FINAL INTERNSHIP REPORT

UNIVERSITY COLLEGE GRONINGEN

September 2020 – January 2021

Emilie Aebischer
degree: Liberal Arts and Sciences
specialization: Humanities
s3698378
e.aebischer@student.rug.nl

supervisor: Simon Friederich

CONTENTS

INTRODUCTION

My internship happened rather accidentally and unexpectedly: I was not planning on doing an internship at the beginning of this year. However, I had some problems in fulfilling the credits I needed at UCG which would match my major/minor and as I was talking about it with my academic director, Simon Friederich, he was kind enough to offer me a great solution: an internship with him. Having only two days left to apply for the internship I honestly did not have much time to prepare myself or think much about what my expectations were beforehand. But this does not mean that I decided to take up his offer just because I needed the credits: not only was I genuinely interested in the topics he was working on, but it also seemed like a great opportunity to get a first insight into the academic world, as I am considering an academic career as a possible future path for me.

TASKS AND ACTIVITIES

Research for the Review of *The Precipice* by Toby Ord

After my first meeting with Simon Friederich , we decided that I would be helping with the writing of a review of *The Precipice: Existential Risks and the Future of Humanity* by Toby Ord, so I first started by finishing reading the latter. My first overall impressions of the book were that although Ord does make some rather compelling arguments, the incredibly high probability he assigns to an existential catastrophe happening this century seems rather unlikely to me. Given that he thinks most of this probability comes from anthropogenic risks I intuitively do not think it very likely that we live in the time period with the highest risk: if we keep making advances in technology and knowledge it seems to me that there will always be risks coming with these advances, risks that may get more and more important as these technologies get more and more powerful. Related to this, I also find the claim that we are among the most influential people that will ever live to be rather astounding: given the sheer enormity of the people that are still to come after us and the fact that we cannot really have any idea about what humanity will look like so far in the future, it seems very improbable that we would be so important. Of course, this does not mean that this is not an important time or that we should not bring more attention to existential risks, just that the claim that we are at such a unique time in humanity's potentially incredibly long future seems rather implausible to me. Finally, Ord loses me a bit when he starts talking about our grand potential and his grand strategy for humanity. It just all seems highly speculative to me, and it reads almost like science-fiction.

Once I was finished with the book, I started looking into how probabilities were assigned to existential risks. Given that existential risks have obviously not occurred before, we cannot base ourselves on past occurrences to calculate a probability. I do not understand everything yet, and this is still a work in progress but I already have a better idea of how it works.

In my second week, I spend a lot of time just looking around on the internet to try and get a grasp of the field. I started by looking at the research and talks on the Global Priorities Institute

webpage and then looked into some of the people that I came across there, such as William McAskill. I also started looking at some posts on the Effective Altruism Forum and created a twitter account in order to follow people who do research in the field of existential risks.

I watched a talk by Christian Tarnsey called *Can we predictably improve the far future?*[1] in which he presents three arguments which give us reason to be pessimistic about our ability to predict the effect our present choices will affect the far future, and three reasons to be optimistic. He basically concludes that more research needs to be done in this field, but that we cannot really be sure for now.

I also came across a debate between William McAskill and Toby Ord on whether we are in fact amongst the most influential people that have/will ever live. McAskill wrote a chapter titled *Are we living at the hinge of history?*[2] In which he presents two arguments against the claim we are among the most influential people that will ever live (HH). In short, his first argument is that our prior in HH should be very low and that the evidence we do not have sufficiently strong evidence in favour of it which would allow us to overcome this low prior. His second argument against HH is an inductive argument which states that: (P1) our influentialness has been increasing over time which is mainly due to increasing knowledge and opportunities (P2) we can expect our knowledge and opportunities to continue over the future (C) we should therefore expect the influentialness of future people to be greater. McAskill then gives two arguments in favour of HH - the fact that we are living on a single planet and that we are living at a time of unusually fast economic and technological progress - but argues that although these might mean that we are living at a highly influential time, it is still unlikely that we are among the most influential people. McAskill has a post[3] about this on the Effective Altruism Forum which prompted a small debate with Ord. Ord agrees with Mcaskill that the evidence is not strong enough if we assign such a low prior. However, he does not agree with McAskill's choice of prior. I have a feeling that this is an interesting and rather relevant discussion, but I still need to look more into it as I do not understand everything yet - I do not know much about how we set priors, or many of the technicalities and laws (such as Laplace's Law of Succession) which they are discussing.

