

# Well-Being Measurements and the Linearity Assumption: A Response to Wodak

Cristian Larroulet Philippi

To cite this article: Cristian Larroulet Philippi (2023): Well-Being Measurements and the Linearity Assumption: A Response to Wodak, Australasian Journal of Philosophy, DOI: [10.1080/00048402.2023.2227642](https://doi.org/10.1080/00048402.2023.2227642)

To link to this article: <https://doi.org/10.1080/00048402.2023.2227642>



© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 03 Jul 2023.



Submit your article to this journal [↗](#)



Article views: 250




View related articles [↗](#)



View Crossmark data [↗](#)



# Well-Being Measurements and the Linearity Assumption: A Response to Wodak

Cristian Larroulet Philippi 

University of Cambridge

## ABSTRACT

Wodak (2019) persuasively argues that we are not justified in believing that well-being measurements are linear. From this, he infers grave consequences for both political philosophy thought experiments and empirical psychological research. Here I argue that these consequences do not follow. Wodak's challenges to the status of well-being measurements do not affect thought experiments, and well-being empirical researchers may be justified in making average comparisons even if their measurements are not linear.

**ARTICLE HISTORY** Received 18 October 2021; Revised 19 July 2022

**KEY WORDS** well-being measurement; scales; thought experiments; Kahneman

## 1. Introduction

The epistemic status of well-being measurements remains contested. Daniel Wodak (2019) persuasively argues that we are not justified in believing that current subjective well-being measurements are quantitative ('linear'). That is, that we lack justification for believing that the intervals—the distance between the subsequent levels—of these measurement scales are of equal size, each representing an equal change in the amount of well-being. This is a plausible claim.

Wodak also argues that this lack of justification has serious consequences for the conclusions typically drawn from both standard political philosophy thought experiments (such as Roger Crisp's) and policy-relevant psychological research (such as Daniel Kahneman's). Here I disagree: the challenges Wodak puts forth to the linearity of actual well-being measurement scales do not affect the validity of those thought experiments, and the linearity assumption is not needed for inferring conclusions such as Kahneman's 'peak bias' thesis.

This does not mean that Wodak's conclusions—about Crisp's thought experiment and Kahneman's peak bias—cannot be supported in other ways. I only wish to challenge Wodak's argument, not these specific conclusions. Perhaps the latter are true; perhaps not. But much more is at stake here. If Wodak's argumentation is correct, important areas of political philosophy and empirical psychology appear in trouble. Therein lies the value of clarifying what follows from a lack of justification for linearity and what does not.

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

## 2. Well-Being Thought Experiments

The literature concerning egalitarianism, prioritarianism, etc., draws heavily on idealized thought experiments that stipulate different levels of well-being. The intuitions some of these experiments aim to pump depend on quantitative comparisons of well-being. Wodak argues that such thought experiments are potentially misleading.<sup>1</sup>

Wodak illustrates his point with Crisp's (2003: 745-6) thought experiment, which involves two groups, and two options: *Equality*, where both groups have a well-being of 50, and *Inequality*, where one group has 10 and the other 90. Both options involve the same average well-being; but in *Equality* all people have 'good lives' while in *Inequality* some have 'much better' and others 'much worse' lives. Readers are expected to prefer Equality. Since utilitarianism cannot justify this preference, some egalitarian principle is called for.

Wodak claims this conclusion does not follow. Readers might interpret these well-being values non-linearly. If so, readers' judgments are about a *different* scenario than what Crisp stipulated—one where even utilitarians could prefer Equality. Hence no egalitarian principle may be needed.

Why might readers interpret well-being values non-linearly despite Crisp's explicit stipulation that these values should be treated quantitatively (that is, that going from 10 to 50 increases as much well-being as going from 50 to 90)? I believe that Wodak's argument can be read in two ways.

First: at moments, the argument seems to be that readers may interpret Crisp's well-being values non-linearly because these values *are themselves* hypothetical outcomes of concrete well-being measurement scales, and such scales are likely to be non-linear. For example, Wodak speaks of Crisp's stipulated numbers as 'hypothetical well-being measurements'. Moreover, Wodak claims: 'we are not justified in believing that [Crisp's] stipulation [of linearity] is true' (2019: 33). Yet this claim only makes sense if we take Crisp's stipulation to be about well-being *measurements* (what Wodak calls 'reported well-being'), not if Crisp's stipulation is about well-being *itself* ('actual well-being'). If Crisp's stipulation concerns well-being itself, facts about well-being measurements (such that they are non-linear) cannot affect the stipulation's truth-value.

