaphor, with providing rules and orientation for good writing style, which presupposes at least implicit grammatical knowledge. That is, normative ethics discusses the procedures or principles to determine what is a good life as a whole or a good action in a particular situation. Virtue ethics is regarded as one of the three major paradigms in normative ethics, and it focuses on virtue and character in these procedures. While

virtue ethics can also be based on other metaethical frameworks, naturalistic virtue ethics is considered here as a virtue-ethical position that is based on neo-Aristotelian naturalism. If we turn from neo-Aristotelian naturalism to naturalistic virtue ethics at this juncture, we have to presuppose, as a formal requirement derived from metaethics, that what can be considered virtuous must be compatible with a life-form description that satisfies the longtermist extension of the sustainability requirement. It follows from the argument of this article that longtermism is not a result of applying virtue ethics to our current situation, such as by reinterpreting old virtues or inventing new virtues in response to current ethical challenges. Rather, longtermism is already implied in the metaethical foundation of naturalistic virtue ethics.

From our metaethical foundation, we can conclude that empathy, for instance, could not be considered a genuine virtue for human beings if developing this virtue were to engender a dynamic that we know will eventually lead to our extinction in the far future. That is, if empathy necessarily implied that empathic persons are ignorant of future threats to the survival of our species because they are solely concerned with their contemporaries, then interpreting empathy as a virtue that is required for humans would be ruled out on metaethical grounds. Only a different interpretation of the practical implications or the essence of empathy could then justify its classification as a virtue. In that sense, the metaethical route to virtue-ethical longtermism suggested here may in fact lead to a reinterpretation of some virtues, though only by clarifying that some interpretations of these virtues are, in fact, "grammatically incorrect." Before I can proceed with further explaining the characteristics of virtue-ethical longtermism, it is necessary to address a challenge to the theory developed thus far, which arises from our knowledge that there are likely hard physical limits for the survival of the human species.

In the third section, I stated that the life-form description of the Stephens Island wrens became internally unsustainable after the arrival of a group of cats on Stephens Island, which led to an update in their natural habitat. Human beings, however, may not only be required to respond to certain existential threats that would have counted as an accident for the members of other species, we can also envision situations where humans are required to prepare responses to predictable, lasting, and irreversible changes of their natural environment in the future, such as those caused by the transformation of the Sun. But would not our knowledge that there are hard physical limits to the existence of human life in our universe, such as the Big Freeze (or heat death of the universe), also make our life-form description internally unsustainable?

At this point, it is important to recall the normative function that the life-form concept has in neo-Aristotelian naturalism, as explained in the third section. The idea behind this concept is not to establish an unachievable ideal. Rather, it aims to bring

ethics down to earth by establishing a connection between normative standards and the nature of the beings who are the subject of evaluation. The normative function of the life-form concept, then, is to provide us with a wider context for determining the most reliable way for the members of a species to attain the goods they need to flourish under normal conditions (see Runge 2024b). If the environment of the Stephens Islands wrens changes in an irreversible and lasting way, it can be reasonable to assume that their natural habitat is updated, given that there may still be evolutionary potential of this species to develop a response to the new challenges of their habitat, such as a gene mutation that reactivates the ability to fly. In this regard, the idea of updating the description of a natural habitat becomes a heuristic tool to facilitate an investigation whether there emerges a new and valid normative standard for that species. Yet, if the species as a whole has no such potential, I believe that the heuristic tool of updating the natural habitat description would lose its normative function.

In response to this challenge, I would like to suggest that when a species with a certain nature has actualized all of its potential to survive under the environmental conditions of our universe, reaching these hard physical limits can no longer be considered an update of the natural habitat. Rather, this environmental change can be considered a threat to the survival of the species that remains external to the life-form description and is, therefore, analogous to an accident in its effects. Just as the mere vision of a trap does not automatically make traps a part of the natural habitat of deer, I think that the mere knowledge of future developments of our environment alone does not make the Big Freeze a normal threat that belongs to our natural habitat. Rather than becoming our natural habitat, the Big Freeze could be conceptualized as the abrogation of the concept *natural habitat*. That is, for a future change in our environment to be considered a normal threat that requires us to respond to it, the currently living generation must also have the potential to influence the ability of future generations to successfully counter such threats, which does not seem to be given in the case of the Big Freeze, but may be possible in the case of the transformation of the Sun.

