Diversity and Inclusion in Existential Risk Studies

SJ Beard and Suzy Levy

UNIVERSITY OF CAMBRIDGE

Contents

Executive Summary 3

Summary of recomendations *4*

Introduction 5

Diversity within the groups we studied *5*

Experiences and Perceptions 8

Implications and Impacts 14

Confidence and Awareness 19

Current Issues and Concerns 25

Correcting under-representation *25*

Resolving pressing challenges *26*

Positive Futures 30

Inclusive leadership *30*

Diversified and professionalized funding *31*

Better representation *31*

A Diversity and Inclusion Strategy of Existential Risk Studies 33

Increasing the representation and inclusion of diverse people and their viewpoints within the field *33*

More fundamentally embracing diversity in how work is constructed and who we 'bring in'. *34*

Recognizing how funding and leadership can limit thinking, how power manifests, and what 'democracy' in the field could look like *34*

Creating opportunities for respectful dialogue around events impacting our field and using them as transparent catalysts for change *35*

More compassion and activism *36*

Appendix – Survey Responses 37

Executive Summary

This report presents the findings of an independent study into Diversity and Inclusion in the emerging transdisciplinary field sometimes known as Existential Risk Studies and offers a series of recommendations for how it can improve. Research for this study was carried out by SJ Beard, a researcher at the Centre for the Study of Existential Risk (CSER), and Suzy Levy, a diversity and inclusion consultant, using a survey and series of in-depth interviews with people who identified with the field; including researchers and operational staff working at CSER, a range of other existential risk research organizations, and related advocacy, support, and funding bodies. Its key findings are that:

- While many researchers within the field reported feeling marginalized or excluded, there was no unified experience of difference and exclusion. There are many and diverse experiences of both marginalization and exclusion, including at the social, cultural, demographic, and intellectual levels, suggesting that target driven approaches to improving diversity, while potentially helpful, are unlikely to be sufficient in themselves and a more detailed focus on the many and varied forces of homogenization and exclusion will be needed.

- The impacts of homogeneity and exclusion within the field are wide ranging and can be subtle and complex. People may remain within the field even though they feel excluded from it. Some common themes include people developing negative attitudes towards the community, struggling to access funding, or becoming unwilling to ask for support from certain bodies, and having to divert resources to deal with problems. These may mean that individuals who experience marginalization or exclusion lose access to social and financial capital or limit the scope of their work. Many researchers feel there is a risk that this this is limiting the growth and creativity of the field.

- While most people in Existential Risk Studies have some awareness of diversity and inclusion, few feel confident in responding to these issues, and the most confident are often also those who have the least direct experience of marginalization or exclusion. Some people's confidence may spring from a simplistic understanding of diversity issues, for instance a focus on numerical representation, while less confident people saw the problem as more systemic and multi-faceted. More confident researchers also focused on the intellectual inclusion of diverse viewpoints and ideas, rather than on how to be inclusive towards different people and their needs.

- The field has been significantly impacted by a range of unfolding high-profile events over the past year that have left many feeling unsure of their position within the field. In responding to these events, it is important that the community pays attention both to the long-term and systemic dynamics that

Executive Summary

drive homogenization and exclusion and to specific instances of bad behaviour. These events also highlight the importance of thinking about both resolving challenging situations and actively supporting those who have been negatively affected by them.

- On the other hand, our research also highlighted the many opportunities for positive and rewarding discussions about building more positive futures for the field. Since Diversity and Inclusion are seldom given much prominence in discussions about the field, people who actively commit to working on these issues often spend much time working alone and against significant challenges. However, there is also joy and purpose to be found in creating spaces where we can bring our whole selves into the discussion and leaders within Existential Risk Studies could easily create many more opportunities for this, while also supporting more constructive and rewarding processes for people to work towards resolving their difficulties and disagreements.

SUMMARY OF RECOMENDATIONS

Drawing on these findings, and in collaboration with stakeholders from across the community, we propose a strategy for improving diversity and inclusion within Existential Risk Studies based around the following key recommendations:

- Increasing the representation and inclusion of diverse people and their viewpoints with part of Existential Risk Studies requires challenging the forces of homogenization and exclusion. This includes addressing the unfair allocation of resources and providing targeted support to specific groups.

- Researchers should embrace diversity and inclusion in how work is constructed and who we ‘bring in’, so that a wider range of perspectives are reflected (including those that may seem critical). This can unlock the potential for this field to act as an agent of positive change in the world.

- The field needs to recognize the ways in which concentrations in funding and leadership can limit thinking, identify how power manifests, and consider what democracy could look like in ERS. This could produce both a fairer distribution of resources and healthier relationships between researchers and their supporters.

- Those with leadership responsibility must provide opportunities for respectful dialogue of ongoing events that are impacting the field where care is taken to ensure participants can speak freely and without fear, and use these to catalyse wider change by supporting those working for resolution and growth.

- Culturally, the field would benefit from a humane approach that recognizes the need for compassion and activism and ensures these are celebrated and rewarded; rather than remaining as additional burdens that often fall disproportionately on those who experience marginalization or exclusion.

SECTION 1

Introduction

This report presents the findings of an independent research study that was undertaken to understand people's experiences around diversity and inclusion within the emerging transdisciplinary field sometimes known as Existential Risk Studies, and how this can be improved. Existential Risk Studies is a relatively recent development, based around a small number of elite research centres such as the Centre for the Study of Existential Risk (University of Cambridge), the Future of Humanity Institute (University of Oxford), the Stanford Existential Risk Initiative (Stanford University), and the Mimir Centre for Long-term Futures (Institutet för framtidsstudier), together with a wider community of affiliated researchers, support organizations (such as 80,000 hours, Legal Priorities Project, Riesgos Catastróficos Globales, and Magnify Mentoring), and funders (such as Open Philanthropy and Effective Giving). While the phrase Existential Risk Studies is more associated with the production of academic research, this study attempts to takes a broad view of the subject based around participants own understanding of their position in relation to the field, although it is likely that this framing influenced participants responses, and may have felt exclusionary in itself. Many participants identified Existential Risk Studies as being closely affiliated, or even part of, the Effective Altruism movement (EA), while some also associated it closely with particular ideas and values, such a transhumanism or 'techno-utopianism'. Several comments reflect these associations, which the authors will neither endorse nor critique.

Diversity and inclusion within this field has been an issue of interests to the lead author of this report, Dr SJ Beard of the Centre for the Study of Existential Risk (CSER), since they entered the field in 2015. In part this reflects their own position as a disabled and transgender researcher and in part an awareness that these issues seemed to receive very little attention by many people in this community. This study was undertaken by Dr Beard working closely with Suzy Levy, an independent diversity and inclusion consultant and founder of The Red Plate. It was made possible when Dr Beard was awarded a Borysiewicz Interdisciplinary Fellowship by the University of Cambridge, which provided support for a novel research study that might not otherwise be undertaken by the researcher.

To understand more about the range of perspectives on issues of diversity and inclusion within the field, the authors surveyed people's experiences and analysed these to produce a viable strategy for improving the field. The study, which obtained ethics clearance from the University of Cambridge School of Arts and Humanities, included an on-line survey that received 85 responses, and a series of 12 in-depth interviews. Interview subjects were selected from the survey respondents to a) produce a representative sample of identities, experiences, and opinions, and b) follow up on specific ideas or suggestions from respondents that were of interest to the authors. Interviews were transcribed and analysed by the report authors to identify key trends and produce a range of quotes to highlight the different perspectives and ideas that were presented. These are presented below, in some cases with limited editing to improve readability and protect the anonymity of respondents.

DIVERSITY WITHIN THE GROUPS WE STUDIED

As part of our survey and interviews we asked people about their identities and relation to marginalized and minority groups. The results of these questions should not be read as providing an assessment of diversity within this community; however, they can help us to understand how individuals in the community think about their own marginalization or minority status.

Of the 84 responders to our survey, 18 said that they definitely did not belong to a marginalized or minority group. Of these only 3 commented on why they felt this way, all saying they were white, cisgender, and male, 2 mentioning they were heterosexual, and 1 mentioning each of the following: being middle aged, non-disabled, not obese, and/or western. 13 participants said they somewhat did not belong to any marginalized or minority group, of whom 7 made further comments. 1 noted that they were Jewish, 1 that they were a woman from a non-traditional academic background, 1 that they came from a working class background, 1 that they were a first generation university graduate, 1 that they were religious, 1 that they only felt marginalized in ways that were 'protected', and 1 saying it depended on the country and environment they were in. 2 participants said they neither belonged nor did not belong to a marginalized or minority group, 1 of whom noted that they were a woman.

26 participants said that they somewhat belonged to a marginalized group, 20 of whom made further comments. Many of these people noted several potentially marginalized identities but among these there were 5 mentions of each of the following: being LGBT+, Hispanic, from a low-income background, and/or mixed race, 4 mentioned being women, 3 that they were neurodivergent, 2 that they were disabled/chronically ill, and 1 mentioning being a non-native user of English, from a 'non-core' academic discipline, and/or politically left wing. Finally, 25 said that they were definitely from a marginalized or minority group with 14 providing additional comments. Of these 6 identified as LGBT, 5 as women, 4 as People of Colour, 3 as Asian, 2 as from low-income backgrounds and/or from immigrant backgrounds, and there was 1 mention of each of the following: being Hispanic, being Chinese, having a mental health condition, and/or being a carer. Many of these people conveyed their identities in much richer terms that we have unfortunately needed to summarize here.

