

What's wrong with human extinction?

Elizabeth Finneron-Burns^{a,b} 

^aUniversity of Warwick, Politics & International Studies, Coventry, UK; ^bInstitute for Futures Studies, Stockholm, Sweden

ABSTRACT

This paper explores what could be wrong with the fact of human extinction. I first present four reasons why we might consider human extinction to be wrong: (1) it would prevent millions of people from being born; (2) it would mean the loss of rational life and civilization; (3) it would cause existing people to suffer pain or death; (4) it would involve various psychological traumas. I argue that looking at the question from a contractualist perspective, only reasons (3) and (4) are admissible. I then consider what implications this limitation on reasons has for the wrongfulness of various forms of human extinction.

ARTICLE HISTORY Received 15 January 2016; Accepted 29 December 2016

KEYWORDS Human extinction; contractualism; T.M. Scanlon; intergenerational ethics; population size

1. Introduction

There is any number of potential future population sizes, one of which is zero. That is, we could decide not to create any more humans at all. There seems to be a fairly widespread thought that it would be undesirable for humans to become extinct as evidenced by the amount of research undertaken that worries about such an outcome. The Future of Humanity Institute at the University of Oxford, for example, states one of its aims to be research into interventions that would reduce the risk of outcomes that 'would end Earth-originating intelligent life' (Institute 2016). Similarly, the Centre for the Study of Existential Risk at the University of Cambridge describes its goal as to 'steer a small fraction of Cambridge's great intellectual resources, and of the reputation built on its past and present scientific pre-eminence, to the task of ensuring that our own species has a long-term future' (Risk 2016). Clearly, a significant quantity of financial and intellectual resources is being invested in the prevention or prolonging of human extinction. Presumably, this is because people see something wrong with the idea that we might cause or allow the human species to become extinct.

CONTACT Elizabeth Finneron-Burns  e.finneron-burns@warwick.ac.uk

© 2017 Canadian Journal of Philosophy

However, there is little work being done to explain what would be wrong with causing or allowing humans to go extinct.

Human extinction could come about in anthropogenic (e.g. climate change or a decision not to procreate) and/or non-anthropogenic (e.g. an asteroid hits the Earth) ways. We can also either directly cause extinction or passively allow it to occur, either intentionally or unintentionally. We could, for example, decide that we no longer want to procreate — perhaps we buy into David Benatar's view that because all lives contain at least a degree of harm, we ought not to procreate at all (Benatar 2006). Eventually, humans would go extinct because we chose to follow a course of action that we knew would and intended to cause extinction. We might also directly contribute to extinction by failing to mitigate climate change even if extinction was not the intended consequence of our actions. Even facing the risk of a completely non-anthropogenic cause of extinction, there may be things that humans can do to prevent it occurring and if we do not, we could be said to have allowed extinction to occur.

Since human extinction could, in many cases, be brought about by human activities or at least passively be allowed to progress, it therefore makes most sense to ask about its potential moral *wrongness*. This paper seeks to answer the question of whether or not it is morally permissible to cause or allow human extinction to occur. Put another way, under what (if any) conditions would people causing or allowing the extinction of the human race be wrong?¹

I will consider the wrongness of human extinction from the perspective of T.M. Scanlon's contractualism (Scanlon 1998). This is for two reasons. The first is that I believe his theory is a convincing account of moral wrongness. Second, the application of contractualism to the area of future generations remains largely unexplored so approaching the question from this angle will also help us learn more about contractualism in an intergenerational context.

The essence of Scanlon's contractualism is contained in his well-known statement: 'an act is wrong if its performance under the circumstances would be disallowed by any set of principles for the general regulation of behaviour that no one could reasonably reject as a basis for informed, unforced general agreement' (Scanlon 1998, 153). We wrong others when we fail to consider their interests in our moral deliberations and do not give them the respect they deserve by virtue of being rational people. This happens when we cannot justify our actions to them using acceptable reasons or when we act according to a principle that they could reject for similarly acceptable reasons.

Scanlon's formulation is a type of person-affecting theory. By person-affecting, I mean that what is important is the effect of principles/actions on persons, rather than the world writ large.² Therefore, acceptable reasons must be, according to Scanlon, *personal*. Personal reasons 'have to do with the claims and status of individuals in certain positions' (Scanlon 1998, 219). In other words, they are grounds 'that a person may find objectionable because of their bearing upon her being able to live a rationally self-governed, meaningful life' (Kumar 2003, 9,

fn. 6). Personal reasons include, *inter alia*, considerations of well-being, rights or fair treatment. In order to be a reason either to justify a principle or to reject one, the reason must be personal — it must have an impact on a person or persons.

