Putting out the dark fire: constraining speculative physics disasters

Anders Sandberg and Forrest Landry

March 23, 2015

Abstract

Various speculative physics disaster scenarios have been proposed, but most can be defeated by the observable lack of large-scale disasters they predict. However, this paper introduces the "dark fire scenario" where baryonic matter is autocatalytically converted into dark matter. If this could happen, the resulting disasters would be quiet and hard to observe. Ruling out dark fire is challenging, although arguments based on charge neutralization and astronomical imprints from early eras appears to do a good job. More generally, the slipperiness of many physics scenarios pose interesting epistemic problems for risk analysis.

1 Introduction

Over the past years there have been some concern that large particle accelerators could produce effects that threaten the Earth. The possibilities have been raised of triggering vacuum decay [34] or production of stable black holes or strangelets [37] that could absorb the planet. There have been several analyses showing this to be unlikely or impossible based on theoretical considerations or empirical constraints [8, 17, 14, 11].

The "cosmic ray argument" is a general argument for the safety of particle accelerators [16]. It its simplest form it states that far more energetic cosmic rays have been hitting the Earth and other celestial bodies for billions of years, so if there was any risk of dangerous collision products they would have had clear, observable effects. In particular, the conversion of stellar bodies into more compact forms would produce supernova-like releases of energy, and the rate of conversion events is hence constrained by observed supernova rates. The basic argument needs some tweaking to deal with the differences from particle accelerators[1] and anthropic considerations[2], but is generally regarded as sound.

[1]The center of mass frame is moving in the cosmic ray case rather than stationary relative to the planet, so products might harmlessly escape the gravitational well.

[2]If observers are precluded by disasters, the only observers will be very lucky ones in disaster-free regions.

The "dark fire scenario" presented here is a new speculative risk scenario that circumvents the cosmic ray argument. In its basic form it states that *under some special conditions high energy interactions can trigger a conversion of ordinary matter into dark matter, and that this process has an autocatalytic effect that leads to a conversion of nearby matter too.* The conversion would be invisible or hard to detect, and hence compatible with current observations. Some motivations for why this scenario is not entirely far-fetched will be discussed in section 3.1. This paper will examine and try to bound the risk of the dark fire scenario.

Why spend effort on problems like these? Existential risk has significant moral priority: it will be more important compared to nearly anything else [4]. Even a tiny chance of losing the astronomically large value of humanity's future requires careful consideration and precaution [2]. Even if one assigns a low probability to a existential risk the enormous value at stake implies that doing some investigation to rule out the risk (allowing more efficient risk mitigation for other, real risks) or to find ways of mitigating the risk, has a large positive expectation. Physics disasters are potentially avoidable.

At the same time there is an opportunity to refine our methods of handling existential risk uncertainty. The dark fire scenario is an interesting example of a "small theory", an exotic and perhaps *a priori* improbable theory predicting high risk and hence posing an epistemic challenge to risk assessment [6]. A particular problem is the looseness of the special condition causing the conversion: since it is not strictly defined it could be arbitrarily exotic, making it arbitrarily hard to rule out beyond gesturing towards its diminishing prior probability. If we can handle scenarios like this well, we will be better prepared for other risk concerns in the future.

2 Constraints

2.1 Theoretical constraints

Theoretical constraints are often very persuasive, since they put the credence in a major physical theory on the line to rule out certain possibilities. Unfortunately there are often issues of whether the theory is correct, whether it has been used to model the situation correctly, and the correctness of the actual calculation [20]. Worse, in the case of dark matter and the dark fire scenario the scenario description is underconstrained: several theoretical arguments hinge on dark matter being of a particular type, something we actually do not know. Hence the strength of theoretical arguments in this case is weaker than when dealing with empirically well-understood domains. However, internal consistency and compatibility with well-tested physical law gives some footing.

2.1.1 The conversion process

The conversion process is by assumption triggered by some condition which makes matter undergo conversion to dark matter. For this to be a risk the

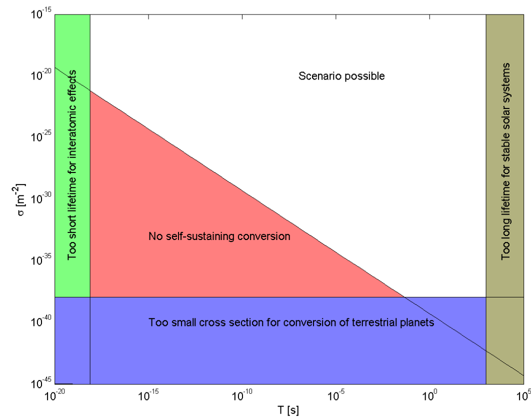

Figure 1: Allowed lifetime/cross section parameter space.

process must be autocatalytic, inducing the condition nearby. This implies certain restrictions: the range of the condition needs to be at least inter-atomic, or converting matter need to move and have enough cross section with normal matter that each converted particle triggers at least another conversion during the time it is "active" before turning into inert dark matter.

