WHAT IS THE UPPER LIMIT OF VALUE?

Anders Sandberg*
Future of Humanity Institute
University of Oxford
Suite 1, Littlegate House
16/17 St. Ebbe's Street, Oxford OX1 1PT
anders.sandberg@philosophy.ox.ac.uk

David Manheim*
1DaySooner
Delaware, United States,
davidmanheim@gmail.com

January 27, 2021

ABSTRACT

How much value can our decisions create? We argue that unless our current understanding of physics is wrong in fairly fundamental ways, there exists an upper limit of value relevant to our decisions. First, due to the speed of light and the definition and conception of economic growth, the limit to economic growth is a restrictive one. Additionally, a related far larger but still finite limit exists for value in a much broader sense due to the physics of information and the ability of physical beings to place value on outcomes. We discuss how this argument can handle lexicographic preferences, probabilities, and the implications for infinite ethics and ethical uncertainty.

Keywords Value · Physics of Information · Ethics

Acknowledgements: We are grateful to the Global Priorities Institute for highlighting these issues and hosting the conference where this paper was conceived, and to Will MacAskill for the presentation that prompted the paper. Thanks to Hilary Greaves, Toby Ord, and Anthony DiGiovanni, as well as to Adam Brown, Evan Ryan Gunter, and Scott Aaronson, for feedback on the philosophy and the physics, respectively. David Manheim also thanks the late George Koleszarik for initially pointing out Wei Dai's related work in 2015, and an early discussion of related issues with Scott Garrabrant and others on asymptotic logical uncertainty, both of which informed much of his thinking in conceiving the paper. Thanks to Roman Yampolskiy for providing a quote for the paper. Finally, thanks to Selina Schlechter-Komparativ and Eli G. for proofreading and editing assistance.

1 Introduction

The future of humanity contains seemingly limitless possibility, with implications for the value of our choices in the short term. Ethics discusses those choices, and for consequentialists in particular, infinities have worrying ethical implications. Bostrom [1] and others have asked questions, for example, about how aggregative consequentialist theories can deal with infinities. Others have expanded the questions still further, including measure problems in cosmology, and related issues in infinite computable or even noncomputable universes in a multiverse.

In this paper, we will argue that "limitless" and "infinite" when used to describe value or the moral importance of our decisions can only be hyperbolic, rather than exact descriptions. Our physical universe is bounded, both physically[1] and in terms of possibility. Furthermore, this finite limit is true both in the near term, and in the indefinite future. To discuss this, we restrict ourselves to a relatively prosaic setting, and for at least this paper, we restrict our interests to a single universe that obeys the laws of physics as currently (partially) understood. In this understanding, the light-speed

* The authors contributed equally in the conception and preparation of the paper.

[1] While cosmology debates some aspects of whether the universe is finite, as we note in the appendix, the various suggested possibilities still admit that the reachable universe is finite.

limit is absolute, quantum physics can be interpreted without multiverses[2], and thermodynamic limits are unavoidable. In addition to those assumptions about the universe, we will assume, based on the overwhelming scientific evidence, that human brains, and those of other beings with moral opinions and values, perform only within the laws of physics. Given that, we also assume that values are either objective functions of the physical world, as posited by Moore [2], or are subjective only to the extent that individual physical brains can conceive of them.

Given these fairly wide boundaries, we argue that there are no infinities that must be addressed for ethical decisionmaking. We do so by establishing concrete bounds on possible sizes of value that can be changed. Even though there are truly mind-boggling numbers involved, these are finite numbers which do not admit the class of question Bostrom and others pose. Given our assumptions, we cannot refute those arguments absolutely, or make the claim that we should assign no probability to such potential value systems. We will, however, make a strong claim that unless our understanding of physics is fundamentally flawed in specific ways, the amount of accessible and achievable value for any decision-relevant question is necessarily finite. We feel that the assumptions are likely enough, or can be modified to be so, that the argument is strong enough to be considered sufficient for resolving the issues for long term consequentialist thinking.

Before addressing fundamental issues about the limit of value, we will address the far easier question of whether there is a limit to economic growth, following and extending Ng's work [3].Based on a few observations about the Milky Way, we find a clear indication that in the short term future of the next 100,000 years, even in the most optimistic case, current levels of economic growth are incompatible with basic physical limits. This has implications for welfare economics and social choice explored by Ng [3], as well as for long term expectations about growth discussed by Hanson [4].

We next use that discussion to motivate questions about whether a more general framework for value allows infinities. After discussing and answering two possible objections to limited value in a finite universe, we outline additional physical limitations to both value and valuing. We then conclude that we can assign a theoretical upper bound to possible value in the physical universe.

2 Economic Growth and Physical Limits

Economic theory, the study of human choices about allocation of scarce resources[3], is useful for describing a large portion of what humans do. This is in large part because it is a positive description, rather than a normative one, and is local in scope. For example, it does not claim that preferences must be a certain way. Instead, economic theory simply notes that humans' values seem to be a certain way. Given some reasonable local assumptions, this can be used to make falsifiable predictions about behavior. Such a theory is by no means universally correct, as noted below, but forms a more useful predictive theory than most alternatives.

Clearly, the arguments and assumptions do not need to extend indefinitely to be useful. For example, economic assumptions such as non-satiation (which Mas-Collel [6] and others more carefully refer to as local non-satiation) will obviously fall apart at some point. That is, if blueberries are good, more blueberries are better, but at some point the volume of blueberries in question leads to absurdities [7] and disvalue. Here, we suggest that there are fundamental reasons to question the application of simple economic thinking about value and growth in value to long-term decisions. This is important independent of the broader argument about non-infinite value, and also both informs and motivates that argument.

Economic growth is an increase in the productive capacity of the economy. Economic growth measures the increase in the ability to produce goods that people derive value from. The above-mentioned locally correct models of human behavior and interaction lead to a natural conclusion that under some reasonable assumptions about preferences, economic growth will continue indefinitely. If there is possible value that can be built via investment of physical or other real resources, humans are motivated to at least attempt to create that value. If growth at some non-nominal rate continues indefinitely, however, this leads to difficult to physically justify results. For example, at a 2% level of real growth, the Gross World Product (GWP) would grow to 10^{860} times current levels in 100,000 years. GWP is currently

[2]This is not a required assumption, though given multiverses, some qualifications on how moral weight or normalization across many-worlds is required to ensure values are not all infinite.

[3]We will not discuss the contentious question of how economics is best defined, a subject of extensive discussion [5].

around \$100 trillion$^{US}_{2020}$ [4], so the total value is $\sim \$10^{874}$ dollars$^{US}_{2020}$. The naive model implies that we can continue to receive positive returns on investment, and humans will in fact value the resulting achievements or goods to that extent.

Note that this is not an argument about the nominal growth rate, but rather the real growth in value. That is, the US dollar may not exist in centuries, much less millennia, galactic years, or aeons. Despite this, if humans survive to continue creating value, the implicit argument of continued growth is that we would find things to do that create more real value in that time.

It is also not an argument that natural limits of the types often invoked in sustainability discussions will *necessarily* stop growth [9] [5]. While there are material limits to the amount of stuff that can be acquired (and, as we will argue below, this does matter for our conclusion) the stuff may be organized into ever better forms. That means that our argument about "Limits to Growth" is both less immediate, and more fundamental than the ecological limits most extensively discussed in economics. [10]

2.1 Economists versus Physicists

> "Scientists have developed a powerful new weapon that destroys people but leaves buildings standing — it's called the 17% interest rate.""
>
> — *Johnny Carson*

As Einstein almost certainly did not say, "compound interest is the most powerful force in the universe." But physicists are careful to limit their infinities so that they cancel. Economists have fewer problems with infinities, so they have never needed a similar type of caution[6].

On the other hand, if the claim that exponential economic growth at a rate materially above zero can continue indefinitely is true, it would indeed need to be the most powerful force in the universe, because as we argue, it would need to overcome some otherwise fundamental physical limits, outlined in the Appendix. This continued growth seems intuitively very implausible, but intuition can be misleading. Still, as will be discussed in more detail below, there are fundamental physical limits to how much "stuff" we can get, and how far we can go in a given amount of time.

2.2 Short-term Limits for Humanity

One initial consequence of the fundamental physical limits outlined in the appendix is the short term expansion of humanity over the next 60-100,000 years. In the best case, humanity expands throughout the galaxy in the coming millennia, spreading the reach of potential value. Despite expansion, the speed of light limits humanity to the Milky Way galaxy during this time frame. The Milky Way is $\sim$100,000 light-years across, and it would take at least on the order of that many years to settle it, with $\sim$60,000 being a lower limit to get to the far end from Earth[7].

[4] We adopt the convention that the ambiguous use of dollars needs to have units properly noted, as should occur everywhere in scientific research for any unit. However, because "dollars" do not have a constant economic value, or even refer to the same currency across countries, the subscript/superscript notation is used to disambiguate. The notation is adapted from Gwern [8].

[5] Daly suggests somewhat informally that "the physically growing macro-economy is still limited by its displacement of the finite ecosphere," in the context of economic versus "uneconomic" growth that creates "risks of ecological catastrophe that increase with growthism and technological impatience." In practice, we agree that sustainable development is a reasonable argument to curtail certain types of economic growth. It seems clear that unsustainable growth which leads to ecological collapse less likely to have unbounded long term potential than the alternative of short-term environmental protection. However, our argument is somewhat more fundamental in nature.

[6] This is not strictly true. Economic endogenous growth models are plagued by finite-time singularities if the feedback from knowledge or other factors to themselves is stronger than linear. Demanding that such factors are never negative and always remain finite force the model to exhibit exponential growth [11]. Others are less concerned about the singularities in the model: "Singularities are always mathematical idealisations of natural phenomena: they are not present in reality but foreshadow an important transition or change of regime" [12].

[7] If we instead consider the short term to stretch slightly longer, we could begin to consider the satellite galaxies to the Milky way, but this still limits us to smaller galaxies that are almost all within 1 million light-years of Earth. From there, there is a notable gap of approximately another 1 million light-years to Andromeda, the nearest major galaxy. For that reason, humanity's potential for expansion is unfortunately somewhat limited over the next 2 million years.

Thankfully, the medium term future looks rosier, since the entire Laniakea Supercluster is within a quarter billion light-years, and all of the Pisces–Cetus Supercluster Complex is accessible to humanity within the next billion years.