I also read a working paper by Andreas Mogensen titled *Doomsday Rings Twice*[4] in which he argues that if we take it to be a priori very unlikely that we are among the most important people who will ever exist and the Hinge of History Hypothesis is true, then we should be rather confident that the human species will be wiped out in the coming centuries. He concludes that even if the Hinge of History Hypothesis is true, we can still refuse the conclusion if we revise our beliefs about our influentialness over the long-term future. Mogensen also formally discusses the arguments, but as with the Ord-McAskill debate, I am currently working on trying to understand his discussion of probabilities.

[1] Tarsney, C. [Centre for Effective Altruism]. (2019). *Can we predictably improve the far future?* [Video file]. Retrieved from https://www.youtube.com/watch?v=W5_bv_IjIcQ&feature=emb_title

[2] MacAskill, W. (forthcoming). Are We Living at the Hinge of History? In J. McMahan, T.Campbell, J. Goodrich & K. Ramakrishnan (Eds.), *Ethics and Existence: The Legacy of Derek Parfit*. Oxford: Oxford University Press.

[3] https://forum.effectivealtruism.org/posts/XXLf6FmWujkxna3E6/are-we-living-at-the-most-influential-time-in-history-1

[4] Mogensen, A. (2019). *Doomsday rings Twice*. GPI Working paper No. 1-2019

I spent a big part of my third week trying to get a better understanding of Bayesian probabilities, specifically how we use Bayes' theorem in order to update our prior confidence in a given hypothesis. This is where Simon Friederich's paper *Choosing Beauty*[5] really helped me a lot because it actually explains the steps explicitly. With this I was able to go back to Magensen's paper *Doomsday Rings Twice* and understand his formal discussion of the Doomsday Argument (DA) and his modification of it which he calls Doomsday Redux (DR) - both arguments are basically the same with the difference that DR is driven by consideration of our importance, not our birthrank as is the case for DA. I have to admit however, that I cannot say I entirely understand yet the criticisms of DA and the reasons for which they do not apply to DR according to Mogensen.

I was also able to better understand where William McAskill and Toby Ord disagree in the assigning of a prior for the Hinge of History Hypothesis (HH). McAskill had argued in *Are we living at the hinge of history?* that we should use a uniform prior given by the self-sampling-assumption (SSA), meaning that we would assign a very low prior probability to HH given that if we do not go extinct the future will plausibly entail a vast number of people: if we were to say that there are a trillion trillion people to come, then our prior confidence that we are among the million most influential people would be one in a million trillion. This obviously sets an extremely high bar for evidence.

Ord agrees that the evidence is not strong enough to overcome such a low prior. However, he believes that a uniform prior is the wrong choice in this case. Indeed, he argues that - given that the domain in this case is one that extends indefinitely in one direction - we should use a prior that diminishes as we move further away. Moreover, we should suspect the likelihood of a century being the most influential to decrease over time since there are certain events such as lock-ins or existential catastrophes which cannot occur a second time. He suggests using something like Jeffrey's prior, where the chance that an event happens in the nth period if it hasn't before is 1/2n+2. This choice of prior gives a prior chance of HH of about 2,5% - a lot higher than McAskill's one in a million trillion.

Ord and McAskill both develop many considerations in support of their choice of prior - given that there is not one choice of prior which is objectively better than the other it comes down to a matter of opinion. I am favouring McAskill's side who thinks that although he agrees a prior which diminishes over time might be better, any prior which makes the claims that (a) there is a single point in time where almost everything about the fate of the universe gets decided and (b) almost no one sees this but a very small fraction of the world is very implausible.