The first condition suggests an exchange force carrier mass $m < \hbar/cd$. For $d = 228$pm (atomic spacing in iron) $m < 865\text{eV/c}^2$. This rules out all known particles except neutrinos, photons, gluons and gravitons. It would be compatible with axions, ultralight scalar dark matter and massless fifth-force carriers, but not with any supersymmetric particles in the MSSM [21].

The later condition implies $\sigma T v > 1/\rho$ where ρ is the particle density of matter (for iron $\rho \approx 6.9 \cdot 10^{30}\text{m}^{-3}$), v the average velocity, σ the cross section and T the active time. If $v \approx c$, $\sigma T > 4.9 \cdot 10^{-40}$. Since reaction products need to have time to reach other atoms, $T > 7.6 \cdot 10^{-19}$s. A typical weak force cross section $\sigma = 10^{-42}\text{m}^2$ implies $T > 490$s. If dark fire occurs naturally, densities of globular cluster cores puts an upper bound $T < 10^7$s and solar systems $T < 10^3$s. Solar system stability hence suggest a cross section on the order of the weak force.

Low cross sections are untenable for a risky particle since it would have time to escape Earth. The relevant factor is whether $2R\sigma\rho$, the number of affected particles when traveling through Earth, is > 1. This happens (for $T > 0.02$s) when $\sigma > 1/2R\rho = 1.1 \cdot 10^{-38}\text{m}^{-2}$.

These bounds do give a somewhat plausible range for T, but does not constrain σ much. Slower particles would further reduce the constraints. However, the dangerous region in the (σ, T) plane is a narrow strip and other considerations may bound it.

2.1.2 Invisibility of conversion

The visible energy release when normal matter is converted to dark matter must be small enough that it does not cause long-distance visible events. Even if a small mass-fraction (above $\approx 1/1000$) is released as visible energy a star-conversion would tend to produce a supernova-like energy release. Supernova rates hence put a limit on the frequency of conversions and mass defect.

Estimated rates of supernovas are on the order of ≈ 3 per century per galaxy [22].

A supernova rate of 1 per 100 years in a normal galaxy implies that the conversion rate (if some supernovas are due to the scenario) is less than 10^{-2} solar masses per year in a galaxy, producing a minuscule and unobservable amount compared to the Milky Way halo over the history of the galaxy: $1.32 \cdot 10^8/10^{12} = 0.000132$.

Truly silent conversion requires that there exist dark matter particle species that are suited to absorb a nucleon or electron mass without releasing detectable decay products. Mass differences must be radiated away as neutrinos or other light weakly interacting particles. The lack of neutrino bursts still constrain the number of stellar-mass occurrences to a few per century[3], or require non-neutrino radiation.

2.1.3 Charge conservation

Dark matter is electrically neutral, while normal matter has charges. The conversion process needs to neutralize these or break charge conservation. Charge conservation is experimentally strongly supported [35]. A neutralization process may act by causing some intermediate charged products that then absorb normal matter (e.g. $p^+ + X^0 \rightarrow Y^+$, $Y^+ + e^- \rightarrow Z^0 + X^0$) but these products need to be stable long enough to allow neutralization. No candidate particles with these properties are known.

It should also be noted that a process capturing electrons will likely release their potential energy, and since the involved particles all are charged it is very plausible that it will be radiated away as photons. For hydrogen this implies about 1 MeV per atom[4], or $1.0 \cdot 10^{14}$J/kg. A stellar conversion would hence produce $2.0 \cdot 10^{44}$J, similar to a typical supernova. It would be bright (and spectrally unusual) enough to be observed over long distances.

Charge neutralization hence may prevent the quiet aspect of the dark fire scenario (and make the cosmic ray/supernova rate argument applicable).

[3] In the 17.3 years Super-Kamiokande has been online the neutrino signature of just one supernova has been detected; since detection outside the Milky Way neighborhood is unlikely, this gives a 90% confidence interval 0.29-27 events per century.

[4] Using the potential difference between the mean electron radius a_0 and the proton Compton wavelength.

2.1.4 Baryon number conservation

Normal matter has positive baryon number, while dark matter is usually assumed to have zero baryon number. Hence the conversion process needs to violate baryon number. R-parity considerations make it plausible that in a supersymmetric theory baryon number is preserved (if it wasn't, proton decay would occur at a high rate). However, baryogenesis does imply baryon number violation at least under some conditions that must have occurred in nature [24], and black holes are suspected of violating baryon number [1][5].

2.1.5 Baryogenesis

The dark fire scenario also needs to explain why baryonic matter did not further decay into dark matter after baryogenesis. A possible answer would likely involve that whatever condition is necessary for conversion, it was not possible during this era despite the high matter density. This likely implies that it cannot just be a thermal issue.

If dark fire did develop in the post-baryogenesis era it would have caused expanding spherical holes with no baryonic matter, expanding until the matter density reached a critical threshold where the expansion ceased. Given that we do not see such holes in the visible universe – intergalactic voids appear to be relatively free from dark matter [32] and can be threaded by sparse filaments of galaxies [38], which would not be the case for holes where all initial matter was dark – we may have a bound the probability of initiation.