The Milky Way local neighborhood masses about 1.5 trillion solar masses (or 3×10^{42} kilograms) [13, 14] within a radius of 652,000 light-years, of which about 6.5% is baryonic matter. There are about 10^{68} cubic centimeters in the Milky Way galaxy alone (if we consider a sphere of diameter 100,000 ly.) And inside of a currently small portion of the space and mass, humanity pursues maximizing value. At this point, the question is how much value is possible.

2.3 Bounds on Short-term Economic Value and Growth

> "We have always held to the hope, the belief, the conviction that there is a better life, a better world, beyond the horizon."
>
> —*Franklin Delano Roosevelt*

Given the above set of bounds, given specific ways to calculate value, we can, *inter alia*, calculate an approximate maximum to total accessible value, and hence the possible growth in economic value. As discussed in the appendix, possible information that can be stored is limited by space and mass.

The volume accessible in our 100,000 year time frame allows us 10^{134} bits of theoretically accessible storage. The 10^{134} bits of storage correspond to $2^{10^{134}}$ possible states, and hence the maximum number we could store in the Milky Way has 3×10^{133} digits[8]. However, short of a misaligned AI which wire-heads into storing the largest possible value in the register containing its value function, it seems unlikely that there is any conception of value that consists solely of the ability to store massive numbers[9]

To consider economic return, we need a baseline for what is being invested. As noted above, current GWP is \$100 trillion$^{US}_{2020}$/year, which can be viewed as an income stream for humanity. Discounting the income at a generous 2%, we find a net present value of human productivity of \$50 quadrillion. Treating the discounted total of human production as an upper bound on how much we can possibly commit to investing now, we ask: how much value can be created in the future?

To create a minimum threshold for value, we consider the value of the universe if converted into some currently expensive substance, say Plutonium-239, which costs around \$5.24$^{US}_{2007}$ per milligram, [15], we find that converting the Milky Way leads to a value of \$$1.5 \times 10^{49}{}^{US}_{2020}$. Given our baseline, this is a return on investment of 3×10^{32}x, which is a huge return, but discounted over the next 100,000 years, this gives a paltry annual return on investment of 0.075%.

But this minimum is pessimistic — surely we can generate more value than just expensive mass with a service economy of some sort. If we consider the value of human productivity, we have a conceptually huge possible space of value that any human can provide to others. Starting with the present levels of productivity, we can very generously assume each human is able to produce \$1 million$^{US}_{2020}$ of value per year. The average human masses 70 kilograms, and we unrealistically ignore the requirement for gravity, air, food, and so on, to find that the 3×10^{42} kilograms of mass allows for 4×10^{40} humans. Assume each creates value, then assume this production starts immediately and accumulates over the next 100,000 years. This gives an upper bound is that the galaxy could produce \$$4 \times 10^{51}{}^{US}_{2020}$ of value in the next 100,000 years, which seems large until we note that it implies an annualized rate of return of 0.08%; far more than our estimate above, but a tiny rate of return.

If we even more generously assume that not only would humans instantly settle the entire Milky Way and convert the entire mass into humans, as above, but that they individually annually produce the Earth's annual Gross World Product (GWP) today, repeating our earlier assumptions, the rate of return reaches 0.1%. We can go further, and even more implausibly claim each can produces a googol dollars$^{US}_{2020}$ of value per year and this accumulates, to reach a 0.3%

[8] This number is exponentially larger than 10^{860} times current GWP (that has just slightly over 860 digits), but still far smaller than many celebrated very large numbers in mathematics, such as Graham's number — which has a number of digits that itself is far larger than can be stored in those 10^{134} bits of storage.

[9] It is, of course, possible to store representations of larger numbers, but these are insufficient for value writ large in various ways, as will be discussed later. For example, floating point numbers can be extended to represent far larger quantities in a given storage volume, but they are not closed under subtraction, and in a typical implementation, very large integers are rounded off. That means that they cannot be used for comparisons where the difference is relatively small. In fact, no encoding scheme using 10^{134} bits can contain, say, the value $2^{10^{134}} + 1$ without losing the ability to store the exact value of some smaller integer. While our large-number storage maximizing AI might be fine turning the universe into storage schemes that allow representing larger numbers, given the discussion below about value as comparisons, this inability is a fundamental issue for not just economic growth, but value-in-general.

annual growth rate. And perhaps we were unfairly pessimistic about our time frame for settling the Milky Way, and use the bare minimum physical limit of 60,000 years — this gives us a still paltry 0.51% rate.

All of this may be argued still to be conservative. Perhaps Hansonian Ems colonize the galaxy[16], taking up far less space, and living faster lives. In the limit, each could constantly be producing things of value for all other Ems to enjoy, allowing for the growth in value to be far higher. Still, and bounds found for a service economy are a function of space and mass growth over time. The way in which available mass increases with time is lumpy, based on locations of mass that are close to our solar system, but is at most similar to the increase in physical space. The lightspeed limitation means that the amount of "stuff" we can acquire from nature will at most grow as $(4\pi c^3/3)\rho t^3 \propto t^3$ [10]

This leads to an inescapable conclusion, that there is at most a polynomial rate of economic growth in the long-term. The available space cannot grow as quickly as any exponential function, so growth in value is guaranteed to be lower than the exponential growth implied by compound interest, at least in the very long term. As the earlier increasingly implausible assumptions suggest, and the rate of growth limitations make even clearer, postulating greater potential value still means that high or even steady low rates of growth would not be possible. Using our narrow economic definition of value, at some point in the near-term (cosmic) future, economic growth will be sharply limited. To posit any greater value, we need to consider the question of value much more broadly. Before doing so, we briefly consider a few implications of the short term conclusions.

2.4 Implications of Short-Term Limits

There are a number of interesting long-term policy implications of the existence of a limit to growth. Critically, many are related to the (not-quite oxymoronic) most immediate long-termist uncertainties.

2.4.1 Discounting

One set of conclusions relate to discounting of the far future, a topic discussed in varying contexts [18]. These implications of the choice of discounting rate range widely, from decisions about personal donations [19], to management theory [20], to climate policy [21].

Our conclusions about limited growth in the cosmological short term provides a much stronger argument for (very) low discount rates than much past work, albeit applicable only when considering longer time scales than even most long-termist policy considers. Applying the conclusions about limited growth to discounting, even over the very-long term, requires care, since different arguments for discounting exist[11] [22]. Specifically, this argument against discounting applies if long-term discounting is primarily based on an arbitrage or alternative investment arguments, where the reason to discount later value is because there is an alternative of investing and receiving a larger amount of capital in the future due to growth. If the argument is based on risk, where the reason to discount future value is because of the possibility that it will not be realized, our argument seems less relevant, through that of Weitzman [23], which argues for low discounts by reasoning over different possible futures, is correspondingly strengthened — and applies over the far shorter term.

2.4.2 Hinge of History

> "We live during the hinge of history... If we act wisely in the next few centuries, humanity will survive its most dangerous and decisive period. Our descendants could, if necessary, go elsewhere, spreading through this galaxy."
>
> *Derek Parfit*, On What Matters, Volume II

Another set of conclusions that can be found from the sharp limit to near-term economic growth relates to "Hinge of History," based on a claim by Derek Parfit [24] about the importance of the near future, which was later put more

[10] Because of the short time frames and local distances being discussed, this can ignore the expansion of the universe. Over longer times scales, as we discuss in the appendix, this further limits it to an asymptotically *finite* amount if the ΛCDM cosmology is a correct description [17].

[11] Note that we do not include equity concerns for discounting [21] because we are considering humanity as a whole, though obviously for policy the equity concerns for discounting can be critical.

pointedly in the above quote. In general terms, the hinge could relate to the recent economic turbulence introduced by Appelbaum and Henderson, [25], and the short term moral opportunities for equality introduced by Head, [26] but is primarily a long-termist idea [27], presaged by Greaves and MacAskill, of "influencing the choice among non-extinction attractor states" over the entirety of the future. [28] MacAskill suggests two worldviews that imply the present is such a hinge, while our exploration implies a third — though unlike those worldviews, our argument does not suggest the hinge is imminent.

The argument is that while there is no guarantee that the upper-bound of long-term value will be reached, or even approached, the current exponential economic growth cannot be maintained. The alternative hypothesis is that even though the world is changing and increasing in value ever more quickly, it will continue to do so indefinitely. Instead, the transition from exponential to polynomial economic growth would imply that a hinge-of-history of a sort must exist, though it may not be in the near future or related to the current slowing of growth, since the necessary timing depends heavily on questions of when the limit will be reached.

This argument for a hinge-of-history rests on the plausible, but not certain, claim that choosing the type of economic growth in the exponential growth phase significantly changes the course of civilization in a way that will not occur afterwards. The weakness in this argument is that at some point after the end of explosive growth, a long-reflection, such at that proposed by MacAskill [29] could still drastically alter trajectory. That is, the limit to growth does not by itself imply that any "hinge" in growth rates leads to irreversible decisions, and a different argument would still apply for why decisions during the hinge would be irreversible, such as MacAskill's two worldviews concerning value-lock-in, or irreversible choices that lead to annihilation.

2.4.3 Economic Singularities

The model above shows that recent growth has been higher than the rates plausible in the long term, and the time frame over which economic growth must drop to a lower rate is a topic for further consideration. This is because economic growth has been, and in the very near-term likely will be, far higher than the long-term economic growth horizon.

The necessity of such a transition also relates to claims of an eventual economic singularity. Such a singularity is already possibly unlikely to occur now, at least based on very-short term economic evidence [30]. But going further, a transition to polynomial growth creates a large but non-exponential limit to the speed of any claimed singularity in the longer term.

3 What is Value?

> "...maybe that means that for civilization, part of civilization is devoted to common sense, thick values of pursuit of art, and flourishing, and so on, whereas large parts of the rest of civilization are devoted to other values like pure bliss... The universe is a big place."
>
> — *Will MacAskill*

So far, the discussion has contained repeated caveats about economic growth and economic value, as distinct from some as-yet nebulous value-in-general. While others have noted the connection, such as Cowen [31][12], we attempt to clarify that concept, and see how it relates to the economic one, and the extent to which it does not.