There are, of course, other reasons people might have for wanting to act in certain ways that do not bear on an impact on persons. Impersonal reasons are one such type. Scanlon understands impersonal reasons to ‘flow from the value of objects themselves, not (at least in the first instance) from anything having to do with my relation to other people’ (Scanlon 1998, 221) and ‘are not tied to the well-being, claims, or status of individuals’ (Scanlon 1998, 219). They may include considerations such as an appeal to the aggregate value of an outcome (e.g. total utility) and the intrinsic value of equality (to name just two examples). In the contractualist formula, since impersonal reasons by definition do not stem from impact on persons, this means that they cannot on their own provide reasons to reject a principle. Whilst Scanlon is not denying that impersonal considerations may be important to people, in order to be relevant to the rejectability of a principle, they must give rise to personal reasons. For example, assuming that non-human animals are not persons, their pain and suffering is not a personal reason to reject a principle permitting it. However, a person could have a personal reason to reject a principle permitting the pain and suffering of animals if it prevented her from living a life consistent with the impersonal values (the well-being of animals) that she finds to be important in her life. So impersonal values cannot on their own provide reasons to reject principles, but they can lead to personal reasons if a principle forbids that person from living a life consistent with those values (Scanlon 1998, 218–223).

2. What could be wrong with human extinction?

Many, though certainly not all, people might believe that it would be wrong to bring about the end of the human species, and the reasons given for this belief are various. I begin by considering four reasons that could be given against the moral permissibility of human extinction. I will argue that only those reasons that impact the people who exist at the time that the extinction or the knowledge of the upcoming extinction occurs, can explain its wrongness. I use this conclusion to then consider in which cases human extinction would be morally permissible or impermissible, arguing that there is only a small class of cases in which it would not be wrong to cause the extinction of the human race or allow it to happen.

2.1. It would prevent the existence of very many happy people

One reason of human extinction might be considered to be wrong lies in the value of human life itself. The thought here might be that it is a good thing for people to exist and enjoy happy lives and extinction would deprive more people

of enjoying this good. The 'good' in this case could be understood in at least two ways. According to the first, one might believe that you benefit a person by bringing them into existence, or at least, that it is good *for that person* that they come to exist. The second view might hold that if humans were to go extinct, the utility foregone by the billions (or more) of people who could have lived but will now never get that opportunity, renders allowing human extinction to take place an incidence of wrongdoing.

An example of this view can be found in two quotes from an *Effective Altruism* blog post by Peter Singer, Nick Beckstead and Matt Wage:

One very bad thing about human extinction would be that billions of people would likely die painful deaths. But in our view, this is by far not the worst thing about human extinction. The worst thing about human extinction is that there would be no future generations.

Since there could be so many generations in our future, the value of all those generations together greatly exceeds the value of the current generation. (Beckstead, Singer, and Wage 2013)

The authors are making two claims. The first is that there is value in human life and also something valuable about creating future people which gives us a reason to do so; furthermore, it would be a very bad thing if we did not do so. The second is that, not only would it be a bad thing for there to be no future people, but it would actually be the *worst* thing about extinction. Since happy human lives have value, and the number of potential people who could ever exist is far greater than the number of people who exist at any one time, even if the extinction were brought about through the painful deaths of currently existing people, the former's loss would be greater than the latter's. Both claims are assuming that there is an intrinsic value in the existence of potential human life. The second claim makes the further assumption that the forgone value of the potential lives that could be lived is greater than the disvalue that would be accrued by people existing at the time of the extinction through suffering from painful and/or premature deaths.

The best-known author of the post, Peter Singer is a prominent utilitarian, so it is not surprising that he would lament the potential lack of future human lives per se. However, it is not just utilitarians who share this view, even if implicitly. Indeed, other philosophers also seem to imply that they share the intuition that there is just something wrong with causing or failing to prevent the extinction of the human species such that we prevent more 'people' from having the 'opportunity to exist'. Stephen Gardiner (2009) and Martin O'Neill (personal correspondence), both sympathetic to contract theory, for example, also find it intuitive that we should want more generations to have the opportunity to exist, assuming that they have worth-living lives, and I find it plausible to think that many other people (philosophers and non-philosophers alike) probably share this intuition.

When we talk about future lives being 'prevented', we are saying that a possible person or a set of possible people who could potentially have existed will now never actually come to exist. To say that it is wrong to prevent people from existing could either mean that a possible person could reasonably reject a principle that permitted us not to create them, or that the foregone value of their lives provides a reason for rejecting any principle that permits extinction.

To make the first claim we would have to argue that a possible person could reasonably reject any principle that prevented their existence on the grounds that it prevented *them* in particular from existing. However, this is implausible for two reasons.

First, we can only wrong someone who did, does or will actually exist because wronging involves failing to take a person's interests into account. When considering the permissibility of a principle allowing us not to create Person *X*, we cannot take *X*'s interest in being created into account because *X* will not exist if we follow the principle. By considering the standpoint of a person in our deliberations we consider the burdens they will have to bear as a result of the principle. In this case, there is no one who will bear any burdens since if the principle is followed (that is, if we do not create *X*), *X* will not exist to bear any burdens. So, only people who do/will actually exist can bear the brunt of a principle, and therefore occupy a standpoint that is owed justification.

