# Wellbeing Distributions: The Democratic Way to Measure Happiness

Population Well-being Lab

UNIVERSITY OF TORONTO

# Wellbeing Distributions: The Democratic Way to Measure Happiness

Anthony McCanny, Sofia Panasiuk, Felix Cheung

TLDR Last month, the What Works Centre for Wellbeing released a report suggesting that when we talk about wellbeing, we should consider not only the average wellbeing in a population but also how that wellbeing is distributed. We strongly support this suggestion but note that choosing a metric for wellbeing inequality shares the same shortcomings as reporting only average well-being: it makes implicit value judgments about what information is important. Seeing as early measures often become widely adopted (e.g. GDP), we believe we have a chance to build a new norm of presenting full wellbeing distributions alongside familiar measures, allowing for individuals and their representatives in government to self-determine what qualities of the wellbeing distribution matter to them.

## 1. ADVANCING WELLBEING INEQUALITY MEASURES

At the Population Well-being Lab, we spend a lot of time talking about wellbeing inequality. When we talk about wellbeing we mean someone's self-reported quality of life including the ability to contribute to the world with purpose and meaning. The reason wellbeing inequality is important is simple: we like to imagine a future world where societies and policymakers use well-being measurements to guide policy, in much the same way that objective indicators like income, education and health are used today. Like most, we tend not to just care about the total or average levels of these resources in society. We also care about their distribution. Most people would prefer equal societies over unequal ones, and we suspect that this preference for equality exists for the distribution of wellbeing as well. Given this ideal, it only makes sense to talk about a society's wellbeing as a distribution. But that is not the norm in practice. In wellbeing policy, the status-quo is to report on the average wellbeing of populations, and when researchers try to understand how policies and events influence the wellbeing of an entire population, they are typically only concerned with the effect on average wellbeing.

But things are changing. The 2023 World Happiness Report released on March 20th includes not only a ranking of countries by average life satisfaction but also a second ranking based on wellbeing inequality. Just three days after that, the What Works Centre for Wellbeing published a report inviting a conversation about the standards of wellbeing inequality measures.

This is an exciting moment to be thinking about wellbeing in policy. The United Nations' 2015 Sustainable Development Goals, explicitly included wellbeing among their metrics. In the years since, New Zealand, Iceland, Canada, Scotland, Wales and Finland have put forward governance frameworks that place population wellbeing and its subjective measures front and center. Not to mention the hundreds of active local governance projects that are using self-reported wellbeing data to guide their policies. This makes it a crucial moment to have conversations about how wellbeing-led policy can best reflect the values of people, and we are glad that thought-leaders like the What Works Centre for Wellbeing have begun this discussion. We are taking this chance to join the conversation, amplifying the call for promoting wellbeing inequality as a key metric, and suggesting a totally new approach. We propose that the best and most democratic metric of wellbeing inequality is actually no metric. Instead we would be well-advised to present full distributions of wellbeing in a population and build the tools to interpret them. The importance of the wellbeing distribution should be central not just to wellbeing policy, but the science that informs it. With this in mind, we will also outline our vision for a distribution-centric approach to wellbeing research.