Before discussing how choices relate to values, we note that our discussion is premised on choice as the central question of ethics. That is, ethics is the study of right and wrong *choices*, and the morality of those *choices*. Outside of a comparison between things, or a decision made about them, "value" has no meaning[13].

[12]Cowen splits the concept, saying that he's interested in "wealth-plus," which he defines as "The total amount of value produced over a certain time period. This includes the traditional measures of economic value found in GDP statistics, but also includes measures of leisure time, household production, and environmental amenities, as summed up in a relevant measure of wealth." But most economists would say that this is what economic value already captures, and the distinction made in Cowen's terminology is a measurement issue, rather than a disagreement about what value is.

[13]This is not a consequentialist claim. Any ethical statement must by definition be a comparison, saying one action (or lack thereof) is allowed, and another is forbidden. Even if moral statements are not factual, they are descriptions of factual scenarios, and short of nihilism, make claims that compare them.

3.1 Values as Choices

> "Give the person what they need — and they will want amenities. Provide him with amenities — he will strive for luxury. Showered with luxury — he will begin to sigh in exquisite. Let him get exquisite — he will crave frenzy. Give him everything that he wishes — he will complain that he was deceived, and that he did not receive what he wanted."
>
> — *Ernest Hemingway*

If there are two items, or two states of the world, and a choice must be made between them, we call the one chosen of higher value than the other. Similarly, if a person is willing to give up one item for another in, for instance, an economic transaction, we say that the one received is higher value to the recipient. Comparisons induce a mathematical "order" of states[14]. For this reason, either preference value or trade value is *at least* an ordinal preference, and any notion of value is comparative, rather than measured.

In this case, it is immediately possible to show that value under this conception in a finite universe is finite. Given a finite set of items or states of the world, it is trivial to see that the most preferred can only be a finite number of steps better than the least preferred. If the accessible universe is finite, as discussed below, it is then clear that the number of steps between possible states is potentially incredibly large, but still finite[15].

But value may be more than this ordinal concept. If we accept the cardinal conception of utility, value may be possible to add and multiply, rather than just compare[16]. If utility is mapped to real numbers, one item can meaningfully be called not just more valuable, but twice as valuable as another thing[17]. One key reason to consider cardinal utility is because it allows comparison of options given preferences with uncertainty about outcomes. That is, a choice may involve uncertainty, in which case the ordinal concept is insufficient.

3.2 Probabilities Require Cardinal Utility

Reasoning about preferences consistently given uncertainty, as introduced by Ramsey [32], requires ordering of preferences over probabilities of outcomes, rather than just outcomes. A decision maker might prefer a 1% probability of outcome A to a 100% probability of outcome B. A simple way to represent this is to assign more than 100 times the value, called utility, to A, then use probabilistic expectation of utility to see that the choice giving a 1% chance of outcome A is preferred. If arbitrary probabilities need to be considered, and we wish to ensure that the preferences being discussed fulfill certain basic assumptions about rational preferences, then cardinal utility, or a structure mathematically identical to it, will be required.

In this way, reasoning and decisions under uncertainty are the conceptual basis for considering utility of outcomes, rather than just atomic comparisons of specific options. And this can lead to problems when we insist on bounds for value. If decision makers consider an arbitrarily small probability of a given outcome preferrable to some other certain outcome, the utility assigned to the improbable outcome must be correspondingly high. To guarantee finitude of utility for a coherent decisionmaker, we need to argue that there is some minimum probability that can be assigned. This is conceptually fraught, but there are several possible responses we will discuss below in 4.2.

[14] We are implicitly ignoring measurability of utility in this discussion, since it is irrelevant once we assume that choices would be made which induce an ordering. Even though an insufficient number of choices are made to determine the utility, and actual measurability is plausibly absent, the argument we present applies to any set of choices that could be made. This makes measurability of utility irrelevant.

[15] While the set of things can be expanded by inventing or making new things, this faces two constraints. First, future time is bounded, as discussed in the appendix, A.2 so only a finite number of new goods can be created. Second, the number of arrangements of matter is finite, so the number of possible goods is limited. There may also be overlap, so that the same atoms participate in two or more valuable things, but is still finite, if exponentially growing.

[16] Mathematically, this is a ring, rather than just an ordered set, because we can define addition and multiplication.

[17] A similar argument does not apply to ordinal utilities — there is no mathematical justification in asserting that if one banana is traded for two apples, the banana is twice as "valuable", since the specific trade implies nothing about general preferences. More precisely, when discussing ordinal value mathematically, the notion of multiplying a position in the order by a number is meaningless.

This cardinal concept of utility also allows other possible objections to the conclusion that a finite universe can only have finite value, including lexicographic preferences, non-doscounted infinite time horizons, and other concerns which we will address.

3.3 Aggregative Ethical Theories and Objective versus Subjective Value

> "For welfare to be finite... the 'amplitude' of welfare cannot be infinite at any particular moment in time, and a life can only have a finite duration of welfare."
>
> — *Siebe Rozendal* [33]

Given our initial claims about the physical universe, we will note that our discussion of finitude of value is independent of a number of important philosophical disputes about ethics, at least in most ethical systems. For instance, whether value is an objective or subjective function of the world does not change whether an upper limit exists since it still needs to be represented[18]. Similarly, aggregative value, where overall value is the sum of the value for each individual, will increase the limit of value being discussed, for example, by multiplying the value limit by as much as the number of possible morally relevant beings which can assign value. Despite this, because the morally relevant beings are physical, and therefore require mass, the number of such morally relevant beings is finite, and therefore so is the total value[19]. Similar arguments for finitude can be made for any other form of value aggregation of which we are aware.

4 Result, Objections, and Responses

Value is finite. That is, in a physical universe that has no infinite physical and temporal scope, no infinities are available to represent infinite value in decision-making processes. Hence, any possible assignment of value used for decision making has to be finite.

It is possible to object to the claim of finitude. We believe that the entire set of possible objections, however, can be answered. Responding to the objections, therefore, is critical to the above claim. We therefore list the key objections, then review and explain them. After each, we will respond, including novel arguments against several such claims.

1. Rejecting (our current understanding of) physics
2. Rejecting preferences, by either
   - rejecting comparability,
   - rejecting finite preferences,
   - rejecting bounded expected utility, or
   - bounding probabilities (possibly via embracing infintesimals as valid probabilities for decisions,)
3. Rejecting ethical theories or embracing nihilism[20]
4. Rejecting the need for accessibility of value for decisions.
5. Rejecting or altering traditional causal decision theories.

4.1 Rejecting Physics

> "It is far better to grasp the universe as it really is than to persist in delusion, however satisfying and reassuring."
>
> —*Carl Sagan*

[18] It may be suggested that value could be purely 'subjective', i.e. independent of even the physical state of the brain of the person whose values are considered. If so, there is no relationship between the world and value, and the "ethics" being discussed does not relate to any decisions which may be made. If, however, ethics does relate to the physical world, then there can be some value assigned to each possible state and/or world-history.

[19] One could imagine an ad-hoc objection assigning moral weight to an infinite number of posited non-physical beings, but this does not change preferences being about physical states, so the resulting infinite value can therefore be mapped to finite numbers. The number of angels dancing on the head of a pin may be infinite, but the value they assign to the pin effectively cannot be.

[20] or perhaps some other non-consequentialist, non-deontological, and non-rights and non-virtue based theory of ethics.

Perhaps our understanding of physics is incorrect. That is, it is possible that our understanding of any of the assumed-correct disciplines discussed here, from cosmology to computation. This is not merely an objection to the authors' personal grasp of the subjects, but a claim that specific premises may, in the future, be found to be incorrect[21].

4.1.1 Pessimistic Meta-induction and expectations of falsification

The pessimistic meta-induction warns that since many past successful scientific theories were found to be false, we have no reason expect that our currently successful theories are approximately true. Hence, for example, the above constraints on information processing are not guaranteed to imply finitude. Indeed, many of them are based on information physics that is weakly understood and liable to be updated in new directions. If physics in our universe does, in fact, allow for access to infinite matter, energy, time, or computation through some as-yet-undiscovered loophole, it would undermine the central claim to finitude.

This criticism cannot be refuted, but there are two reasons to be at least somewhat skeptical. First, scientific progress is not typically revisionist, but rather aggregative. Even the scientific revolutions of Newton, then Einstein, did not eliminate gravity, but rather explained it further. While we should regard the scientific input to our argument as tentative, the fallibility argument merely shows that science will likely change. It does not show that it will change in the direction of allowing infinite storage. Second, past results in physics have increasingly found strict bounds on the range of physical phenomena rather than unbounding them. Classical mechanics allow for far more forms of dynamics than relativistic mechanics, and quantum mechanics strongly constrain what can be known and manipulated on small scales[22].

While all of these arguments in defense of physics are strong evidence that it is correct, it is reasonable to assign a very small but non-zero value to the possibility that the laws of physics allow for infinities. In that case, any claimed infinities based on a claim of incorrect physics can only provide conditional infinities. And those conditional infinities may be irrelevant to our decisionmaking, for various reasons.

4.1.2 Boltzmann Brains, Decisions, and the indefinite long-term

One specific possible consideration for an infinity is that after the heat-death of the universe[23] there will be an indefinitely long period where Boltzmann brains can be created from random fluctuations. Such brains are isomorphic to thinking human brains, and in the infinite long-term, an infinite number of such brains might exist [34]. If such brains are morally relevant, this seems to provide a value infinity.

We argue that even if these brains have moral value, it is by construction impossible to affect their state, or the distribution of their states. This makes their value largely irrelevant to decision-making, with one caveat. That is, if a decision-maker believes that these brains have positive or negative moral value, it could influence decisions about whether decisions that could (or would intentionally) destroy space-time, for instance, by causing a false-vacuum collapse. Such an action would be a positive or negative decision, depending on whether the future value of a non-collapsed universe is otherwise positive or negative. Similar and related implications exist depending on whether a post-collapse universe itself has a positive or negative moral value.

[21] This is different from a broader and fundamental possible argument, which is that science has no final conclusions which can be relied on for absolute moral claims. We reject this as morally irrelevant, since our discussion is about decisions which are made in reality. Given that, objections about the impossibility of certainty are also implicitly rejected by argument about the limits to probabilities.