Second, existence is not an interest at all and a possible person is not disadvantaged by not being caused to exist. Rather than being an interest, it is a necessary requirement in order to have interests. Rivka Weinberg describes it as 'neutral' because causing a person to exist is to create a subject who can have interests; existence is not an interest itself.³ In order to be disadvantaged, there must be some detrimental effect on your interests. However, without existence, a person does not have any interests so they cannot be disadvantaged by being kept out of existence. But, as Weinberg points out, 'never having interests itself could not be contrary to people's interests since without interest bearers, there can be no 'they' for it to be bad for' (Weinberg 2008, 13). So, a principle that results in some possible people never becoming actual does not impose any costs on those 'people' because nobody is disadvantaged by not coming into existence.⁴

It therefore seems that it cannot be wrong to fail to bring particular people into existence. This would mean that no one acts wrongly when they fail to create another person. Writ large, it would also not be wrong if everybody decided to exercise their prerogative not to create new people and potentially, by consequence, allow human extinction.

One might respond here by saying that although it may be permissible for one person to fail to create a new person, it is not permissible if everyone chooses to do so because human lives have value and allowing human extinction would be to forgo a huge amount of value in the world. This takes us to the second way of understanding the potential wrongness of preventing people

from existing — the foregone value of a life provides a reason for rejecting any principle that prevents it.

One possible reply to this claim turns on the fact that many philosophers acknowledge that the only, or at least the best, way to think about the value of (individual or groups of) possible people's lives is in impersonal terms (Parfit 1984; Reiman 2007; McMahan 2009). Jeff McMahan, for example, writes 'at the time of one's choice there is no one who exists or will exist independently of that choice for whose sake one could be acting in causing him or her to exist ... it seems therefore that any reason to cause or not to cause an individual to exist ... is best considered an impersonal rather than individual-affecting reason' (McMahan 2009, 52).

Another reply along similar lines would be to appeal to the value that is lost or at least foregone when we fail to bring into existence a next (or several next) generations of people with worth-living lives. Since *ex hypothesi* worth-living lives have positive value, it is better to create more such lives and worse to create fewer. Human extinction by definition is the creation of no future lives and would 'deprive' billions of 'people' of the opportunity to live worth-living lives. This might reduce the amount of value in the world at the time of the extinction (by killing already existing people), but it would also prevent a much vaster amount of value in the future (by failing to create more people).

Both replies depend on the impersonal value of human life. However, recall that in contractualism impersonal values are not on their own grounds for reasonably rejecting principles. Scanlon himself says that although we have a strong reason not to destroy existing human lives, this reason 'does not flow from the thought that it is a good thing for there to be more human life rather than less' (104). In contractualism, something cannot be wrong unless there is an impact on a *person*. Thus, neither the impersonal value of creating a particular person nor the impersonal value of human life writ large could *on its own* provide a reason for rejecting a principle permitting human extinction.

It seems therefore that the fact that extinction would deprive future people of the opportunity to live worth-living lives (either by failing to create either particular future people or future people in general) cannot provide us with a reason to consider human extinction to be wrong. Although the lost value of these 'lives' itself cannot be the reason explaining the wrongness of extinction, it is possible the knowledge of this loss might create a personal reason for some existing people. I will consider this possibility later on in section (d). But first I move to the second reason human extinction might be wrong *per se*.

2.2. It would mean the loss of the only known form of intelligent life and all civilization and intellectual progress would be lost

A second reason we might think it would be wrong to cause human extinction is the loss that would occur of the only (known) form of rational life and

the knowledge and civilization that that form of life has created. One thought here could be that just as some might consider it wrong to destroy an individual human heritage monument like the Sphinx, it would also be wrong if the advances made by humans over the past few millennia were lost or prevented from progressing. A related argument is made by those who feel that there is something special about humans' capacity for rationality which is valuable in itself. Since humans are the only intelligent life that we know of, it would be a loss, in itself, to the world for that to end.

I admit that I struggle to fully appreciate this thought. It seems to me that Henry Sidgwick was correct in thinking that these things are only important insofar as they are important to humans (Sidgwick 1874, I.IX.4).⁵ If there is no form of intelligent life in the future, who would there be to lament its loss since intelligent life is the only form of life capable of appreciating intelligence? Similarly, if there is no one with the rational capacity to appreciate historic monuments and civil progress, who would there be to be negatively affected or even notice the loss?⁶

However, even if there is nothing special about human rationality, just as some people try to prevent the extinction of nonhuman animal species, we might think that we ought also to prevent human extinction for the sake of biodiversity. The thought in this, as well as the earlier examples, must be that it would somehow be bad for the world if there were no more humans even though there would be no one for whom it is bad. This may be so but the only way to understand this reason is impersonally.