In addition, we asked those we interviewed a range of questions about different aspects of their identity and which of these they felt were most relevant to conversations around diversity and inclusion. To prevent participant identification, we will only discuss the five categories here that people felt were most important. Of 13 participants, 7 felt that their ethnic group was relevant and our group comprised 9 white and 5 non-white individuals (with one participant identifying as both); 6 felt that gender identity was important and our group consisted of 11 cisgender people and 2 non-cisgender people; 5 felt that sex was important, with our groups consisting of 7 males and 6 females. 5 also felt that social class was important, we failed to ask about this directly but 7 individuals volunteered information about it with 5 indicating privilege, 1 poverty, and 1 a background that would be considered privileged in their home community but not in richer parts of the world (on the other hand we did ask people about their educational background but got rather different results, all participants said they were university graduates and 7 said they had PhDs, three also mentioned attending private education, while two others also talked about this later in the interview, but many did not mention pre-university education, 1 said they were there first in their family to attend university). Finally, 4 indicated that disability was important, and our group consisted of 1 person with a long-term illness, 3 with neurodivergences, and 9 non-disabled people. In general, there was no strong relationship between whether a person belonged to a minority group and whether they thought that characteristic was important to our discussion, although this was truer for race and sex (which were more or less equally split) than for gender identity, disability, and social background (where people belonging to marginalized groups were more likely to say these were important).

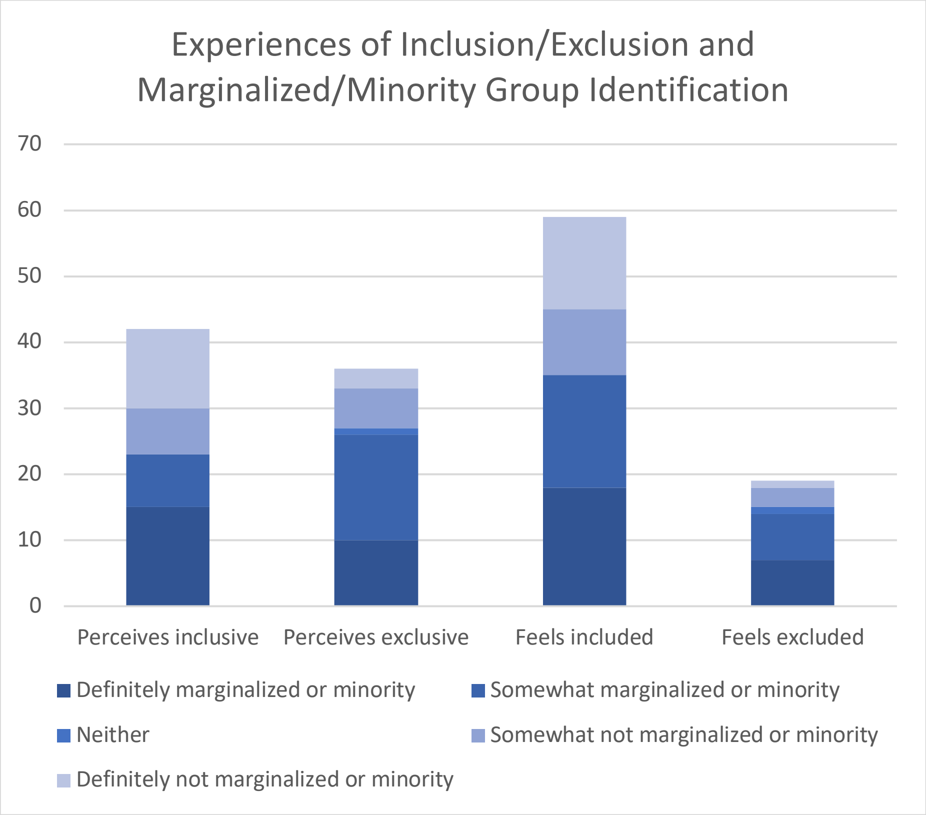
While not providing much evidence about the underlying diversity of Existential Risk Studies as a field, these results highlight how different people perceive their own contribution to diversity. Different people may see the same characteristic of their identity, such as being a woman or from a poor socioeconomic background, as either definitely marginalized or as largely non-marginalized, and people also had very different ideas about which demographic characteristics were most relevant to discussions around diversity and inclusion. Indeed, several survey respondents and interview participants said that they were conscious that within Existential Risk Studies they were often viewed as representatives of marginalized or minority communities but that they felt the only reason they were part of this field to begin with was that they were not, in fact, representative of those communities but rather had had a relatively very privileged background that had made this sort of career possible for them.

There is also clearly a very large range of factors that people see as potentially making them from a marginalized or minority background. This suggests that even if one had the resources to undertake a full census of the Existential Risk Studies community to collect data about people's demographic characteristics and identity, this might not provide all the information needed either to fully understand it's diversity or to ascertain whether it was inclusive of people from marginalized and minority backgrounds.

SECTION 2

Experiences and Perceptions

When we asked survey respondents Whether they perceived the field of Existential Risk Studies to be inclusive or exclusive, and whether they themselves felt included or excluded from it, and compared this to how much they saw themselves as belonging to one or more marginalized or minority groups we found the following trends.



In general, survey respondents from all groups were more likely to say they personally felt included in the field rather than excluded from it (although this may be unsurprising, given that this was a survey most likely to reach, and be of interest, to people who self-identified with the field). However, the proportional differences remain striking:

- people who said they definitely did not belong to a marginalized or minority group were over 10 times more likely to say they personally felt included (14 responders) rather than excluded (1 responder);
- those who said they somewhat did not belong to a marginalized or minority group were 3 times more likely to feel included (10 responders) than excluded (1 responder);

- finally, those who somewhat or definitely felt they did belong to a marginalized or minority group were roughly 2.5 times as likely to feel included (35 responders) than excluded (15 responders).

Turning from people's personal experiences to their perceptions of the field: people who definitely saw themselves as not belonging to one or more marginalized or minority groups were 4 times more likely to perceive the field as generally inclusive (12 responders) rather than exclusive (3 responders). On the other hand, people who saw themselves as somewhat belonging to one or more marginalized or minority groups were twice as likely to see the field as *exclusive* (16 responders) rather than *inclusive* (8 responders). Both the people who definitely saw themselves as belonging to one or more marginalized or minority groups and those who saw themselves as somewhat not belonging to any such groups, were more evenly split between perceiving the field as generally inclusive and exclusive.

On the face of it, this latter trend may seem contradictory, or at least inconsistent, why would people who only somewhat identity with a minority or marginalized group view the field as more exclusionary than those who definitely identified with such a group? However, when we look at the comments left by survey responders these results begin to make more sense. Survey participants who identified as somewhat belonging to a minority or marginalized group and who saw the community as exclusionary, generally pointed to issues of social exclusion, for instance:

- "Ideologically, there is the hegemonic view of the techno-utopians which will coerce anyone with a critical view",

- "I feel that there is a lot of "groupishness" in the Oxford part of the community",

- "It is difficult to get involved unless you know someone who is already an active participant in the community",

- "The few women are noticeable when there are symposiums and so on and at some point (eventually, after a couple of days at the same site) the men start presuming we (women) are there for any other purpose than work", and

- "I think there is a lot of in-group signalling RE: language, terms, and occasionally assumptions around world view models, etc".

On the other hand, survey participants who saw themselves as definitely belonging to a marginalized or minority group were more likely to evaluate Existential Risk Studies relative to other communities they had experienced, with several indicating that they thought it performed well by comparison:

Experiences and Perceptions

- “It has all the usual passive issues of a community that under-represents some groups”,
- “This is a field that often takes place, with discussions at and with people from, top-tier universities not exactly known to be inclusive in the first place”.
- “Most exclusionary aspects come from background society and academia being exclusionary to a large extent”.
- “I don't think it's worse than your typical "community" and it seems to be trying to address the issues”,
- “I think people have been pretty friendly & welcoming, particularly as I don’t come from the traditional academic background”.

In the follow up interviews we asked participants about both their good and bad experiences of the field. Thinking about their positive experiences, many felt very positive about the field in general, although in many cases this was only partial or caveated; one said they found their entry into the field “hugely encouraging” and another said they had been “pretty welcomed as an individual”. Several also noted signs of the field developing in promising directions, for instance one participant observed that “in recent years there has been a bit more of a concerted effort to have more women and non-white people in the field” while another noted that “I think the existential risk community has a lot of people who are neurodivergence so I feel that there is a very high degree of acceptance of that”. Two participants also noted positive experiences around affirmative action, from which they felt they had benefited. One participant saw this in a purely positive light saying.

> *“I am pretty sure that I am more likely to be asked to be an MC for major events or to interview someone because I am a woman of colour who is comfortable doing these things. I am very aware that this is because of a focus on diversity, and I am extremely happy about that”.*

while another offered a more nuanced view “if I am careful not to let it harm my self-confidence, I do feel like some of these characteristics have enabled me to do things like be on panels and give talks that maybe I would not have had.” On the other hand, several participants who identified themselves as not belonging to any marginalized or minority group expressed unease about what their positive experiences might imply about the community, for instance noting that:

> *“My own experience as a white male in a field that is mostly white male is there is a sense that I fit in. But any time that this thought crosses my mind my primary thought is to try and use my own position to help support a wider range of people getting involved so that people who are not like me feel more comfortable getting involved.”*

Another participant noted that they didn't have much to say on the topic and felt that "my absence of reflection about this probably reflects that fact that I belong only to privileged categories, and so I have the luxury of not thinking about this."