I also read Nick Bostrom's *Vulnerable World Hypothesis*[6]. I found his argument that the stabilizing of certain vulnerabilities would require ubiquitous-surveillance-powered preventive policing and effective global governance - a scenario which we usually consider to be dystopian - to be rather compelling. It is not really something that I would have thought

[5] Friederich, S. (2017). Choosing Beauty. *Logique et Analyse, 60*(240), 449-463. doi: 10.2143/LEA.240.03254092
[6] Bostrom, N. (2019). The vulnerable world hypothesis. *Global Policy, 10*(4), 455–476. doi : 10.1111/1758-5899.12718

about before. The thought that some existential risks would require a state of affairs which Ord considers to be an existential risk itself - namely a dystopia - is not a pleasant one. Although I wish I had an answer for Bostrom I am not quite sure how to argue against his point.

Writing the Review

After it was decided that I should start off with writing the review, I mostly spent time on drawing up a structure for the review. Going over my notes on the material I have read I first tried to pinpoint the main points I wanted to address in the section with our reflections/criticisms on Ord can came up with the following broad points:

- The claim that we are now at a precipice and that this is a highly unique feature of this century is astonishing: it does not seem to make much sense if we consider the potentially astronomical duration of humanity that Ord thinks we might reach.
- This is linked to the fact that Ord seems rather dramatic in many of his considerations. The high probabilities he assigns to risks such as AI or Pandemics do not seem very plausible, whereas his very optimistic conception of what we can do and what the future will hold are very speculative.
- In his insistence on the fact that we should keep all of our options open, Ord overlooks the fact that the stabilization of some risks might actually require us to restrict the possibilities open to humanity.

I then spent quite a lot of time trying to get a structure for the summary – I suppose that this is a good exercise for me as I have always found it difficult to try to concisely summarize an entire book. I took a rather systematic approach by simply going over each chapter and my notes on it and writing down the gist of it in a couple of bullet points. This allowed me to get a better grasp on the book as a whole and the flow of Ord's argumentation. Having not much experience with writing reviews, I have to admit that I was a little bit intimidated by the task which means that I was a bit more hesitant and slower at starting the actual writing than I usually am – when writing an essay I always take a little while once I am done with most of my research before I am able to put down my thoughts in writing but I struggled a bit more here than usually.

I started off the following week by finally writing down a big part of the review draft. I am not really happy with the result, mainly because I don't think that it reads really well but it is a start and I think that with some work it can turn into something good. My main problem is that I am not entirely sure what the review should look like so I don't feel very confident in what I have so far.

With the writing of my essay for Existential Risks I have also been grappling with a lot of questions which I have not entirely resolved yet around the issue of our limited knowledge when it comes to the future. Of course, I don't think we cannot say anything about the future but I feel like uncertainty is a bigger factor than it often plays in longtermist accounts. I am just not sure how big of a factor it is and what exactly kind of implications it would have. These

are just questions running through my mind which seem rather important to me but that I don't have a good grasp on yet.

Further Research into the Effective Altruism Community

Once the review was almost done, I started looking into work by Hilary Greaves. I read a paper she wrote with MacAskill called *The case for strong longtermism*[7] and watched a talk she gave about it[8]. In it they formulate the case for axiological strong longtermism which states that the best option is the one whose long-term effects are the best, instead of its immediate effects. They identify the main challenge to longtermism as being what they call the washing out hypothesis: the worry that the effect of our actions decay over time at such a rate that the effects of our action are mainly restricted to the short-term. I thought that this is a very interesting objection because I had mostly considered that we might not be able to know what the effects of our actions will be on the long-term future, and not that they might end up having no effect on it. But Greaves and MacAskill claim that while the washing-out hypothesis might be true for some trivial actions it does apply to many decision situations because there exist actions which we know can have long-term effects, such as accelerating progress and mitigating existential risk. Greaves and MacAskill make their case for axiological strong longtermism in terms of "a total utilitarian axiology" and "an expected utility treatment under uncertainty" but argue that it still holds up to plausible deviations of these theses. They end by suggesting that deontic strong longtermism might follow from axiological longtermism.

I have also been reading a paper by Greave titled *Population Axiology*[9] and started reading a paper she wrote with Ord titled *Moral uncertainty about population axiology*[10]. I am still working my way through them though: given that this is a completely new field to me, I don't have a complete grasp on the different theories in population axiology which Greaves lays out in the former. Although I think I understand the rough outline of the theories I am still working on getting a clear picture of their implications and problems that come with them.