At this point, we have arrived at the conclusion that virtue-ethical longtermism implies that the human life-form describes the most reliable way to attain the goods that human beings need to flourish, under the (metaethical) precondition that the life-form description, as a whole, is also compatible with our long-term survival as a species, as far as the fate of future generations is in our hands and as far as we can know that it is in our hands. This requires from us, at least implicitly, to "positively influence the long-term future," and since satisfying this requirement is a precondition for deriving any other normative contents, it can also be considered "a key moral priority," which was our initial definition of longtermism provided by MacAskill. However, given the differing understandings of the grammar of goodness,

virtue-ethical longtermism diverges significantly in practical implications from consequentialist longtermism, as I will briefly explain in the next section.

8 Differences Between Virtue-Ethical and Consequentialist Longtermism

In consequentialism, longtermism emerges from a calculation: Reducing existential risks to our long-term survival, even to an infinitesimal degree, may in some or even all situations have a higher expected value than even the most effective ways to do good in the short term, such as saving lives through disease prevention or diminishing the suffering of animals (see Greaves and MacAskill 2021). According to neo-Aristotelian naturalism, however, these calculations are based on a misunderstanding of the grammar of goodness, as explained in the second section.

If the human species becomes extinct because of the egoism, gluttony, and greed of the present generation, virtue-ethical longtermism would suggest that the ethical issue is not that the pleasure that future generations could have experienced outweighs the pleasure that our generation has gained through our choice to live in an unsustainable way. The number of individuals who find their lives worth living, whether in the future or in the present, cannot be considered a species-independent good that can somehow be aggregated. If a lion lives his life in accordance with his carnivorous nature, this may result in an overall greater balance of pain over pleasure in the world. Nevertheless, lions who "teach their cubs to kill" (Foot 2001, 15) may still be considered good lions, and this could hold true even if they were to acquire rationality in their future evolutionary history. Similarly, the character trait of compassion would be considered a natural defect in lions and not aligned with the most reliable way to attain the goods that are relevant for their flourishing. According to naturalistic virtue ethics, the extinction of our species in the far future, caused by the character dispositions of our generation, would be bad because it indicates that many people must have been vicious, and even if some of the members of this generation were virtuous, they apparently lived in dysfunctional societies. Not undermining the ability of future generations to survive is relevant to us not because those generations might be larger or happier, and the future therefore more valuable than the present. Rather, it becomes relevant to us because it is a precondition to realize our own virtuous, good life and to qualify ourselves for attaining *eudaimonia* – the flourishing and happiness that comes from living a virtuous life. As a consequence, virtue-ethical longtermism, in contrast to consequentialist longtermism, is a minimal and a humanist longtermism, as defined at the beginning of this article.

Act consequentialism, the most popular approach within the consequentialist paradigm, has a focus on individual actions, such as the highest intrinsic value that can be generated by a single donation. This action-focused perspective, however, tends to separate the action from its meaning in the wider context, which is the life-form of the human species. From a virtue-ethical perspective, act consequentialism thereby creates a distorted picture of a good human life, as if it consisted merely of a sequence of actions, with each of these actions, taken in isolation, producing the best consequences. If a compassionate, empathic person encounters a homeless person on the street, she may know that giving a certain amount of money not to him but to an effective charity organization would ultimately help more homeless persons in similar circumstances. This calculation, however, does not take into account that one does not become (or stay) a compassionate and empathic person by merely setting up an automatic monthly donation to a charity. A narrow focus on quantifying and maximizing abstract values (such as the number of fulfilled lives) may eventually lead one to become an unempathetic person that is guided by a cold and calculating rationality.

From the perspective of neo-Aristotelian naturalism, consequentialist approaches to longtermism underestimate that a certain form of instrumental rationality may itself negatively interact with human nature. Consequentialism seems to operate under the implicit assumption that there is a neutral, impartial, and general rationality that, ideally, all rational agents should apply. In theory, if one acknowledges that our life-form must be internally sustainable, also in the long term, it may sound rational to focus on reducing existential risks by making effective donations. According to neo-Aristotelian naturalism, however, the form of rationality that can be considered an expression of goodness in the case of humans is, in a way, also dependent on our species' biological nature (Foot 2001, 14). This article has mainly focused on reconstructing the longtermist implications of the basic assumptions made by neo-Aristotelian naturalism. The focal point for developing the virtue-ethical dimension of virtue-ethical longtermism then becomes the idea that humans need a specifically human rationality.

