2021-22 Research Agenda and Context

Michael Plant, Clare Donaldson, Barry Grimes, Joel McGuire

April 2021

Happier Lives Institute

Contents

Our vision and mission 3

Research agenda and context 5

Area 1: Foundational research into the measurement of well-being 7

1.1 The nature of well-being 7

1.2 The measurement of well-being 9

1.3 Developing the 'WELLBY' approach: issues with putting changes to quality and quantity of life into a single metric 15

Area 2: Applied research to identify and evaluate the most cost-effective ways to increase well-being 20

2.1 Understanding the causes and correlates of subjective well-being 20

2.2 Cause area analysis 23

2.3 Using SWB to compare the cost-effectiveness of highly-regarded health and development interventions used in low-income countries 25

Area 3: Understanding the wider global priorities context 27

3.1 Longtermism 27

3.2 Animal welfare 29

Concluding remarks 29

Our vision and mission

Let us state the obvious: happiness is important. Even if it is not the sole thing of importance, it is still important. Not only do we seek happiness for ourselves, but we hope others experience it too. In fact, many of us devote considerable time and effort to reducing misery and creating joy for the world at large. Last, but not necessarily least, we want our governments to keep us safe and create the conditions that allow us all to thrive.

These statements of the obvious lead us to ask a question whose answer is not at all obvious: if we want to improve global well-being and, further, do so by as much as possible, what should we do? The world faces many challenges. Resources are scarce. We cannot solve every problem, at least not immediately. Therefore, if we want to have the biggest impact on others' lives, we must prioritise.

Few actors (individuals or institutions) give serious and explicit consideration to how to use their resources to help others by as much as possible. One reason, among many, is that prioritisation is an off-puttingly hard task that requires us to tangle with a knot of complex theoretical and empirical questions. Examples include: What is well-being? How should it be measured? What increases well-being? How should we weigh the interests of those who are alive today against those that may come after us?

Attempts to engage in *global priorities research*, in order to cut through this knot, are nascent. They are associated, in large part, with the *effective altruism* movement, which exhorts actors to do the most good they can with their spare resources. The effective altruism project requires little motivation: helping others is good; helping others by more is better. The movement has produced and pushed forward novel and exciting ideas about how to do good. One example, among many, is that if citizens of wealthy countries want to help others through their charitable giving, they can usually do a lot more good by supporting organisations that provide highly cost-effective, and evidence-based, health and economic interventions to those in absolute poverty, rather than if they donate domestically.

Quite independently of the growth of effective altruism, the last two decades has seen an explosion of research in economics and psychology into *subjective well-being* (SWB), self-reported measures of happiness, life satisfaction, and meaning. Efforts to collect large-scale datasets of self-rated quality of life only started around sixty years ago. Interest began to grow when it became clear that large increases in economic prosperity over time were associated with modest, if any, improvements in subjective well-being—the so-called 'Easterlin Paradox.' More generally, it's become clear that there are systematic differences between what we expect will increase our happiness, and what does in reality.

Subjective well-being measures are now being taken seriously by citizens and policymakers as a complement—if not an alternative—both to gross domestic product as a measure of overall social progress and a means of assessing the effectiveness of individual government policies. They hold the promise of allowing us to measure, with scientific rigour, what makes people's lives go well, after which we can work out what should be done to most impactfully help those lives go better. The use of self-reported happiness, life satisfaction, and meaning, represents a substantial stride forward over our current best methods for working out what helps others. The status quo approach is to use some combination of crude and objective approximations for well-being, such as wealth or health, as well as decision-makers' intuitive judgments of what makes life go well for others.

While these two lines of thought—prioritising the most effective ways to help others and working out what helps people by using individuals' own assessment of their subjective well-being—have a natural and obvious affinity, little work has been done so far to fuse them together. The effective altruism community has not (yet) attempted to evaluate its top recommendations using subjective well-being; subjective well-being researchers have, in just the last few years, begun to evaluate what should be done, but have primarily focussed on policy making in high-income countries. All told, efforts to identify the most cost-effective ways to make people happier from an impartial, global perspective have barely begun.

The Happier Lives Institute (HLI) was founded in 2019 to help address this gap: to undertake and promote rigorous global priorities research using the lens of subjective well-being. We expect some combination of academic and independent research is the optimal mix for this task, with specific outputs depending on factors such as the topic and intended audience. Our firm belief is that a better understanding of how to measure and increase well-being is not only achievable, but can be successfully communicated to decision-makers, public and private, who will then take actions that increase it. The result should be a substantial, wide-ranging, and long-lasting improvement in the experience of life, one that will not be achieved without deliberate effort.

Given all this, our vision and mission are as follows:

Vision: a world where everyone lives their happiest life.

Mission: to conduct and promote clear, useful, and rigorous research into how best to measure and increase global well-being.

Research agenda and context

Our mission is broad and draws on research across multiple academic disciplines, primarily economics, philosophy, and psychology, but also others, such as neuroscience and medicine. The purpose of this document is two-fold.

First, to set out the wider research context that is relevant to this mission: below we specify three research areas. Given the interdisciplinary nature of this project, we expect interested readers and researchers familiar with one field, say philosophy, to be unfamiliar with the topics and existing literature in another, say economics; issues which may appear new and intractable in one discipline may be considered 'old hat' in another and it is important to know this. Therefore, in an effort to bridge such gaps, this document aims to give the reader a lay of the intellectual landscape. Of course, we do not have the capacity to address all these questions in each research area, nor is further work on each of them all equally important.

The second, and more important, purpose is to articulate, within each research area, where additional research seems more (or less) useful, and therefore what our research agenda is for the next one to two years.[1]

We hope this information is useful for fostering coordination with researchers interested in this area; **if you are one of those researchers, we strongly encourage you to contact us**.

In what follows, we set out, in some depth, the three research areas that comprise our research context, and our research agenda within each area. Before doing this, we provide a summary of what these are. In the summary, we also note our recent work, so readers can see how our research has developed.

Our main current focus, and where the majority of our effort will go, is Area 2.3: using subjective well-being scores to compare the cost-effectiveness of highly-regarded health and development interventions used in low-income countries.

[1] For clarity, we can formalise the counterfactual cost-effectiveness of research as follows:

((value per unit of resource counterfactually reallocated) * (expected total of resources reallocated due to the research))/total cost to produce research.

In theory, this neatly generates a cost-effectiveness figure for each research item. In practice, supplying reasonable numbers for the equation is hard to do.

Research Area 1: Foundational research into measuring well-being

Specifically, this concerns issues related to determining the value of outcomes in terms of subjective well-being metrics.

Research priorities:

- Examining how best to convert between different SWB, as well as other, measures (1.2.1)
- Investigating how to compare existence to non-existence using subjective well-being scales: determining the 'neutral-point' (1.3.1)

Recent prior work:

- Understanding the nature and plausibility of life satisfaction theories of well-being (1.1)
- Investigating the comparability of subjective self-reports, e.g. of happiness (1.2)

Research Area 2: Applied research to identify and evaluate the most cost-effective ways to increase well-being

Research priorities:

- Estimating, in terms of SWB, the impact of potentially highly-effective interventions, including: psychotherapy for common mental disorders; cataract surgery for blindness; deworming tablets to improve lifelong earnings (2.3)
- Setting out how different moral assumptions—about what well-being is, the badness of death, and population ethics—alter those cost-effectiveness estimates and may alter the priorities (2.3; 1.3).