The next week I continued looking into Hilary Greaves' work which mainly discusses uncertainty when it comes to population axiology and the long-term effects of altruist interventions. In her paper *Population Axiology,* she outlines the main theories in population axiology and shows that none is immune to serious objections: in fact, some impossibility theorems show that, for various sets of desiderata, it would be impossible to have a population axiology which fulfils all of the listed desiderata. Greaves suggests that if – as it seems to be the case – no theory can be fully satisfactory, then the choice between theories becomes a matter of what one is the least unwilling to give up.

---

[7] Greaves, H. & McAskill, W. (2019). *The case for strong longtermism.* GPI working paper No. 7-2019

[8] Greaves, H. [Global Priorities Institute – University of Oxford], (2019). *The case for longtermism* [video file]. Retrieved from: https://www.youtube.com/watch?v=Wz8lgjBLTpI&feature=emb_imp_woyt

[9] Greaves, H. (2017). Population axiology. *Philosophy Compass, 12*(11). doi : https://doi.org/10.1111/phc3.12442

[10] Greaves, H. & Ord, T. (2017). Moral uncertainty about population axiology. *Journal of Ethics and Social Philosophy, 12*(2). doi: https://doi.org/10.26556/jesp.v12i2.223

In light of this, Greaves wrote the paper *Moral uncertainty about population axiology* with Toby Ord in which they try to tackle the question of how to make decisions in certain domains where our actions have an effect on the population in light of axiological uncertainty. The question is thus which effective axiology one should use when making decisions. The approach that Greaves and Ord explore in the paper is the expected moral value (EMV) approach to axiological uncertainty. The latter aims at taking the same approach as for empirical uncertainty, namely using an effective axiology that "corresponds to the ordering of alternatives according to their expected value". This means that EMV ranks the available options according to the breadth of support they get from different theories as well as the magnitude each theory assigns to the stakes. Applying EMV to three different scenarios – adding a single person, extinction risk, and space colonization – Greaves and Ord show that in each case, when considering an "appropriately specified large-population limit", the alternative supported by Critical Level views such as the Total View have the highest expected moral value and swamp all other population axiologies.

I also watched a presentation given by Greave titled *Evidence, cluelessness and the long-term*[11] in which she discusses the limits of evidence when it comes to cost-effectiveness analyses. Most of the expected value of altruist interventions comes from the long-term future, but the evidence taken into account in cost-effectiveness analyses only tracks near-term effects, leaving us relatively clueless about the effect of our interventions on the further future. Greave outlines several responses to this cluelessness but states that she leans towards the longtermist response – meaning that instead of focussing on short-term interventions, we should focus on the kinds of interventions which we know will have a certain effect on the future. Of course, there too is much uncertainty, which is why Greaves stresses the need for global prioritites research.

Researching Sustainability

This week I attended Jonathan Symons' talk on ecomodernism. This was very interesting for me because it is not a view that I have encountered before: most of the people in my circles who are engaged against climate change stand quite firmly against industrialism and generally think in terms of degrowth and going back to nature. I cannot say that I personally have a set opinion on the question, but it was interesting to hear about a different view, and I think that Jonathan Symons definitely made some relevant arguments. I read his essay *In defense of Greta Thunberg*[12] and after the lecture I also went over some parts of his book *Ecomodernism: Technology, Politics and the Climate Crisis*[13] – his call for "state-funded, democratically controlled innovation" is certainly appealing and I am planning to take the time to read his whole book sometime when I am on holidays or have less to do for my classes.

---

[11] Greaves, H. [Centre for Effective Altruism], (2020). *Evidence, cluelessness and the long term* [video file]. Retrieved from: https://www.youtube.com/watch?t=123&v=fySZIYi2goY&feature=emb_imp_woyt
[12] https://thebreakthrough.org/journal/no-11-summer-2019/in-defense-of-greta-thunberg
[13] Symons, J. (2019). *Ecomodernism, technology, politics and the climate crisis.* Cambridge: Polity Press

I then started reading essays on the topic of sustainability, most of which were part of the *Routledge Handbook of the History of Sustainability*[14] edited by Jeremy L. Caradonna. In the second chapter, *Sustainabilty: a New Historiography,* Caradonna gives a brief overview of the different currents out of which grew an identifiable and publicly discussed concept of sustainability. He also lays out the 4 main ideas that historians of sustainability analze:

- The idea of the interconnectedness of human society, the ecomomy and the natural environment,
- The idea that there are certain ecological limits within which human societies must operate if they wish to persist into the far future,
- The idea of future-oriented planning: the inter-generational component,
- The idea that industrial society must get away from the logic of the big and the centralized if it is to survive in the long term.