When we asked people whether they had ever felt excluded or marginalized within the field, the answers were more varied, but every participant was able to describe at least one instance of this. For some, exclusion took a very social form, for instance one participant noted how:

"I don't dine indoors when COVID incidents are high because it could finish me up; so, I once asked if we could hold an event outdoors... I got so much shit from colleagues for asking for that one time, and I have never asked since."

Another reported getting the sense that "because I didn't go to an elite university, I wasn't included in certain discussions or considered for certain job roles", while a third felt that "I have been overlooked for promotion, even though my outputs, policy and other impacts have been above expectations." Several participants reported experiences of exclusion around their membership and support for the Effective Altruism community. For instance, one participant noted that "if you don't agree with lots of EA's methods or use the language conventions they expect, then people just start to not invite you or include you in things and that has a negative career impact because you meet a lot of relevant people through EA" while another explained how:

"At certain events that have been organized by Effective Altruism I have felt very different. I think this is very largely a question of age but being non-white and not from the UK is also a factor and I have also felt different when it comes to diversity of thought and that has certainly felt uncomfortable."

On the other hand, another participant noted that "I feel like I have usually had a pretty strong sense of belonging because I got involved in the wider Effective Altruism community early on and feel like I have developed a lot of credibility that way." Several participants initially expressed uncertainty about whether they had ever experienced marginalization or exclusion before reflecting further and deciding that on at least one issue, such as age, disciplinary background, or not having a PhD, they had had felt different or marginalized to at least some extent. Other participants felt it have been a very central part of their experience with the field of Existential Risk Studies, one responded to the question of whether they had ever felt excluded or marginalized "Yes, enormously and pervasively throughout my entire career thus far in existential risk!"

Several participants felt that it was not only themselves as individual's who were being excluded from the community, but also their views and ideas. As one participant put it "there have been several occasions where I felt there have been either professional or personal repercussions for speaking about views or beliefs that are not part of the norm… the onus and burden of proof was placed on me, and I was always going to lose because people in senior positions had different beliefs, and this wasn't a fair debate.' Another participant felt that the field was not open to female viewpoints in particular noting that "existential risk reduction is something that follows from the transhumanist movement and I feel like the transhumanist movement is very male dominated, especially when talking about births and optimising humans", while another noted how perspectives from the global south were typically either dismissed or co-opted by colleagues, noting how the field "fails to consider more universal human values like values that are inclusive of the global south or … frames something as novel just because Peter Singer or Derek Parfit or Toby Ord or Will MacAskill said it but that in the global south feels very normal."

CONCLUSION

A key lesson from these findings and experiences is that there simply is no unified experience of difference and exclusion within the field. It is possible for more marginalized community members to feel less excluded than less marginalized ones, whether because they have more negative experiences of other communities to compare it with, because they benefit more from affirmative action, or for any number of other reasons. Similarly, it is possible for even the most privileged members of the community to identify at least some times when they have felt marginalized or excluded, or to feel that their own inclusion is somehow problematic or difficult; yet it generally felt like people saw this as an experience that made them feel isolated and uncomfortable rather than using it as an opportunity to work with others against exclusion.

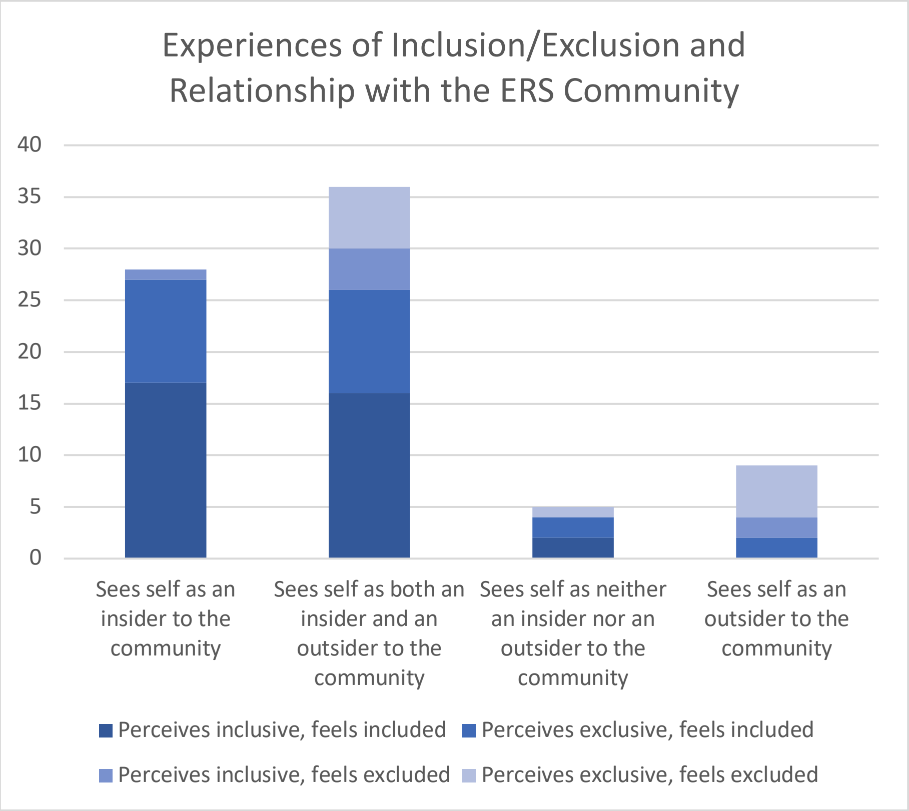
More importantly still, there are many and diverse experiences of both marginalization and exclusion, including at the social, cultural, demographic, and intellectual levels. It is common within Existential Risk Studies, as with many other fields and communities, to seek to resolve issues around diversity and inclusion with target initiatives, committees, organizations, and programmes. However, while unquestionably needed, these can run the risk of homogenizing people's diverse experiences of difference and exclusion in order to produce one size fits all solutions.

Based on the findings of this study, we recommend that initiatives should focus on the many and varied forces of homogenization and exclusion that operate within this field. As identified by survey and interview participants, these clearly include in-group/out-group dynamics, academic norms and paradigms, ideologies, personal networks and relationships, existing social injustices and oppressions, prejudiced attitudes like racism, sexism, and ableism (whether explicitly or implicitly held), and career expectations; although this is certainly not an exhaustive list. In some cases, such as academic norms and paradigms, there may be a trade-off between being less homogenous or exclusive and potentially being less coherent and rigorous (although this is certainly open to challenge), while in other cases, such as background social injustices, there may be little that the community, on its own, can achieve (although that does not mean we should not try). Importantly however, these forces affect everyone in our community, and they do so in different ways, as the diversity of experiences reported by participants clearly attests.

SECTION 3

Implications and Impact

People's experiences of diversity and inclusion impacted on their relationship with the field of Existential Risk Studies in a number of ways. When we asked about these two things in our survey, we got the following results:



One initial interpretation of this graph could be that most people in most groups responded that they either saw themselves as an insider to the community or at least as both an insider and an outsider. However, this would be too hasty since survey respondents were self-selecting and thus more likely to be people with some level of investment in the field and its future, favouring people who saw themselves as at least partial insiders to the field.

One interesting, but hardly surprising, result is that none of the 12 responders who said that they perceived the field as exclusive and felt excluded from it saw themselves as a straightforward insider to the community (although the majority didn't see themselves as a straightforward outsider either), while none of the 45 who both saw the community as inclusive and felt included in it saw themselves as a straightforward outsider (with the vast majority, 43, seeing themselves as either a straightforward insider or both an insider and an outsider). This highlights how important diversity and inclusion is to the future of the field; it seems that whenever people experience the field as inclusive, they seek to belong inside it, but that when they experience the field as exclusive, even if they maintain some kind of relationship with the field, people never feel like they can be a straightforward insider.

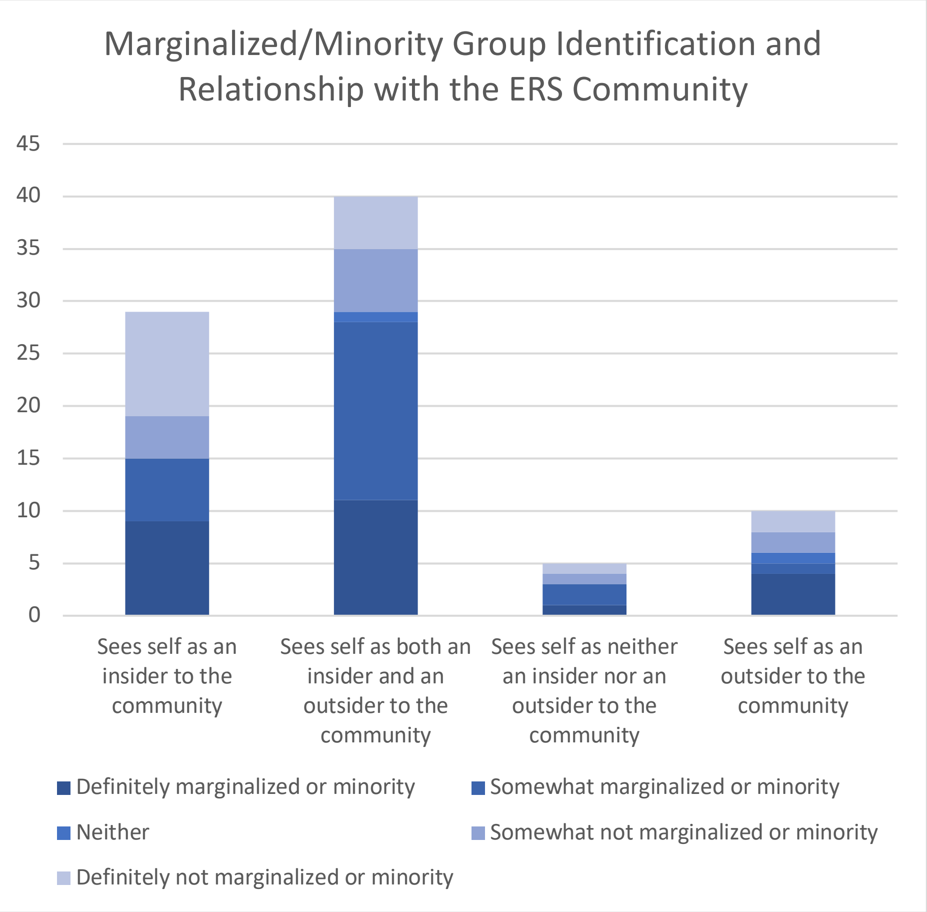
Another revealing finding is the large number of responses who indicated an uncertain, contested, or ambiguous relationship with the field. 52% of responders to this question (41 out 78) indicated that they were either both an insider and an outsider or neither an insider nor an outsider, rather than saying that they were definitely one or the other. Similarly, 40% of responders indicated that there was a difference between their general perception of the field's exclusivity and their own experiences, saying either that while they perceived the field as generally exclusive, they felt personally included in it or that while they perceived the field as generally inclusive, they felt personally excluded from it. This finding highlights the importance of qualitative research to understand the nature of people's experiences and relationships, which may well not be captured by simple surveys or censuses.