Recent prior work:

- A meta-analytic review of the impact of cash transfers, in low income contexts, on SWB (2.3)
- Using SWB to estimate the moral weights of averting deaths and reducing poverty (2.3)
- A problem area report into alleviating pain (2.3.2)

Research Area 3: Understanding the wider global priorities context

Research priorities:

- Exploratory research into the plausibility and implications of the *longtermist* paradigm, the idea that the primary determinant of the value of our actions today is how those actions influence the very long-run future (3.1).

Area 1: Foundational research into the measurement of well-being

Fundamental to any effort to increase well-being is a theoretically sound and empirically robust account of what well-being is and how it should be measured.

1.1 The nature of well-being

In order to measure well-being, we must first establish what it is. Following Parfit (1984), philosophers standardly differentiate three accounts of *well-being*, that is what makes someone's life go well for them: (1) *hedonism,* well-being consists in happiness, i.e. overall positive conscious states; (2) *desire-fulfilment theories*, well-being consists in getting what you want; (3) *objective list theories,* well-being may consist in happiness and/or satisfied desires in addition to other 'objective' goods, e.g. wisdom, love, friendship, and autonomy. As one might expect, the theoretical pros and cons of these three accounts have been heavily discussed in academic philosophy (Crisp, 2008); see here for our short, non-technical summary of this debate. Given this, we do not consider academic philosophy work into well-being to be, in general, high-impact. Is this presumptuous? Are there new perspectives we are overlooking?

Curiously, however, social scientists' efforts to gauge subjective well-being (SWB) measures have proceeded somewhat in parallel to these philosophical discussions. SWB is often taken to have three measurable components: an experiential component (sometimes 'affective', 'happiness', 'hedonic'), an evaluative component (a judgement of how life, or some part of it, is going; sometimes 'cognitive'), and a 'eudaimonic' component (an assessment of meaning or purpose in life) (OECD 2013). While experiential measures seem to fit well within hedonism, it is less obvious how evaluative or eudaimonic measures connect to the canonical philosophical accounts of well-being. We are, or have been, more optimistic about work bringing SWB measures into theoretical alignment with coherent philosophical accounts of well-being.

The most commonly used SWB question is life satisfaction, an evaluative measure. This is typically found by asking, "Overall, how satisfied are you with your life, nowadays" (0 - 10). This, combined with doubts about life satisfaction's relationship to extant theories of well-being, motivated our earlier work: Plant (2020a) investigated the nature and plausibility of life satisfaction theories of well-being. Plant argued that life satisfaction theories are best understood as a type of desire-fulfilment theory in disguise and then pressed two serious objections to such a view. One objection is that the view implies (counter-intuitively) that most animals cannot have well-being as they cannot make an overall assessment of their lives. Given the objections, Plant concluded that life satisfaction theories are not plausible candidates for a theory of well-being. On the basis of this

analysis, we are not currently prioritising additional research on the nature of and plausibility of life satisfaction theories, although we expect this debate will continue. Can life satisfaction theories be rescued from such criticisms? Is this the right way to understand them?

Further work could examine the nature of the relationship between objective list theories of well-being, eudaimonic measures of SWB, and the concept of a meaningful life, as well as their importance.[2] For instance, the connection between Aristotle's original conception of eudaimonic, and so-called 'eudaimonic' measures of well-being, is open to debate; see Vitterso (2016) for a recent collection of some issues in this area.

For clarity, as the name of this organisation, as well as the introduction indicates, we are most sympathetic to the view that happiness is what ultimately matters. Given that, and that there is not very much research into meaning being undertaken, further research into meaning does not seem very urgent.

It is worth emphasising, however, that our research is not *just* of interest to card-carrying hedonists. We intend to present, where possible, what the research indicates best improves each of happiness, life satisfaction, and meaning; on any plausible view of well-being, these are either contributors or constitutes of well-being, and so crucial information for decision-makers.

Selected academic literature:

- Crisp, R. (2006). Hedonism reconsidered. Philosophy and Phenomenological Research, 73(3), 619–645.
- Crisp, R. (2008). Well-being. Stanford Encyclopedia of Philosophy.
- Haybron, D. M. (2016). Mental State Approaches to Well-Being. In M. D. Adler & M. Fleurbaey (Eds.), The Oxford Handbook of Well-Being and Public Policy (Vol. 1).
- Parfit, D. (1984). Reasons and persons. Oxford University Press.
- Plant, M. (2020a) Life Satisfaction and its Discontents. HLI Working Paper
- Nussbaum, M. C. (2012). Who is the happy warrior? Philosophy, happiness research, and public policy. International Review of Economics, 59(4), 335–361.
- Sumner, L. W. (1996). Welfare, happiness, and ethics. Clarendon Press.
- Wolf, S. R., & Koethe, J. (2010). Meaning in life and why it matters. Princeton University Press.
- Vitterso, Joar. (2016). Handbook of Eudaimonic Well-Being. Edited by Joar Vittersø. International Handbooks of Quality-of-Life. Cham: Springer International Publishing.

[2] We note accounts of meaning may suffer the same challenge noted for life satisfaction theories, namely that many animals seem incapable of evaluating their lives as meaningful.

1.2 The measurement of well-being

1.2.1 Can subjective states, for instance happiness, be measured in theory?

A long-standing worry is that it is not possible, even in theory, to measure subjective states such as happiness or life satisfaction. The recent view in philosophy of science seems to be that subjective well-being is, in principle, measurable in just the same way that other *latent* (i.e. unobservable) constructs, such as intelligence and personality, are (Anger, 2011, 2013; Alexandrova and Haybron, 2016). This consensus relies on the *construct validation* theory of measurement. Thus, claiming that subjective well-being is not measurable because construct validation is false would lead to the arguably implausible conclusion that huge swathes of social science, which also rely on construct validation, are mistaken.

Given this issue is foundational, we thought it merited double-checking ourselves (HLI, 2020a). Our update was that the construct validation approach seems unavoidable and unobjectionable. To explain, in brief, the idea behind construct validation for latent phenomena is this. Although the *constructs* (particular phenomena or attributes) are latent, you assume there are *measures*, ways to elicit the observable indicators of the construct. To check if the measures are *valid*—that is, successfully measure the underlying constructs—social scientists engage in a process of *construct validation*, where the measure of the construct is tested to ensure that it behaves *in the way we think it should*, given the researchers' existing understanding of the topic. As a simple example, if you expect more intelligent people to earn more on average, and that IQ scores measure intelligence, then higher IQ scores should be associated with higher earnings. Whether a construct is valid is ultimately determined by a judgement, as opposed to a test, when looking at the full sweep of evidence. It wasn't clear to us what other method one would use to reasonably assess if a measure was any good.

Part of our internal confusion on the topic stemmed from the fact that social scientists talk about different types of validity (e.g. 'face', 'content', 'discriminant', 'convergent', etc.) and it wasn't clear how they fitted together. The report clarified this by tracing the intellectual history of such terms (HLI, 2020a). Roughly, the concept of validity has evolved over time, but the overarchingly important sense of validity is construct validity (does a measure behave as expected?) and the other types of validity just test aspects of that.