If initiation has probability p per unit volume during an early matter dominated era at proper time t, there should be about $p[a(t_{now})/a(t)]^3$ holes per unit volume at present. Assuming a scale factor $a(t) \propto t^{2/3}$ and requiring that there is less than one hole in our Hubble volume,

$$p < \frac{3[t_{now}/t]^2}{4\pi\chi_0^3}$$

where $\chi_0 = 14$Gpc. If $t = 10^{-12}$s (early electroweak symmetry breaking) then $p < 5.6 \cdot 10^{-22}\text{m}^{-3}$, while a later $t = 1$s era gives $p < 5.6 \cdot 10^{-46}\text{m}^{-3}$. The first case corresponds to a temperature $T = 10^{16}$K, the second $T = 10^{10}$K.

Expansion must cease at some lower density limit. For dark fire to be a problem at molecular matter densities the finishing time has to be after roughly 1,000 seconds after the big bang when the density dropped below this level.

The co-moving size of the hole is $\chi = 3c(t_2^{1/3} - t_1^{1/3})$ where t_1 and t_2 the proper times where dark fire begins and ends, respectively. This then scaled up by a factor $a(t_{now})/a(t_2)$ to current proper distance. For $t_1 \approx 0$ and $t_2 = 1000$s the hole size is $5.2 \cdot 10^{19}\text{m} = 1.6$kpc.

[5] Since decaying black holes would be able to emit dark matter, they might seem to be a good candidate for the mediating particle in the dark fire scenario. However, this possibility does not work in the scenario because of the basic cosmic ray argument, theoretical considerations that rule out dangerous black hole remnants in accelerators, and the fact that black hole accretion/decay is radiatively noisy.

This appears to be a strong constraint on dark fire scenarios: unless some hard-to-meet condition is introduced, the probability of triggering it even at very high energy density must be extremely small to fit astronomical observations.

2.2 Empirical constraints

If matter is converted to dark matter at some rate, then the total amount of matter will decline exponentially: $M(t) = M(0)e^{-\lambda t}$ and the dark matter content increases with an equal amount[6]: $D(t) = D(t_0) + M(t_0)(1 - e^{-\lambda})$, where t_0 is the time when the initial balance is set. The ratio $M(t)/(D(t) + M(t)) = e^{-\lambda t}/(D(t_0)/M(t_0) + 1)$ declines exponentially over time. At present $M(t_{now})/(D(t_{now}) + M(t_{now})) = 0.16$. If conversion happens at an appreciable rate over cosmological timescales there should hence be noticeable differences between the present and past ratio.

The main source of information on early universe dark matter content would be patterns in the cosmic microwave background radiation and remnants of big bang nucleosynthesis.

2.2.1 CMB and structure formation

The pattern of the cosmic microwave background indicates the state of dark matter relative to matter inhomogeneity in recombination epoch; if the dark matter content has changed significantly since then there should be signs here. Unfortunately most studies appear to have assumed a constant ratio and fitted the data to that.

If the dark fire scenario is true, then dark matter can form late. This has been studied for the case of decaying scalar fields in the era between nucleosynthesis and recombination, concluding that it may be possible to fit observed data to a fraction of dark matter forming in the relatively early part of the interval (around 17 years of age), but that later formation is largely ruled out by observations [26]. They also point to the need to run N-body models to see what observable microstructure effects it has: a changing dark matter content would likely have noticeable effects on structure formation.

2.2.2 Nucleosynthesis

Big bang nucleosynthesis is sensitive to the degrees of freedom in the early universe: the observed content of heavier primordial elements constrains some of the physics beyond the standard model. It looks that during the $t \sim 0.1 - 10^4$sec era not much beyond known relativistic degrees of freedom (photons, electron, positrons, the three left-handed neutrinos plus baryons) was present and interacting with the element formation. In particular, when dark matter was still assumed to be baryonic, attempts to reconcile element abundances with baryons dominating the mass budget consistently failed [18]. This failure

[6] We here assume there is no loss into energy, which would contradict the quietness assumption of the scenario

provides compelling reason to suspect that the dark matter content during the early universe was already close to the present level.

Another relevant observation is that if there exist new charged massive particles they would catalyze nucleosynthesis, which place a bound on their abundance and lifetime (section 6 in [18]). This is relevant due to the assumed intermediary step in handling charge conservation in the dark fire scenario: the Y^+ particles would have Y^- antiparticles that could catalyze nuclear reactions. For abundances to fit observation short-lived Y-particles need to have been extremely abundant (essentially making up nearly all matter), and while more long-lived particles could have been rarer (a fraction of 10^{-5}) their lifespans would have been very long, making their present undetectability hard to explain. It is very likely that further investigation of the necessary properties of dark fire intermediate particles will show that they are hard to reconcile with nucleosynthesis.