[22] Of course, some results may find looser rather than stricter bounds. Despite this, even if we conclude that most specific currently known limits will be rejected at some point, this does not go far enough to imply that no such limits exist, and the central claim of this paper remains true.

[23] If there is no universe-ending Big Rip or the cosmological constant is negative enough to cause recollapse. At least the latter is disfavored by current cosmological observations. The former has no theoretical or empirical support. See also the Appendix. The nature of the heat-death does not matter much for the argument: the classic idea was a state of minimum free energy, while the modern is an equilibrium state of maximum entropy, or a "freeze" state where individual particles remain isolated at finite (microscopic) temperature. In either case random thermal fluctuations will occur briefly bringing it away from equilibrium from time to time. There may be a causal effect of our actions on the post heat-death state, but no action now can determine a post-heat death event.

Despite the caveat, however, a corresponding (and less limited) argument can be made about decisionmaking for other proposed infinities that cannot be affected. For example, inaccessible portions of the universe, beyond the reachable light-cone, cannot be causally influenced. As long as we maintain that we care about the causal impacts of decisions, they are irrelevant to decisionmaking.

4.2 Rejecting preferences

It is possible to reject the claims of relevant finitude by dispensing with one of the various required aspects of preferences needed for decisionmaking[24].

4.2.1 Rejecting Comparability

It may be objected that perhaps value is not finite because comparison is impossible, or alternatively, that some things are "infinitely valuable" on their own. Or perhaps humans can assign values in ways that are incompatible with finite value[25]. We discuss both, in sections 4.2.1 and 4.2.2, and reject them as untenable.

To address the first, we note the discussion in philosophy about whether values can be incomparable — that is, given two items or states of the world, neither is better. Chang's work [35, 36] makes a compelling argument rejecting incomparability, which view we would adopt for this paper. However, even without that, this incomparability argument is less than fatal to our claim. This is because incomparability still leads to a partial ordering of value, rather than a total ordering. That is, in a universe with positive value on bananas and blueberries, it is still the case that two blueberries are better than one, and two blueberries and a banana are better than one blueberry and a banana, even if we reject any possibility that the two can be compared. This leads to a large number of partial orderings of preferences, but any claims made about full orderings will apply to each partial ordering. For that reason, an analogue of any argument we provide will exist even if values are incomparable, and non-comparability alone does not allow for infinite value.

The alternative objection is where one item is "infinitely better" than another, and is thus incomparable in a different sense. These lexical preferences, as they are called, are not commensurate with any other value; most people would consider taking 2 bananas for one blueberry, but is seems at least arguable that there is no number of bananas many of them would take in exchange for not staying alive[26]. This idea of lexical preferences will be dealt with formally and in general below.

To address the second point, that humans might have an intrinsic ability to assign infinite values, we need to address what the assignment of human values means. One key question is what preferences are coherent, or valid, and a second is how these relate to decision making. There is a significant philosophical literature on whether infinities are coherent or logically possible, from Aristotle's rejection of "actual infinities", to recent work on infinite ethics [37]. We do not address these points, and limit ourselves to whether there are morally relevant physical infinities.

Given that, we must return to a central assumption we have made about values, that they must be morally relevant, i.e. make a difference in some ethical comparison or decision. This will be discussed further after considering lexicographic preferences.

[24] Aside from the obvious but ineffective method of rejecting the requirement for coherency or consistency, since doing so, and allowing utilities that do not conform to the required characteristics of rationality makes any discussion of maximum "utilities" irrelevant.

[25] For example, due to infinitesimal probability assignments.

[26] This argument cannot be used to justify claims of specifically exponential economic growth, since that relies on the claim that by investing resources now, the choice will lead to greater value in the future by enabling that growth. However, if a lexically preferable outcome can be purchased or created with money that can be invested, the analogue of economic growth has a utility function which is discontinuous, not growing exponentially.

4.2.2 Rejecting Finite Preferences

So-called lexicographic preferences consider some states infinitely better than other states[27]. There are two approaches that would justify a lexicographic claim, one intrinsic, and one based on probability. The intrinsic justification is that there are incomparably better states. For example, a negative utilitarian could argue that any state in which there is no suffering is infinitely preferable to any state that contains suffering. Compatibly, the probabilistic justification is that no probability of one state is sufficiently low that it would not be preferred. In this model, a negative utilitarian could say that any finite probability of more suffering is worse than a guarantee of less suffering[28].

If such preferences exist, they are typically claimed to lead to the impossibility of representing preferences as a real-valued utility function [38]. That is, if one item is "infinitely better," and preferences are cardinal[29], the claim is that we cannot bound utility to any finite value at all. We argue that as long as goods or states of the universe are finite, as occurs in a fixed volume of space with fixed total mass, this is untrue. This is based on a constructive proof, shown below.

As an example, we can consider a finite universe with three goods and lexicographic preferences $A \succ B \succ C$. We denote the number of each good N_A, N_B, N_C, and the maximum possible of each in the finite universe as M_A, M_B, M_C. Set $M = \max(M_A, M_B, M_C)$[30]. We can now assign utility for a bundle of goods

$$U(N_A, N_B, N_C) = N_C + N_B(M + 1) + N_A(M^2 + 1).$$

This assignment captures the lexicographic preferences exactly[31]. This can obviously be extended to any finite number of goods N_n, with a total of $N = \max(n)$ different goods, with any finite maximum of each[32].

As the most extreme possible example, assume our social welfare function has a lexicographic preference for filling the Milky Way with hedonium A over hedonium B, B over C, etc. We could still bound the number n of different such "goods" that could plausibly be lexicographically preferred, and the number M which could be made in the universe, to derive a bound of $2 \times M^{n+1}$. Even if the number of lexicographically preferred goods is enormous, it is bounded by the physically limited arrangement of matter that is possible, giving a still finite, if even more unimaginably large number.

To extend this logic to address probabilities, we must consider the assignment of probabilities and assignment of utility, which we do below. Before doing so, however, we will justify a claim underlying our argument.

4.2.3 Rejecting Bounded Expected Utility

> "We have therefore to consider the human mind and what is the most we can ask of it."
>
> —*Frank Ramsey*

[27] Etymologically, this comes from the idea of a lexicographic order, which generalizes the notion of alphabetic ordering. In an alphabetized list, any word starting with the letter "A" is lexicographically prior to any word starting with "B". Similarly, any world with a lexicographically preferred good is always better than one without. This is equivalent to saying that no matter what else occurs, that world is better. As we will show, however, lexicographic preferences do no necessarily imply actual infinities.

[28] We do not address the interesting but unrelated case where a negative utilitarian might have preferences that include trading off amounts and probabilities of suffering, though this might also involve claimed infinities, as they are addressed with the same argument as is used for other cases below.

[29] If preferences are ordinal, this just requires placing lexicographic preferred goods above less preferred ones, so the objection is irrelevant.

[30] This will be a huge number, of course. As an illustrative example, bananas are approximately 150 grams each, so the Milky Way would has $M_{Bananas}$ of 2×10^{43}, for normal sized bananas. Blueberries are around half a gram, leading to $M_{Blueberry} = 3 \times 10^{45}$.

[31] In the previous footnote's banana-blueberry universe, someone with a lexicographic preference for bananas over blueberries who assigns blueberries value 1 would assign value 3×10^{45} to a banana.

[32] Per the previous footnote, many believe that human lives are lexicographically superior to bananas. As the 2nd century Jewish saying notes, "Whoever saves a single human life, it is as if they have saved a whole world," (Sanhedrin 4:5) which presumably is even more true if the world that is saved in entirely filled with bananas. But representing the value of infinitely valuable (presumably happy) human lives does not require use of infinity. In fact, the by-assumption infinite value of a human life can be represented as being at most 2×10^{43}+1 times the value of a banana, or around 6×10^{88}+1 times the value of a blueberry. In a blueberry-banana-human value universe, infinitely valuable human lives are much better than blueberries, but mathematically still not even a googol times better, much less infinitely so.

Given that we conclude even lexicographic preferences are finite, might a person still assign infinite value to some outcome? That is, if utility is not only an ordering of states, but a function, is it coherent for a person to insist on discontinuities, where they assign five times as much utility to an apple as to a banana, and infinitely more utility to remaining alive as to dying? It is possible to argue that value-in-general is different than utility, but to the extent that the value is used for decision, we need some way to choose, and to be coherent, this method must compare states. Since we assume all states must be *comparable* using a (perhaps non-VNM-like) utility function, it still seems that value is bounded by the ability of the valuer to make decisions and to consider the different outcomes.

A utility function, in the decision theoretic or economic sense, is invariant to affine transformations. That is, multiplying every value by 2, or adding 17 to each utility, does not change the preferences that the utility function describes. But placing anything as infinitely valuable is a lexicographic preference, and for utility functions, the exact location of the lexicographic preference is irrelevant — as long as the order is preserved. This is true even when allowing for truly different experienced utility. If two humans both experience utility from a good, but (as an extension of Nozick's monster, [39]) one of them has a qualitatively infinitely better experience, we can treat their value as a lexicographic one. But this only implies that the earlier construction of a finite representation of lexicographic preferences captures all decision relevant factors, even infinite value. We therefore conclude that in a finite universe, any choices that are made can be reduced to perhaps incomprehensibly large but necessarily finite comparisons. This demonstrates that given physical finitude, ethics overall cannot be changed solely by claimed infinities in preferences between outcomes, at least before accounting for probabilities.

4.2.4 Bounding Probabilities

> "...it was just very very very big, so big that it gave the impression of infinity far better than infinity itself."
>
> —*Douglas Adams*

As noted above, any act considered by a rational decision maker, whether consequentialist or otherwise, is about preferences over a necessarily finite number of possible decisions. This means that if we restrict a decision-maker or ethical system to finite, non-zero probabilities relating to finite value assigned to each end state, we end up with only finite achievable value[33]. The question is whether probabilities can in fact be bounded in this way.

We imagine Robert, faced with a choice between getting $\$1^{US}_{2020}$ with certainty, and getting \$100 billion$^{US}_{2020}$ with some probability. Given that there are two choices, Robert assigns utility in proportion to the value of the outcome weighted by the probability. If the probability is low enough, yet he chooses the option, it implies that the value must be correspondingly high.