Is the conclusion that construction validation is unavoidable and unobjectionable correct, or too bold? Is there anything more to be said on this topic?

Selected academic literature:

- Angner, E. (2013). Is it possible to measure happiness?: The argument from measurability. European Journal for Philosophy of Science, 3(2), 221–240.
- Angner, E. (2011). Are subjective measures of well-being 'direct'? Australasian Journal of Philosophy, 89(1), 115–130.
- Alexandrova, A. (2012). Well-Being as an Object of Science. Philosophy of Science, 79(5), 678–689.
- Alexandrova, A. (2016). Is well-being measurable after all? Public Health Ethics, 10(June), 1–15.
- Alexandrova, A., & Haybron, D. M. (2016). Is Construct Validation Valid? Philosophy of Science, 83(5), 1098–1109.
- Hausman, D. (2015). Valuing health: Well-being, freedom, and suffering. OUP.

Relevant informal literature:

- HLI (2020a), Review of the validity of subjective well-being measures, HLI internal report

1.2.2 Given subjective well-being can be measured in theory, how good are our current measures? How can we convert within and between SWB and other measures?

On the construct validation framework, a measure is deemed *valid* if it succeeds in capturing the underlying construct it is supposed to be capturing. One might accept that subjective well-being can be measured in theory but deny that the current measures are, in fact, valid. Social scientists often tend to claim the balance of evidence shows measures of subjective well-being are valid because they behave in the way we expect them to. For example, we expect richer people to be more satisfied, at least up to a point, and that is what the data shows, indicating that the purported measures of life satisfaction do measure life satisfaction; see OECD (2013) for an excellent overview.

Broadly, we are satisfied that the various purported measures for each component do capture that component, e.g. that life satisfaction and the Cantril Ladder, two evaluative measures, both convey information about how individuals judge their own lives (OECD 2013).

We are also satisfied that evaluative measures of well-being are reasonable, but non-ideal, proxies for experienced measures of well-being, i.e. those of happiness. The two share a moderate correlation, the same things seem to increase happiness and life satisfaction, but some things are more important for one than the other, e.g. income matters more for life satisfaction than happiness (Boarini et al., 2012). Hence, we can generally assume life satisfaction and happiness scores will

indicate the same things are good(/bad) but may differ over the overall priority ranking. This needs to be borne in mind.

There are, however, open questions.

First, what would be the theoretically ideal measure for each account of well-being, and which of the current options is closest in each case? Usually, measures are described as valid or not. But presumably validity admits of degree; measures can be more or less accurate. To illustrate, the Positive and Negative Affect Scale (PANAS), the Day Reconstruction Method (DRM), and the Experience Sampling Method (ESM) are all regarded as validated measures of happiness. Should we say they are all equally good as measures of happiness as each other? In which case, what would we do if they disagreed?

Second, and relatedly, how different are the results for various measures, both among measures of a given component and across the components? And how can we convert between them and other proxies for well-being, e.g health measures? This matters because SWB is often measured in different ways and, if we want to know what the result would be in terms of a particular measure, we have to know how to convert or adjust between the measures.

Given the prominence of life satisfaction data (a measure of how individuals judge their lives) and our primary interest in happiness (how good/bad individuals feel during their lives) the priority question is: How best can we convert from the former into the latter?

We are aware of some efforts to convert not just between measures of SWB, but between SWB and other metrics, such as various health scores, including Quality- and Disability-Adjusted Life Years (QALYs and DALYs) (Layard, 2016). We are not aware of anything reasonably comprehensive.

Third, which measure of each component should be used in practice? Plausibly, the most accurate measures of well-being, and thus those we would ideally use given unlimited resources, are not ideal in reality, given limited resources. This might be because more accurate data-collection methods are more effortful for subjects, and so less practical to use. For instance, the Experience Sampling Method, which requires subjects to say how they feel many times a day, requires more work than using the Positive And Negative Affect Scale, which asks them about a range of recent emotions and can be done in one go. To be able to make recommendations about which measures are mostly practically useful, information on the trade-offs and adjustments is needed.

Selected academic literature:

- Diener, E., Lucas, R., Schimmack, U., & Helliwell, J. (2010). Well-Being for Public Policy. In Well-Being for Public Policy.
- Diener, E., Inglehart, R., & Tay, L. (2013). Theory and Validity of Life Satisfaction Scales. Social Indicators Research, 112(3), 497–527.
- Dolan, P., & White, M. P. (2007). How Can Measures of Subjective Well-Being Be Used to Inform Public Policy? Perspectives on Psychological Science, 2(1), 71–85.
- Dolan, P., Peasgood, T., & White, M. (2008). Do we really know what makes us happy? A review of the economic literature on the factors associated with subjective well-being. Journal of Economic Psychology, 29(1).
- Dolan, P., & Metcalfe, R. (2012). Valuing Health. Medical Decision Making, 32(4), 578–582.
- Layard, R. (2005). Happiness: lessons form a New Science. London: Allen Lane.
- Layard, R. (2016). Measuring wellbeing and cost-effectiveness analysis using subjective wellbeing. What Works Centre for Wellbeing
- Mukuria, C., & Brazier, J. (2013). Valuing the EQ-5D and the SF-6D health states using subjective well-being: A secondary analysis of patient data. Social Science & Medicine, 77, 97–105.
- Mukuria, C., Rowen, D., Peasgood, T., & Brazier, J. (2016). An empirical comparison of well-being measures used in the UK (Vol. 2017). Vol. 2017.
- Pavot, W. (2018). The Cornerstone of Research on Subjective Well-Being: Valid Assessment Methodology. Handbook of Well-Being, 83–93.
- Plant, M. (2019). Doing Good Badly? Philosophical Problems Related to Effective Altruism. D. Phil. Dissertation, University of Oxford, chapter 4.
- OECD. (2013). Guidelines on Measuring Subjective Well-being.

Relevant informal literature:

- Foster, D. (2020), Health and happiness research topics, Effective Altruism forum post series

1.2.3 To what extent are happiness (and other subjective) scores comparable?

It is very common for people to give numerical ratings of their subjective experiences. For instance, people often score their happiness, life satisfaction, job satisfaction, health, pain, the movies they watch, and so on, on a 0-10 scale.

A long-standing worry about these self-reports is whether the numbers are comparable and so mean the same thing to different people at different times. For example, if two people say they are 5/10 happy, can we assume they are as happy as each other? More technically, the main (but not only) question is whether subjective scales are *cardinally comparable*: does a one-point change, on a given scale, represent the same size change for different people and at different times?

Opinions are split, largely on disciplinary lines: psychologists are sympathetic to the cardinality assumption, economists are suspicious of it (Ferrer-i-Carbonell and Frijters, 2004). Only a handful of papers have investigated various aspects of this issue, despite its fundamental importance—for discussion of this, see Kristoffersen (2011), Stone and Kruger (2018). However, determining the truth of the matter is difficult because it isn't clear exactly which assumptions are required, in theory, for cardinal comparability and whether, in reality, they hold.