Turning then to people's description of their relationships with the field, survey responders who said they were neither an insider nor an outsider commented on seeing themselves as working on existential risk, or related issues, but not from within the community; for instance, once said "I am a veteran to the issue of existential risk, but find myself new to the community that focuses on existential risk" while another noted that "I work in a related field but do not directly study existential risks." People who saw themselves as both an insider and an outsider gave a wider range of responses, for instance at least two respondents commented that they felt this way for each of the following reasons:

- they worked on AI risk but not existential risk more generally;
- they were interested in existential risk but felt geographically isolated due to being outside of Oxbridge/the bay area;
- they had studied the subject but did not feel able to build a career working on it;
- they worked at existential risk organizations but not researching existential risk;
- they are involved in Effective Altruism generally;
- they are perceived negatively by others due to their work and beliefs; and
- they felt unsure about what the field of existential risk studies was or whether it really existed.

Implications and Impact

When we compared people's statements about their relationship with the field with their identification with one or more marginalized or minority groups, we got the following results:



The main trends in this chart are that 59% of those who felt they identified as definitely not belonging to a marginalized or minority group felt they were straightforwardly insiders (10 responders out of 18), the only group where a majority saw their relationship with the field as straightforward, while 65% of those who identified as somewhat belonging to a marginalized or minority group felt they were both insiders and outsiders (17 responders out of 26), potentially reflecting the complex experiences of marginalization and exclusion experienced by this group, as described in the previous section. Other groups were more evenly distributed, although it is worth noting that 40% of those who saw themselves as straightforwardly outsiders also identified as definitely belonging to a marginalized or minority group (4 out of 10); although this was less than 1/6th of all those who identified as definitely belonging to a marginalized or minority group overall (4 out of 25).

Interview subjects were able to provide far more detailed comments about the specific impacts that their unique experiences around diversity and inclusion had on their work and relationship with the field. Impacts ranged from personal challenges (funding, lack of promotion, stress, and tension) to more fundamental concerns with the field itself. At the more personal level, one participant said that their experiences had "resulted in some stress and tension in my life and career" while another believed they had "faced professional repercussions in terms of not being promoted" and "know of certain bodies who would not hire me despite my considerable academic output." Another participant felt that "I would be less inclined to join certain discussion spaces, like an event, conference, or Q&A and fear that if I did, I would just be making myself look silly or something."

Many participants indicated that their experiences had altered their perceptions of the field and its work. For instance, one said they had:

> *profound practical and ideological concerns with the field; it seems to have a very cis, white, middle class, elitist, joint to it, which is also quite unappealing as a vehicle to get onto to explore any topic. I have stopped engaging the community very much … I just don't have time for these people, except some of the more marginalized members.*

Other participants made similar comments, for instance that one noted that their experiences had "prompted a lot of reflection about the extent to which I want to associate with portions of the field", another said they had "nudged my views on how much I trust the conclusions of the research", while another said "the more I have been thinking about these issues the more I have realized that there are important problems that have to do with culture in this community that prevent it from being as good as it could be."

Another common theme was how people's experiences of marginalization or exclusion related to their ability to get funding. One participant said that "[my institute has] to work a lot harder to make my profile look good than with other people because we know that they might face some obstacles getting funding for my work because of [my ethic and educational background]", comments which another participant echoed when talking about setting up a programme to work on outreach with a global community that is currently under-represented within the field. Other participants focused more specifically on issues that they perceived with the Effective Altruism community, as a prominent funder for this work. For instance, one noted how:

> *I think I was lucky in that if I didn't have friends who were well versed in EA and didn't know how to play the game then I probably wouldn't have got funding in the first place. I had a lot of applications that went in but then failed because they didn't have the right layout and language. Then I changed the language but kept the concepts and work I was describing the same and all of a sudden, I started getting funding, which felt kind of horrible!*

A third participant noted how "at this stage, I first of all would not want to take funding from EA sources, largely due to integrity but also I feel I have been blacklisted from the main sources of funding in this space" while another commented that “I have developed a healthy, or maybe even an unhealthy, degree of scepticism towards EA having previously seen it as something benign, or even something aspirational”.

Several participants noted a desire to change and improve the field as a significant impact of their experiences of marginalization and exclusion, with some even seeing this as a positive thing that inspired them in their work. One participant identified themselves as “part of a small-p protestant branch within Existential Risk Studies that is trying to expand it beyond this very narrow origin” another took a similar view:

> *I even feel more encouraged to work on [issues around transhumanism and population ethics] because, for a lot of these things, there is a more inherent female perspective to be had, and so it would be of value for me to work on them. However, it feels a bit lonely to be the only person talking about this and not having a group of people who share this perspective with me.*

Others, on the other hand, saw this work in a more negative light, for instance one said that “I would say these experiences have made me somewhat critical of the field, especially when it fails to consider more universal human values like values that are inclusive of the global south.”

CONCLUSION

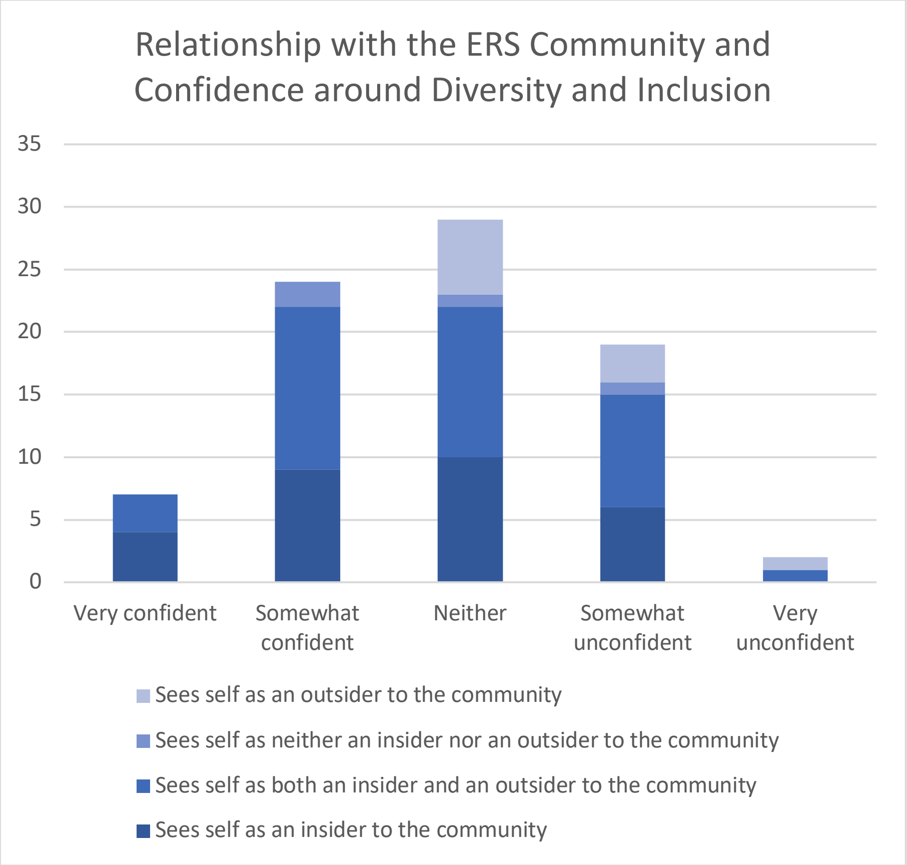
A key lesson from these findings and experiences is that the impacts of homogeneity and exclusion within the field can be subtle and complex. People who have experienced exclusion, or who perceive the field as exclusive, may remain involved with it in a variety of ways, meaning that they may still be a visible presence (for instance as administrators, collaborators, or supporters) that create an illusion of inclusivity. However, it still seems that many people experience exclusion or marginalization and that this can prevent them from feeling like true insiders within the field or from participating as much as they otherwise might.