2.2.3 Galaxy rotation

If the dark matter content of galactic halos has changed, rotation curves of early galaxies should be different from rotation curves of present galaxies.

Unfortunately the observational evidence is somewhat scattered. [25] looks at z=1 galaxies and found them to be fairly like present ones. In fact, they seem to have *more* dark matter than present ones, explained in the paper as them having formed from heavier halos than similar-looking present galaxies (i.e. an observation selection effect). [27] seems to get about normal dark matter content for z=2. However, an earlier galaxy (z=2.38, 3 byr after the big bang) had low levels of dark matter [13], and [31] did not see evidence for dark matter within 12 kpc of a z=4.9 galaxy. But these papers are based on individual galaxies which may introduce a measurement bias: these could be exceptional galaxies that are easy to observe.

2.2.4 Structure formation

Structure formation models are sensitive to the type and amount of dark matter. Hence it should in principle be possible to run these models with time-varying matter/dark matter ratios and see what the highest conversion rate compatible with current observation is.

In particular, dark matter appears to be non-self interacting while baryonic matter is [15], and baryonic matter can radiate away kinetic energy through electromagnetism. This makes it naturally more clumpy. If there was more baryonic matter in early eras that was converted to dark matter, we should expect to see more rapid dense structure formation than current models assume. In addition, if converted matter does not get a significant thermal kick from the conversion process, dark matter halos formed from converted baryonic matter would be on the same scale as the structures they formed from, producing significantly denser halos than in models where cold dark matter is constant.

2.2.5 MACHOs

While MAssive Compact Halo Objects (MACHOs) have declined as candidates for the galactic and cosmological dark matter in favor of weakly interacting particles, the dark fire scenario would produce MACHO-like objects. A planet or star converting to dark matter would remain gravitationally bound unless the energy release was significant (and hence very likely to cause a noticeable phenomenon)[7]. The dark matter particles would approach viralisation after some time, forming a cusp density profile ($\rho \propto r^{-\alpha}$) with a dense tightly bound core. This kind of object would be potentially detectable due to microlensing.

The EROS collaboration concluded that MACHOs cannot contribute more than 8% to the mass of the galactic halo [12]. Hence if *all* of the MACHOs were dark fire remnants the conversion rate would be 3.6-18 solar masses per year[8].

It might be argued that halo structures are not the remnants of dark fire outbreaks. If stars converted we should also expect the distribution of converted stars to fit the stellar distribution, forming a dark disk of similar size to the visible disk. This is however hard to reconcile with the observed rotation curves (best explained by an extended spheroidal dark matter halo) unless the fraction of converted stars was small compared to the visible matter fraction.

Another constraint on stellar or planetary conversion would be the detectability of dark matter annihilation from the dense clump, since the annihilation rate scales as the square of the density. A star-sized clump would have a dark matter density some 10^{24} times greater than the the density in a spiral galactic core, and would hence produce copious amounts of gamma rays despite its smaller mass (a mere factor of 10^{11} or so): this appears hard to hide from existing searches, unless one postulates that the conversion process will not produce anti-dark matter or that it is rare enough that these dense clumps get disrupted [28] quickly enough not to be noticeable.

2.3 Anthropic selection

The safety record of particle physics and stability of our close astronomical environment gives some evidence for lack of risk. Unfortunately this may be merely an *apparently* safe situation since our existence is predicated on no disaster having occurred: past safety is a less strong argument for future safety than normal if observer selection effects play a role [5]. However, such effects can also be used to make risk estimates.

One way of bounding the risk from vacuum decay is to note that Earth is a fairly-late forming planet and we occur fairly late in the expected span of the

[7] If a fraction f of the mass-energy of a spherical mass is released as kinetic energy, it will be approximately enough to disrupt it if $f > 3GM/5Rc^2$. For a sun-like star this fraction is $f \approx 1.2 \cdot 10^{-10}$. Undetectably quiet conversion events (less luminous than a small supernova) hence likely imply smaller f and hence bound remnants.

[8] This is higher than the present star formation rate (0.68 to 1.45 solar masses per year [23]). One can hence argue that the observed star formation rate bounds the conversion rate: a higher conversion rate would have left no stars in the disk. However, since the star formation rate [30] and possibly the conversion rate are time varying this is a rather weak constraint.

biosphere. Had the universe been very risky most observers would be found close to the start, on early planets where evolution was unusually fast [33]. This places a bound on how likely bodies are to be converted ($< 10^{-9}$ per year).

We may still worry that we occur suspiciously early in the stelliferous era, which began a few hundred million years after the big bang but continues for at least a trillion years. However, the star formation rate has been declining rapidly since the beginning and is now only at 3% of the maximum; only 5% more stars will ever form [30]. This places us very late in the ordering of solar systems, reducing the likelihood that any process removes habitable solar systems.