In recent work, we have investigated this topic. Plant (2020b) attempts what we believe is the first comprehensive overview of the topic. Plant notes there are four individually necessary and sufficient conditions for cardinal comparability of subjective data: (1) phenomenal cardinality (subjective states like happiness are perceived in units), (2) linearity (individuals perceive the units of the scale convey equal changes), (3) intertemporality (each individual uses the same end-points of the scale over time), (4), interpersonality (different individuals use the same end-points of the scale at a time). (2) – (4) are about how people interpret subjective scales.

Plant also offers a novel hypothesis for why people might be trying to interpret subjective scales in a cardinally comparable way. Following philosopher of language, Paul Grice (1989), language use is a cooperative endeavour where we try to make ourselves understood. As the meaning of subjective scales is vague and individuals want to be understood, scale interpretation can be understood as what is known in economic game theory as a search for a 'focal point' (or 'Schelling point'), a default solution chosen in the absence of communication (Schelling, 1960). Plant surveys the current empirical literature relevant to each condition and tentatively concludes each condition is met and hence subjective data should be interpreted as cardinally comparable, at least until and unless other evidence suggests otherwise. Plant also suggested some further tests. Since the writing of Plant (2020b), research has indicated there may be individual differences in scale use (Benjamin et al., MS) and that individuals alter their scale use over time (Kaiser, 2020), although it is unclear how substantial the deviations from cardinality would be.

At present, given the scant evidence base, we strongly encourage further empirical tests of the various conditions. These could shed new light on whether the conditions are met, or, if they aren't, the extent to which they aren't, and how to correct for it. For instance, if we knew that, on average, people living in Germany used a scale 10% 'taller' than people living in France we could correct for that to make their answers comparable. While we don't plan to do this work internally,

we would be excited to explore (further) collaborations with empirical researchers working on this topic.

Selected academic literature:

- Benjamin, D. J. et al. (manuscript in preparation) Adjusting for Scale-Use Heterogeneity in Self-Reported Well-Being.
- Ferrer-i-Carbonell, A, and Frijters, P. (2004). "How Important Is Methodology for the Estimates of the Determinants of Happiness?*." The Economic Journal 114 (497): 641–59.
- Grice, Paul. (1989). Studies in the Way of Words. Harvard University Press.
- Kaiser, C. (2020) Using memories to assess the intrapersonal comparability of wellbeing reports. EconStor Preprints 226218, ZBW - Leibniz Information Centre for Economics.
- Kristoffersen, I. (2017). The Metrics of Subjective Wellbeing Data: An Empirical Evaluation of the Ordinal and Cardinal Comparability of Life Satisfaction Scores. Social Indicators Research, 130(2), 845–865.
- Kristoffersen, I. (2011). The Subjective Wellbeing Scale: How Reasonable is the Cardinality Assumption? In Economics Discussion / Working Papers. The University of Western Australia, Department of Economics.
- Krueger, A. B., & Schkade, D. A. (2008). The reliability of subjective well-being measures. Journal of Public Economics, 92(8–9), 1833–1845.
- van Praag, B. M. S. (1991). Ordinal and cardinal utility. An integration of the two dimensions of the welfare concept. Journal of Econometrics, 50(1–2), 69–89.
- van Praag, B. M. S. (1993). The Relativity of the Welfare Concept. In The Quality of Life (pp. 362–385).
- Ng, Y. (1997). A case for happiness, cardinalism, and interpersonal comparability. The Economic Journal, 107(445), 1848–1858.
- Ng, Y. (2008). Happiness studies: Ways to improve comparability and some public policy implications. Economic Record, 84(265), 253–266.
- Plant, M. (2019). Doing Good Badly? Philosophical Problems Related to Effective Altruism. D. Phil. Dissertation, University of Oxford. Chapter 4
- Plant, M. (2020b). "A Happy Possibility About Happiness (And Other Subjective) Scales: An Investigation and Tentative Defence of the Cardinality Thesis." Happier Lives Institute working paper.
- Schelling, Thomas C., (1960). The Strategy of Conflict. Massachusetts: Harvard University Press.
- Stone, Arthur, and Alan Krueger. (2018). "Understanding Subjective Well-Being." In For Good Measure: Advancing Research on Well-Being Metrics Beyond GDP. OECD, edited by Joseph E. Stiglitz, Jean-Paul Fitoussi, and Martine Durand. OECD.

- Taylor, T. (2014). Adaptation and the Measurement of Well-being. Ethics and Social Welfare, 8(3), 248–261.

1.3 Developing the 'WELLBY' approach: issues with putting changes to quality and quantity of life into a single metric

It's common to measure impact using objective measures, such as health or wealth. But what ultimately matters, we believe, is the effect on how people think or feel, that is, on their subjective well-being. On the basis of our understanding of the topics above, we conclude that measures of SWB are valid and comparable. As such, SWB scores offer an invaluable tool: a common currency in which to measure the impact different interventions have on *quality* of life, for instance those which alleviate poverty, enhance education, or improve mental health.

By combining information about how much well-being has changed with objective information about *duration*, we can *quantify* the value of outcomes in terms of a common currency, *Well-being-Adjusted Life-Years*, or 'WELLBYs.' We might specify an outcome which raised life satisfaction by 1-point on a 0-10 scale for one year is worth '1 WELLBY'. Hence, another outcome which raised life satisfaction by 0.5 points for 2 years would also be worth 1 WELLBY, and so on.

We consider translating the value of different actions into WELLBYs, so we can then work out the most cost-effective ways to benefit individuals, to be at the core of what we do. Our projects mostly either improve this method or apply it.

To be clear, the WELLBY approach is not theoretically novel. Victorian ethicists and economists such as Jeremy Bentham and Frances Edgeworth would immediately recognise the idea behind it (Bentham, 1789; Edgeworth, 1881). Putting the theory into practice is, however, new , and attempts to do so have only begun in the last few years (Bronsteen, Buccafusco, and Masur, 2012; Frijters and Krekel, 2019; Frijters et al. 2019; Layard et al., 2020; Plant, 2019),

Clearly, this is a major empirical task, an issue discussed further in area 2, particularly 2.3. In many cases, we expect to generate the first estimates of the impact of different interventions using SWB.

However, it is not only an empirical project. If we want to produce a single, all things considered, cost-effectiveness number in terms of WELLBYs we have to make some philosophical choices, namely[3]:

[3] A further issue, and one we don't discuss, pertains to theories of value aggregation. For instance, do we opt for utilitarian theory of aggregation, where the best outcome is the one with the highest unweighted sum of well-being, or (say) a prioritarian function, which gives more weight to increasing the well-being of the worse off? See Holtug (2015) for a good summary of the options and issues. The utilitarian approach is the standard one used by SWB research, i.e.

- What is the preferred account and measure of well-being? For instance, does happiness or life satisfaction matter more?
- How do we compare improving lives to extending lives? In other words, what is the right account of the *badness of death*?
- How do we compare improving or extending lives with creating or averting new lives? In other words, what is the right account of *population ethics*?[4]
- Where on SWB scales is the 'neutral point' equivalent to non-existence? This is needed to compare improving lives to either extending lives or altering the number of lives.

For an informal discussion of how views of population ethics and the badness of death may alter prioritisation decisions, see Plant (2016), and Cotra (2016) for a reply, as well as HLI (2020b).

These issues—with the exception of choosing between theories and measures of well-being, which we have already discussed (1.1, 1.2)—are elaborated in the next three subsections.