Individual experiences around the impacts of exclusion also vary. However common themes include people developing negative attitudes towards the community as a whole, struggling to access funding or becoming unwilling to ask for support from certain bodies, and feeling the need to work on making the field better. Together these create a situation in which individuals who have experienced marginalization or exclusion come to lose access to both social and financial capital while also feeling prompted to do additional work, which may itself be unsupported or seen as unwanted or controversial. Therefore, even though these experiences of the impacts of homogeneity and exclusion are varied and unique, we can see how they can combine into a general barrier that makes it hard for individuals to progress within the Existential Risk Studies community.

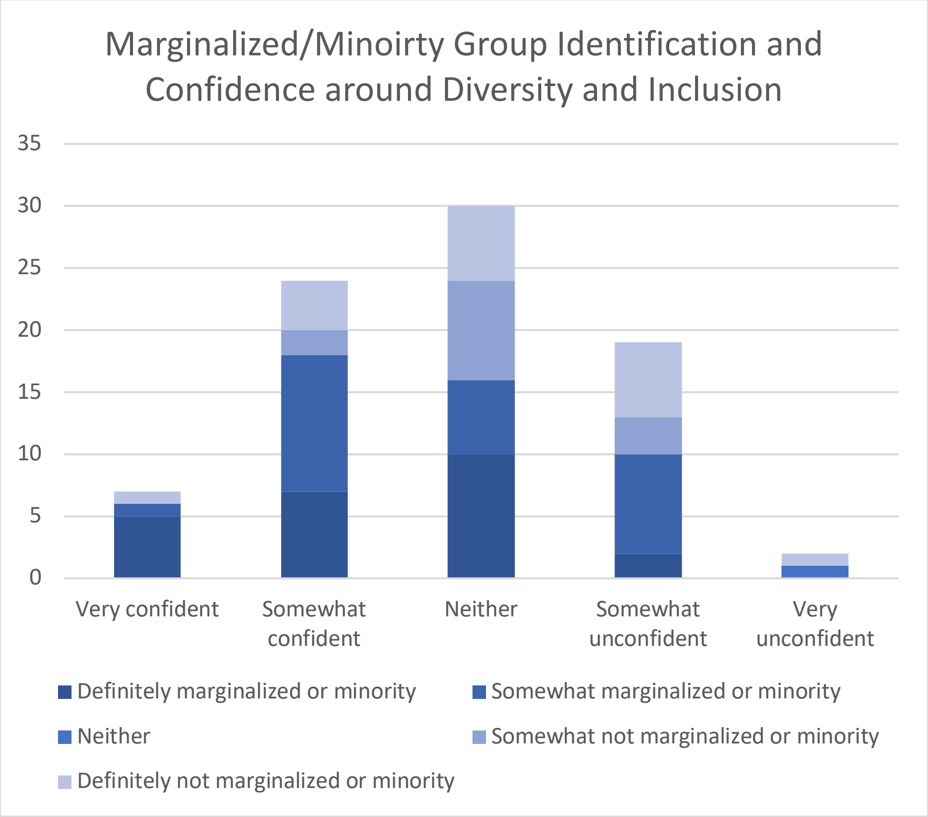
SECTION 4

Confidence and Awareness

The final question in our diversity and inclusion survey asked, “how confident do you feel in your understanding of, and ability to respond to, issues affecting marginalized and minority groups in the community of Existential Risk Studies?” We have plotted people’s response to this question against all the other factors considered thus far below.

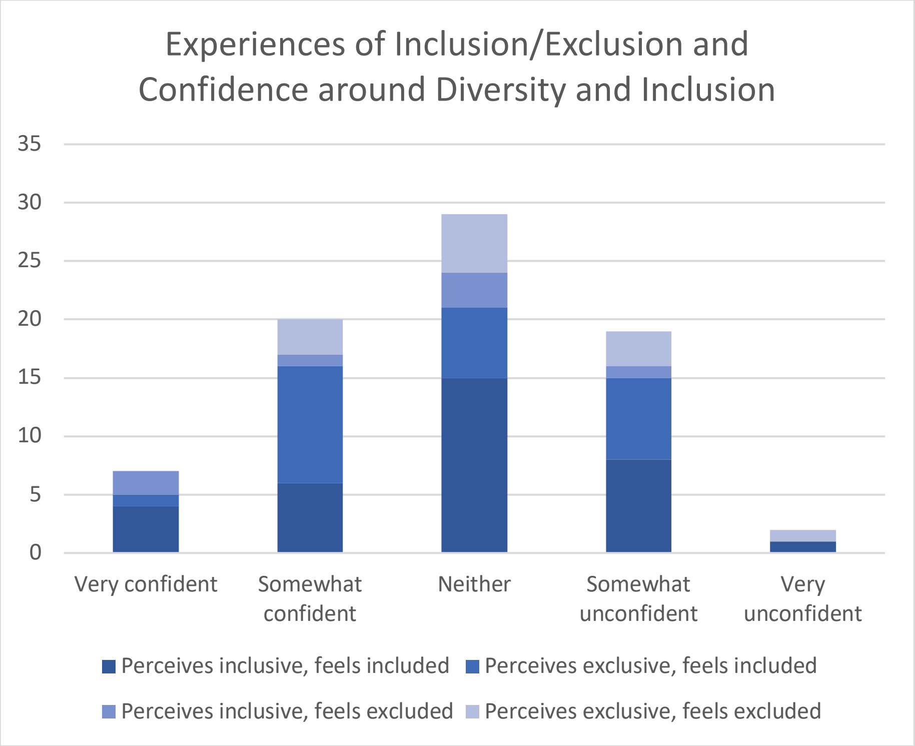


This chart shows a surprisingly strong relationship between people’s confidence in dealing with issues around Diversity and Inclusion and their perception of their own relationship with the field of Existential Risk Studies. In theory, one would not expect these things to be strongly related, yet nobody who saw themselves as straightforwardly an outsider to the community felt confident in dealing with these issues and nobody who saw themselves as straightforwardly an insider felt very unconfident. Furthermore, there is at least some indication that those who saw themselves as insiders felt more confident than those who did not. On the one hand, this could be taken as a sign that confidence around diversity and inclusion is seen as important and conferring insider status in the community; however, on the other hand, it could simply be that people who perceived themselves as insiders felt generally more confident about issues in general (whether or not this was reflected in their actual abilities) and/or were more likely than insiders who did not feel confident to complete this survey.



This chart shows that people who identify with marginalized and minority groups also feel somewhat more confident about issues around Diversity and Inclusion, although this relationship is not very strong. For instance, 5 out of 7 people who felt very confident said they definitely belong to a marginalized or minority group, while only 2 of the 19 people who said they felt somewhat unconfident (and neither of the two people who said they felt very unconfident) said they belonged to such a group.

Confidence and Awareness



Finally, this graph shows that there is, surprisingly perhaps, very little relationship between people's experiences of diversity and inclusion and their confidence in these issues. Across all three graphs we can see that the most popular answer among both individuals as a whole and most sub-groups within our survey population was that they were neither confident nor unconfident, with a few cases where individual groups tended towards being somewhat confident (those who perceived the field as exclusive but felt included, identified somewhat with a marginalized or minority group, and felt they were both an insider and an outsider to the field), in all cases representing a group where people expressed a somewhat nuanced or sophisticated identity.

This question also received fewer written comments from survey participants, while several who commented expressed uncertainty about how to evaluate their confidence. Of the comments received some themes included feeling more confidence in one domain than others (e.g. "very confident in describing issues that directly affect me, but I couldn't say much on e.g. race beyond the usual truisms"), an awareness that people understand these issues in different ways (e.g. "while I can make a full-throated classically liberal/libertarian defence of individual diversity, this is often at odds with the discourse currently used for discussing marginalized and minority groups"), and a desire for further work in this area (e.g. "I would feel more confident if there was more research on the topic").

In order to provide further specificity on participants confidence in issues around diversity and inclusion we asked interview subjects in which ways they include considerations of diversity and inclusion in their work and why they did so (or didn't do so).

Perhaps the most significant way in which people sought to bring diversity and inclusion into their work was by featuring, and engaging with, a wider range of viewpoints and perspectives. As one participant put it:

> *In general, I think that when working on catastrophes it is very easy to understand what elites think but very hard to think what common people think, it's what I call the 1% view of catastrophe. However, we are feeding into this problem when we repeatedly engage with elites in our work and amplify their voices.*

People sought to bring different perspectives into their work in a variety of ways. This included relatively simple and tangible things, for instance "make a spreadsheet of participants and code things like their racial and gender identity", "the inclusion of junior collaborators in the project or taking appropriately issues of diversity into account", or "[for a] hiring round, we explicitly said that apart from some countries we legally could not hire from everyone was eligible to apply, regardless of characteristics." For other's however the focus was on less tangible, deeper, forms of inclusivity. One participant said that "in terms of framing existential risk, I try to frame it in a way that speaks to lowest common denominators, for example parenting and disability or gender identity" while another said that "I am constantly scouring different fields to find perspectives that may have value in being applied to existential risk". Still, several participants said that they felt it was more important to include considerations of diversity and inclusion in some projects than others: for instance, in foresight exercises, research activism, or work on collapse; but not in technical AI safety research.

Apart from viewpoints however, people seemed to have less to say about diversity and inclusion of people. One participant talked about the need to avoid extractive research and cultural appropriation and their "strict rule about not allowing students to study countries they are not from, except under very specific circumstances". While another spoke approvingly about what their organization did to make people feel welcome "when people start working, at least at my organization, we have plenty of policies with leadership from specific staff… who make an effort to highlight different things that people might celebrate at the organization so that they felt comfortable being themselves and things like that" while another participant spoke approvingly at schemes run in another organization on being inclusive towards LGBT and disabled people. Several female participants, and one male participant, spoke about their experience supporting colleagues with issues around sexual harassment and the awareness this had given them of imbalances of power within the community.