However, Crisp's stipulation concerns actual well-being, not any concrete measurement of it. Crisp makes explicit that '[n]o commitment to precise measures or to any particular view of welfare itself is intended by the use of numbers' in his thought experiment (2003: 746). Thus, this version of the argument—which takes the likely non-linearity of current well-being measures to cast doubt on the linearity of the thought experiment's well-being values—equivocates.<sup>2</sup>

A second version draws from Wodak's discussion of psychophysical scales.<sup>3</sup> These scales represent how experimental subjects perceive physical quantities such as weight, sound intensity, etc. Psychophysicists claim that these scales are logarithmic. For example, subjects tasked with assigning numbers to different physical magnitudes (say, sounds of different intensities) provide numbers that are logarithmically related to these physical magnitudes. Wodak takes these results to bear both on the

---

<sup>1</sup> More precisely, Wodak claims that they are either misleading or redundant. Here, I focus on the misleading side of the dilemma.

<sup>2</sup> Daniel Wodak clarified that this is not the intended argument (personal communication).

<sup>3</sup> I thank the referees and Wodak for correcting previous misunderstandings on my part about this argument.

linearity of actual well-being measurements and on the fallibility of Crisp's thought experiment. I reconstruct his argument as follows:

- (1) Psychophysical scales are logarithmic.
- (2) *Similar class of phenomena*: Psychophysical scales represent how experimental subjects report *subjective phenomena*, just like subjective well-being scales.
- (3) *Shared key property*: Psychophysical scales are logarithmic because the range of magnitudes we can perceive (say, from the barely audible to the loudest sound) is immense, making it cumbersome to represent this range using numbers linearly. This, says Wodak, also holds for 'well-being. The spectrum from indifference to noticeable pain to agony is vast' (2019: 31). Both in psychophysics and well-being, therefore, the need for compression leads to the use of logarithmic scales.

(1–3) are taken to support:

- (4) Subjects report well-being logarithmically, thus well-being measurement scales are logarithmic (hence, non-linear).

Moreover,

- (5) *Common mechanism*: Whatever leads people to report well-being non-linearly may also lead people to imaginatively engage with well-being values deployed in thought experiments non-linearly.
- (6) *Opacity*: The usage of numerals connotes linearity. Even if readers treat well-being numbers non-linearly, they might still think they are treating them linearly.

Wodak concludes:

- (7) Thought experiments that use quantitative well-being comparisons (such as Crisp's) are potentially misleading: unbeknownst to them, readers may imagine *not* the scenario that was stipulated, involving linearity, but one involving non-linearity.

In what follows, I will not contest (7). There may be good reasons for (7)—firstly, as Wodak argues, Crisp's description of the thought experiment is quite thin, which makes intuition pumping unreliable since readers' imagination is not constrained; secondly, as (6) implies, readers might be unreflective about their numerical reasoning; thirdly, (7) may be the best explanation for why, when asked in surveys to think through different well-being scenarios numerically, some people fail to take the stipulated values seriously;<sup>4</sup> etc. These are important, empirical issues to settle. But they do not concern actual well-being measurement scales. My claim here is that Wodak's argument concerning the warrant of the linearity of actual well-being measurement scales does not provide additional support to (7).

---

<sup>4</sup> I thank a referee for suggesting this hypothesis.

First, pace (1), that ‘psychophysical scales are logarithmic’ is less straightforward than presented. For instance, some researchers quoted by Wodak claim that subjects report logarithmically when the number of answer categories is large (50), but *report linearly* when that number is less than 10 (Poulton 1979). This matters—most subjective well-being scales have 11 or fewer answer categories. Hence, even if premises (3) and (4) are correct—and so, features of psychophysical scales speak to features of well-being scales—the argument would not establish non-linearity in well-being scales with few categories (which are the most common scales).

What about (2) *Similar class of phenomena*? This premise overlooks a crucial difference between psychophysical and subjective well-being scales: a difference that challenges the idea that the functional form of the former speaks to that of the latter. Subjective well-being scales correlate subjects’ verbal reports with their subjective experience. To take two prominent examples, these measures ask subjects to report how satisfied they are with their lives from 0 to 10 (*Life Satisfaction scale*), or to report whether they feel ‘very happy’, ‘pretty happy’, or ‘not too happy’ (*Happiness scale*; these categories are then coded numerically). So, if there is a logarithmic relationship here, it is between verbal reports and subjective experience. In contrast, psychophysical scales correlate verbal reports with *external physical magnitudes*, not with subjective experiences themselves. Hence, subjective well-being and psychophysical scales have different kinds of relata (verbal reports and subjective experience versus verbal reports and external physical magnitudes). To see why this matters, note that there are several possibilities for interpreting the logarithmic association (sometimes) found in psychophysics: either as (i) representing how subjects’ reports relate to their inner subjective experience (which requires assuming a linear relationship between the physical magnitudes and the subjective experiences), as (ii) how subjects’ subjective experience relate to the physical magnitudes (assuming a linear relationship between subject’s reports and their subjective experiences) or, as (iii) the result of two other functional forms that when *combined* produce a logarithmic relation. This very issue has been hotly disputed since Fechner’s times (Briggs 2022: 40-41). Crucially, this dispute shows that *even if* psychophysical scales are logarithmic, what is logarithmic need not be the relation between verbal reports and subjective experience (that is, the relation represented in subjective well-being scales). Thus, the fact that both scales somehow involve ‘subjective phenomena’ is not a strong reason to think that functional forms travel from psychophysical scales to well-being scales.