In chapter 7, *Eternal Forest, Sustainable Use,* Ulrich Grober situates the origin of the word "sustainable" in German forestry. In the second part of the 17th century, many regions of Europe were increasingly worried about experiencing a severe wood shortage: the idea of a perpetual forest, a forest which would provide a steady yield over an infinite number of years, became a central concern in forestry. The term "nachhaltig" was coined for the first time by Hans Carl von Carlowitz in his *Sylvicultura Oeconomica* in 1713. The overarching argument von Carlowitz was trying to make, is that the consumption of wood should remain whithn the limtis of what "der Wald-Raum zu zeugen un zu tragen vermag". His term "nachhaltig" not only provided the blueprint for the modern word "sustainability" but according to Grober also provided the triple bottom line of sustainability: a consideration of ecology, economy (a morally based economy in harmony with the natural environment) and social justice (responsibility for future generations).

In the 4th chapter of the *Routledge Handbook of the History of Sustainability* titled *Understanding Sustainabilty through History*, Joseph A. Tainter argues that against the commonly held assumption, it is in general impossible for a society to voluntarily reduce its consumption of resources over the long term by studying the cases of the Roman and the Byzantine Empire. Tainter argues that problem-solving generates complexity and increase in complexity is costly because it takes energy to maintain a system away from equilibrium: because human societies are becoming ever more complex as a result of problem-solving it would therefore be impossible to reduce resource consumption as it will need more and more energy to sustain said societies.

I decided to read Ulrich Grober's book *Sustainability: a Cultural History*[15] as well as Caradonna's book *Sustainability: a History*[16] because these were cited in most of the papers I came across which discussed the history of the concept of sustainability. This allowed me to get more in-depth knowledge on the history of the concept and gave me a solid foundation. However, it did take me more time to get through the books than I thought it would, which means that the drawback was that I was not able to get as much of a variety of viewpoints as I would perhaps have wanted if I had had more time. Nevertheless, I was able to prepare a

---

[14] Caradonna, J. (2017). *Routledge Handbook of the History of Sustainability*. London: Routledge
[15] Grober, U (2012). *Sustainability: a cultural history*. Totnes: Green Books
[16] Caradonna, J. (2014). *Sustainability: a history.* New York: Oxford University Press

presentation for Simon Friederich and Jonathan Symons which gave an overview of the origin and the development of the word.

The Question of Effective Altruism and Sustainability

After I was done with the topic of the history of sustainability, I started looking into the question of why the effective altruism community – and particularly longtermism – does not seem to identify itself with the field of sustainability. My hunch is that in the end it might come down to their conception of growth and progress. So while there is actually considerable overlap between the two fields – especially when it comes to more concrete matters such as the kinds of policies that they would recommend and technologies they seek to develop – and sustainability might be considered a condition for longtermism, longtermists generally favor growth, which is a much more contentious issue amongst sustainists. Typically the industrial revolution for example is seen in a more positive light by longtermists, while sustainists are more prone to argue its negative consequences outweigh the positives. Of course this is very caricatural, but I think there is some truth to it. Ultimately, I would argue that the two sides have somewhat different goals: as the name indicates sustainability is principally about sustaining – meaning ensuring the continuation –, whereas longtermism has a perhaps slightly more ambitious focus on improving and maximizing good. These observations are however still very tentative and I'm still trying to figure out much of what I've said here, but these are my current thoughts on the question.