As I stated in the preceding section, if empathy were to imply ignorance concerning the future, it could, in theory, be ruled out as a virtue for humans with reference to the longtermist extension of the sustainability requirement. However, if we conceive of empathy not as some kind of cardinal virtue that becomes the measure of all our actions, but merely as one of the many virtues that are part of the human life-form description, we can acknowledge that empathy might contribute to develop a rationality with a proper, humane focus. From our knowledge of human history, we have reason to believe that a free-floating rationality that is not rooted in a person possessing the full range of virtues tends to inadvertently engender destructive dynamics, even when starting with initially good intentions, such as

"doing good better." The logic of instrumental rationality, of quantification and maximization, has been described by many authors, such as Horkheimer and Adorno (2002), as the source of a dialectic that has led not only to technological progress and human domination over nature but also to human regression and the destruction of the conditions of life on Earth. While this goes beyond the scope of this paper, it seems reasonable to suggest that merely switching to the altruistic side of the same logic is not the most promising approach to solving the crises that the egoistic side of that logic has created in the first place (see also Adams et al. 2023).

In conclusion, to achieve the long-term survival of our species, it may be a more reliable bet to accept that some kind of present-focused empathy is, among other virtues, an integral part of our life-form description, as well as a noninstrumental rationality that focuses not only on reducing existential risks but also on the value of certain virtues, such as empathy, for their own sake. Having such a specifically human rationality may ultimately be the more appropriate and sustainable way for us to survive as a species, although many individual actions that can be justified by this approach might be seen as ineffective from the perspective of, say, Kant's Saturnians, for whom instrumental rationality might be a more suitable mode of thinking with less problematic side effects (see Foot 2004, 13). As a consequence, virtue-ethical longtermism does not focus on individual actions and their potential to reduce existential risk but derives its normative standard from a life-form description that, taken as a whole, has to be the most reliable bet to enable our long-term survival as a species. From this perspective, there may be valid reasons for empathic, virtuous persons to give money to a homeless person, at least in some situations, rather than always giving it to organizations that aim to reduce existential risks. These actions, however, derive their ethical meaning and value – their status as an expression of human goodness – from the wider context of the life-form description, not from the consequences of the individual action. Ultimately, it is not only the character traits and the form of rationality of the individual that must be compatible with our long-term survival as a species, but also the form of society, since the form of society influences and shapes the character traits and the form of rationality of its members.

The longtermist extension of the sustainability requirement may challenge certain interpretations of character traits as having the status of a virtue. Nevertheless, valuing some present-focused virtues such as empathy for their own sake may open up space for a nonmaximizing, nonquantifying logic that could potentially lead to a halt in the accelerating destructive dynamic of modern capitalism – arguably the most significant threat to the long-term survival of our species we currently face. Having a focus in ethics on the cultivation of the virtues may be essential for establishing a more appropriate, humane mode of rationality, which also implies a critical stance toward forms of society that tend to erode the virtues (see MacIntyre

1981). Although positively influencing the survival of future generations is of great importance in virtue-ethical longtermism, the proposed approach can still be considered a genuine virtue ethics that is deeply rooted in the Aristotelian tradition.

9 Conclusions

This article has proposed a metaethical route from neo-Aristotelian naturalism to virtue-ethical longtermism. It has argued that the metaethical structure of neo-Aristotelian naturalism inherently implies a requirement of internal sustainability of the life-form as a formal condition for the validity of natural-historical statements. Since the life-form description of the human species involves abilities that lay the fate of future generations into our hands, the currently living generation must consider more than just existential threats to our own survival in order to satisfy the sustainability requirement. Rather, we may also be required to prepare responses to existential threats that will only affect future generations. Consequently, in the case of the human species, the sustainability requirement attains a longtermist extension. If virtue ethics is based on this interpretation of neo-Aristotelian naturalism, then our understanding of the virtues must be aligned with a minimal and humanist version of longtermism. Nevertheless, virtue-ethical longtermism remains a genuine virtue ethics that has a focus on cultivating the virtues as the only reliable way to achieve the good life of the agent.

Acknowledgment: I would like to thank my esteemed colleague Lukas Tank for his insightful comments on an early draft of this paper and for initially bringing longtermism to my attention. I am also grateful to the two anonymous reviewers of the journal for their valuable feedback. Most importantly, I express my sincere gratitude to the guest editor of this issue, Stefan Riedener. His exceptional dedication and genuine effort to understand my perspective – not just to critique it but to support me in addressing potential weaknesses – were crucial in helping me develop a much better version of this article.

References

Adams, C. J., A. Crary, and L. Gruen. 2023. “Future-Oriented Effective Altruism: What’s Wrong with Longtermism?” In *The Good It Promises, the Harm It Does: Critical Essays on Effective Altruism*, edited by C. J. Adams, A. Crary, and L. Gruen, 265–9. Oxford: Oxford University Press.