As a first argument, imagine Robert rationally believes there is a probability of 10^{-100} of receiving the second option, and despite the lower expected dollar value, chooses it. This implies that he values receiving \$100 billion$^{US}_{2020}$ at approximately 10^{100}x the value of receiving $\$1^{US}_{2020}$. While this preference is strange, it is valid, and can be used to illustrate why Bayesians should not consider infinitesimal probabilities valid[34].

To show this, we ask what would be needed for Robert to be convinced this unlikely event occurred. Clearly, Robert would need evidence, and given the incredibly low prior probability, the evidence would need to be stupendously strong. If someone showed Robert that his bank balance was now \$100 billion dollars$^{US}_{2020}$ higher, that would provide some evidence for the claim—but on its own, a bank statement can be fabricated, or in error. This means the provided evidence is not nearly enough to convince him that the event occurred[35]. In fact, with such a low prior probability, it seems plausible that Robert could have everyone he knows agree that it occurred, see newspaper articles about the fact,

[33] For those decision-makers who have other value systems, the earlier discussion suffices, and probabilities do not enter the discussion.

[34] We are grateful to Evan Ryan Gunter for suggesting several points we address in this section.

[35] One could argue that Robert's goal is not to have the state of receiving \$100 billion$^{US}_{2020}$, but rather the state of believing that he received the money. If so, of course, the relevant probability to assess is not that he would receive the money - and if he assigns a probability of 10^{-100} to that, he is severely miscalibrated, at least about the probability of delusions. Despite this, the below arguments still apply, albeit with a different referent event.

and so on, and given the low prior odds assigned, still not be convinced. Of course, in the case that the event happened, the likelihood of getting all of that evidence will be much higher, causing him to update towards thinking it occurred.

A repeatable experiment which generates uncorrelated evidence could provide far more evidence over time, but complete lack of correlation seems implausible; checking the bank account balance twice gives almost no more evidence than checking it once. And as discussed in the appendix, even granting the possibility of such evidence generation, the amount possible is still bounded by available time, and therefore finite.

Practically, perhaps the combination of evidence reaches odds of 10^{50}:1 that the new money exists versus that it does not. Despite this, if he truly assigned the initially implausibly low probability, any feasible update would not be enough to make the event, receiving the larger sum, be a feasible contender for what Robert should conclude. Not only that, but we posit that a rational decision maker should know, beforehand, that he cannot ever conclude that the second case occurs[36].

If he is, in fact, a rational decision maker, it seems strange to the point of absurdity for him to to choose something he can never believe occurred[37], over the alternative of a certain small gain.

Generally, then, if an outcome is possible, at some point a rational observer must be able to be convinced, by aggregating evidence, that it occurred. Because evidence is a function of physical reality, the possible evidence is bounded, just as value itself is limited by physical constraints. We suggest (generously) that the strength of this evidence is limited to odds of the number of possible quantum states of the visible universe — a huge but finite value[38] — to 1. If the prior probability assigned to an outcome is too low to allow for a decision maker to conclude it has occurred given any possible universe, no matter what improbable observations occur, we claim the assigned probability is not meaningful for decision making. As with the bound on lexicographic preferences, this bound allows for an immensely large assignment of value, even inconceivably so, but it is again still finite.

The second argument seizes on the question of inconceivability, without relying on Bayesian decision theory or rationality. Here we appeal to an even more basic premise of expected value, which is needing a probability assignment, or a value assignment at all. If Robert cannot conceive of the probability, he cannot use it for computations, or make decisions as if it were true. The question at this point is whether he can conceive of infinitesimal probabilities.[39]

We have been unfortunately unable to come up with a clear defense of the conceivability of infinities and infinitesimals used for decisionmaking, but will note a weak argument to illustrate the nonviable nature of the most common class of objection. The weak claim is that people can conceive of infinitesimals, as shown by the fact that there is a word for it, or that there is a mathematical formalism that describes it. But, we respond, this does not make a claim for the ability to conceive of a value any better than St. Anselm's ontological proof of the existence of God. More comically, we can say that this makes the case approximately the same way someone might claim to understand infinity because they can draw an 8 sideways — it says nothing about their conception, much less the ability to make decisions on the basis of the infinite or infinitesimal value or probability.

Finally, we can also appeal to what Aaronson calls the Evolutionary Principle, which states that "knowledge requires a causal process to bring it into existence." [40] If moral statements and values are truth-apt, any value, or probability, which is found in moral epistemology or in an individual's preferences requires that some physical process led to the

[36]Perhaps he can accept the result with less convincing evidence. One might argue that if every conceivable result of having the money occurs, he might as well accept it as having occurred. In that case, however, the odds he assigned to the possibility are not actually 10^{-100}, which is verified by the fact that less than the corresponding amount of evidence effectively convinced him.

[37]If he is not, in fact, a rational Bayesian, and his probability assignment was a statement of preference rather than an estimate, it is a lexicographic preference rather than a probability, and can be discussed as above.

[38]About $\exp(10^{123}) \approx 10^{10^{4.3\times10^{122}}}$.

[39]When reasoning about a probability like 10^{-100} we can use mathematical methods to reach reliable conclusions, e.g. that 10^{-99} is 10 times more likely, despite not having any intuition about the value itself. This ability to place concepts into lawful relations to each other relies on the existence of representations that can be manipulated. The need to represent the relations applies even if consideration must be outsourced to formal methods rather than intuitive comprehension. In fact, given any number of possible states in a universe, the number of possible states is the maximum number of distinct values which can be represented. By the pigeon-hole principle, the probability of at least one state must be lower than the smallest discretely representable value in the system. As the number of possible quantum states of the universe suggests, there are probabilities which cannot be explicitly represented using any finite system, but they will not be relevant for decision-making.

assignment of moral value. This argument is potentially incompatible with moral non-cognitivism, but even ethical subjectivism requires individual value judgements to occur, and these are subject to the same physical constraints due to being a result of a causal process.

The relationship of value to probability is itself closely related to the relationship of outcomes to value. That is, there must be a physical or cognitive process that arrives at the decision of what value to assign to an outcome. Any probability assignment, per the Evolutionary Principle, is a function of the computation available. And given the earlier-discussed limits on storage and on computation, assigning a probability value X to a state will be limited. Even if we assume that the entire universe's computational capacity is available, there is some immensely large but finite number that can never be found[40].

4.3 Rejecting Ethical Theories

However, value might be fundamentally different than we assume. We begin by looking at consequentialist version of the argument, then briefly address other moral claims.

We noted above that one of the arguments about bounding probability, and value, does not work given moral non-cognitivism. We can make a similar claim about moral realism, where perhaps value is in fact inherent in objects in a physical sense. After all, while the location of an object or its temperature can be represented, they are also inherent in the object (or at least inherent in the relationship between the object and the surroundings).

However, we again appeal to the question of decision making. Even for moral realists, either this value can be directly experienced or it cannot. In the latter case we still need to represent our estimates of the value, and these representations will be subject to the earlier bounds on physical reality. In the former case we need to be able to compare values to each other. Either this occurs through comparing mental representations of the actual value experiences (necessarily bounded), or we directly compare the intrisic values without any representational intermediary — but the comparison requires some minimal computation to occur outside of the objects. In either case, a clear bound exists on what value is possible.

Alternatively, we can consider ethical-theory objections, rather than the meta-ethical ones above. We assume in the discussion a utilitarian or at least consequentialist viewpoint. This is in large part because the question of finitude of value is most clearly relevant in that frame. Despite this, other theories face similar limits. Deontological and rights-based theories are faced with a finite number of possible actions which have moral value, and the earlier arguments for comparability and finitude would still apply.

4.4 Accessibility

Bostrom's discussion of infinite ethics is premised on the moral relevance of physically inaccessible value. That is, it assumes that aggregative utilitarianism is over the full universe, rather than the accessible universe. This requires certain assumptions about the universe, as well as being premised on a variant of the incomparability argument that we dismissed above, but has an additional response which is possible, presaged earlier. Namely, we can argue that this does not pose a problem for ethical *decision-making* even using aggregative ethics, because the consequences of any ethical decision can have only a finite (difference in) value. This is because the value of a moral decision relates only to the impact of that decision. Anything outside of the influenced universe is not affected, and the arguments above show that the difference any decision makes is finite.

We argued earlier that Boltzmann brains are inaccessible, since our actions do not impact the distribution of random matter after the heat death of the universe. This relies on a different type of inaccessibility, since our actions can have an impact, but one that is fundamentally unpredictable — making us morally clueless [41] in an even stronger sense than complex cluelessness [42]. Still, any solution to cluelessness seems to leave inaccessible impacts morally irrelevant [43], and this would apply even more strongly to our case.

[40] In computer science, infinities of a certain type are limited to non-halting programs, and these programs do not return a value before the end of the universe. For that reason, conceivable infinities are only ever potential, rather than actual, in an interesting return to an Aristotelian dichotomy about infinities.

4.5 Rejecting or Altering Decision Theory

Another approach to avoiding finitude is to question not preferences, but decision theory itself. There are discussions like [44] which consider decision theories that would allow for causal relationships with entities outside of the reachable universe in various ways, such as those discussed by Yudkowsky [45]. This alone does not imply infinite value. However, there are some proposed cosmologies in which these decision theories imply infinite value is possible[41]. For example, this would be true if we accept the Mathematical Universe Hypothesis proposed by Tegmark[46].

This conclusion goes further than most proponents of such theories would argue, and farther than is required for the purposes of the current argument — it says that ethics, rather than just decision making, should be based on these theories. In fact, most of the arguments in favor of non-causal decision theories are based on the consequentialist claims that these decision rules perform better in some situations. For this reason, the use of such theories to reject the type of consequentialism that justified them is not inconsistent, but seem a bit perverse.

Not only that, but Stoeger [47] points out that the universes with infinite value proposed by Tegmark are both unreachable, and unfalsifiable. Despite all of this, if we consider value aggregated over the multiverse in ways that do not renormalize to finite measure, we can be left with infinities. And as with rejecting physics, if we assign any finite positive probability to this being true, we are potentially[42] left with decisions that have infinities in their value.

Another key point about decision theory can be used to address the argument about potential infinities, related to our discussion of accessibility. That is, if we assign a small but non-zero value to physics being incorrect in ways that allow, say, reversing entropy, and infinite value is possible, all infinities are still limited to this possible universe, and decisions must be made on that basis.