Since we are concerned with wrongness rather than badness, we must ask whether something that impacts no one's well-being, status or claims can be wrong. As we saw earlier, in the contractualist framework reasons must be personal rather than impersonal in order to provide grounds for reasonable rejection (Scanlon 1998, 218–223). Since the loss of civilization, intelligent life or biodiversity are per se impersonal reasons, there is no standpoint from which these reasons could be used to reasonably reject a principle that permitted extinction. Therefore, causing human extinction on the grounds of the loss of civilization, rational life or biodiversity would not be wrong.

2.3. Existing people would endure physical pain and/or painful and/or premature deaths

Thinking about the ways in which human extinction might come about brings to the fore two more reasons it might be wrong. It could, for example, occur if all humans (or at least the critical number needed to be unable to replenish the population, leading to eventual extinction) underwent a sterilization procedure. Or perhaps it could come about due to anthropogenic climate change or a massive asteroid hitting the Earth and wiping out the species in the same way it did the dinosaurs millions of years ago. Each of these scenarios would

involve significant physical and/or non-physical harms to existing people and their interests.

Physically, people might suffer premature and possibly also painful deaths, for example. It is not hard to imagine examples in which the process of extinction could cause premature death. A nuclear winter that killed everyone or even just every woman under the age of 50 is a clear example of such a case. Obviously, some types of premature death themselves cannot be reasons to reject a principle. Every person dies eventually, sometimes earlier than the standard expected lifespan due to accidents or causes like spontaneously occurring incurable cancers. A cause such as disease is not a moral agent and therefore it cannot be wrong if it unavoidably kills a person prematurely.

Scanlon says that the fact that a principle would reduce a person's well-being gives that person a reason to reject the principle: 'components of well-being figure prominently as grounds for reasonable rejection' (Scanlon 1998, 214). However, it is not settled yet whether premature death is a setback to well-being. Some philosophers hold that death is a harm to the person who dies, whilst others argue that it is not.⁷ I will argue, however, that regardless of who is correct in that debate, being caused to die prematurely can be reason to reject a principle when it fails to show respect to the person as a rational agent.

Scanlon says that recognizing others as rational beings with interests involves seeing reason to preserve life and prevent death: 'appreciating the value of human life is primarily a matter of seeing human lives as something to be respected, where this involves seeing reasons not to destroy them, reasons to protect them, and reasons to want them to go well' (Scanlon 1998, 104). The 'respect for life' in this case is a respect for the person living, not respect for human life in the abstract. This means that we can sometimes fail to protect human life without acting wrongfully if we still respect the person living. Scanlon gives the example of a person who faces a life of unending and extreme pain such that she wishes to end it by committing suicide. Scanlon does not think that the suicidal person shows a lack of respect for her own life by seeking to end it because the person whose life it is has no reason to want it to go on. This is important to note because it emphasizes the fact that the respect for human life is person-affecting. It is not wrong to murder because of the impersonal disvalue of death in general, but because taking someone's life without their permission shows disrespect to *that person*. This supports its inclusion as a reason in the contractualist formula, regardless of what side ends up winning the 'is death a harm?' debate because even if death turns out not to harm the person who died, ending their life without their consent shows disrespect to that person.

A person who could reject a principle permitting another to cause his or her premature death presumably does not wish to die at that time, or in that manner. Thus, if they are killed without their consent, their interests have not been taken into account, and they have a reason to reject the principle that allowed

their premature death.⁸ This is as true in the case of death due to extinction as it is for death due to murder.

However, physical pain may also be caused to existing people without killing them, but still resulting in human extinction. Imagine, for example, surgically removing everyone's reproductive organs in order to prevent the creation of any future people. Another example could be a nuclear bomb that did not kill anyone, but did painfully render them infertile through illness or injury. These would be cases in which physical pain (through surgery or bombs) was inflicted on existing people and the extinction came about as a result of the painful incident rather than through death. Furthermore, one could imagine a situation in which a bomb (for example) killed enough people to cause extinction, but some people remained alive, but in terrible pain from injuries.

It seems uncontroversial that the infliction of physical pain could be a reason to reject a principle. Although Scanlon says that an impact on well-being is not the only reason to reject principles, it plays a significant role, and indeed, most principles are likely to be rejected due to a negative impact on a person's well-being, physical or otherwise.

It may be queried here whether it is actually the involuntariness of the pain that is grounds for reasonable rejection rather than the physical pain itself because not all pain that a person suffers is involuntary. One can imagine acts that can cause physical pain that are not rejectable — base jumping or life-saving or improving surgery, for example. On the other hand, pushing someone off a cliff or cutting him with a scalpel against his will are clearly rejectable acts. The difference between the two cases is that in the former, the person having the pain inflicted has consented to that pain or risk of pain. My view is that they cannot be separated in these cases and it is involuntary physical pain that is the grounds for reasonable rejection. Thus, the fact that a principle would allow unwanted physical harm gives a person who would be subjected to that harm a reason to reject the principle.