## 2. WHY WELLBEING DISTRIBUTIONS?

### 2.1. CONSERVATION OF INFORMATION

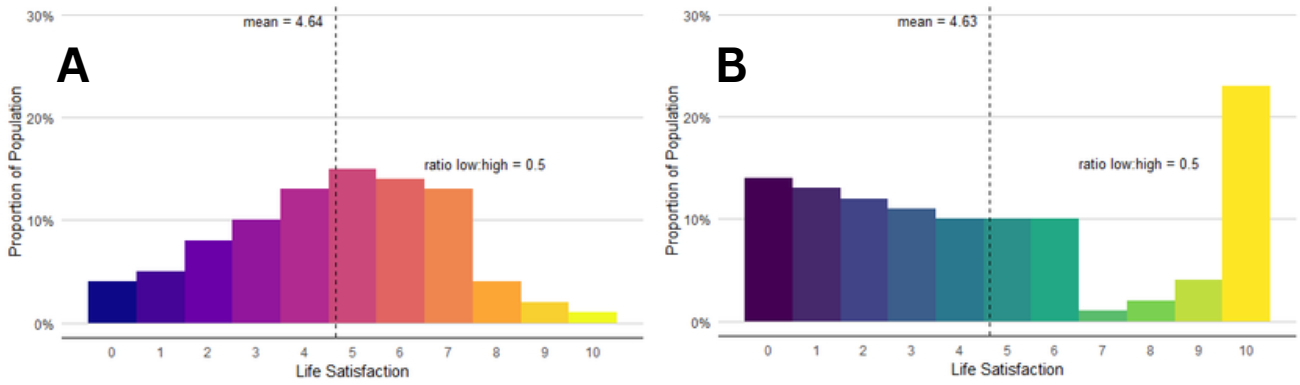
Any metric of inequality only tells part of the story. In their report, What Works suggests four candidate measures of wellbeing inequality:

1. Ratio of Low/High: The ratio of the proportion of people with low wellbeing to the proportion of people with high wellbeing. Low wellbeing in their example is defined as a life satisfaction score of less than or equal to 4/10 while high wellbeing is defined as a score of greater than or equal to 7/10. A life satisfaction score of 10/10 indicates that someone is completely satisfied with their life.
2. Standard deviation: The most common measure of spread in science: the average of the distance of each person's life satisfaction from the average satisfaction, squared.
3. Gini Coefficient: A common measure of income inequality that measures the distance from a society where resources are equally distributed.
4. Shannon Index: A measure of biodiversity in ecology that measures not so much the spread of a value but rather the diversity of wellbeing values in a population.

While each of these has its strengths and weaknesses, what should be clear is that by adopting any of them, we lose information about the underlying distribution. To see how this works, in the following sections we will show how two very different distributions can look exactly the same if all we knew about them was the average wellbeing and a measure of wellbeing inequality.

## 2.2. RATIO OF LOW/HIGH

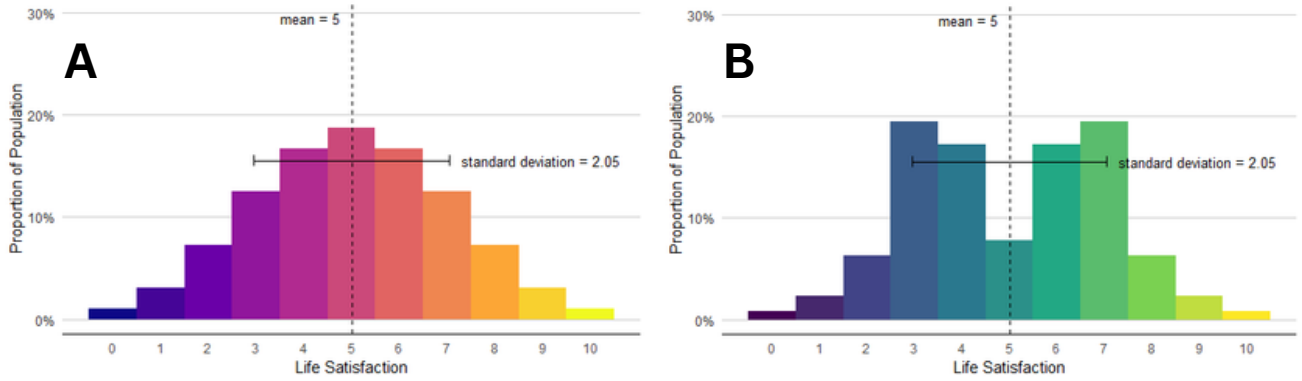
If there are many more people who are self-reporting wellbeing at the bottom of the spectrum compared to the top, it suggests that wellbeing in the population is unequal and being concentrated among the few. However, by grouping together all the people with a reported life satisfaction of 4 and below, and a life satisfaction of 7 and above we lose a lot of information about variation within those categories. Many people intuitively feel like there is a large difference in self-reporting a 4 and a 0, for example. Below we present two hypothetical populations that have nearly identical average population wellbeing and nearly identical measures of ratio of low to high wellbeing.