I decided to look into the definitions given to effective altruism by different organizations and academics – McAskill being the main academic working on a definition of effective altruism which is used for example by 80000hours. Here are a couple definitions:
- *Centre for effective altruism[17]:*
  "Effective altruism is about using evidence and reason to figure out how to benefit others as much as possible, and taking action on that basis"
- *McAskill – The definition of effective altruism[18]:*
  "Effective altruism is:
  (i)   the use of evidence and careful reasoning to work out how to maximize the good with a given unit of resources, tentatively understanding 'the good' in impartial welfarist terms, and
  (ii)   the use of the findings from (i) to try to improve the world."
- *McAskill – Effective altruism[19]*
  "As defined by the leaders of the movement, effective altruism is the use of evidence and reason to work out how to benefit others by as much as possible, and the taking action on that basis. So defined, effective altruism is a project, rather than a set of normative commitments. It is both a research project — to figure out how to do the

[17] https://www.centreforeffectivealtruism.org/
[18] https://drive.google.com/file/d/1rQu75k8uMFpdsp1y3JWlHP6kev3T-97N/view
[19] https://globalprioritiesinstitute.org/william-macaskill-effective-altruism/

most good — and a practical project, of implementing the best guesses we have about how to do the most good."

- *effectivealtruism.org*[20] :
  "[Effective altruism] is a research field which uses high-quality evidence and careful reasoning to work out how to help others as much as possible. It is also a community of people taking these answers seriously, by focusing their efforts on the most promising solutions to the world's most pressing problems."
  "three main areas: alleviating global poverty, improving animal welfare, and trying to influence the long-term future."

Looking at how effective altruism is defined, I think that it becomes clear that effective altruism and the field of sustainability are different in kind: sustainists make normative claims, whereas effective altruism is a non-normative project. As is also pointed out by McAskill, effective altruism makes claims about how to bring about the most good, but it does not tell us what we ought to do or that we should want to do the most good possible. Furthermore, although effective altruism is often thought to be mainly about alleviating poverty and it actually does have three main areas (animal welfare, global health and development and existential risk), this is only due to the fact that these are the areas in which it was identified that the most good could be done: effective altruism does not start off with a central focus. Sustainability on the other hand does have a much narrower focus and makes a normative claim: we ought to meet the needs of the present without compromising the future, we ought to use our resources in a sustainable manner.

Another factor in explaining why effective altruists do not seem too concerned with sustainability could be the three criterions they commonly use in order to decide which issues to focus on: importance, tractability and neglectedness. Sustainability is perhaps not exactly a neglected issue: although people involved in the climate change movement might argue that it is still too neglected, it does receive an important amount of attention and has many people – be it scientists, policymakers, etc. – working on it. I think that effective altruists might think climate change receives enough attention as it is, and that their attention would be better used on other, more neglected issues.

However, although it is perhaps not the most prominent issue, there are effective altruists working on climate change. Johannes Ackva gave a talk titled *An update to our thinking on climate change*[21] in 2020, in which he gives a brief account of the effective altruism community's outlook on climate change. Interestingly, the word "sustainability" does not come up once in his talk. Ackva tells us that "climate change is mostly about solving an energy problem", with climate change being part of a triple challenge together with air pollution and energy poverty. I think that this is perhaps where a major difference between sustainists and the effective altruism community lies: sustainists usually consider climate change a symptom of our capitalist consumer society that is not functioning properly, whereas effective altruists see it primarily as question of energy consumption. I think that even though effective altruists do believe that the way people are living their life today is immensely problematic – hence

[20] https://www.effectivealtruism.org/articles/introduction-to-effective-altruism/
[21] Ackva, J. [Centre for Effective Altruism] (2020). *An update to our thinking on climate change* [video file]. Retrieved from: https://www.youtube.com/watch?v=yfBzoCPQ8Kc&feature=youtu.be

their advocacy for giving to charities – they do not approach climate change as inherently linked to a malfunctioning society. This is why effective altruists consider the top charity dealing with climate change to be the Clean Air Task Force, which focuses on neglected technologies, innovation and advocacy: reducing our consumption or returning to a "simpler way of life" are not really considered as serious option by effective altruists. The way forward is not in reducing our energy consumption, but by finding technologies that will help us reduce the carbon emissions while still consuming as much (and even more) energy as we do today.