Another argument may be that we should expect ourselves to be close to the median observer in the history of mankind. While this leads to the familiar Doomsday argument (especially when connected to population growth), in this case we may take it to mean that if there is a physics risk ending humanity we should expect to be halfway through humanity's history rather than right at the end, and we should hence assign a fairly low risk per year from (current) particle physics. However, this category of arguments are strongly controversial and depend on nontrivial choices of reference class (maybe only observers of LHC physics experiments count, rather than all humans?) and choices between the self-indication assumption and the self-sampling assumption.

2.4 Summary

The constraints on the conversion process appear unavoidable, yet does not constrain the dark fire scenario as much as we may wish. Charge conservation is on firmer grounds, and together with big bang nucleosynthesis may expose inconsistencies in the scenario (but this needs to be worked out). The invisibility assumption is flexible enough to be unconstrained. Baryon number violation is unfortunately allowed by other parts of our current scientific model. The lack of dark matter holes appears to be a very strong constraint. Empirically, it does not appear that there is much evidence in favor of an increasing dark matter content in the universe, but to truly demonstrate this we need to go back to the original studies and fit the data (or run N-body simulations). Nucleosynthesis places some nontrivial constraints on the scenario. Galaxy rotation curves are inconclusive. Anthropics provides some evidence that it is unlikely there is significant conversion.

These constraints are each relatively underwhelming – they generally reduce the probability of the dark fire scenario, but does not do so in a watertight fashion. Their benefit is that they are independent: while some may fail, there are multiple reasons to be skeptical of the scenario. The best argument is the lack of early dark fire outbreaks. However, the slipperiness of the scenario allows it to evade this constraint by invoking the exotic conditions required for an outbreak: it can simply be something that does not occur during the big bang or later.

3 Philosophical discussion

3.1 Unfalsifiability and priors for risk

The assumption that special conditions are needed to trigger the dark fire scenario makes it unfalsifiable. One can always imagine peculiar conditions that have occurred so rarely that there cannot be any observational evidence for conversion.

Unfortunately, while unfalsifiability is very bad for a scientific theory, it does not rule out risks. Somebody may have poisoned your tea with an untraceable (in your current environment) poison: the fact that you cannot rule out this possibility doesn't mean you should discount it if you have other reasons to fear poisoning. The relevant issue is whether we think we have any reasons independent of the risk scenario in question to believe there is a risk.

In our case the most obvious such reason is the Fermi question ("where are they?"): the absence of detectable extraterrestrial intelligence adds at least some credence to the possibility that there exist self-created existential dangers to intelligent life. Unfortunately this is also a weak constraint since alternative explanations (extreme rarity of life, the zoo hypothesis, the simulation argument...) are possible and could contain most of the probability.

We also have historical experience with bad outcomes occurring because of human activities creating special conditions (whether unusual ecological states, high population densities, nuclear chain reactions or stable toxins), giving us at least a weak prior that exotic actions can be risky. Human experience also shows that we tend to be overconfident and not in the habit of looking for dangers of an abstract type – there is no built-in instinct for that, despite a tendency to look for other types of risk.

How much does it matter that the dark fire scenario is a specific physics experiment risk scenario rather than a possible implication from existing theoretical physics? Vacuum decay, strangelets and black holes were conceptualized as a result of other research: they were introduced for reasons other than being risks. While dark matter production is a major theoretical and experimental issue in physics, autocatalytic dark matter production does not seem to have been suggested before. Hence we may argue that it is more like worrying about the tea and coming up with a poisoning scenario than first recognizing that there exist enemies and then deducing they might have the capability to poison it.

The possibility that baryonic matter is unstable is however not too far-fetched: supersymmetric theories have proton decay as a common implication, and over very long timespans protons might decay due to tunneling into black holes that then evaporate [9]. Conversion to dark matter is similarly not too far-fetched given that most of the matter already is dark. Dark matter also has significantly more entropy ($\approx 10^{87}$ in total) than baryonic matter ($\approx 10^{81}$) [10], so if there exist a possible transition we should expect an entropic force driving conversion[9]. One could also argue that since Fermi's golden rule make

[9]However, as shown in [10], the entropy contribution to the universe from black holes and

transition rates (to a first approximation) proportional to the density of final states, if there are many of possible end states (presumably various states of rather light dark matter) then we have a reason to think the conversion is likely if/when it is allowed.

3.2 Context

An obvious target for doubt is the ill-defined non-thermal context of the scenario. Particle physics is mediated through basic interactions, typically seen as the vertices of Feynman diagrams. There is no apparent room for any "context" to enable the conversion. However, bare Feynman diagrams do hide a fair bit of extra considerations. For example, many particle interactions are disallowed if the energy is not high enough. More non-trivially, that interactions occur locally does not imply non-local features of the system cannot influence them. Consider the Aharonov-Bohm solenoid effect, where the presence of a magnetic field in a region where the particle wave function is negligible causes a macroscopically observable phase shift. In principle the whole interaction can be described in terms of local photon-particle interactions, but in practice the only reasonable description is looking at the macroscopic system. Another elementary example is a convex lens focusing a light beam: each of the local interactions between photons and electrons is totally innocuous, yet the whole system concentrates energy into a small region (and, at least in principle, for sufficient intensities, could produce black holes [36]).