It's clear that the average wellbeing and the ratio of low/high wellbeing have not told us everything there is to know about these two populations. Would we consider these two populations to have a comparable state of wellbeing? Would we be indifferent between these two states for our own society? We think this is unlikely. Many will consider Population B to have a greater deal of inequality; where nearly a quarter of people are perfectly satisfied with their lives while more than half of individuals are suffering below the midpoint of the scale.

## 2.3. STANDARD DEVIATION

The advantage of the standard deviation over the ratio method above is that it does not group together any large sections of the wellbeing scale when considering inequality. However the standard deviation is a symmetric measure of inequality. Meaning that being a certain distance above and below the mean is weighted the same when calculating standard deviation. Because of this there are many distributions that will have the same mean and standard deviation as we see in these two distributions below.

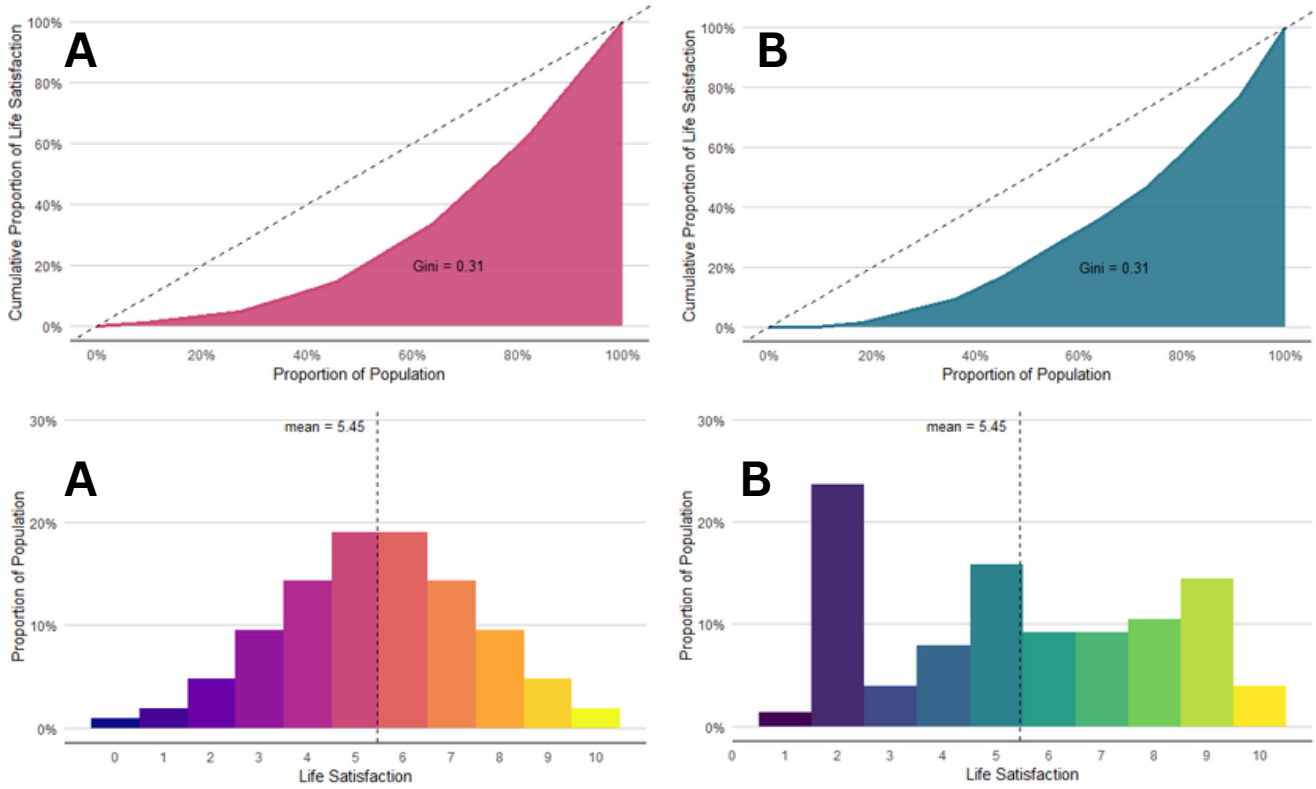



Intuitively, many people may feel like Population B is more unequal than Population A and yet if we used average and standard deviation of wellbeing as our metrics, they would appear identical.