Unfortunately, the internship came to an end before I was able to fully explore the question of sustainability and the effective altruist community. I also lost some time when I started looking into the question because I was not entirely sure what I was looking for, which means that I lost some time reading articles that were not directly relevant. However, I think that I did end up on the right track, with several factors potentially explaining the separation between the two communities. I mainly looked into the work of effective altruists however, and if I had more time, I would want to look more into works on sustainability as I only have a rather superficial knowledge of the work that is being done in the field today. I therefore think I might have an idea of why effective altruism does not speak much of sustainability, but I have not looked into what sustainists would think of the effective altruism community.

REFLECTIONS ON THE FUTURE

Simon Friederich also encouraged me to take some time allocated for the internship to think about what I want to do once I am finished with my bachelor at UCG. I am trying to figure out what I want to do but it is difficult because I can become interested in pretty much anything within the Humanities if I have to work on it for a course – or internship in this case. The problem is that I find it hard to know if it is something that I would actually want to do further than that. The main reason I have History and Religious studies in mind is that those are subjects that I spend time on in my free time as well by reading books or listening to podcasts for example – I think I might therefore have more intrinsic motivation to study them but I also know that I can get pretty passionate about a lot of different things. Before this internship, I had never seriously considered studying philosophy as main subject: although I have always enjoyed philosophy, I have always considered it to simply be complementary to my main subject. However, I have now started to properly consider it as an option but am still unsure if it is really the path I would want to follow. And if I was to decide to study Philosophy, I am also hesitant about what area of philosophy I would want to study. In any case, I am actually seriously considering the philosophy master in Groningen, and given that the deadline for applications is quite late I still have some time to try to make up my mind.

I think that a lot of the uncertainty I have at the moment stems from the fact that I feel like UCG did not allow me to really explore the directions I would have wanted to. I am certainly not regretting my choice of coming to UCG, but UCG does not offer many options when it comes to humanities courses, which means that I often simply had to take whatever was available and not necessarily what I would have wanted to. This is of course not only a bad thing: it also means I got to explore areas that I would not have thought to otherwise – I

certainly would not have done as much philosophy as I have done now. But it also means that I am still wondering about whether I would not rather study a more historical discipline.

Over the Christmas holidays I spent quite a lot of time looking at different masters. What I noticed, is that I often do not meet the requirements: for many philosophy masters I do not have the necessary amount of credits in philosophy, for history masters I do not have the necessary amount of credits in history, and so on. Combined with the fact that I am currently not really feeling like starting a master right away – especially if I am unsure of my choice – this made me think about adding a semester or two studying at the bachelor level. I am currently considering simply studying at the bachelor level for a semester or two at a university in Germany or Austria. I was supposed to go to Vienna for an exchange before it got cancelled because of corona, and I still would like to study for a little while in a German-speaking country. My mother is from Germany and I speak German as my mother tongue, but my academic and written German is pretty bad, which is why I would for example be too insecure to start a master in German. I therefore think that studying for a while in a German-speaking country would be a good solution to give me a little bit more time to think about what I want to do, perhaps explore a different direction, gain more credits, and improve my German. Given that the deadlines are very late in Germany, I can always decide for this last minute if I have not found something I would rather like to do.

CONCLUSION

This internship made me discover the effective altruism community and the work associated with it, which introduced me to new challenges I had not really thought too much about before, such as existential. In some ways it even changed how I think about my own future. Indeed, although I am not entirely sure where I would situate myself with regard to effective altruism, it has made me think more about what it is that I would like to achieve with my own work. The question of how to bring about good was not really one I considered to be prior to my choice of future career: it was rather something to think about once I had made my choice. Although I do not know yet how it will influence my choice, I am now definitely thinking about it, about what it is that I want to achieve. What exactly will come out of it is something that I will only be able to tell in hindsight.

Finally, I believe that this internship has taught me a lot. It was challenging because Simon Friederich was encouraging me to work rather independently by giving me quite a lot of freedom when it came to what I wanted to focus on during my research as well as giving me responsibility when writing the review for example: I often felt a little bit insecure, but by managing to go through the different tasks I think that it helped me gain more confidence in my capabilities and the expression of my opinions. I also got a first – very positive – glance into the academic world and had the great opportunity to have a first publication. I can definitely say that I very much enjoyed the internship and am grateful for what it has brought me.