Participants gave a range of reasons for why they thought diversity and inclusion were important. These included:

- experience working in low-resource settings;
- personal experiences of inclusion and a desire to pay it back/forward;
- personal experiences of exclusion and a desire to fix this for others;
- the creativity of working with new perspectives and ideas;
- the need to engage constructively with the critical perspectives of others;
- the exposure to new ideas that comes from interdisciplinarity and social media;
- engaging with a wider range of policymakers and influencers; and
- fear that by being exclusionary and homogeneous the field is missing out on a lot of talent and expertise.

However, they also spoke of many reasons that make this work difficult. For instance, the 'epistemic injustice' that centres some voices while excluding others and makes the pronouncement of some privileged people sound like objective facts while dismissing those of less privileged people as subjective or political. Another said that focusing too much on diversity and inclusion would be seen as a distraction by certain funders and lower the chance of receiving support. Several respondents indicated that they were either unsure of the value of diversity and inclusion for their own sake, especially when dealing with technical problems such as those in AI safety, or that they had experienced it being done lazily and in ways that made work worse. However, many simply said that they lacked the time or resources to give these issues the attention they thought they deserved, for instance one participant regretted that "I am not as thoughtful as I could be about making sure things like materials and slides are accessible and the reason for that is that they are usually very last minute things that I put together", another felt that "one thing that makes it more difficult is that my network in this field is relatively limited", and a third felt that "for me to do more would require way more understanding of what is going on in other countries."

One participant, in particular, spoke movingly of their sense that they had only been able to become confident in issues around diversity and inclusion through trial and error:

> *I think I am relatively good at building some aspects of diversity and inclusion into my work, probably from being a relative dickhead in the early years of my research as well as trial and error in trying to be more empathetic and do better research. I was*

probably pretty awful when I started out but now, I feel that trying to give opportunities, mentor, and specific support to students from lower income countries is helping me develop.

In environments where diversity and inclusion are often not made priorities this kind of experience can sadly be inevitable, and yet we also inhabit situations in which making mistakes can lead to very negative personal consequences for researchers. The consequence of this is, sadly, very often that people develop risk averse mindsets of not even trying because they feel they do not have the support required to succeed.

CONCLUSION

A key lesson from these findings is that while most people in Existential Risk Studies, or at least among those who responded to this survey, have some awareness of diversity and inclusion, very few seem to feel very confident in responding to these issues. Where confidence does exist it seems to be among those who feel most secure as insiders within the field; however, it also focuses more on the intellectual inclusion of diverse viewpoints and ideas than on how to be inclusive towards different people and their needs. This feels like an urgent need for further work, both in helping people to understand how they can be more inclusive in the work that they do and ensuring that they have the support to do this and to learn and grow through doing so. It also seems like the more we can have marginalized people who also feel like they are insiders for the field, the more that they will be able to confidently share their knowledge and experience with others for the benefit of the entire community.

SECTION 5

Current Issues and Concerns

Interview participants spoke about a wide range of concerns with the field of Existential Risk Studies and the community around it. These can broadly be classified as falling into two groups, correcting the historical under-representation of particular groups and responding to current challenges and controversies.

CORRECTING UNDER-REPRESENTATION

In the first group of challenges, participants spoke about a wide range of under-represented communities and how this negatively impacted on the field. Several participants spoke about the under representation of women both as something that lead to the exclusion of important voices and perspectives (for instance one remarked that “sometimes I feel that women are being talked about rather than discussed with”) and also as a causal factor in wider cultural issues within the field (for instance another suggested that “my impression is that if there are too few women in the room then there is a certain atmosphere or jargon that can emerge and that I don’t appreciate”). Another concern that many participants expressed was that the community was very geographically focused on a few small areas in wealthy countries and that anyone who was not located in one of these centres faced significant barriers to entry.

> *The most outstanding feature of this community that drives non-diversity is the geographical concentration of research. We have a very strong Oxbridge community and something similar in the [San Francisco] Bay Area and a few other places. If one could make this a more truly global field, then I think that could contribute in a very productive way to making the field more diverse in various ways. Both intellectually diverse and in respect to people’s background and so on.*

Other participants explicitly focused on the exclusion of people from poorer countries; “I feel that the field is very much based in the global north and very hermetic towards considering international applications or mechanisms that might assist in reducing existential risk.” Interestingly characteristics like race, ethnicity, and religion were discussed far less than this geographical exclusion, at least in terms of under representation, although for some participants it seemed like this may be because the under-representation of non-white people within the community, and how bad this is, was taken as a given (although the issue was discussed in relation to eugenics, see below). On the other hand, several participants focused specifically on the issue of class and educational background; one participant noted that “almost everyone comes from an economically privileged background and has come through elite institutions” and believed this was the reason the community so infrequently considers issues around inequality and injustice. Another, non-white, participant feared that people's reluctance to talk about class, in particular, could mean that we missed many important intersectionalities "because [coming into this community] just like I did would still limit us greatly to those few brown people, women, or anyone else who is different, coming from an elite institution, speaking well, and knowing how to behave at a formal dinner".

Participants expressed different views about two other marginalized or minority groups. As one participant put it "compared to the population at large the field is probably over representative in terms of cognitive diversity and in terms of LGBTQ+ people", this is a sentiment SJ has heard from others and it was echoed by at least one survey respondent who said that the field was "Inclusive along some dimensions -- neurodiversity, LGBT+, etc. -- but generally exclusive along others -- class, ethnicity, philosophical/methodological views". However, interview subjects who belonged to these communities suggested that the field may not be as inclusive as others assume. For instance, one participant noted that acceptance was dependent upon still being able to achieve against conventional notions of excellence:

> *When people say "so long as you are doing good work it doesn't matter what you are" that can be kind of liberating, because at least you feel like people aren't going to be whispering about you. However, it's also kind of the minimum; we [autistic people] don't all fit into the highly functional groupings like the autistic computer genius you know.*

Similarly, on LGBT issues, people still expressed a feeling of isolation, as one LGBT participant put it "Not seeing anyone like me was just scary and strange; often if you go into a space as a trans person and there are lots of non-trans people there you just go straight to any other trans people and stick with them". One point to note in relation to both neurodiversity and LGBT identity is that both groups are very small, only a few percent of the general population, so that even if a group is overrepresented in this community to some extent it is still easy to feel isolated in the absence of full inclusion and support, because they still remain a small minority.

RESOLVING PRESSING CHALLENGES

In their assessment of the greatest challenges currently facing the field of Existential Risk Studies, interview participants converged on two key issues, power, and money. As one participant put it:

> *The field is politically and economically dominated by a few doners who all have politically and economically aligned values, and lots of money, and a few scholars who are also aligned with the same values that the key funders have.'*

This economic situation manifests in a wide range of homogenizing and exclusionary pressures within the field, as the following quotes from participants describe:

- "I think the field is dramatically and unfairly unequal. To me this is most blatant and abrasive when it comes to which scholars are funded and celebrated within the field."
- "There's very little consideration of issues around inequality. I think there is, interestingly, more consideration around racial and sexual inequality then there is around economic inequality."
- "If you are not on board with the language and the ideology that are most appealing to certain demographics and to quantitative facing people, then you are just less likely to get funding."
- "You are just much less likely to be obsessed with IQ or on board with transhumanism if you are not male, cisgender and white. A lot of it is psychological, if you don't agree with lots of EAs methods, or use the language conventions they expect, then people just start to not invite you or include you in things and that has a negative career impact."

These forces may be acting to make Existential Risk Studies homogenous and exclusive, even against the wishes of many of the people in the field. With a few notable exceptions there were few instances of people describing overt prejudice or intentional exclusion; however, many participants appeared to share the view that some people were simply never ging to make it in this area of research and that they privately believed that, often unspoken, dynamics of power and influence were behind this fact. They felt that these tended to shape things so that the researchers who made it were also those who were most closely aligned with the interests of the most elite and powerful people within the community, and even more so the people and institutions who funded it.

These dynamics were also somewhat apparent in discussions of the two instances in which people did point to more concrete instances of 'bad behaviour'; however, these also deserve specific attention. The first of them related to a controversy around comments made by Nick Bostrom, which came to light during the process of conducting this study. Nick Bostrom has been, in many ways, the most senior and influential researcher in the field. He also has a long track record of talking about 'dysgenic pressures', meaning evolutionally forces that might be reducing human intelligence and thus, on his view, humanity's potential. Several researchers have expressed concern that this was being used as cover for eugenical ideas and one of these researchers uncovered a post Nick Bostrom made in the 1990s in which this was much more apparent, including the use of racial slurs. This discovery led to Nick Bostrom posting an apology and justification for his previous post in January 2023, which many people felt was inadequate and only made it look as if he still had, or at least was open to, eugenicist and/or racist beliefs. Many interview participants commented on this controversy, both in terms of the comments themselves and the reaction from the community. One, non-white, participant told the following story about their experiences responding to the comments:

> *I was more upset about this than other people around me. For the first time this felt like if anyone was going to say something about this then I was who that was going to be.... My experience of writing something like that was that I found it really hard because I rarely engage in internet discourse, and it is scary to have opinions on the internet at the best of times.... So, then I posted it and I generally felt good about how my post was received.... [When discussing the comments from other members of the community] .... If someone had written something as inflammatory as what Bostrom wrote but was about a particular sexual orientation I don't think the community would have engaged in as much "but maybe there is something in what he is saying..." and that would be good, because we shouldn't engage in speculation around bigoted opinions about why some sexual orientations might be 'worse' than others. It should be fine to just say "this is bunk and we don't discuss it".*

Two participants who were white also commented that they felt they were in a position to respond to the controversy (or even had a responsibility to do so) because they were relatively secure in their position in the field, as one of them said "I felt able to criticise his apology because I am not working in the same office and don't think I will ever be employed by him and also because I have a reasonable degree of security in the field; However, I am certainly aware of others who did not feel able to do the same." For balance, however, we should also note that one participant felt that they were not able to express their views because, even though they were in a more secure position, they wished to defend Bostrom and this would have been too controversial: "I think it's a witch hunt; I think that Bostrom has nothing to apologise for and I think that the whole things is a disgrace... However, this is a topic where I feel very reluctant to speak publicly because I sense that these views are not welcome".