Of course the mere fact that a principle causes involuntary physical harm or premature death is not sufficient to declare that the principle is rejectable — there might be countervailing reasons. In the case of extinction, what countervailing reasons might be offered in favour of the involuntary physical pain/death-inducing harm? One such reason that might be offered is that humans are a harm to the natural environment and that the world might be a better place if there were no humans in it. It could be that humans might rightfully be considered an all-things-considered hindrance to the world rather than a benefit to it given the fact that we have been largely responsible for the extinction of many species, pollution and, most recently, climate change which have all negatively affected the natural environment in ways we are only just beginning to understand. Thus, the fact that human extinction would improve the natural environment (or at least prevent it from degrading further), is a countervailing reason in favour of extinction to be weighed against the reasons held by humans

who would experience physical pain or premature death. However, the good of the environment as described above is by definition not a personal reason. Just like the loss of rational life and civilization, therefore, it cannot be a reason on its own when determining what is wrong and countervail the strong personal reasons to avoid pain/death that is held by the people who would suffer from it.⁹ Every person existing at the time of the extinction would have a reason to reject that principle on the grounds of the physical pain they are being forced to endure against their will that could not be countervailed by impersonal considerations such as the negative impact humans may have on the earth.

Therefore, a principle that permitted extinction to be accomplished in a way that caused involuntary physical pain or premature death could quite clearly be rejectable by existing people with no relevant countervailing reasons. This means that human extinction that came about in this way would be wrong.

There are of course also additional reasons they could reject a similar principle which I now turn to address in the next section.

2.4. Existing people could endure non-physical harms

I said earlier than the fact in itself that there would not be any future people is an impersonal reason and can therefore not be a reason to reject a principle permitting extinction. However, this impersonal reason *could give rise to* a personal reason that is admissible. So, the final important reason people might think that human extinction would be wrong is that there could be various deleterious psychological effects that would be endured by existing people having the knowledge that there would be no future generations. There are two main sources of this trauma, both arising from the knowledge that there will be no more people.

The first relates to individual people and the undesired negative effect on well-being that would be experienced by those who would have wanted to have children. Whilst this is by no means universal, it is fair to say that a good proportion of people feel a strong pull towards reproduction and having their lineage continue in some way. Samuel Scheffler describes the pull towards reproduction as a 'desire for a personalized relationship with the future' (Scheffler 2012, 31). Reproducing is a widely held desire and the joys of parenthood are ones that many people wish to experience. For these people knowing that they would not have descendants (or that their descendants will endure painful and/or premature deaths) could create a sense of despair and pointlessness of life.

Furthermore, the inability to reproduce and have your own children because of a principle/policy that prevents you (either through bans or physical interventions) would be a significant infringement of what we consider to be a basic right to control what happens to your body. For these reasons, knowing that you will have no descendants could cause significant psychological traumas or harms even if there were no associated physical harm.

The second is a more general, higher level sense of hopelessness or despair that there will be no more humans and that your projects will end with you. Even those who did not feel a strong desire to procreate themselves might feel a sense of hopelessness that any projects or goals they have for the future would not be fulfilled. Many of the projects and goals we work towards during our lifetime are also at least partly future-oriented. Why bother continuing the search for a cure for cancer if either it will not be found within humans' lifetime, and/or there will be no future people to benefit from it once it is found? Similar projects and goals that might lose their meaning when confronted with extinction include politics, artistic pursuits and even the type of philosophical work with which this paper is concerned. Even more extreme, through the words of the character Theo Faron, P.D. James says in his novel *The Children of Men* that 'without the hope of posterity for our race if not for ourselves, without the assurance that we being dead yet live, all pleasures of the mind and senses sometimes seem to me no more than pathetic and crumbling defences shored up against our ruins' (James 2006, 9).

Even if James' claim is a bit hyperbolic and *all* pleasures would not actually be lost, I agree with Scheffler in finding it not implausible that the knowledge that extinction was coming and that there would be no more people would have at least a general depressive effect on people's motivation and confidence in the value of and joy in their activities (Scheffler 2012, 43). Both sources of psychological harm are personal reasons to reject a principle that permitted human extinction. Existing people could therefore reasonably reject the principle for either of these reasons. Psychological pain and the inability to pursue your personal projects, goals, and aims, are all acceptable reasons for rejecting principles in the contractualist framework. So too are infringements of rights and entitlements that we accept as important for people's lives. These psychological reasons, then, are also valid reasons to reject principles that permitted or required human extinction.