Existing academic literature:

- Bentham, J. (1789). An Introduction to the Principles of Morals and Legislation.
- Layard, R. et al. (2020) When to release the lockdown A wellbeing framework for analysing costs and benefits CEP Wellbeing Policy Group. CEPOP49.
- Bronsteen, J, Buccafusco, J., and Masur, J. S (2012). "Well-Being Analysis vs. Cost-Benefit Analysis." SSRN Electronic Journal, January.
- Edgeworth, F. Y. (1881). Mathematical Psychics. London: Kegan Paul.
- Frijters, P. and Krekel, C. (2019) A Handbook of Wellbeing decision-making in the UK.
- Frijters, P. et al. (2020) 'A happy choice: wellbeing as the goal of government', Behavioural Public Policy. Cambridge University Press (CUP), 4(2), pp. 126–165.
- Holtug, N. (2015) 'Theories of Value Aggregation', in Hirose, I. and Olson, J. (eds) The Oxford Handbook of Value Theory. Oxford University Press, pp. 267–284.
- Plant, M. (2019). Doing Good Badly? Philosophical Problems Related to Effective Altruism. D. Phil. Dissertation, University of Oxford. Chapter 7.

all the data are summed together when assessing overall changes. We don't discuss this issue because (1) we are sympathetic to the utilitarian theory of aggregation anyway and (2) attempting anything else would be extremely impractical. It would involve getting access to, then reanalysing, any study of interest ourselves, rather than using the crude average already given.

[4] Some readers may wonder how all this fits together. Population ethics concerns the issues that arise when the number of individuals who ever live, their identities, and their levels of life-time well-being vary. The standard unit of aggregation in population ethics is the lifetime well-being of individuals. Discussions in population ethics thus usually leave open how the lifetime well-being level of individuals is determined. Different accounts of the badness of death offer differ answers over exactly how individual's lifetime well-being levels are to be determined.

Informal literature:

- Plant, M. (2016) Are You Sure You Want To Donate To The Against Malaria Foundation?, Effective Altruism Forum.
- Cotra, A. (2016) AMF and Population Ethics, The GiveWell Blog.
- HLI (2020b) Moral weights. Happier Lives Institute

1.3.1 Comparing existence to non-existence: where, on SWB scales, is the 'neutral point'?

The most common measure of SWB is life satisfaction, which is usually measured on a 0-10 scale. Which point on the 0-10 scale is equivalent in value for someone as non-existence? What is the correct means for determining this? As noted, this is relevant for comparing improving lives to extending or altering the number of lives. The following explanation is drawn from HLI (2020b).

One might assume that the middle point on the scale—5/10—is the neutral point. SWB researchers sometimes treat the mid-point of SWB scales as where someone is neither satisfied nor dissatisfied, or neither happy nor unhappy, e.g. Diener et al. (2018). This has the controversial implication many of those in developing countries have lives 'not worth continuing'—average life satisfaction in (e.g.) Kenya is 4.4.

However, a neutral point of zero is difficult to reconcile with widely-held intuitions. It implies that a life filled with agony is always worth continuing (at least, considering just the person whose life it is) even if the individual themselves would prefer to die.

Getting the neutral point wrong means resources will be spent inappropriately, either putting too much, or too little, weight on averting deaths vs improving lives, with obvious implications for private philanthropy and public policy.

There is almost no work discussing how, in principle, to determine where the neutral point should be (HLI, 2020b). One implicit assumption is that it is wherever individuals say it is. It's not obvious this is the right approach and it generates *prima facie* oddities: if two people assess their quality of life as the same, but only one of them can be saved, the decision turns on a seemingly arbitrary decision about where they perceive the neutral point to be for themself.

We think conducting work here is an internal priority, at least to sketch some possible approaches, evaluate them, and set out a way forward. This is not least because our initial work into the cost-effectiveness of life-extending vs life-improving interventions is highly sensitive to the choice of a neutral point (HLI, 2020b).

Existing academic literature:

- Diener, E. et al. (2018) 'Revisiting "Most People Are Happy"—And Discovering When They Are Not', Perspectives on Psychological Science. SAGE PublicationsSage CA: Los Angeles, CA, 13(2), pp. 166–170.
- Frijters, P. (1999). Explorations of welfare and well-being. Thela Thesis Amsterdam.
- Peasgood, T., Mukuria, C., Karimi, M., & Brazier, J. (2018). Eliciting preference weights for life satisfaction: A feasibility study.

Existing informal discussion:

- Foster, D. (2019). Health and happiness: Some open research topics. Effective Altruism Forum
- HLI (2020b) Moral weights. Happier Lives Institute

1.3.2 Comparing the value of improving lives to extending lives: different views on the badness of death

Comparing improving to extending lives requires taking a stand between different accounts of the *badness of death*. See Gamlung and Solberg (2019) for a recent collection of essays.

Perhaps the standard view of the badness of death is *deprivationism*, where the badness of someone's death, for them, is the amount of well-being they would have had, if they had lived. This is a product of the number of years they would have lived multiplied by their net well-being (their level of well-being above the neutral point).

Deprivationism is not the only view of the badness of death. Another is the *Time-Relative Interest Account* (TRIA), where the badness of death is a function of the future well-being the person would have had combined with also their psychological connection to their future self at later stages. TRIA captures the common intuition that it's better to save 20-year-olds than 2-year-olds, even though the latter would live longer, because the 2-year-old has a much weaker interest in continuing to live.

A third view is *Epicureanism*, where someone's death is not bad for them. This is perhaps motivated by the sense that existence and non-existence cannot properly be compared in value for someone.

To be clear, these are views about the badness of *death* for someone. On each view we can still count the badness of *dying* for someone, e.g. if painful, and the effect of a person's death on others.

Comparing the value of some life-improving vs life-extending interventions will depend quite substantially on which view is taken. Given there is already an existing philosophical literature, we

do not plan to prioritise original academic work evaluating these views; however, do plan to summarise, in accessible language, the existing arguments for and against the various options to inform decision-makers facing these choices. One open question which may be of practical importance is how exactly to fill in the details of TRIA—there are multiple ways to make precise the view such that it captures the intuition that saving adults is more valuable than saving young children (HLI, 2020b).

Examples of relevant literature:

- Gamlund, E. and Solberg, C. T. (2019) Saving people from the harm of death. OUP.
- Liao, S. M. (2007) 'Time-Relative Interests and Abortion', Journal of Moral Philosophy, 4(2), pp. 242–256.
- HLI (2020b) Moral weights
- Rubio, D. (forthcoming). Death's Shadow Lightened. In Sara Bernstein & Tyron Goldschmidt (eds.), Non-being: New Essays on the Metaphysics of Non-existence. Oxford, UK

1.3.3 Comparing the value of improving lives to creating lives: population ethics

Lots of questions in ethics concern fixed-populations, that is, where the number of people we affect, and who they are, is fixed. Sometimes, however, we encounter variable-population cases, where our actions may determine not just who gets born, but how many people get born, and as well as how their lives go. *Variable population ethics*, hereafter, just *'population ethics'*, concerns the special issues that arise when the number of individuals who ever live, their identities, and their levels of life-time well-being vary (Greaves, 2017). If we want to compare the value of improving the lives of current people to that of adding (/averting) new lifes, we must take (implicitly or explicitly) a stand on population ethics.