## 2.4. GINI COEFFICIENT

The Gini coefficient is like a fairness score for wellbeing distributions. It's a number between 0 and 1, where 0 means everyone has the same level of wellbeing, and 1 means one person has the highest possible wellbeing while others have none. By looking at this score, we can see how equally or unequally wellbeing is spread across a population. In the curved diagrams below you can see the colored curves that describe the cumulative distribution of wellbeing throughout society. A perfectly equal society would have a perfectly straight diagonal line along the dashed path; the Gini coefficient measures the white area between this perfectly equal distribution and the real distribution in the population.

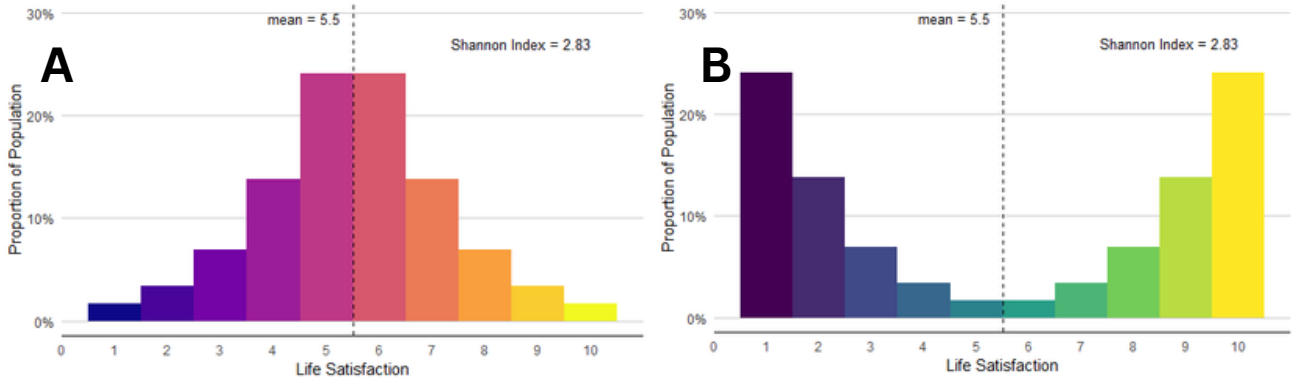
There are, however, different ways to achieve the same Gini coefficient. As you can see in the example below, two very different distributions can have both the same Gini coefficient and the same mean.



### 2.5. SHANNON INDEX

The Shannon Index is often used by ecologists to measure the diversity of species within a region. In this way, the Shannon Index stands out from the other measures by not being a measure of diversity over ordered, cardinal numbers, but the diversity of distinct wellbeing states in society.