The other controversy that several participants commented about during the interviews concerned allegations of sexual harassment in the field. This has been an issue of longstanding concern but has been raised to prominence in recent months by an article in TIME that reported on a number of specific allegations within the Effective Altruism community. Some participants saw this as more of an institutional framing within the field, as one argued "as the TIME magazine article on EA and related communities has shown, I think more progress on structures for reporting and resolving cases of sexual harassment and sexual assault would be needed". However, others felt it was more cultural "I would like to see people become more educated and for this to shift the discourse norms a bit … after the TIME article I encouraged more of my guy friends to publicly say this was terrible because I felt a lot of my women friends were nervous that it had mostly been women saying how terrible it was". It was not the goal of this report to make specific recommendations around sexual harassment; although it is clearly a huge barrier to diversity and inclusion it is also a problem that deserves more specific attention and that neither the authors of this report or our methodology (which was agreed prior to the emergence of these controversies) are best suited to trying to resolve.

CONCLUSION

A key lesson from these findings is the importance of paying attention both to long term and systemic dynamics that drive homogenization and exclusion and also to specific instances of bad behaviour. These events also highlight the importance of not only thinking about resolving challenging and controversial situations but also actively supporting those who have been negatively affected by, and/or who are already seeking to resolve, them. We shall therefor close this section with another quote from one of our participants, who has been in both of these situations, and that we heartily endorse.

> *Basically, my view is that we should have focused on diversity 10 years ago and the consequences of not doing that are things like the TIME article. So, if you don't want us to face challenges like that in the future then we should change what we are doing now.*

SECTION 6

Positive Futures

We asked all participants to consider possible futures for the Field of Existential Risk Studies, in which the field had become a lot more diverse and inclusive, and to think about how this might have happened and what the field could have gained as a result. Their responses can be brought together under three headings, inclusive leadership, diversified funding, and better representation, and we will let their voices, and the contributions that might exist between them.

INCLUSIVE LEADERSHIP

“The field should actively embrace diversity and inclusion, especially in positions of power and leadership, and this should be more than nominal support but involve active initiatives to ensure people from many different backgrounds all have good opportunities.” “Initiatives like the ILINA fellowship are where we need to go - we must help new people get a voice in these fields.” “These initiatives should be designed based on a good understanding of what might work well, to the extent that anyone knows what that is.” “To enable more racial diversity, we are also going to need to promote more methodological diversity and champion different forms of knowledge making, rather than requiring everyone to use the same forms of modelling, statistics, and philosophical discourse.”

“A first step is trying to take account of the implicit norms, biases, and rhetoric of the field as it stands and provide a clean demographic breakdown for what the field is currently like. I think there is a real supremacy of rationalist modes of thought that is not acknowledged enough, and I say that as someone who does mathematical modelling as part of my research so I’m not just a ‘qually’ dissing on ‘quants’ here.” “On the norms, I think a lot of this requires top-down change so that organizations will have different organizational practices, like better conflict of interest procedures or stricter rules on the interface between grant making and personal relationships.”

“I imagine some kind of reckoning with the leadership of the field. Bostrom has been a key figure shaping the field for 20 years now and has made some very unfortunate and racist comments, and in apologising for it only made it worse. I guess we need to do a little more thinking about whether we really think dysgenic pressures, or overpopulation and underpopulation (not to mention the makeup of those populations) are existential risks and how they compare with other risk drivers like nuclear war. However, I have less of a sense of what needs to change with regard to this compared with how we can improve recruitment and retention of women and non-white people.” “The cynic in me says an external shock, something like another crypto crash or funder collapse that both induces further soul searching and gives a competitive advantage to researchers who are not part of the establishment around the Techno-Utopian Approach, forcing a more pluralistic and democratic approaches to existential risk".

DIVERSIFIED AND PROFESSIONALIZED FUNDING

"Diversified funding sources so not everyone has to go through EA." "Creating funding sources for people who are in fields that have obvious implications for existential risk but don't actively work on these questions, like synthetic biology or the burgeoning field of geoengineering. That would provide opportunities and incentives for more people to do this work and thus broaden the epistemic status and trans-disciplinary nature of this field." "I think that funders are likely to become more willing to employ standard due diligence requirements for institutions – that could include publishing more of their internal stuff and having more transparent recruitment processes." "I imagine Existential Risk Studies as almost turning into a kind of area-studies, but where the area is the terra nucleus on the other side of human extinction."

BETTER REPRESENTATION

"The 'what' might have happened is that there would be a lot more people who look like me in the field." "The problem is that the current makeup of the field is very un-diverse. Solving that would mean ending up in a field where groups were represented proportionally, so there would be 50% women and other groups represented, perhaps in proportion to the British population (for a British organization) or maybe even the global population, and certainly a lot more representation from the global south. How would we get there? Certainly, by paying a lot of attention to it in recruitment, putting out adverts in places for those kinds of audiences, continuing to do stuff around internships and other things to support people earlier on in their careers." "Inclusive summer research programmes like they have at GPI would be awesome, although I know that money is an issue. Advertising research positions specifically for existential risk in the global south and having partnerships with universities in the global south that involve heavily subsidised visitors and a much bigger dialogue." "I would imagine that it would also involve dealing with issues around retention. For instance, getting better structures to deal with sexual harassment and assault and maybe getting more support for parents, something that I think will be more of a problem we have to grapple with in the coming years."

"Speaker events that centre diverse perspectives on existential risk and paying people." "Prioritizing, or even simply recognizing, non-English contributions would be a big start." "Having events in less formal settings.... Not that doing things in the pub is any better but rather simply having a variety of spaces where people do their networking. Having childcare provision. Making a variety of spaces inviting to people. I think that could help. Getting talent from universities we might not have heard of and who speak languages we might not know." "It should also be the case the people from all different backgrounds will actively feel welcome and will feel supported and will have pathways to success within the field." "I would be happy to see more gender parity within the field and I think that might have some knock on consequences of changing norms so that it is also more welcoming for women but I am not entirely sure I expect that to actually reduce the frequency of people being egregiously treated poorly or if that is just a consequence of humans being humans and there will always be people in the community who do bad things. So, in part you can have better representation, in part better norms, in part you need specific enforcement mechanisms. In the space of gender parity, I would like to see greater focus from organizations on outreach and inclusion."

CONCLUSION

A key lesson from this section was simply how fun it was to listen to people imagining positive futures for our field, and to place their contributions side by side to see the breadth and creativity of ideas put forward. Since Diversity and Inclusion are often treated as a 'nice to have' and not given much resource and space, people who actively commit to working on them often spend a long time working alone and against significant challenges. However, there is also joy and purpose to be found in creating spaces where people can bring their whole selves into the discussion. We imagine the above as emerging from a brainstorming session in which many people feel free to propose what they think might work to make our field better without having to justify everything they say or worrying that if they suggest something then it will be up to them alone to make it happen. Such sessions are not hard to run, yet they will only succeed if there is enough commitment and resource to build on their findings. As a field we should be seriously thinking about how to make that possible!

SECTION 7

A Diversity and Inclusion Strategy for ERS

Drawing on the findings of this report and the suggestions of survey responders and interview participants, we produced a series of recommendations to improve the field of Existential Risk Studies, which we have developed further through conversations with stakeholders across funding, the university, and mentorship within the field. Behind all these recommendations is the belief that leaders and colleagues need to see inclusion as their role, have capability to recognise when exclusion is likely or is at play, and have the skills to create a more inclusive environment. Exclusion doesn't happen in a vacuum; people are part of the inclusion/exclusion process and therefore part of the solution. We all have a role to play in the inclusion of others.

INCREASING THE REPRESENTATION AND INCLUSION OF DIVERSE PEOPLE AND THEIR VIEWPOINTS WITHIN EXISTENTIAL RISK STUDIES

Diversifying ERS cannot be achieved without diversifying the range of people who actively contribute to the field. As one interviewee noted; "If the demographics of the community were different, that would have a huge flow through effect on a bunch of norms." Suggestions for improving this included:

- "Continued progress on creating channels for more non-white people, women, and people from a variety of socioeconomic and educational backgrounds."
- "We should have specific spaces for discussions for people like me, who bring a different background, to answer other types of questions related to ERS."
- "[people should be] more welcoming toward a variety of intellectual perspectives; not just in the sense of letting people present their perspective but also seriously engaging with those perspectives and really trying to push their own boundaries, in terms of learning the essence of other perspectives, embracing them, and making the field intellectually pluralistic."

However improved representation does not mean treating anyone as merely representing particular groups. People can have very different experiences of belonging to a community and may experience different levels of marginalization or exclusion. Furthermore, many people who enter this field are not going to be 'representative' of the groups they come from, as ERS remains elitist in its institutional and geographical location, educational and academic expectations, and the communities it focuses on influencing and engaging with. Hence, this recommendation is more about challenging the under-representation and exclusion of people and their viewpoints.