Traditional expected-value decision theory is often interpreted to require risk-neutrality. This means that a single infinity will dominate any decision calculus. Many of the arguments for risk-neutrality, such as arbitrage and exploitation of repeated chances, fall apart in the current scenario. For example, if risk-neutrality is based on the possibility of arbitrage, where a risk-neutral participant in a market can receive free money by taking and perfectly hedging a risk, this becomes impossible when the risk is a single binary question which cannot be hedged. The same is true for the argument from repeated chances. A person might prefer $100 with certainty to a 60% chance of $200, but if they believe that this and similar choices will occur again in the future, the choosing the riskier option each time becomes more and more attractive, as the expected value remains the same but the risk of losing overall decreases with each additional bet. This clearly cannot apply to a single possibility about the question of which physical laws obtain in the universe.

However, a rational actor might choose to embrace a regret-minimization approach[43]. In this case, the regret from not maximizing the small probability of infinite return is infinite. We note, however, that key justifications of regret minimization involve arguments from long-term results that we reject above, while others are game-theoretic and do not apply here[50].

If we consider uncertainty over ethical theories, then given the standard metanormative theory of maximizing expected choiceworthiness, [51] we would apply the arguments above. One key criticism of that approach, however, is that it requires intertheoretic unit comparisons, and per Greaves and Cotton-Barratt, [52] this leads to a number of issues pointed out by Dai [53]. If we choose an alternative metanormative approach to address this, we my be able to reject possible infinities due to moral uncertainty even more simply. In Greaves and Cotton-Barratt's moral parliament, using bargaining theory, the problem of infinities being assigned some nonzero probability is addressed in a straightforward way, as by design no ethical theory can hijack the decision.

Note that an implicit conclusion from the assumption of infinite possible value is that moral progress is unbounded. Of course, that implies that any finite value achieved, however large, is an exactly nil fraction of possible value. In contrast

[41] In mainstream cosmological theories, there is a single universe, and the extent can be large but finite even when considering the unreachable portion (e.g. in closed topologies). In that case, these alternative decision theories are useful for interaction with unreachable beings, or as ways to interact with powerful predictors, but still do not lead to infinities.

[42] It is of course possible to embrace all of these claims, but still find that for other reasons, such as choice of the theory of ethics, infinities do not apply.

[43] A rational actor can do so not as a failure or accommodation due to biases, [48] but as an alternative axiomatic framework [49].

to this, if value can be taken to be finite, moral progress is limited to a finite value, but progress is meaningful, in the sense that we can approach that maximum[44].

5 Conclusions

> "I see the world being slowly transformed into a wilderness; I hear the approaching thunder that, one day, will destroy us too. I feel the suffering of millions. And yet, when I look up at the sky, I somehow feel that everything will change for the better, that this cruelty too shall end, that peace and tranquility will return once more."
>
> —*Anne Frank*

The above argument leads to the clear conclusion that humanity's best current understanding of physics implies that possible value is finite. Despite the usefulness of infinities in mathematics, physics, and even in discussions of preferences, given humanity's current understanding of physics we have shown that the morally relevant universe is finite, and can have only finite value. Of course any human reasoning is fallible, and any probability that this argument is wrong would lead to an expected infinite value, and lead to a Pascal's-wager-like obviation of any comparative value. Short of that, however, we can safely conclude that in this universe, abiding by the currently understood physical laws, moral value is, and will always be, finite.

To reject this claim, a few choices are available. First, one could rejecting our current understanding of physics, and insist that modern physics is incorrect in very specific ways. Second, one can reject values and decision theory in very specific ways, such as rejecting comparability, relying on non-cognitivism or embracing infintesimals as valid probabilities for decisions, or embracing non-causal models for decision theory as the basis of ethics and simultaneously rejecting accessibility of value. Lastly, one could choose nihilism, or some nontraditional ethical theory designed to avoid finitude.

None of these is unreasonable. However, we caution that each allows for infinite value only conditional on a variety of assumptions laid out in the paper.

Without these, our universe, and any universe with similar physical laws, has at most finite value for any moral actor. The peculiar nature of the infinite means that *any* finite value of the universe, no matter how large, as a fraction of infinity is exactly zero. Considered not as a fraction of infinity, of course, the immensely large physical limits do not preclude, and in fact imply, the existence of possible value far beyond that which humans currently imagine. Rejecting infinite values, and the various paradoxes and dilemmas they implicate, allows us to focus on considering what values should be pursued, and how best to reach the paradise that the future can become.

[44] cf. MacAskill's argument that "the vast majority of my expectation about the future is that relative to the best possible future we do something close to zero. But that's cause I think the best possible future's probably some very narrow target.... how much better could the world be? I don't know, tens of times, hundreds of times, probably more. In the future, I think it'll get more extreme." [54]

References

[1] Nick Bostrom. Infinite ethics. *Analysis and Metaphysics*, 10:9–59, 2011.

[2] GE Moore. The conception of intrinsic value. ed. moore, ge philosophiscal studies (253-275), 1922.

[3] Y-K Ng. Should we be very cautious or extremely cautious on measures that may involve our destruction? *Social Choice and Welfare*, 8(1):79–88, 1991.

[4] Robin Hanson. Long-term growth as a sequence of exponential modes. *Working manuscript*, 2000.

[5] Roger E Backhouse and Steven G Medema. Retrospectives: On the definition of economics. *Journal of economic perspectives*, 23(1):221–33, 2009.

[6] Andreu Mas-Colell, Michael Dennis Whinston, Jerry R Green, et al. *Microeconomic theory*, volume 1. Oxford university press New York, 1995.

[7] Anders Sandberg. Blueberry earth. *arXiv preprint arXiv:1807.10553*, 2018.

[8] Gwern (Pseudonym). Inflationadjuster. `https://www.gwern.net/Inflation.hs`, 2019–2020.

[9] Herman DALY. Three limits to growth. *Mother Pelican*, 10(10), 2014.

[10] Donella H Meadows, Dennis L Meadows, Jorgen Randers, and William W Behrens. The limits to growth. *New York*, 102(1972):27, 1972.

[11] Hendrik Hakenes and Andreas Irmen. On the long-run evolution of technological knowledge. *Economic Theory*, 30(1):171–180, 2007.

[12] Anders Johansen and Didier Sornette. Finite-time singularity in the dynamics of the world population, economic and financial indices. *Physica A: Statistical Mechanics and its Applications*, 294(3-4):465–502, 2001.

[13] Laura L Watkins, Roeland P van der Marel, Sangmo Tony Sohn, and N Wyn Evans. Evidence for an intermediate-mass milky way from gaia dr2 halo globular cluster motions. *The Astrophysical Journal*, 873(2):118, 2019.

[14] TK Fritz, A Di Cintio, G Battaglia, C Brook, and S Taibi. The mass of our galaxy from satellite proper motions in the gaia era. *arXiv preprint arXiv:2001.02651*, 2020.

[15] Reading list: Bernstein, jeremy. plutonium, 2007.

[16] Robin Hanson. *The Age of Em: Work, Love, and Life when Robots Rule the Earth*. Oxford University Press, 2016.

[17] Stuart Armstrong and Anders Sandberg. Eternity in six hours: Intergalactic spreading of intelligent life and sharpening the fermi paradox. *Acta Astronautica*, 89:1–13, 2013.

[18] Hilary Greaves. Discounting for public policy: A survey. *Economics & Philosophy*, 33(3):391–439, 2017.

[19] William MacAskill. When should an effective altruist donate? 2016.

[20] Cliff Landesman. When to terminate a charitable trust? *Analysis*, 55(1):12–13, 1995.

[21] Christian Azar and Thomas Sterner. Discounting and distributional considerations in the context of global warming. *Ecological Economics*, 19(2):169–184, 1996.

[22] Marc Fleurbaey and Stéphane Zuber. Discounting, risk and inequality: A general approach. *Journal of Public Economics*, 128:34–49, 2015.

[23] Martin L Weitzman. Why the far-distant future should be discounted at its lowest possible rate. *Journal of environmental economics and management*, 36(3):201–208, 1998.

[24] Derek Parfit. *Reasons and persons*. OUP Oxford, 1984.

[25] Richard P Appelbaum and Jeffrey Henderson. The hinge of history: Turbulence and transformation in the world economy. *Competition & change*, 1(1):1–12, 1995.

[26] Ivan L Head. *On a hinge of history: the mutual vulnerability of South and North*. University of Toronto Press, Toronto, Ont., CA, 1991.

[27] William MacAskill. Are we living at the most influential time in history?, 2019.

[28] Hilary Greaves and William MacAskill. The case for strong longtermism. Technical report, Global Priorities Institute Working Paper Series. GPI Working Paper, 2019.

[29] Toby Ord. *The precipice: existential risk and the future of humanity*. Hachette Books, 2020.

[30] William D Nordhaus. Are we approaching an economic singularity? information technology and the future of economic growth. Technical report, National Bureau of Economic Research, 2015.

[31] Tyler Cowen. *Stubborn attachments: a vision for a society of free, prosperous, and responsible individuals*. Stripe Press, 2018.

[32] Frank P Ramsey. Truth and probability" later reprinted in he kyburg and he smokler eds. *Studies in Subjective Probability*, 1926.

[33] Siebe T. Rozendal. *Uncertainty About the Expected Moral Value of the Long-Term Future*. PhD thesis, University of Groningen, Faculty of Philosophy, 2019.

[34] Sean M Carroll. Why boltzmann brains are bad. *arXiv preprint arXiv:1702.00850*, 2017.

[35] Ruth Chang. *Incomparability and practical reason*. PhD thesis, University of Oxford, 1997.

[36] Ruth Chang. *Making comparisons count*. Routledge, 2014.

[37] Adrian W Moore. *The infinite*. Routledge, 2018.

[38] Amartya Sen. Utilitarianism and welfarism. *The Journal of Philosophy*, 76(9):463–489, 1979.

[39] Robert Nozick. *Anarchy, state, and utopia*, volume 5038. New York: Basic Books, 1974.

[40] Scott Aaronson. Why philosophers should care about computational complexity. *Computability: Turing, Gödel, Church, and Beyond*, pages 261–328, 2013.

[41] James Lenman. Consequentialism and cluelessness. *Philosophy & public affairs*, 29(4):342–370, 2000.