The Shannon Index will be highest when there is an equal spread of wellbeing states throughout the population, and lowest when everyone has the same wellbeing value. The unintended outcome of the Shannon Index's insensitivity to the actual numerical wellbeing values is that half the population having a life satisfaction of 4 and the other having a life satisfaction of 6, will generate the same Shannon Index as half the population being a 0 and half a 10. The Shannon Index only captures the difference of the real distribution from a completely flat distribution, not where in the distribution this inequality occurs as you can see from the two "identical" populations below:




Based on these examples it's clear that if we adopt any of these four measures of inequality, or any other single metric, we necessarily lose a significant amount of information about the underlying distribution of wellbeing in society.

## 3. VALUE FREE MEASUREMENT

Losing information about a wellbeing distribution by choosing a specific metric is not inherently a problem. In fact, we argue that it is essential in the process of decision-making: accountable decision-making needs metrics, and choosing a metric necessarily highlights the information that the decision-maker finds important and implicitly hides the information that isn't. This must be done. However, the question arises as to who should be making this decision. It seems inappropriate for any researcher or policy worker to assume this role, as choosing a metric involves a value judgment. When the discarded information involves morally relevant factors like goodness, equality, or justice – aspects of individuals' lives deserving moral consideration – this loss of data can lead to unwarranted 'value paternalism'.

For example, using the mean of the wellbeing distribution imposes a value of averagism, which suggests that what matters is how the average person in society is doing. Similarly, using any measure of wellbeing dispersion, such as the Shannon Index, implies a value of egalitarianism, where the primary concern is the equal wellbeing of all individuals, by some mathematical definition. Neither, nor both, of these values fully encapsulate what many would consider an ideal society. Relying solely on these metrics forces us to focus on just a narrow dimension of wellbeing, one that we suspect is not aligned with many intuitions about a thriving society.

Perhaps this is best seen by looking at the examples we've laid out above:

In the first example, why would many people intuitively think that Population B is worse off than Population A, even though the mean and ratio of low to high wellbeing are identical? Perhaps these individuals greatly value the reduction of suffering and thus prioritize minimizing the number of people with wellbeing lower than 4 (or some other threshold). To them, Population B, with a higher number of suffering individuals, is evidently worse off. This value is not explicitly captured by any of the wellbeing inequality metrics we have examined thus far.

In the second and third examples we suspect that many people would perceive Population B as less equal than Population A. This might be because their definition of equality encompasses a judgment about the disparity between the best- and worst-off individuals. In both cases, opting for Society B when Society A is available might appear immoral, as some people benefit in Society B while others are disadvantaged. As we have seen, it is challenging for a single, simple inequality metric to encapsulate this value.

To this end, we suggest a simple, yet not-altogether-obvious practice: when presenting information about the distributions of wellbeing within a population to an audience of the public or policy-makers we forgo simplifying metrics as much as possible and present full graphical presentations of these distributions, as we have done here for hypothetical distributions.

As you may have experienced reading through this report, it seems that people generally have a good instinct for interpreting visual distributions of wellbeing values. But it can also be difficult to make judgments between visual rankings due to the complexity of the information that they share. In order to make rankings and comparisons between full distributions we can employ another opportunity that this moment grants us: technological tools can make it easier to interpret and compare distributions. The simplest versions of these tools would be dashboards that would allow individuals to choose between metrics of inequality, like the ones we've listed here, to rank and compare distributions, but the possibilities extend much further. Tools could allow individuals to specify their own weightings over wellbeing levels, letting them indicate how much importance to put on each part of the distribution. These tools could also allow people to combine a complex set of values–say prevention of suffering, equality, and total wellbeing–in a personalized way.

This puts the choice of the right metric in the hands of the public and elected policy-makers, precisely where democratic values suggest it should be.

# 4. WHY NOW?

The device you are reading this on likely has a keyboard with the QWERTY layout. In the 1870s, the QWERTY layout was a clever solution to a pressing problem: the jamming of typewriter arms when adjacent keys were struck in rapid succession. By scattering commonly used letters across the keyboard, the layout reduced key jams. Despite its inherent inefficiencies, this seemingly arbitrary arrangement of keys became the enduring standard into the digital age—typists became accustomed to it, training materials and typing schools reinforced its use, and manufacturers followed suit. This is the process we call path dependence, a phenomenon where initial decisions, once established, create self-reinforcing patterns that are resistant to change.