Bostrom, N. 2013. “Existential Risk Prevention as Global Priority.” *Global Policy* 4 (1): 15–31.

Cadilha, S., and S. G. Vaz. 2023. “Prospection as a Sustainability Virtue: Imagining Futures for Intergenerational Ethics.” *Zeitschrift für Ethik und Moralphilosophie* 6 (2): 293–309.

Chappell, R. Y. 2019. “Overriding Virtue.” In *Effective Altruism: Philosophical Issues*, edited by H. Greaves, and T. Pummer, 218–26. Oxford: Oxford University Press.
Foot, P. 2001. *Natural Goodness*. Oxford: Clarendon Press.
Foot, P. 2004. “Rationality and Goodness.” *Royal Institute of Philosophy Supplement* 54: 1–13.
Geach, P. 1956. “Good and Evil.” *Analysis* 17 (2): 32–42.
Greaves, H., and W. MacAskill. 2021. “The Case for Strong Longtermism.” In *Global Priorities Institute Working Papers Series 5*. https://globalprioritiesinstitute.org/wp-content/uploads/The-Case-for-Strong-Longtermism-GPI-Working-Paper-June-2021-2-2.pdf (accessed June 27, 2024).
Higgins, P. J., J. M. Peter, and W. K. Steele. 2001. *Handbook of Australian, New Zealand and Antarctic Birds, Vol. 5: Tyrant Flycatchers to Chats*. Melbourne: Oxford University Press.
Horkheimer, M., and T. W. Adorno. 2002. *Dialectic of Enlightenment: Philosophical Fragments*. Stanford: Stanford University Press.
Hursthouse, R. 1999. *On Virtue Ethics*. Oxford: Oxford University Press.
Jonas, H. 1984. *The Imperative of Responsibility: In Search of an Ethics for the Technological Age*. Chicago: University of Chicago Press.
Kant, I. 2008. *Universal Natural History and Theory of the Heavens*. Arlington: Richer Resources Publications.
Lewontin, R. C. 2000. *The Triple Helix: Gene, Organism, and Environment*. Cambridge: Harvard University Press.
MacAskill, W. 2022. *What We Owe the Future: A Million-Year View*. London: Oneworld.
MacIntyre, A. 1981. *After Virtue: A Study in Moral Theory*. Notre Dame: University of Notre Dame Press.
MacIntyre, A. 2016. *Ethics in the Conflicts of Modernity: An Essay on Desire, Practical Reasoning, and Narrative*. Cambridge: Cambridge University Press.
Marra, P. P., and C. Santella. 2016. *Cat Wars: The Devastating Consequences of a Cuddly Killer*. Princeton: Princeton University Press.
McDowell, J. 1995. “Two Sorts of Naturalism.” In *Virtues and Reasons: Philippa Foot and Moral Theory: Essays in Honour of Philippa Foot*, edited by R. Hursthouse, G. Lawrence, and W. Quinn, 149–79. Oxford: Clarendon Press.
Ord, T. 2020. *The Precipice: Existential Risk and the Future of Humanity*. New York: Hachette Books.
Runge, R. F. 2023. *Eine kritische Theorie der Tugendethik*. Frankfurt: Campus.
Runge, R. F. 2024a. “Nachhaltiger Konsum als Ideal der naturalistischen Tugendethik: Zum normativen Status und der konflikthaften Dimension der Nachhaltigkeit.” *Zeitschrift für Wirtschafts- und Unternehmensethik* 25 (1): 56–78.
Runge, R. F. 2024b. “Neo-Aristotelian Naturalism, Local Ethical Supervenience, and the Beneficial Character of Virtue.” *Zeitschrift für Ethik und Moralphilosophie* 7 (1): 23–46.
Slote, M. A. 2007. *The Ethics of Care and Empathy*. London: Routledge.
Tarsney, C. 2023. “The Epistemic Challenge to Longtermism.” *Synthese* 201 (195). https://doi.org/10.1007/s11229-023-04153-y.
Thompson, M. 2004. “Apprehending Human Form.” In *Modern Moral Philosophy*, edited by A. O’Hear, 47–74. Cambridge: Cambridge University Press.
Thompson, M. 2008. *Life and Action: Elementary Structures of Practice and Practical Thought*. Cambridge: Harvard University Press.
Voorhoeve, A. 2003. “The Grammar of Goodness: An Interview with Philippa Foot.” *The Harvard Review of Philosophy* 11: 32–44.
Williams, B. 1981. “Persons, Character, and Morality.” In *Moral Luck: Philosophical Papers, 1973-1980*, edited by B. Williams, 1–19. Cambridge: Cambridge University Press.