Premise (3) *Shared key property* also aims at justifying the claim that the functional form of psychophysical scales speaks to that of well-being scales. It claims that the explanation for why psychophysical scales are logarithmic—the vast range of perceivable magnitudes leads to compression—also holds for well-being scales. However, the alleged explanation Wodak cites for loudness scales being logarithmic comes from a passage where the author (Gelfand 2009) is *explaining the scientific notation* used for sound intensity. (The section is titled ‘Decibel Notation’.) That is, Gelfand is explaining why *scientists* prefer representing the physical quantity ‘sound intensity’ in a logarithmic metric; not explaining why *experimental subjects’* reports are logarithmically related to absolute sound intensity. For all we know from Wodak’s argument, then, the vast range of tones is not an explanation of why experimental subjects report loudness logarithmically (if or when they do). Hence, premise (3) does not support the claim that the vast range of pains leads to logarithmic well-being scales.

All said, the plausibility of (1), (2) and (3), and their capacity to support (4) remain dubious. Since (5) and (6) cannot warrant (7) on their own, the argument fails. Thus, the challenges that Wodak puts forth to the linearity of actual well-being measurement scales—by drawing parallels between psychophysical and actual subjective well-being measurement—do not undermine the validity of thought experiments. Perhaps Crisp’s thought experiment is indeed misleading; but Wodak’s argument regarding our lack of confidence in the linearity of actual well-being measurements does not establish this.

### 3. Empirical Well-Being Research

Wodak argues that the lack of justification for the equal-intervals assumption also questions the usage of well-being measurements to ground conclusions in policy-relevant psychological research. He acknowledges that not all such conclusions depend on the equal-intervals assumption; but he claims that some important conclusions do depend on it.

Wodak illustrates his argument with Kahneman’s (2011) ‘peak bias’ thesis.<sup>5</sup> This thesis is inferred by drawing from patients’ minute-by-minute reports of pain during colonoscopies, using a scale that goes from ‘0’ (‘no pain at all’) to ‘10’ (‘intolerable pain’). Kahneman contrasted what he conceives to be patients’ “objective” total pain—the sum of each patient’s report—with patients’ retrospective judgments of how painful the colonoscopy was overall. Patients with higher peaks of reported pain judged their overall experience as worse, even if their total pain was less. Therefore, Kahneman concluded, patients have peak bias.

Wodak (2019: fig. 3) challenges this conclusion using data from two representative patients: patient B has a peak of 8 and a total pain of 68.1, and patient C has a peak of 9 and a total pain of 27.2. If patient C judges her overall experience worse than patient B, Kahneman would infer that C has peak bias. Wodak would not, because he claims that ‘the case for PEAK BIAS goes through only if we assume linearity in particular—that is, that the difference between 9 and 8 is the same as between 7 and 8, and so on’ (39).

This claim is incorrect. We do not need the equal-intervals assumption for C’s total pain to be less than B’s. Imagine, for instance, that Kahneman’s scale gets substantially wrong only the 9<sup>th</sup> interval, such that, say, the difference between 9 and 8 is not the same as the size of the other intervals, but rather it is ten times bigger than the other intervals: C’s total pain (36.2) would still be much less than B’s (68.1). One can add variation in the relative sizes of the other intervals by say, stipulating that the 5<sup>th</sup> and the 6<sup>th</sup> interval are twice as large as the previous ones, and arriving at the same conclusion. More generally, there are many possible sets of differences between 9 and 8, 8 and 7, and so on, that result on C’s total pain being less than B’s. Thus, the inference to peak bias is substantively more robust than Wodak suggests. Indeed, in Wodak’s example, the scale must be radically skewed to overturn the conclusion.<sup>6</sup>

My goal here is not to defend peak bias. It is to show that Wodak’s main claim—that peak bias can *only* be inferred assuming linearity—is incorrect. A much wider set of

<sup>5</sup> Peak bias is one part of Kahneman’s ‘peak-end bias’ thesis.

<sup>6</sup> As Wodak says, a logarithmic scale would overturn peak bias. Such a scale is radically skewed: the increase in pain that takes us to report ‘9’ versus ‘8’ on this scale is ten thousand times the increase that takes us to report ‘5’ versus ‘4’. Less extremely skewed scales can also overturn peak bias in this example.

possible assumptions supports peak bias in Wodak's example. (Which of these assumptions are plausible, is another matter.)