3. Implications

I have shown that two of the four reasons I suggested earlier could be the basis for reasonably rejecting principles permitting human extinction. I have rejected:

- (a) It would prevent many billions of happy people from being born.
- (b) It would mean the loss of the only form of intelligent life and all civilization and intellectual progress would be lost.

and accepted:

- (c) Existing people would endure physical pain and/or painful and premature deaths.
- (d) Existing people would endure psychological traumas such as depression and the loss of meaning in their pursuits and projects.

Since the two accepted reasons will or will not apply depending on the way in which extinction came about, the implication of my argument is that whether or not contractualism supports principles that would lead to human extinction depends primarily on the way in which extinction would come about, and more specifically, its effects on current people.

This means that human extinction could only be wrong insofar as it negatively impacts already existing people's interests — either through the pain and premature death or the fact that people know that it is going to occur (thus causing psychological distress. Furthermore, as I said earlier in the discussion of physical pain and premature death, these negative impacts become relevant reasons to reject a principle when they are unwanted negative impacts.

I now want to briefly consider what this means about what types of extinction could or could not be considered wrong by contractualist standards whilst keeping in mind two important points: (i) contractualism is only concerned with what we owe to each other; and (ii) extinction could be wrong only when existing people have one of the two personal reasons listed above to reject it.

The first point (i) rules out deeming a completely non-agential and/or unavoidable cause of extinction as being wrong. For example, an asteroid hitting the Earth would not be wrong since an asteroid is not a moral agent and nothing it does can be considered right or wrong, even though it would presumably cause pain to the people it killed. It could still be regrettable, or the situation/event considered bad because it caused a large amount of pain to those affected, but it would not be *wrong* because it is not part the moral realm of what we owe to each other and not performed by a moral agent.

The second point (ii) also means that if a principle permitting or allowing extinction had no involuntary negative impacts on actual current people's interests, it would not be rejectable, and the resulting extinction would not be wrong. Even if such an extinction happened in a painful way, voluntarily incurred pain is not necessarily a reason for rejection because a person does not have to reasonably reject a principle even if it objectively lowers their well-being (Scanlon 1998, 213–218). If there were a way of causing human extinction that were universally voluntary — viz. by all those who exist/will exist as the extinction occurs — and involved no other infringements of rights or autonomy, a principle permitting it could not be reasonably rejected. It is possible to imagine a physically painless way of causing extinction, but I struggle to imagine a principle permitting human extinction that would be universally voluntary; nonetheless it is theoretically possible and if it were possible in reality, it would not be wrong.

As I said earlier, if extinction came about in a way that is involuntarily physically or psychologically painful (including infringements on autonomy) it could be reasonably rejected. As I also said earlier, both physical and psychological pain affect a person's well-being and are therefore reasons to reject a principle. This means that cases of extinction caused by painful, involuntary, nuclear holocaust would clearly be wrong. Furthermore, if, for example, people were

involuntarily sterilized (even if painlessly), they could still reasonably reject this on the grounds of either psychological harm or an infringement of autonomy (since it was involuntary). So extinction caused by involuntary physical or psychological painful causes could also be reasonably rejected and would therefore be wrong.

So too would cases in which we simply failed to prevent human extinction despite knowing that it would cause harm to the people who exist when it came about, even if they do not exist at the present time. An example of this could be an asteroid that is hurtling towards Earth. Once it hits, it will prematurely kill everyone alive at the time. Luckily, we have discovered how to divert asteroids in advance so that they do not make contact with our planet in the future. It would be wrong for us not to divert the asteroid since whoever will exist at the time of the extinction would have a strong reason (based on physical pain, premature death and psychological harms they would suffer) to reject a principle that permitted a failure to divert the asteroid.

One final form of extinction to be considered is one that were caused painlessly and without anyone knowing about it (thus also negating any psychological harms). The person enacting such a policy would be showing disrespect to those who were impacted since the agreement in contractualism is only hypothetical. It is not necessary that the negatively impacted parties know that they are being negatively impacted for the action to be wrong. As Rahul Kumar explains, permitting others to make decisions about how another person's body may be used is grounds for reasonable rejection, regardless of whether or not harm or the risk of harm is being imposed because 'each individual has good reason to want this kind of decision-making discretion to be solely her own (regardless of whether or not her body would in fact be used by others were it (to even in part be) ceded)' (Kumar 2015, 38). In any case, there would have to be very strong, personal, reasons in favour of the sterilization in order for it to overcome the possible objections of the people being affected.

A related form of extinction that has been suggested to me is the following.¹⁰ Suppose the current generation were somehow altered somehow as to be able to bear children, but these offspring will all be fertile. It was suggested that this would be neither a psychological nor physical harm for the current generation and would therefore not be wrong according to my formulation. I think this form of extinction could still be rejected for the following reasons. First, if the current generation is caused to have infertile babies involuntarily, this would be objectionable for the same reasons I have just discussed. Second, I disagree that there would be no psychological harm as the current generation will still be alive to see that their children are all infertile and will therefore be subject to the same difficulties as discussed in section (d) of the paper. Third, more contestably, even if the alteration were voluntarily undertaken by the current generation, I wonder whether there might be an objection from the subsequent infertile generation that their life choices were constrained significantly by others, creating a reason

for rejection similar to that described by Kumar. Although the current generation may not experience any physical or psychological harms from extinction, who is to say that the next one will not either? So, this case of extinction could not be said to be truly voluntary as a large proportion of people who will be affected by it have not agreed.