[42] Hilary Greaves. Xiv—cluelessness. In *Proceedings of the Aristotelian Society*, volume 116, pages 311–339. Oxford University Press, 2016.

[43] Andreas L Mogensen. Maximal cluelessness. *The Philosophical Quarterly*, 2020.

[44] Benjamin A Levinstein and Nate Soares. Cheating death in damascus. *The Journal of Philosophy*, 117(5):237–266, 2020.

[45] Eliezer Yudkowsky and Nate Soares. Functional decision theory: A new theory of instrumental rationality. *arXiv preprint arXiv:1710.05060*, 2017.

[46] Max Tegmark. Is "the theory of everything" merely the ultimate ensemble theory? *Annals of Physics*, 270(1):1–51, 1998.

[47] William R Stoeger, GFR Ellis, and U Kirchner. Multiverses and cosmology: philosophical issues. *arXiv preprint astro-ph/0407329*, 2004.

[48] David E Bell. Regret in decision making under uncertainty. *Operations research*, 30(5):961–981, 1982.

[49] Peter C Fishburn. *The foundations of expected utility*, volume 31. Springer Science & Business Media, 2013.

[50] Eric Pacuit and Olivier Roy. Epistemic foundations of game theory. 2015.

[51] William MacAskill, Krister Bykvist, and Toby Ord. *Moral Uncertainty*. Oxford University Press, 2020.

[52] Hilary Greaves and Owen Cotton-Barratt. A bargaining-theoretic approach to moral uncertainty'. *Unpublished ms, August*, 2019.

[53] Wei Dai. Is the potential astronomical waste in our universe too small to care about? 2014.

[54] Robert Wiblin and Keiran Harris. Will MacAskill on the moral case against ever leaving the house, whether now is the hinge of history, and the culture of effective altruism, 2020.

[55] J Richard Gott III, Mario Jurić, David Schlegel, Fiona Hoyle, Michael Vogeley, Max Tegmark, Neta Bahcall, and Jon Brinkmann. A map of the universe. *The Astrophysical Journal*, 624(2):463, 2005.

[56] Tamara M Davis and Charles H Lineweaver. Expanding confusion: common misconceptions of cosmological horizons and the superluminal expansion of the universe. *Publications of the Astronomical Society of Australia*, 21(1):97–109, 2004.

[57] Michael T Busha, Fred C Adams, Risa H Wechsler, and August E Evrard. Future evolution of cosmic structure in an accelerating universe. *The Astrophysical Journal*, 596(2):713, 2003.

[58] Matt Visser. Jerk, snap and the cosmological equation of state. *Classical and Quantum Gravity*, 21(11):2603, 2004.

[59] Michael J Mortonson, David H Weinberg, and Martin White. Dark energy: a short review. *arXiv preprint arXiv:1401.0046*, 2013.

[60] Fred C Adams and Gregory Laughlin. A dying universe: the long-term fate and evolutionof astrophysical objects. *Reviews of Modern Physics*, 69(2):337, 1997.

[61] Freeman J Dyson. Time without end: Physics and biology in an open universe. *Reviews of Modern Physics*, 51(3):447, 1979.

[62] Lawrence M Krauss and Glenn D Starkman. Life, the universe, and nothing: Life and death in an ever-expanding universe. *The Astrophysical Journal*, 531(1):22, 2000.

[63] Jacob D Bekenstein and Marcelo Schiffer. Quantum limitations on the storage and transmission of information. *International Journal of Modern Physics C*, 1(04):355–422, 1990.

[64] Raphael Bousso. A covariant entropy conjecture. *Journal of High Energy Physics*, 1999(07):004, 1999.

[65] Raphael Bousso, Éanna É Flanagan, and Donald Marolf. Simple sufficient conditions for the generalized covariant entropy bound. *Physical Review D*, 68(6):064001, 2003.

[66] Raphael Bousso. The holographic principle. *Reviews of Modern Physics*, 74(3):825, 2002.

[67] Jacob D Bekenstein. Universal upper bound on the entropy-to-energy ratio for bounded systems. *Physical Review D*, 23(2):287, 1981.

[68] Horacio Casini. Relative entropy and the bekenstein bound. *Classical and Quantum Gravity*, 25(20):205021, 2008.

[69] Ram Brustein and Gabriele Veneziano. Causal entropy bound for a spacelike region. *Physical Review Letters*, 84(25):5695, 2000.

[70] Eanna E Flanagan, Donald Marolf, and Robert M Wald. Proof of classical versions of the bousso entropy bound and of the generalized second law. *Physical Review D*, 62(8):084035, 2000.

[71] Norman Margolus and Lev B Levitin. The maximum speed of dynamical evolution. *Physica D: Nonlinear Phenomena*, 120(1-2):188–195, 1998.

[72] L Mandelstam and IG Tamm. The uncertainty relation between energy and time in non-relativistic quantum mechanics. In *Selected Papers*, pages 115–123. Springer, 1991.

[73] Diego Paiva Pires, Marco Cianciaruso, Lucas C Céleri, Gerardo Adesso, and Diogo O Soares-Pinto. Generalized geometric quantum speed limits. *Physical Review X*, 6(2):021031, 2016.

[74] Manaka Okuyama and Masayuki Ohzeki. Quantum speed limit is not quantum. *Physical review letters*, 120(7):070402, 2018.

[75] Qiaojun Cao, Yi-Xin Chen, and Jian-Long Li. Covariant versions of margolus-levitin theorem. *arXiv preprint arXiv:0805.4250*, 2008.

[76] Sebastian Deffner and Steve Campbell. Quantum speed limits: from heisenberg's uncertainty principle to optimal quantum control. *Journal of Physics A: Mathematical and Theoretical*, 50(45):453001, 2017.

[77] Seth Lloyd. Computational capacity of the universe. *Physical Review Letters*, 88(23):237901, 2002.

[78] Stephen DH Hsu and David Reeb. Black hole entropy, curved space and monsters. *Physics Letters B*, 658(5):244–248, 2008.

[79] Yehoshua Bar-Hillel and Rudolf Carnap. Semantic information. *The British Journal for the Philosophy of Science*, 4(14):147–157, 1953.

[80] Roman Krzanowski. What is physical information? *Philosophies*, 5(2):10, 2020.

[81] John D Norton. The impossible process: Thermodynamic reversibility. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 55:43–61, 2016.

[82] Paul Erker, Mark T Mitchison, Ralph Silva, Mischa P Woods, Nicolas Brunner, and Marcus Huber. Autonomous quantum clocks: does thermodynamics limit our ability to measure time? *Physical Review X*, 7(3):031022, 2017.

Appendix A — Bounds on the Physical Universe

> "Space is big. You just won't believe how vastly, hugely, mind-bogglingly big it is. I mean, you may think it's a long way down the road to the chemist's, but that's just peanuts to space."
>
> —— *Douglas Adams*

If the accessible physical universe is unbounded, the assumption which leads to our conclusion is incorrect. As we outline below, it seems difficult to make this claim. Earlier, we considered the short term future of humanity expanding throughout the Milky Way Galaxy. This rested on assuming finite time, and therefore finite available space. Now, we consider the longer term future, and point to fundamental limits that will apply over the full lifespan of the universe.

A.1 The Accessible Universe is Neither Eternal, nor Infinitely Large

> "Infinity itself looks flat and uninteresting. Looking up into the night sky is looking into infinity — distance is incomprehensible and therefore meaningless."
>
> *—Douglas Adams*

The physical extent of the universe could, conceivably, be infinite. Unfortunately, even in this case physics limits the reachable portion of the universe to necessarily be finite. The reason is the accelerating expansion of the universe that not just moves remote galaxies away from us but moves them at such a speed that most can never be reached even if we expand from Earth at lightspeed. While the observable universe is approximately 46.5 billion light-years in radius and increasing in size [55], the *reachable* universe is limited to inside the cosmological event horizon 14.5 billion light-years away and decreasing in size. [56] No material resources outside this distance can be acquired, nor can we causally affect such resources or places.

The current distance to the event horizon is

$$\chi = c \int_{t_{now}}^{\infty} \frac{dt}{a(t)}$$

where $a(t)$ is the scale factor of the universe. If the equation of state parameter $w = p/\rho$ of the dominant component of the universe[45] is $w < -1/3$ then the integral converges and the amount of ever accessible matter is bounded by $(4\pi/3)\chi^3\rho_0$ where ρ_0 is the current matter density. [57]

For dark energy $w = -1$, but in the past radiation- ($w = 1/3$) and matter-dominated ($w = 0$) eras have occurred. They are unlikely to recur since dark energy appears to be dominant and growing. Quintessence theories allow for time-varying w, but there is neither any evidence nor any counter-evidence for them. Determining the higher order terms of the equation of state (that would allow detecting a time-varying w) through observation may turn out to be infeasible for the foreseeable future [58]. Similarly there are a number of alternative explanations of the accelerating expansion, but the accepted mainstream model is ΛCDM with horizons. Empirically, $w = -0.98 \pm 0.06$, well away from the $-1/3$ boundary [59].

Even if there were no event horizon, at any finite time the total space that could be settled would still be finite. Infinite resources are only obtainable in the limit even in welcoming cosmologies. Given that supposition, we would need to argue that physics guarantees that time for value-related activities itself is finite. We do so now.

A.2 Temporal Limits

> If your time to you is worth savin'
> Then you better start swimmin' or you'll sink like a stone
> For the times they are a-changin'
>
> *—Bob Dylan*

[45] Typically, the lowest w in a mixture of components will tend to dominate the expansion at large time.

Our current understanding of cosmology is incomplete, but current theories agree that even if the universe is temporally infinite in some sense, there is a point at which we reach either universal heat death, or a big freeze. Long before this proton decay would remove any baryonic matter (the only form we know can produce computation and life), and black hole evaporation later removes all other forms of concentrated matter, leaving individual elementary particles causally disconnected in expanding space[46] [60]. In either case, at some theoretically determinable point in the future, the state of the universe will not be affected by any current actions. Hence, value-related activities will necessarily be limited to end before this point.