The reason the Aharonov-Bohm effect and the lens work is of course that they are highly non-arbitrary configurations. Most configurations of matter and energy likely have no distinct effect in these cases. One may hence argue that the most likely situation will not produce a context that has a strongly specific effect. Unfortunately the situation in a particle accelerator is a fairly atypical environment that has deliberately been designed to cause certain interactions that are normally uncommon. The argument can still be made that since the accelerator is not set up to make dark fire occur it is a priori unlikely (just as an optical lens is not going to show the Aharov-Bohm effect). This, however, requires the assumption that the interactions necessary for dark fire are uncorrelated with the interactions the accelerator has been designed to promote: this whole line of reasoning hinges on claims about the dissimilarity of an unknown physical theory to known physics. It hence seems more robust to argue against it on other grounds.

3.3 The envelope of nature

One can argue that a "principle of plenitude" applies to the universe: if it can happen, it will happen (somewhere, somewhen). If we do not observe self-reinforcing macroscale disasters like vaccum decay, dark fire or replicating von Neumann probes we have a strong reason to suspect they cannot occur.

---

horizons is far larger. Transitions to black holes are obviously far more attractive if they occur.

Leaving aside the complications of observability and correcting for anthropic bias, this kind of argument seems reasonably solid. If you do something that is within the envelope of what happens in the universe normally and there are no observed dangerous processes linked to it, then this activity is likely safe. Given that the activity is already happening it can be viewed as just as benign as the normal activity of the universe.

There is nevertheless a problem with this argument: it is possible for humans to generate conditions that have not occured at any point earlier.

An example is the creation of the first experimental Einstein-Bose condensates in 1995 (and possibly superfluid helium in 1938). Up until that point there had not been any environments cold enough to achieve this state of matter: assuming no aliens beat us to it, this was the first time in the history of the universe when this phenomenon occured.

As far as we are aware, there was never any concern this could pose a risk, nor any credible theory for how. There was no evidence for risk from the past. There was no cause for alarm given the known laws of physics. Yet this lack of evidence does not argue against risk either: a person naive to nuclear physics withdrawing a control rod from a reactor may get a nasty surprise when the reaction shifts from subcritical to supercritical. We might have been similarly naive to some aspect of (say) quantum gravity that made EBCs collapse into black holes or Q-balls[10].

This consideration amplifies the Fermi question issue. We know there is some probability that intelligence gets reliably wiped out. We also know intelligence is likely to explore conditions not found in the natural universe. So a potential explanation could be that there is some threat in this exploration.

Estimating how risky it is to generate an outside of the envelope phenomenon is hard but not entirely impossible. After all, the EBC itself was successfully predicted. The prediction was based on extrapolating known physics to a situation outside past situations. Nuclear wars may not be a physical phenomenon that has ever occurred before, but we can predict nontrivial consequences such as nuclear winter with some confidence. The epistemically troubling possibilities are the possible existence of new physics that we do not have experience with, and the endless number of possible hypotheses that cannot be ruled out.

3.4 Ultraviolet cut-offs

Are there principled reasons to ignore exotic "small theories" [6]?

[10] As far as we know, Q-balls (solitons in minimal supersymmetric theories consisting of coherent states of squarks, sleptons and the Higgs fields) have not been studied much as a threat from physics experiments. They would, if they existed and were stable, disintegrate and absorb nucleons [19], and would split into further balls above a certain mass [29]. The main argument against Q-ball risk – besides the lack of direct observational evidence for them – is that cosmic Q-balls moving at $10^{-3}c$ would lose negligible kinetic energy when passing through a planet or star. However, neutron stars would be rapidly converted if big Q-balls existed, giving a bound on their mass and prevalence, especially if they are seen as a dark matter candidate [29].

Risk scenarios in domains of ignorance are vulnerable to a symmetry argument. For every scenario that a certain condition is dangerous there is another scenario that the condition is perfectly safe. If both are equally complex and equally likely a priori there is no reason to think there is anything to worry about – until some consideration breaks the symmetry. This can be a difference in complexity (e.g. if there are many more ways something could be risky than safe), likeliehood, or a major utility difference (e.g. existential risk vs. business as usual).

Most peculiar conditions are also very complex: Occham's razor militates against assigning them high probability. In fact, one can make a somewhat stronger argument. As we increase the allowed complexity of our risk scenarios the number of possibilities increase at least exponentially. Unless one believes that the total risk is *completely* dominated by bizarre complex risks, the sum of the probabilities of risks of a certain complexity range must stay bounded. This implies a prior assigning a roughly exponentially decreasing probability to complex risks.

There is a link here to the "Pascal's mugging" scenario, where a sequence of ever more unlikely but increasingly high-value possible actions are presented. A rational expectation maximizing agent in this situation would have to perform actions with exceedingly low probability of success [3]. However, one way out is to argue that the utility at stake is bounded. In the case of existential risk the total value of the future is very large but finite (at least to most consequentialists) [2], so possibilities below a certain threshold can be ignored or at least postponed until more pressing issues have been dealt with [7].