In most cases of potential human extinction, then, it looks as though contractualism is consistent with the intuition I laid out at the start — that there would be something wrong with causing or allowing human extinction. Interestingly, however, the wrongness is not grounded in the forgoing of potential future lives, but in the interests of the people who would exist at the time of the extinction.

However, we cannot ignore the two cases in which extinction would not be wrong:

- (i) When the cause of extinction is completely non-agential and/or unavoidable (e.g. an unavoidable asteroid); and
- (ii) When the extinction is universally (by all affected) voluntary.

Is it a problem that we cannot explain the wrongness of human extinction in these scenarios? I don't think so. It is not clear in (i) how an unavoidable natural disaster could plausibly ever be described as wrong when 'wrong' seems in almost every moral theory to mean something that was done through human choice when there was an alternative (however, viable) option available.

What of (ii) though? It is possible that the extinction in (ii) could be impersonally *bad* since it prevents new people from being enjoying happy lives and destroys the only form of intelligent life, but this is not the same as being wrong. Those unconvinced by the contractualist framework to begin with may want to throw the baby out with the bathwater and declare that this is evidence that contractualism simply cannot supply an adequate response to the problem of human extinction since it cannot explain the wrongness of all of its possible instances.

Those of us more sympathetic to the theory have two options. The first is to bite the bullet and accept that (ii) is not wrong, despite an apparent intuition to the contrary. Our intuition sets are often inconsistent and all theories have bullets to bite; contractualism is not unique in this regard. Furthermore, wrongness has to do with relations between persons and how we ought to treat each other and in these cases, no one is being treated unfairly or harmed in any way, so perhaps it is simply the case that there is no wrongdoing here.

The other option might be to claim that actually there are people being disrespected, harmed or treated unfairly: past people. It might be argued that allowing human extinction frustrates past people's interests by not continuing their projects (either the project of continuing the human species itself or other projects that can only continue by virtue of people existing to continue them). As I said earlier, Scanlon allows that any past, current, or future person constitutes a standpoint towards which justification makes sense. If it is in fact

possible to frustrate a past person's interest, this might be a personal reason to reject principles that allow human extinction, even if every currently existing person were in favour. Or, perhaps, a current person may have made a promise to their grandmother (for example) to carry on the family name. The promisee may have a posthumous claim to the promiser keeping their promise. I am cautiously sceptical of this possible solution, primarily because even if past people's interests can constitute personal reasons in this way, it seems unlikely that they would outweigh current people's much weightier interest in personal autonomy and the reason to make one's own decisions about one's body. However, further analysis of how contractualism might address moral obligations to the past is certainly warranted and if it can help.

Regardless of which option one chooses, the analysis in this paper has put us in the position to start the comparison of contractualism against other theories that have something to say about human extinction. In almost every case, we have seen that it would be wrong to cause human extinction, and the rare scenarios in which it would not be do not preclude regretting that extinction has occurred, even if it were not *wrong*.

Notes

1. Throughout I will discuss causing and allowing extinction interchangeably. Although there may be reasons in some cases why causing an outcome is worse than allowing it to occur, I do not find the distinction important in the context of human extinction.
2. Note that this does not entail the sometimes-included claim that an act can only be wrong if it makes someone *worse off*.
3. This was first proposed by (Narveson 1967).
4. Also, depending on how much weight we assign to the 'good' of existence, to say that a possible person could reasonably object to not being created would make every action or principle wrong if it outweighed the burdens we would have by not creating 'them'. Any time we were not actively engaged in creating a person we would be acting wrongly. In fact, even if we were trying to create as many people as possible, simple biological facts make it impossible to bring into existence every possible person. Every sperm and egg combination represents a possible person, but women can only conceive and birth one combination every nine months. There will therefore always be thousands or even millions of possible people that we do not create even if we were trying to create as many as possible. If we act wrongly when we fail to bring a possible person them into existence, then we will be constantly acting wrongly. There is little use for a position in which it is impossible to avoid wrongful action.
5. Sidgwick writes: 'I think that if we consider carefully such permanent results as are commonly judged to be good, other than qualities of human beings, we can find nothing that, on reflection, appears to possess this quality of goodness out of relation to human existence, or at least to some consciousness or feeling ... Still, as soon as the alternatives are clearly apprehended, it will, I think, be generally held that beauty, knowledge, and other ideal goods, as well as all external material things, are only reasonably to be sought by men in so far as

they conduce either (1) to Happiness or (2) to the Perfection or Excellence of human existence.'