More importantly, the point goes well beyond this specific example. The justification of total and average comparisons—of which Wodak's example is just one case—does not collapse when we are not justified in the equal-intervals assumption. Despite what standard methodology (and Wodak) assume, substantially weaker assumptions regarding intervals' relative sizes can justify such comparisons (for discussion, see Larroulet Philippi 2021, *forthcoming*). To illustrate, Easterlin's (1974, table 2) famous happiness comparisons across income groups used a 3-points happiness scale. To infer that the richest group is on average happier than the second richest group we do not need equality of intervals, but only that the first interval is not eight (or more) times larger than the second interval. Other comparisons require substantially weaker assumptions. The recent *Handbook for Wellbeing Policy-Making's* comparison of Denmark's versus UK's life satisfaction (Frijters and Krekel 2021: fig. 2.3) only requires that the first three intervals are not 86 times larger than the last seven intervals. In sum, Wodak's argument, which focuses on linearity, fails to identify the epistemic burdens for total and average comparisons correctly.

#### 4. Wodak's Proposal

Yet Wodak's concern goes beyond researchers lacking justification for assuming linearity. His worry is compounded by the *representational fallacy*: researchers automatically assuming linearity merely because well-being scales use numbers. The more prevalent this fallacy, the more well-being scales are misused. Hence Wodak (2019: 43) proposes: 'When we do not have evidence to support linearity, change how the scale is represented: replace numerals with letters'. Thereby the scale only represents rank-order information—the representational fallacy is avoided.

I agree with Wodak: representing all scales with numbers can increase researchers unreflectively assuming linearity, which can increase misuse. However, we ignore how prevalent the fallacy is, thus how impactful the proposal would be. Clearly not *all* average comparisons are made because of unreflectively assuming linearity. For example, many researchers *explicitly* justify such comparisons by drawing on specific statistical arguments (Wodak 2019: 40).

Regardless, as stated, the proposal seems too blunt. Not 'having evidence to support linearity' is compatible with having enough evidence supporting weaker but still valuable propositions about intervals' differences (such as 'all intervals are of the same order of magnitude'); or with having evidence that supports that intervals' differences conform roughly to a logarithmic scale. Why would we want to represent these two cases with letters? Just as numbers suggest cardinality, letters suggest mere ordinality. But the two examples just given do not constitute strict ordinality. Wodak's proposal risks generating a reverse fallacy; more nuanced proposals are needed.

Summing up, although Wodak's (2019) proposal raises important concerns about well-being measurements, neither his alleged consequences nor his proposed solution follows straightforwardly from the lack of justification for linearity and the representational fallacy.

## Acknowledgments

Thanks to Anna Alexandrova, Lukas Beck, Jacob Stegenga, Daniel Wodak, and two anonymous referees for helpful comments.

## Disclosure Statement

No potential conflict of interest was reported by the author(s).

## Funding

This paper is based on research that was funded by Gates Cambridge Trust.

## ORCID

Cristian Larroulet Philippi  <http://orcid.org/0000-0001-5793-4670>

## References

- Briggs, Derek (2022) *Historical and Conceptual Foundations of Measurement in the Human Sciences*. Routledge.
- Crisp, Roger (2003) 'Equality, Priority, and Compassion', *Ethics* **113**: 745–63. doi:10.1086/373954.
- Easterlin, Richard (1974) 'Does Economic Growth Improve the Human Lot?', in Paul David and Melvin Reder, eds., *Nations and Households in Economic Growth*: 89–125. Academic Press.
- Frijters, Paul, and Christian Krekel (2021) *A Handbook for Wellbeing Policy-Making*. Oxford University Press.
- Gelfand, Stanley (2009) *Hearing: An Introduction to Psychological and Physiological Acoustics*, 5th edition. Informa Healthcare.
- Kahneman, Daniel (2011) *Thinking, Fast and Slow*. Farrar, Straus and Giroux.
- Larroulet Philippi, Cristian (2021) 'On Measurement Scales: Neither Ordinal nor Interval?', *Philosophy of Science* **88**: 929–39. doi:10.1086/714873.
- Larroulet Philippi, Cristian (forthcoming) 'Against Prohibition (Or, When Using Ordinal Scales to Compare Groups is OK)', *British Journal for the Philosophy of Science*. doi:10.1086/721759.
- Poulton, E. C. (1979) 'Models for the Biases in Judging Sensory Magnitude' *Psychological Bulletin* **86**: 777–803. doi:10.1037/0033-2909.86.4.777.
- Wodak, Daniel (2019) 'What If Well-Being Measurements Are Non-Linear?' *Australasian Journal of Philosophy* **97**: 29–45. doi:10.1080/00048402.2018.1454483.