The (mathematically) simplest view of population ethics is *totalism*, where the value of an outcome is the sum of lifetime well-being for everyone who ever lives. On totalism, adding happy new lives is good. In contrast to totalism are *person-affecting views*, which hold, in slogan form "morality is about making people happy, not about making happy people." There are a number of ways to make this precise, but the gist is that adding new (happy) lives is neutral in value.[5] There are many other views, but these need not detain us here.

Clearly, how one thinks about this choice will affect the value of interventions that change the number of people who are created, such as family planning interventions that reduce fertility rates.

[5] More technically, that it has no value, i.e. its value is undefined, rather than that it has a value, and that value is zero.

Non-obviously, this may affect the value of life-extending interventions. If parents desire a specific number of adult children, reducing child mortality may correspondingly reduce female fertility.

As for the badness of death, this area also has a substantial academic literature already; we expect to present the views, their implications, and the existing arguments, but not prioritise novel investigation ourselves.

Examples of relevant literature:

- Greaves, H. (2017) 'Population Axiology', Philosophy Compass
- Gustaf. A.. 'Population Ethics: The Challenge of Future Generations'. Manuscript in preparation
- Parfit, D. (1984) Reasons and Persons. Oxford: Oxford University Press

Area 2: Applied research to identify and evaluate the most cost-effective ways to increase well-being

Fulfilling our mission requires determining which actions, for given actors, are the most cost-effective means of increasing well-being. To identify potential priorities, it helps to have a broad understanding of what sort of things impact well-being, and what the barriers are to realising those improvements. From here, it is then possible to narrow down to specific options, then evaluate these for their cost-effectiveness. The following research topics in area 2 progress from quite broad 'background' issues to more specific ones about the cost-effectiveness of particular interventions.

To restate our comment made in the introduction: **our current focus, and where the majority of our effort will go, is area 2.3**: using subjective well-being scores to compare the cost-effectiveness of highly-regarded health and development interventions used in low-income countries.

2.1 Understanding the causes and correlates of subjective well-being

There is now a profusion of research into the causes and correlates of subjective well-being. What are the latest summaries about what sort of things impact well-being, and by how much (Dolan et al. 2008)? Are there particular sets of evidence that are more important to be aware of than others?

Decisions in health prioritisation seem to be quite influenced by the *Global Burden of Disease* report, which identifies how many Disability-Adjusted Life-Years (DALYs) are lost to various diseases. Would it be helpful to construct a similar 'Sources of Lost Worldwide Happiness' or 'Global Burdens of Suffering' report, which identified the lost well-being from various sources (GBD, 2018)?

It seems possible to identify fairly general mechanisms about how subjective well-being functions, such as *hedonic adaptation*, getting used to life changes (Luhmann et al. 2012); *social comparison,* how we assess our lives and experiences in comparison to relevant others (Alderson and Katz-Gerro, 2016); biases in *affective forecasting*, ways in which we mispredict how the future will feel to ourselves and others (Wilson and Gilbert, 2005). What is the latest on these? To what extent might they inform thinking about where the priorities might lie?

Can evolutionary theory offer novel insights by providing an account of what the evolutionary purpose of valenced psychological states (i.e. happiness) is (Graham and Oswald, 2010; Perez-Truglia, 2012; Rayo and Becker, 2007)? Equally, what, if any, decision-relevant information can be gained from understanding of the neuroanatomy of valence states (Kringelbach and Berridge, 2010)?

One long-running topic of interest is the relationship between happiness and income. A common line of thought is that there is no need for governments to focus specifically on well-being: we can just grow the economy and well-being will take care of itself. However, this view is in tension with the so-called 'Easterlin Paradox' (Easterlin 1974; 2016; Kaiser and Vendrik, 2016; Stevenson and Wolfers, 2008). The paradox is the finding that, at least in rich countries, increasing wealth over time does not seem to increase aggregate subjective well-being, even though richer people are happier than poorer people. Is the paradox true and, if so, what does it imply for national policies? How mysterious is it that well-being would not increase in aggregate despite the improvements in many non-pecuniary aspects of life in these countries, such as greater health provision?

We do not plan to conduct original academic research or produce systematic reviews in this area; such work less directly advances our understanding of how best to increase well-being. However, we are likely to produce short reviews both to deepen our own understanding and communicate these issues to our audience.

Example of relevant literature:

- Alderson, A. S. and Katz-Gerro, T. (2016). Compared to Whom? Inequality, Social Comparison, and Happiness in the United States. Social Forces, 95(1), 25–54
- Boarini, R. et al. (2012) What Makes for a Better Life?, OECD Statistics Working Papers.
- Clark, A. E. (2016) 'Adaptation and the Easterlin Paradox', in. Springer Japan, pp. 75–94.

- Clark, A. E. et al. (2018) The origins of happiness: the science of well-being over the life course.
- Diener, E., Lucas, R. E. and Napa-Scollon, C. (2009). Beyond the Hedonic Treadmill: Revising the Adaptation Theory of Well-Being. In: E. Diener (ed.) The Science of Well-being. Springer
- Dolan, P., Peasgood, T. and White, M. (2008) 'Do we really know what makes us happy? A review of the economic literature on the factors associated with subjective well-being', Journal of Economic Psychology, 29(1), pp. 94–122.
- Easterlin, R. A. (1974) 'Does economic growth improve the human lot? Some empirical evidence', Nations and households in economic growth, 89, pp. 89–125.
- Easterlin, R. A. (2016) 'Paradox Lost?', SSRN Electronic Journal. doi: 10.2139/ssrn.2714062.
- Frederick, S. and Loewenstein, G. (1999) 'Hedonic Adaptation', in Well-being: The foundations of hedonic psychology, pp. 302–329.
- Graham, L. and Oswald, A. J. (2010) 'Hedonic capital, adaptation and resilience', Journal of Economic Behavior and Organization. North-Holland, 76(2), pp. 372–384.
- GBD 2017 Disease and Injury Incidence and Prevalence Collaborators, S. L. et al. (2018) 'Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017.', Lancet (London, England). Elsevier, 392(10159), pp. 1789–1858.
- Kaiser, C. and Vendrik, M. (2018) Different Versions of the Easterlin Paradox: New Evidence for European Countries. 11994.
- Kringelbach, M. L. and Berridge, K. C. (2010) 'The functional neuroanatomy of pleasure and happiness.', Discovery medicine. NIH Public Access, 9(49), pp. 579–87.
- Layard, R. (2005) Happiness : lessons from a new science. Allen Lane.
- Luhmann, M. et al. (2012) 'Subjective well-being and adaptation to life events: a meta-analysis.', Journal of personality and social psychology, 102(3), p. 592.
- Rayo, L. and Becker, G. S. (2007) 'Evolutionary efficiency and happiness', Journal of Political Economy, 115(2), pp. 302–337.
- Stevenson, B. and Wolfers, J. (2008) Economic Growth and Subjective Well-Being: Reassessing the Easterlin Paradox. Cambridge, MA.
- Perez-Truglia, R. (2012) 'On the causes and consequences of hedonic adaptation', Journal of Economic Psychology, 33(6), pp. 1182–1192.
- Wilson, T. D. and Gilbert, D. T. (2005). Affective Forecasting. Current Directions in Psychological Science, 14(3), 131–34

- Yudkin, D. A., Liberman, N., Wakslak, C. and Trope, Y. (2016). Measuring Up to Distant Others: Expanding and Contracting the Scope of Social Comparison. SSRN Electronic Journal

2.2 Cause area analysis

With a broad appreciation of what affects human well-being, the further level of analysis is focusing on more specific problems with the aim of identifying the best ways to make progress on each.