We mention this as a cautionary tale. Economic history is full of stories of standards adopted early on in an industry or government that stick, not because of their merits, but because they got their first.

There is a noteworthy example even in the field of wellbeing metrics: Gross Domestic Product was adopted in the US during World War II not for its inherent fitness as a measure of societal good, but partly because a measure was needed that rewarded the ramping up of the industrial war machine. Once this need had passed, US power in the post-war period, ideological factions and inertia let it take hold as the key metric in measuring the health of society.

Wellbeing governance is gaining traction around the world, but if it is to become a widely used paradigm, as its proponents hope for, then we are in the very early days of its story. This means we must be aware of the specter of path-dependence: the decisions that researchers and policy-makers make now may have long-lasting impacts. We urge, then, that we cannot settle for a measure which is "good enough", or as good as the standards in measuring other sorts of inequality. By choosing an effective and democratic standard for talking about well-being inequality now, we have the potential to do a lot of good.

# 5. WHY NOT?

There are some noteworthy arguments about the drawbacks of using full distributions which we would like to discuss next. Metrics like GDP are appealing for their simplicity and convenience in policy decisions. However, relying solely on such metrics can lead to limited interpretations and obscure crucial information.

Full wellbeing distributions, while potentially harder to interpret, offer a wealth of information that empowers individuals to form their own judgments about the state of affairs. The primary goal of a wellbeing metric should be to inform, and distributions provide the maximum amount of information while maintaining interpretability. Single measures of inequality can still be derived from these distributions, and presented alongside them, offering a familiar and accessible perspective alongside the comprehensive data. This hybrid approach allows us to avoid the pitfalls of oversimplification and better understand the nuances of wellbeing, while maintaining accessibility.

A second concern is that using full distributions may seem unfamiliar, discouraging adoption, and the discussion of wellbeing inequality in policy circles. Measures like Gini (widely used in economics literature) are familiar and might be circulated with more ease. However, novel measures can quickly gain traction when researchers, policymakers, and stakeholders find it easy to recognize their value. For example, the Human Development Index rapidly gained momentum as an alternative to GDP for assessing a country's progress, due to its common sense appeal to objective and widely agreed upon measures of human flourishing. If wellbeing researchers, policymakers and other stakeholders see the merit in using full distributions, and are persuaded by the intuitive appeal to complexity in a matter like population wellbeing, adoption could even be aided by taking an uncommon approach. Starting can be difficult though, so we again suggest the hybrid approach where common measures like Gini and average wellbeing are presented alongside distributions to help interpretation and facilitate adoption.

One exception that may be taken with the argument we have laid out is that in the examples above, the distributions for Society B are highly unlikely to occur in a real population. The risk then is that information loss does not occur or not at a significant level among realistic distribution. We choose extreme distributions on purpose so that differences can be easily visually distinguished, but loss of information from a single metric still happens among a more plausible set of distributions. We suggest without argument that this loss is at a level relevant for policy, which is often concerned with modest improvement, even if these losses are not dramatic enough to make a good example here.

If it were the case that most populations had roughly normally distributed well-being, then the mean and a single measure of well-being inequality would capture *all* relevant information about well-being in a population and our argument would be invalid. But briefly, we suggest that it is easy to find examples of non-normally distributed populations (Afghanistan in 2022 is an extreme example) and easy to consider policies which would result in non-normally distributed populations (tax breaks for the most well-off is an example), so at least in certain cases, loss of information from adopting simple inequality metrics is a real concern. Without knowing which cases these are in advance, it is best to make full well-being distributions a part of the research, communication and policy toolkits. However, more careful research into the extent to which well-being distributions are normal, and under what conditions, is needed.