A.2.1 Evading spatial or temporal finitude

These assumptions could be invalidated if, for example, faster than light (FTL) travel is possible (the human domain could expand beyond the cosmological event horizon) or radically different cosmological theories were true. However, the consequences of FTL include time-travel with the concomitant trouble with causality, CTC-based hypercomputing[47], and the difficulty of defining *when* value exists, or whether the value is caused by an action that occurred at a different time. While this cannot be ruled out, the consequences for value theory are much more problematic than a finite limit to value.

Alternatively, if dark energy is absent, intergalactic settlement could continue indefinitely, acquiring a slowly diverging amount of matter until the time limits set in. A world without proton decay is certainly conceivable, but to allow structure to persist indefinitely, temperatures need to decline indefinitely. There have been proposals for evading the heat death of such an expanding universe by hibernating longer and longer periods, exploiting the ever colder environment for a diverging amount of computation with a finite energy budget [61]; this is not compatible with accelerating expansion, which causes a finite temperature horizon radiation that makes indefinite information processing with error correction impossible for finite-resourced civilizations [62]. The cosmology could also be closed or have a Big Rip singularity, producing an even more definite endpoint.

Universes where indefinite settlement is possible require that either we must be empirically wrong about the accelerating expansion, or that an as-yet-unknown physical phenomena change w in the future (and henceforth maintain $w > -1/3$), *and* in either case, that proton decay and all other late-era structure-disrupting phenomena predicted from current theories must all be wrong. One can never rule out radically different physical theories as alternatives to the mainstream model, but their prior probability does not appear high, especially since several independent properties of physics need to conspire to allow indefinite settlement.

The remaining unaddressed question is whether finite time and finite space still allow relevant infinities. As explained below, the physical limits on storage make this impossible, and this will allow us to consider the remaining objection to finite value.

A.3 Physical limits to storage

In addition to limitations on the size of the universe, there are also fundamental limits on physical information storage. The most obvious limit is on the mass or energy used to encode information — but volume matters too.

While the efficiency of current magnetic storage is about 1 million atoms per bit, DNA storage can achieve 32 atoms per bit and in principle one could store one bit per atom (for example by using ^{12}C and ^{13}C atoms in a diamond lattice). Information can also be stored as radiation, an example being light circulating in long delay-lines. These limits depend on the types of information-carriers available[48].

[46] Long before this point there will be no possibility of there existing any sentient beings to consider the possible value, but perhaps those beings have preferences about later outcomes regardless of that fact.

[47] Which would, by allowing sending information back in time, allow to always find the action with the highest measured value. Whether this solves ethics or merely makes implementing ethical systems trivial (at a slight cost of the concept of free will) may be debated.

[48] For example, if we assume all of the baryonic mass in the Milky Way is converted to carbon atoms 9.3×10^{66} bits could be stored. Were the whole mass converted to light in delay-lines across the galaxy, the storage capacity would be $(R/c)(E/2\pi\hbar) = 1.3 \times 10^{105}$ bits (based on [63, eq. 112]).

For fundamental reasons, it is believed that we cannot store more than $10^{66}\,\mathrm{bits\,cm^{-3}}$. This is based on the covariant entropy bound that states that the maximum amount of information that can be stored within a region is bounded by

$$I \leq \frac{A}{4A_{Pl}} = \frac{c^3 A}{4\hbar G},$$

where A is the area of the region [64, 65, 66][49].

There exist a rich flora of such entropy bounds. The "classic" bound is the Bekenstein bound for the information inside a spherical region of radius R and energy E [67], which bounds it as

$$I \leq \frac{2\pi RE}{\hbar c}.$$

Since it was proposed in 1981 it has held up well despite much effort to produce a counterexample, and it has been proven that a version holds in any relativistic quantum field theory [68]. The Bekenstein bound implies a capacity of 2.5×10^{106} bits in the Milky Way and 8.3×10^{121} bits in the reachable universe. Brustein and Veneziano propose another one that is the geometric mean of the Bekenstein and the covariant bound [69].[50]

Generally these bounds are closely related to the generalised second law of black hole thermodynamics [70]. A heuristic argument for why such bounds are very plausible and appear unavoidable given known physics can be found in `https://www.scottaaronson.com/blog/?p=3327`. Basically, quantum fields in a bounded region with enough spatial variation to encode much information have greater energy and hence greater gravitational mass, and black hole formation around the region places an upper limit on this capacity.

While we may quibble about which bound is most accurate, physicists would generally agree that the amount of storeable and retrievable information in a finite volume with finite energy is finite. Were it not so, then one could exploit the storage capacity to run Maxwell's demon to provide perpetual motion[51].

A.4 Maximum computations / value over time

The physical bound on value might be argued to relate to the amount of computation that is possible, rather than the maximum storage. Given the temporal and physical bounds above, however, this too is strictly finite.

There exist limits on how fast distinguishable states (i.e., information) can be changed into other states (i.e., information processing). The Margolus-Levitin bound [71] states that a system with mean energy $\langle E \rangle$ cannot move to another orthogonal state in less time than

$$\tau_{ML} = \frac{\pi\hbar}{2\langle E \rangle}.$$

This bound implies a bound per quantum bit of 6×10^{33} operations per second per joule. Given a finite time and finite energy there will be a finite amount of computation. A related limit is the Mandelstam-Tamm limit linked to total energy [72]; such limits generalize in quantum mechanics [73], classical mechanics [74], and curved spacetimes [75]. These (quantum)limits can be derived straight from the formalism of quantum mechanics [76], and to evade them one needs to evade quantum mechanics.

[49] This is slightly oversimplifying things: the bound is on the information across the inward light sheet from a particular instant of the boundary. For practical purposes here it corresponds to the spacelike interior.

[50] That these bounds involve a spatial factor may inspire the hope that the expansion of the universe would enlarge the storage capacity. While the total amount of information that could be stored across the universe does increase over time, the accessible amount from any given point unfortunately declines: the distance to the event horizon shrinks as time goes by and more and more remote memory storage units disappear.

[51] Since the demon could retain infinite information, it does not have to pay a negentropy cost to erase past data and could hence persistently produce a thermodynamic disequilibrium from which energy can be extracted, contra the arguments due to Szilard, Landauer and Bennett. A world running on continuous physics might allow *potentially* arbitrarily dense information storage, but would still not allow *actual* arbitrarily dense storage due to noise. For example, the Planck scale does not (contrary to many popularizations) indicate that physics is discrete on sufficiently small scales, merely that as-yet unknown quantum gravity will be needed to describe processes below this scale. If measurement and manipulation below this scale is not possible then physics could be truly continuous yet only finite amounts of information can be stored by us.

This leads to Seth Lloyd's estimate "The universe can have performed no more than 10^{120} ops on 10^{90} bits." [77] Again, the exact numbers are less relevant to the general argument than that they are bounded: achieving unbounded computation requires unbounded space, time, and energy[52].

Clearly, information storage would be a fundamental limit on compound interest in a given currency, since fiat currency must be tracked somehow[53]. This isn't sufficient to limit value in a more general sense, however. Instead, we ask if information storage can limit value writ large[54]. For any consequentialist value system, this depends on preferences about the state of the system, and if the number of possible states is finite, so is the number of steps in a comparative value ordering.

A.5 Exponential or Polynomial Growth?

The above discussion assumes that value is related to physical limits, and these limits differ markedly in how they increase over time and space[55]. Available space grows only polynomially with time and eventually more slowly, with matter/energy following suit. However, this allows at least initially exponential growth of storage. The number of possible states that can be stored is multiplied by 2 for each additional bit that can be stored, and the highest storeable number doubles: there is an exponential growth of the maximum representable value if it is just represented as that number, as the number of bits used increase. This may suggest exponential growth of maximal value with energy and space. Even if cosmological expansion makes the expansion of value-representing systems slow down to an exponentially declining trickle the representable value can grow at least linearly until the last matter is collected.

If time is included, e.g. by not representing the value explicitly but making it available for comparison through potentially long computations, then it can grow exponentially in time. Just mapping some static representation x to a computation $C(x)$ that may be compared to other things is not enough since there are only a given number of bits N to represent x and there will be at most 2^N possible x or $C(x)$. However, if we include a clock time we can define the computational object $C(x,t)$ in such ways that it can represent values that are larger than the one represented by $C(x,t-1)$ and yet comparable to other objects $C(y,t)$. This in principle allows exponential growth of value until either the clock runs out[56] or the conditions for computation being possible cease.

If the value is just determined by t, for example making t-values lexically higher than $t-1$-values, it is possible to "cheat" by just setting the clock to the maximal possible t. To avoid this, the computation needs to depend in a nontrivial way on previous steps so that the quickest way of reaching the ultimate timestep is to perform all computations as fast as they can be done. This replaces the external clock with an internal computational state. Since dissipative operations are the one cost that will eventually run down any energy reservoir this may lengthen survival significantly, especially since it might be performed by a non-dissipative quantum computation that is only limited by rare tunneling errors (that occur exponentially rarely as a function of the height and width of the energy barriers used, in turn proportional to the energy and mass available). Error correction, however, is dissipative: one cannot survive indefinitely on a finite amount of energy or negentropy at finite temperature if errors are corrected.

It hence looks like space, mass/energy, and time each allow exponential representation of value — up to a limit set by cosmology and the physics of computation.

---

[52] Or at least energy, for special cases such as Frank J. Tipler's collapsing Omega Point cosmology. Theoretical physicists has also proposed exotic "Bekenstein's Monster" states where infinite information is stored in finite volume; such configurations appear to inevitably evolve into black holes before any information can escape or be used [78].

[53] This once again justifies the earlier argument about clear physical limits on economic growth, since any finite amount of information storage implies that there will still be some maximum rate of interest in a finite universe within finite time that could apply.

[54] Note that it may not seem obvious that all information in a philosophical sense requires a physical medium. We admit that care is needed to ensure that one does not naively over-interpret "information" in the sense used in physics and information theory to encompass all meanings of the term. [79] However, it seems clear, as Krzanowski argues, that even if a form of information does not exist in a physical sense, that form must still depend on information which does physically exist, and is therefore addressed in our argument. [80].

[55] We are grateful to Adam Brown for suggesting this question and initial thoughts about answers.

[56] Clocks are necessarily physically irreversible and hence dissipative and would have a finite state space to represent the time [81, 82].

The functional form of possible value does not change the argument for finitude, but materially impacts our expectation for the actual value over time, and is a critical moral question for the long term.