2.2.1 Cause prioritisation methodology

Given there are lots of problems out there, it would be helpful if there were a method that allowed us to prioritise between them without, somehow, getting into the 'nitty gritty' of quantifying the cost-effectiveness of interventions.

A popular claim among members of the effective altruism community is that different problems should be prioritised according to an assessment of their *scale*, *neglectedness*, and *tractability*, which can be used as heuristics (MacAskill, 2015; 2018); further, *cause prioritisation* and *intervention evaluation* should be understood as genuinely separate steps that can be done in turn (Dickens, 2016; Wiblin, 2016). Cotton-Barratt (2016) proposes formalising scale, neglectedness, and tractability as three factors that combine to determine the marginal cost-effectiveness of further resources to a problem.

However, if we understand 'causes' as problems and 'interventions' as solutions, it becomes unclear how we can evaluate the cost-effectiveness of problems somehow 'in the abstract', in a way that is prior to, and distinct from, assessing particular solutions to those problems (Plant 2019). In this case, the purported distinctiveness of 'cause prioritisation' from 'intervention evaluation' breaks down; we no longer seem to have a method of prioritising between causes that avoids us getting into the 'nitty gritty' of assessing interventions to them. One potential way to proceed is by engaging in *cause mapping*, where, in short, we try to set the actions one could take to solve a problem, the potential obstacles for each action, and a further assessment of how each obstacle might be addressed; the result is a semi-exhaustive set of options for further investigation (Plant 2019).

In some ways, this is a very disappointing, but perhaps unsurprising, result: we do not have a particular method for quickly evaluating various problems. Rather, we must simply make intuitive judgments about how good the solutions are, then evaluate some of them in greater depth. We do not plan to investigate cause prioritisation methodology further as it is unclear how progress could be made.

Is this analysis mistaken? Is it possible, after all, to evaluate 'causes' independently of evaluating specific interventions? What short-cuts, if any, are available for this?

Existing academic literature:

- MacAskill, W. (2015). Doing Good Better. Faber & Faber.
- MacAskill, W. (2018). Understanding Effective Altruism and Its Challenges. In The Palgrave Handbook of Philosophy and Public Policy (pp. 441–453).
- Plant, M. (2019). Doing Good Badly? Philosophical Problems Related to Effective Altruism. D. Phil. Dissertation, University of Oxford.

Existing informal literature:

- 80,000 Hours. (2019). One approach to comparing different problems in terms of impact.
- Cotton-Barratt, O. (2016). Prospecting for Gold.
- Dickens, M. (2016). Evaluation Frameworks (or: When Importance / Neglectedness / Tractability Doesn't Apply).
- Halstead, J. (2019). The ITN framework, cost-effectiveness, and cause prioritisation - EA Forum.
- Open Philanthropy Project and Karnofsky, H. (2014). Narrowing down U.S. policy areas.
- Wiblin, R. (2016). The Important/Neglected/Tractable framework needs to be applied with care.

2.2.2 Problem area investigations

As a way of identifying promising solutions to a given problem, it seems useful to gain an overview of that problem and understand it from various angles. Questions whose answers may be relevant include: How many people does the problem affect and by how much? What is being done about this problem at the moment? What could be done? What is getting in the way of useful solutions being used? Various examples of 'cause area' or 'problem area' problems from the effective altruism community are given below. Various organisations, such as 80,000 Hours, Founders Pledge, and the Open Philanthropy Project, produce a range of these for the benefit of donors.

As noted, members of the effective altruism community often claim that focusing on poverty and physical health in low-income countries will be the most effective way to help people alive today (MacAskill, 2015). This seems plausible, but is not certainly the case. We plan to evaluate a range of causes, using the lens of subjective well-being, to determine the most impactful ways to do good.

We have recently produced a problem area report on pain (HLI, 2020c). We are currently working on a similar problem area report into mental health.

How useful is it to conduct this sort of overview research? If it is useful, what would be appropriate directions for further work?

Existing (mostly informal) literature:

- 80,000 Hours. Problem profiles
- Founders Pledge. Research Reports
- Happier Lives Institute. (2020c). Problem area report: pain
- MacAskill, W. (2015) Doing Good Better. Faber & Faber.
- Plant, M. (2017). What are the best ways to improve world happiness? Talk EAGlobal London 2018
- Plant, M.. (2018). Cause profile: mental health.
- Plant, M.. (2019). *Doing Good Badly? Philosophical Problems Related to Effective Altruism*. D. Phil. Dissertation, University of Oxford. See chapter 6.
- Open Philanthropy Project. Focus areas
- Whittlestone, J. (2017). Animal Welfare
- Whittlestone, J. (2017). The Long-Term Future
- Whittlestone, J. (2017). Global Health and Development

2.3 Using SWB to compare the cost-effectiveness of highly-regarded health and development interventions used in low-income countries

The question of ultimate interest, and that preceding topics inform, is: what are the most cost-effective ways to increase well-being? Evaluating this requires examining particular options, looking at the evidence, and making estimates. Efforts to assess the cost-effectiveness of outcomes in terms of SWB are in their infancy, having only begun in the last 10 years; in many cases, we expect our estimates to be the first (Bronsteen, Buccafusco, and Masur, 2012; Frijters and Krekel, 2019; Frijters et al. 2019; Layard et al., 2020; Plant, 2019).

Our immediate priority is assessing the cost-effectiveness of several health and development interventions in low-income countries. We focus on these because the effective altruism-aligned charity evaluator GiveWell, has suggested that, amongst all the options available to charitable donors, the most impactful per dollar are certain health and development interventions; their specific recommendations include life-saving anti-malarial bednets and giving large sums of cash to very poor people. These are 'atomic' or 'micro' interventions in the sense they impact one person at a time, as opposed to 'systemic' or 'macro' interventions, such as a public policy change, which work across an entire society.

While these are certainly plausible candidates for the best interventions, these have not yet had their impact assessed in terms of subjective well-being; nor, either, has it been set out how their cost-effectiveness may be sensitive to different moral assumptions (as noted in 1.3 above). Therefore, the natural next step is to estimate the cost-effectiveness of these interventions in SWB using available evidence, and presenting the results given different moral views.

The further step is to analyse other 'atomic' interventions which seem plausibly highly-effective but have not yet been considered top priorities by those in the effective altruism community (GiveWell, 2021). On this list are psychotherapy for depression and cataract surgery for blindness. Focusing on other atomic interventions allows an 'apples-to-apples' comparison: it should be unambiguously clear if new priorities emerge. If they do and our research is used (and, further, we are correct!) this should result in tens or hundreds of millions of dollars a year going to more effective interventions.