6. Of course existing people might be negatively affected by the knowledge that the progress they made would be lost, but this is a matter of psychological harm covered in section (d) and not to do with the value of civilization/rationality *per se*.
7. For a good summary of the debate, see (Bradley 2009; Taylor 2013).
8. Note that the fact that a person has a reason to reject a principle does not mean that the principle is necessarily *reasonably rejectable*. If there are stronger reasons in favour of the principle, then it might not be rejectable despite the objections. Killing in self-defence is an example in which the reasons in favour of permitting the death might outweigh the reasons against it.
9. Göran Duus-Otterström suggests to me the following reply. He points out, correctly, that although impersonal reasons cannot on their own provide reasons for rejecting a principle, they may create other personal reasons. He cites Crake in Margaret Atwood's novel *Oryx & Crake*. Crake embarks on a genocidal mission to rid the Earth of human beings in order to benefit the environment. Duus-Otterström suggests that the impersonal reason to preserve the environment is so important to Crake's ability to live a fulfilling life that it creates a personal reason for Crake's programme. However, even if one wants to accept that Crake's reason is personal, it must still be weighed against the personal reasons of those who would be killed as a result of his programme. A person's reason not to be involuntarily killed must surely outweigh another person's reason to want to preserve the environment.
10. Thank you to the anonymous reviewer for this suggestion.

Acknowledgments

I would like to thank Dan Butt, Simon Caney, Göran Duus-Otterström, Simon Gardiner, Clare Heyward, Martin O'Neill, Melinda Roberts, Anthony Skelton, the anonymous reviewer for the *Canadian Journal of Philosophy* and the members of the Nuffield Political Theory Workshop and Institute for Futures Studies' philosophy seminar for helpful comments on earlier drafts of this paper.

Funding

This research was undertaken with the support of the Social Sciences Research Council of Canada and the Swedish Research Council.

Notes on contributors

Elizabeth Finneron-Burns is a Teaching Fellow at the University of Warwick and an Affiliated Researcher at the Institute for Futures Studies in Stockholm. Her primary research interest is our obligations to future generations, including who and how many people we create, and what kind of lives we leave them. She can be contacted at e.finneron-burns@warwick.ac.uk.

ORCID

Elizabeth Finneron-Burns  <http://orcid.org/0000-0002-5632-7665>

References

- Beckstead, Nick, Peter Singer, and Matt Wage. 2013. "Preventing Human Extinction." In *Effective Altruism*. http://effective-altruism.com/ea/50/preventing_human_extinction/.
- Benatar, David. 2006. *Better Never to Have Been*. Oxford: Clarendon.
- Bradley, Ben. 2009. *Well-Being and Death*. Oxford: Oxford University Press.
- Gardiner, Stephen M. 2009. "A Contract on Future Generations." In *Intergenerational Justice*, edited by Lukas H. Meyer, and Axel Gosseries, 77–118. Oxford: Oxford University Press.
- Institute. 2016. Future of Humanity. Accessed December 4, 2015. <http://www.fhi.ox.ac.uk/research/research-areas/>.
- James, P. D. 2006. *The Children of Men*. Vintage Books: Random House.
- Kumar, Rahul. 2003. "Reasonable Reasons in Contractualist Moral Argument." *Ethics* 114: 6–37.
- Kumar, Rahul. 2015. "Risking and Wronging." *Philosophy and Public Affairs* 43 (1): 27–51.
- McMahan, Jeff. 2009. "Asymmetries in the Morality of Causing People to Exist." In *Harming Future Persons: Ethics, Genetics and the Nonidentity Problem*, edited by Melinda A. Roberts and David T. Wasserman, 49–68. New York: Springer.
- Narveson, J. 1967. "Utilitarianism and New Generations." *Mind* 76 (301): 62–72.
- Parfit, Derek. 1984. *Reasons and persons*. Oxford: Clarendon Press.
- Reiman, Jeffrey. 2007. "Being Fair to Future People: The Non-Identity Problem in the Original Position." *Philosophy and Public Affairs* 35 (1): 69–92. doi: 10.1111/j.1088-4963.2007.00099.x.
- Risk. 2016. Centre for the Study of Existential. Accessed December 4, 2015. <http://cser.org/about/>
- Scanlon, T. M. 1998. *What We Owe to Each Other*. London: Belknap Press of Harvard University Press.
- Scheffler, Samuel. 2012. *Death and the Afterlife*. Oxford: Oxford University Press.
- Sidgwick, Henry. 1874. *The Methods of Ethics*. 7th ed. London: Macmillan.
- Taylor, James Stacey. 2013. *The Metaphysics and Ethics of Death*. Oxford: Oxford University Press.
- Weinberg, Rivka M. 2008. "Identifying and Dissolving the Non-Identity Problem." *Philosophical Studies* 137 (1): 3–18.