One can argue for bounded rationality to set the threshold: there is an infinite set of possibilities but we only have finite time, intelligence and effort to spend on them. So we should spend our available effort rationally to bring down the expected risk, starting from the least complex and most likely scenarios and moving outwards until we run out of time. There will be possibilities we have not considered, but we were rational in ignoring them since they were too a priori unlikely to matter. However, this leaves us with a meta-problem of estimating the right priority ordering among little understood scenarios.

Is the dark fire scenario a complex one? The basic conversion process is not too dissimilar to other physical interactions. The standard cosmological model(s) contains some rather intricate effects like the Higgs mechanism, inflation and the various kinds of dark matter. It seems that the basic dark fire scenario is relatively simple compared to theoretical physics theories that are generally regarded as reasonable. Catalytic particle interactions occur and can convert macroscopic amounts of matter, for example the CNO cycle in stellar fusion. A priori, the existence of an exothermic process that can catalytically convert protons into alpha particles seems fairly implausible[11].

[11]One can of course make an anthropic argument that something like the CNO cycle has to exist, since otherwise observers would not exist. However, this is no reason to believe the fusion pathways of stars have to have a catalytic loop. It might be enough to have steps like the triple-alpha process enabling the slow accumulation of heavier nuclei. In addition, the CNO cycle is mainly burning hydrogen, not building heavier elements. If the catalytic step

4 Conclusion

How much effort is rational to put into analyzing a scenario like this?

One way of reasoning would be to consider an acceptable risk level and estimate how many independent arguments are needed to reach it. Consider a risk less than 10^{-30} (which starts to be an acceptable risk for the sizes of potential future generations discussed in [2]). In this case, if each argument has 1% chance of being wrong, we need 15 independent arguments to feel confident that the chance of the scenario is less than 10^{-30}. In this paper we have perhaps a dozen, but few of them appear 99% certain.

It appears likely that developing more generic methods to establish the safety of systems like physics experiments would be effective. Instead of arguing against a specific family of scenarios like this, broader considerations could handle many scenarios. For example, the reasoning in [33] can rule out a large set of natural physics risks with certain properties, without regard to their particular details.

In the end, what matters is not putting out a particular (dark) fire, but constructing better overall fire safety.

4.1 Acknowledgments

We wish to thank Michael Vassar for bringing us together and participating in the discussion leading up to the paper, and Daniel Dewey for useful discussion and ideas.

References

[1] Adams, F. C., Kane, G. L., Mbonye, M., & Perry, M. J. (2001). Proton decay, black holes, and large extra dimensions. International Journal of Modern Physics A, 16(13), 2399-2410.

[2] Bostrom, N. (2003). Astronomical waste: The opportunity cost of delayed technological development. Utilitas, 15(03), 308-314.

[3] Bostrom, N. (2009). Pascal's mugging. Analysis, 69(3), 443-445.

[4] Bostrom, N. (2013). Existential risk prevention as global priority. Global Policy, 4(1), 15-31.

[5] Ćirković, M. M., Sandberg, A., & Bostrom, N. (2010). Anthropic shadow: Observation selection effects and human extinction risks. Risk analysis, 30(10), 1495-1506.

has an anthropic role, it would be through the higher fusion efficiency of larger stars enabling more heavy nuclei to form. However, it would again seem entirely possible that the right resonances could enable a non-catalytic fusion chain leading to heavy nuclei in such stars.

[6] Ćirković, M. M. (2012). Small Theories and Large RisksIs Risk Analysis Relevant for Epistemology?. Risk Analysis, 32(11), 1994-2004.

[7] Cotton-Barratt, O. (2015). Allocating risk mitigation across time, Future of Humanity Institute, Technical report #2015-2 `http://www.fhi.ox.ac.uk/Allocating-risk-mitigation.pdf`

[8] Dar, A., De Rújula, A., & Heinz, U. (1999). Will relativistic heavy-ion colliders destroy our planet?. Physics Letters B, 470(1), 142-148.

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

[10] Egan, C. A., & Lineweaver, C. H. (2010). A Larger Estimate of the Entropy of the Universe. The Astrophysical Journal, 710(2), 1825.

[11] Ellis, J., Giudice, G., Mangano, M., Tkachev, I., & Wiedemann, U. (2008). Review of the safety of LHC collisions. Journal of Physics G: Nuclear and Particle Physics, 35(11), 115004.

[12] EROS Collab., AA 469, 387 (2007);

[13] Genzel, R., Tacconi, L. J., Eisenhauer, F., Schreiber, N. F., Cimatti, A., Daddi, E., ... & Mignoli, M. (2006). The rapid formation of a large rotating disk galaxy three billion years after the Big Bang. Nature, 442(7104), 786-789.

[14] Giddings, S. B., & Mangano, M. L. (2008). Astrophysical implications of hypothetical stable TeV-scale black holes. Physical Review D, 78(3), 035009.