This involves more than just inviting people to participate, it requires finding ways to target the distribution of resources to people from marginalized groups, to support them and their work and give them the independence needed to develop and present their unique perspectives.

MORE FUNDAMENTALLY EMBRACING DIVERSITY IN HOW WORK IS CONSTRUCTED AND WHO WE 'BRING IN'.

We can also embrace the global and radical perspectives needed to think about existential and global catastrophic risk as a vehicle for thinking about diversity and inclusion in more fundamental ways. As one interviewee put it:

- "It could be that ERS is in a particularly good position to do something about [inclusivity] because we are concerned about HUMAN extinction. How do we keep doing the field and writing the papers with so little diversity? Everyone seems to agree it is an issue, but they still keep doing it."

Not only does this create opportunities for assessing exclusion and injustice within existential risk research, but also for existential risk organizations to take on leadership within their broader institutional or community settings (such as the Effective Altruism movement or the University of Cambridge).

However, we need to be careful to not create an illusion that the field is more diverse or inclusive than it is, or of feeding into elite saviour narratives. Diversity and inclusion will mean different things to different people and creating spaces in which people feel comfortable to challenge one another is difficult and requires leadership; especially as it appears that there are still many people in ERS, including some with more seniority and security, who do not share this ambition. Actively supporting those who challenge or disagree with us is hard but important and this is one reason why diverse leaders and funders are crucial for the field to achieve its lofty ambitions.

RECOGNIZING HOW FUNDING AND LEADERSHIP CAN LIMIT THINKING, HOW POWER MANIFESTS, AND WHAT 'DEMOCRACY' IN THE FIELD COULD LOOK LIKE

The issue of who funds the field, and how this impacts the people in it and the work that they do, was raised by many survey respondents, interviewees, and stakeholders.

- "Changing the political economy of the field, making it less dependent upon key funding sources and industry affiliations."

- “You are far more likely to be included in key decisions, promoted, and receive funding if you align with EA/techno-utopian values.”

- “Funding seems like a major problem -- this is still a weird area that even at elite universities can be hard to fundraise for; it’s even harder in more marginalised settings.”

Of course, it is not surprising to find researchers who feel that the funding landscape they face is unfair; however, in this case we found that even researchers who had secure funding and people who were actively involved in making funding decisions felt there were issues with homogenized and ideologically driven funding allocations.

While many participants raised specific concerns around the Effective Altruism movement and its associated charities; other sources of funding, including non-EA aligned philanthropy and more standard academic funding sources, also have their own problems. We thus need to be open and supportive of people finding opportunities to fund their work that suit their own situation, while also supporting those who wish to challenge funders to change their practices to make them more diversified and less exclusionary. It is likely better for the field to rely on a diverse range of funding sources than to seek to make a small number of existing funders more ‘ideal’ in their decisions. This is also something that individual researchers have more control over as we can actively seek to cultivate and support new funding relationships, rather than hoping that the same few supporters will keep financing our work indefinitely,

CREATING OPPORTUNITIES FOR RESPECTFUL DIALOGUE AROUND EVENTS IMPACTING OUR FIELD AND USING THEM AS TRANSPARENT CATALYSTS FOR CHANGE.

As already mentioned, this research was undertaken during a period when several high-profile events impacted the field, including the uncovering of racist writing by a research leader and accusations of sexual misconduct. While this research has not attempted to respond directly to these events it has uncovered a certain tension, with some members of the community apparently wishing such negative publicity would pass away quickly while others felt this was an opportunity to initiate long needed changes.

- “I hope, and I think it is not unlikely, that last year’s scandals and problems that have been revealed in the community will lead to a significant shake up of the people who have been directing EA, and the Centre for Effective Altruism, for a while. I think it would be good for many of those currently in active positions to find other avenues of work and for other people to take their places. That can also contribute.”

- “Also, to take the field more broadly, as the TIME magazine article on EA and related communities has shown, I think more progress on structures for reporting and resolving cases of sexual harassment and sexual assault would be needed.”

Ensuring that people have opportunities to freely air their concerns and adequate mechanisms for working towards resolution are vital for the future health of the community. As we saw in section 4, we need to combine thinking about controversial topics, and how they are handled, with supporting those who have been most negatively impacted or are doing most to work for change, while also paying attention to the systemic issues that often lie behind such controversies.

ERS both explores potentially catastrophic events and the systemic forces causing them, and we should take this perspective into how we think about our own field. We also need to be aware that scandal or controversy likely arises because of failures of leadership to take opportunities for earlier resolution of difficult issues. Thus, the work we do now could save everyone, but especially the most marginalized, from having to deal with negative situations in the future.

MORE COMPASSION AND ACTIVISM

Our final recommendation goes beyond ‘the community’, to focus on those individuals who are already working hard to make a positive difference. This human element is easily missed but is also invaluable in achieving change. As one participant noted, a more positive future for ERS.

- “…would need more compassion as a field. I know that a lot of people do try to practise generous charitable lives. However, it doesn’t seem to be working out. Moving away from dispassionate rationalism and taking more account of proximity ethics, kindness, and compassion, as soft as they all sound, could make a tangible practical difference to the field.”

Compassion is more than just being emotionally supportive, although that is important and often overlooked in competitive ‘rationalist’ discourse. It requires actively seeking to understand what makes people feel uncomfortable and excluded, and trying to be part of the solution that will change this for them. This requires supporting people who wish to be an active part of improving their community, recognizing the labour this involves, and compensating them appropriately for that. Strong leadership for diversity and inclusion requires ensuring that these topics are never left for people to pursue in their ‘spare time’.

As we saw in section 5, collectively working for, and celebrating, positive change can be a rewarding and joyous process; but only when the work recognised and supported. Unfortunately, at present it is too often left as an additional burden on the already marginalized and excluded. It is time for that to change!

Appendix – Survey Responses

Question 1) How do you see yourself in relation to the community of Existential Risk Studies?

- I see myself as an insider to the community – 29
- I see myself as both an insider and an outsider to the community - 40
- I see myself as neither an insider nor an outsider to the community - 5
- I see myself as an outsider to the community - 10

Question 2) Do you see yourself as belonging to one or more minority or marginalized groups?

- Definitely no - 18
- Somewhat no - 13
- Neither yes nor no - 2
- Yes somewhat - 26
- Yes definitely - 25

Question 3) How inclusive or exclusive do you perceive the community of Existential Risk Studies to be?

- I perceive the community as generally exclusive and feel excluded from it – 12
- While I perceive the community as generally exclusive, I personally feel included in it - 24
- While I perceive the community as generally inclusive, I personally feel excluded from it - 7
- I perceive the community as generally inclusive and feel included in it - 35

Question 4) How confident do you feel in your understanding of, and ability to respond to, issues affecting marginalized and minority groups in the community of Existential Risk Studies?

- Very confident - 7
- Somewhat confident - 24

Appendix – survey responses

- Neither confident nor unconfident - 30
- Somewhat unconfident - 19
- Very unconfident - 2

CROSS TABULATIONS

Questions 1 and 2

	Definitely marginalized or minority	Somewhat marginalized or minority	Neither	Somewhat not marginalized or minority	Definitely not marginalized or minority
I see myself as an insider to the community	9	6	0	4	10
I see myself as both an insider and an outsider to the community	11	17	1	6	5
I see myself as neither an insider nor an outsider to the community	1	2	0	1	1
I see myself as an outsider to the community	4	1	1	2	2

Questions 1 and 3

	Perceives inclusive, feels included	Perceives exclusive, feels included	Perceives inclusive, feels excluded	Perceives exclusive, feels excluded
I see myself as an insider to the community	17	10	1	0
I see myself as both an insider and an outsider to the community	16	10	4	6
I see myself as neither an insider nor an outsider to the community	2	2	0	1
I see myself as an outsider to the community		2	2	5

Questions 1 and 4

	Very confident	Somewhat confident	Neither	Somewhat unconfident	Very unconfident
See self as an insider to the community	4	9	10	6	0
Sees self as both an insider and an outsider to the community	3	13	12	9	1
Sees self as neither an insider nor an outsider to the community	0	2	1	1	0
Sees self as an outsider to the community	0	0	6	3	1

Questions 2 and 3

	Perceives inclusive, feels included	Perceives exclusive, feels included	Perceives inclusive, feels excluded	Perceives exclusive, feels excluded
Definitely marginalized or minority	12	6	3	4
Somewhat marginalized or minority	6	11	2	5
Neither	0	0	0	1
Somewhat not marginalized or minority	6	4	1	2
Definitely not marginalized or minority	11	3	1	0

Appendix – survey responses

Questions 2 and 4

	Very confident	Somewhat confident	Neither	Somewhat unconfident	Very unconfident
Definitely marginalized or minority	5	7	10	2	0
Somewhat marginalized or minority	1	11	6	8	0
Neither	0	0	0	0	1
Somewhat not marginalized or minority	0	2	8	3	0
Definitely not marginalized or minority	1	4	6	6	1

Questions 3 and 4

	Very confident	Somewhat confident	Neither	Somewhat unconfident	Very unconfident
Perceives inclusive, feels included	4	6	15	8	1
Perceives exclusive, feels included	1	10	6	7	0
Perceives inclusive, feels excluded	2	1	3	1	0
Perceives exclusive, feels excluded	0	3	5	3	1

CENTRE FOR THE STUDY OF
EXISTENTIAL RISK

UNIVERSITY OF
CAMBRIDGE