# 6. A VISION FOR THE FUTURE

Metrics of wellbeing need to strike a balance between providing as much information as possible while still remaining useful. Ultimately, any single metric will obscure important value-laden information. We hope that this short piece proved the usefulness of wellbeing distributions in communicating information beyond what's available and allowing for a more democratic interpretation of a society's wellbeing depending on one's value system. Paradigm changes come along seldomly, and wellbeing governance shows promise in being a new one. This means that the decisions we make now can have a big impact on the state of our world. Explicitly centering wellbeing as a policy indicator is already likely to be a huge improvement in aligning governance to the needs of people, but the establishment of a new standard and new working methodologies also gives us a chance to improve the decision-making tools and communication tools used in governance. We need not perfectly switch-out GDP and income measures of success with wellbeing, but instead we can contribute to a future in which the values of people, communities and nations are explicitly incorporated into the way decisions are made.

Throughout this piece we have stated repeatedly that the values of researchers or policy professionals should not be the ones guiding how wellbeing is aggregated in a population. But we have said little about what values should be chosen to guide. We envision an expanded democratic process where the public is encouraged to share how they value the distribution of wellbeing and representative groups of people are surveyed to build an empirical understanding of what distributions should be prioritized within the constraints of limited resources.

However even if governments had this wellbeing mandate, they would have little knowledge about how to pursue one wellbeing distribution over another. The research into determinants of and interventions on wellbeing has mostly focused on what can be done to impact the mean of wellbeing within a population, but surprisingly little is known about how policy can impact the whole distribution. The wellbeing equality project relies upon distributional analysis of wellbeing becoming a norm in wellbeing research.

There is good news: the statistical tools exist and much usable data has been collected. What is needed is for individual researchers to begin building the skills to perform distributional analysis and the courage to try it to start building a new research norm. Only then will governments and philanthropists have the tools to make decisions between interventions and the wellbeing distributions they will produce.

Our vision for the future involves the widespread use of wellbeing distributions, rather than single metrics, to inform wellbeing policy. However, we are uncertain about the manner in which we get there. Here are questions that will serve as starting points for collaboration toward this goal:

- What does desirable societal wellbeing look like? Perfect equality? No suffering?
- How do we ask an individual to share what they value in the distribution of wellbeing in society?
- What qualities does the perfect tool for interpreting wellbeing distributions have?
- How can we make thinking about distributions more common in wellbeing research?
- How do we begin collecting democratic data on values over wellbeing distributions?

These questions and more are ripe domains for collaborations between researchers, policymakers and citizens to develop metrics that truly reflect our concerns for societal wellbeing. We are interested in your thoughts and continuations of this conversation. We welcome anyone interested to get in touch on collaborating on developing, reshaping and actualizing a vision for a world of wellbeing governance. There are many questions to answer, much work to do and lots of things to build on the path to a fair, just and flourishing world.

## ABOUT THE POPULATION WELL-BEING LAB

This report was written by members of the Population Well-being Lab at the University of Toronto. The lab focuses on research, teaching, and advocacy related to the determinants, consequences, and policy relevance of a satisfying, purposeful, and engaging life. The lab takes an interdisciplinary perspective and uses diverse methodological and statistical techniques to examine population well-being in relation to pressing global issues (e.g. sociopolitical unrest, economic justice, and major population events). Find out more on our lab website: https://cheung.artsci.utoronto.ca/

## SUGGESTED CITATION

McCanny, A., Panasiuk, S., & Cheung, F. (2023, April 19). Wellbeing Distributions: The Democratic Way to Measure Happiness. Retrieved from osf.io/hg2a8

This report draws on research supported by the Social Sciences and Humanities Research Council and the Canada Foundation for Innovation. This report was undertaken, in part, thanks to funding from the Canada Research Chairs Program. The views expressed in this report are solely our own.