[15] Harvey, D., Massey, R., Kitching, T., Taylor, A., & Tittley, E. (2015). The nongravitational interactions of dark matter in colliding galaxy clusters. Science, 347(6229), 1462-1465.

[16] Hut, P. & M. J. Rees (1983). "How Stable Is Our Vacuum." Nature 302(5908): 508-509.

[17] Jaffe, R. L., W. Busza, et al. (2000). "Review of speculative "disaster scenarios" at RHIC." Reviews of Modern Physics 72(4): 1125-1140

[18] Jedamzik, K., & Pospelov, M. (2009). Big Bang nucleosynthesis and particle dark matter. New Journal of Physics, 11(10), 105028.

[19] Kusenko, A., Kuzmin, V., Shaposhnikov, M., & Tinyakov, P. G. (1997). Experimental signatures of supersymmetric dark-matter Q-balls. arXiv preprint hep-ph/9712212.

[20] Ord, T., Hillerbrand, R., & Sandberg, A. (2010). Probing the improbable: methodological challenges for risks with low probabilities and high stakes. Journal of Risk Research, 13(2), 191-205.

[21] Particle Data Group. (2014). July 2014 Particle Physics Booklet. http://pdg.lbl.gov/2014/download/rpp-2014-booklet.pdf

[22] Raffelt, G. G. (2011). Supernova neutrino observations: What can we learn?. Nuclear Physics B-Proceedings Supplements, 221, 218-229.

[23] Robitaille, T. P., & Whitney, B. A. (2010). The present-day star formation rate of the milky way determined from spitzer-detected young stellar objects. The Astrophysical Journal Letters, 710(1), L11.

[24] A. D. Sakharov (1967). "Violation of CP invariance, C asymmetry, and baryon asymmetry of the universe". Journal of Experimental and Theoretical Physics 5: 247., republished as A. D. Sakharov (1991). "Violation of CP invariance, C asymmetry, and baryon asymmetry of the universe". Soviet Physics Uspekhi 34 (5): 39293. Bibcode:1991SvPhU..34..392S. doi:10.1070/PU1991v034n05ABEH002497.

[25] Salucci, P., Swinbank, A. M., Lapi, A., Yegorova, I., Bower, R. G., Smail, I., & Smith, G. P. (2007). The baryonic and dark matter properties of high-redshift gravitationally lensed disc galaxies. Monthly Notices of the Royal Astronomical Society, 382(2), 652-656.

[26] Sarkar, A., Das, S., & Sethi, S. K. (2014). How Late can the Dark Matter form in our universe?. arXiv preprint arXiv:1410.7129.

[27] Schreiber, N. F., Genzel, R., Lehnert, M. D., Bouch, N., Verma, A., Erb, D. K., ... & Sternberg, A. (2006). SINFONI integral field spectroscopy of z 2 UV-selected galaxies: rotation curves and dynamical evolution. The Astrophysical Journal, 645(2), 1062.

[28] Schneider, A., Krauss, L., & Moore, B. (2010). Impact of dark matter microhalos on signatures for direct and indirect detection. Physical Review D, 82(6), 0635

[29] Shoemaker, I. M. (2009). Conversion of baryonic fermions into squarks in neutron stars from supersymmetric dark matter Q-balls. Physical Review D, 80(3), 031702.

[30] Sobral, D., Smail, I., Best, P. N., Geach, J. E., Matsuda, Y., Stott, J. P., ... & Kurk, J. (2012). A large Hα survey at z= 2.23, 1.47, 0.84 and 0.40: the 11 Gyr evolution of star-forming galaxies from HiZELS. Monthly Notices of the Royal Astronomical Society, sts096.

[31] Van Starkenburg, L., Van Der Werf, P. P., Franx, M., Labb, I., Rudnick, G., & Wuyts, S. (2008). Dynamical properties of a large young disk galaxy at z= 2.03. Astronomy & Astrophysics, 488(1), 99-112.

[32] Sutter, P. M., Lavaux, G., Wandelt, B. D., Weinberg, D. H., & Warren, M. S. (2014). The dark matter of galaxy voids. Monthly Notices of the Royal Astronomical Society, stt2425.

[33] Tegmark, M., & Bostrom, N. (2005). Astrophysics: Is a doomsday catastrophe likely?. Nature, 438(7069), 754-754.

[34] Turner, M. S., & Wilczek, F. (1982). Is our vacuum metastable?.

[35] W.-M. Yao et al. (Particle Data Group), Tests of Conservation Laws. J. Phys. G 33,1 (2006) (URL: http://pdg.lbl.gov)

[36] Yu, P. (2011). Dynamical formation of black holes due to the condensation of matter field. arXiv preprint arXiv:1105.5898.

[37] Witten, E. (1984). Cosmic separation of phases. Physical Review D, 30(2), 272.

[38] Zitrin, A., & Brosch, N. (2008). The NGC 672 and 784 galaxy groups: evidence for galaxy formation and growth along a nearby dark matter filament. Monthly Notices of the Royal Astronomical Society, 390(1), 408-420.