In broad terms, our method is as follows. We select interventions that seem intuitively promising to us or others. We survey the literature to find relevant studies that measure impact in terms of SWB (or something we can convert into SWB). We then create models of the average total effect over time and the cost, using a Monte Carlo simulation (rather than point estimates) to account for uncertainty. We build several such models to account for different moral assumptions, as noted in 1.3.

To create these cost-effectiveness analyses requires putting several jigsaw pieces together. So far, we have already conducted, in collaboration with two external academics, a meta-analytic review of the impact of cash transfers for SWB (McGuire, Bach-Mortensen and Kaiser, 2020).

We have also developed a method for comparing improving lives to extending lives using units of subjective well-being, including how this is sensitive to two different views of the badness of death (HLI, 2020b). Further research will evaluate other interventions, expand the sensitivity of our analysis to incorporate different moral views, and refine the estimates.

A substantial literature on these interventions is not provided below; part of the project is to find it. GiveWell's intervention reports are a useful resource.

Attempting this cost-effectiveness analysis raises various methodological issues *in addition* to those already discussed in area 1 above, namely:

- How do we combine various studies, which have different locations, size, outcomes measure, etc. into a single estimate? What sort of quantitative reductions are justified for various qualitative differences?
- What is the most defensible way to account for total per-person effects over time? Interventions plausibly have some impact for several years, but studies generally only measure outcomes directly after the intervention.

We welcome collaboration with researchers who can help improve our analysis.

In the years to come, we want to expand our analysis in two directions. First, to systemic interventions that private donors could fund, such as public health campaigns to prevent, rather than treat, mental illness. Second, to consider which public policies governments should adopt, given limited budgets, to most effectively increase the well-being of their populations.

Relevant literature:

- Bentham, J. (1789). An Introduction to the Principles of Morals and Legislation.
- Bronsteen, J, Buccafusco, J., and Masur, J. S (2012). "Well-Being Analysis vs. Cost-Benefit Analysis." SSRN Electronic Journal, January.
- GiveWell. (2021). Intervention Reports
- Layard, R. et al. (2020) When to release the lockdown A wellbeing framework for analysing costs and benefits CEP Wellbeing Policy Group. CEPOP49.
- McGuire, J., Kaiser, C. and Bach-Mortensen, A. (2020) The impact of cash transfers on subjective well-being and mental health in low- and middle- income countries: A systematic review and meta-analysis. Happier Lives Institute working paper
- Frijters, P. and Krekel, C. (2019) A Handbook of Wellbeing decision-making in the UK.
- Frijters, P. et al. (2020) 'A happy choice: wellbeing as the goal of government', Behavioural Public Policy. Cambridge University Press (CUP), 4(2), pp. 126–165
- Plant, M. (2019) Doing Good Badly? Philosophical Problems Related to Effective Altruism. D. Phil. Dissertation, University of Oxford.
- HLI (2020b) Moral weights. Happier Lives Institute

Area 3: Understanding the wider global priorities context

Our research implicitly aims to improve the lives of people living now or in the near future, i.e. the next 100 years or so. Alternatives to this are, schematically, to focus on benefitting (1) sentient life over the long-run and (2) non-human animals (hereafter 'animals') in the near-term. Hence, if and when we determine the most cost-effective means of helping people in the near-term, we can still ask: How important is this, relative to the alternatives? This is the subject of area 3.

3.1 Longtermism

A recent and increasingly popular idea in global priorities research is *longtermism*, the view that the primary determinant of how much good our actions will have are their effects over the very-long

run, rather than within our lifetimes (Greaves and MacAskill, 2020). If true, may imply a radical reorientation of priorities for society at large, including the Happier Lives Institute.

Greaves and MacAskill (2020) have recently articulated the case for 'strong longtermism' and argued that it is true on a wide variety of empirical and ethical assumptions. This raises the question: under which assumptions, or sets of assumptions, longtermism is not true? Subsequent work could then ask: How plausible are these (set of) assumptions? This is a primarily theoretical task we think merits closer examination.

One, amongst many, specific angles to explore is the significance to longtermism of *worldview diversification*, the idea that altruistic agents should diversify their resources across different interventions or areas depending how strongly they believe in each 'worldview' (Karnofsky, 2016). If this were correct, it would imply agents should allocate some, but not all, of their resources to both near-term and long-term projects. Worldview diversification is sometimes regarded as a candidate solution to *moral uncertainty* (the problem of what we ought to do when we don't know what we ought to do); however, it's attracted almost no attention in the philosophical literature on moral uncertainty. Is there a plausible justification for worldview diversification? What would it imply, practically?

Supposing longtermism is true, what are its implications? How reasonable is HLI's current project—researching how to measure and increase well-being, aimed at impact in the near-term—from the perspective of longtermism? Presumably one valuable, although not necessarily urgent, longtermist project would be determining how to increase the quality of lives over the long-term. Can an understanding of SWB literature inform that task and, if so, how valuable might this analysis be? A preliminary thought is that, as noted above, economic and technological growth may do little to raise aggregate well-being, and hence making lives go better will require deliberate study and intervention.

Examples of relevant literature:

- Bostrom, N. (2003) 'Astronomical waste: The opportunity cost of delayed technological development', Utilitas, 15(03), pp. 308–314.
- Beckstead, N. (2013) On the overwhelming importance of shaping the far future. PhD Thesis. Rutgers University.
- Greaves, H. and MacAskill, W. (2019) The case for strong longtermism. Global Priorities Institute Working Paper 7–2019.
- Karnofsky, H. (2016). Worldview Diversification. Open Philanthropy Project blog

3.2 Animal welfare

A natural challenge with doing the most good, where animals feature, is our ability to be able to make unit comparisons of welfare changes between different species. For instance, how do we determine the level of suffering of animals in factory farms compared to both each other and to the level of happiness (or suffering) that humans experience? Note, the challenge of interspecies comparisons also arises for longtermists when comparing current humans to sentient computers or genetically-modified descendants of humans.

If we assume subjective well-being scores are the least worst measure of well-being for humans, what does this imply, if anything, for whether and how to compare humans to animals, who cannot make such self-reports? Perhaps it will be illuminating to consider under what assumptions it is reasonable to make cardinal comparisons among those able to give self-reports (see 1.2.3) and then assess whether those assumptions persist across species.

Examples of relevant literature:

- Browning, H. (2020) Assessing Measures of Animal Welfare. Working paper
- Charity Entrepreneurship. (2018). Is it better to be a wild rat or a factory farmed cow? A systematic method for comparing wild animal welfare. Blog post
- Ng, Y.-K. (1995) 'Towards welfare biology: Evolutionary economics of animal consciousness and suffering', Biology and Philosophy, 10(3), pp. 255–285.
- Norwood, F. B. and Lusk, J. L. (2011) Compassion, by the Pound: The Economics of Farm Animal Welfare, Compassion, by the Pound: The Economics of Farm Animal Welfare. Oxford University Press.

Concluding remarks

We want to see a world where everyone lives their happiest life. To do that, we conduct and promote research into how best to measure and increase global well-being. This document has set out the broad range of topics this research draws on and, further, has specified what we take the priorities to be for ourselves and fellow researchers who share our goals.

If you have any comments or questions on the Research Agenda, or you would like to work on any of these issues, we strongly encourage you to get in touch at hello@happierlivesinstitute.org.

Oh, and, whoever you are, we hope you have a happy day.