



Does “Better Than” Have a Common Scale?

Erich H. Rast¹

Accepted: 29 May 2024
© The Author(s) 2024

Abstract

Ordinal, interval, and ratio scales are discussed and arguments for the thesis that “better than” comparisons reside on interval or ratio scales are laid out. It is argued that linguistic arguments are not conclusive since alternative rank-based definitions can be given, and that in general “better than” comparisons do not have a common scale type. Some comparison dimensions reside on ratio scales, whereas others do not show any evidence of lying on a scale stronger than an ordinal scale.

Keywords Good · Better than · Value theory · Measurement theory

1 Introduction

On occasion, the idea has come up to use measurement theory to determine the scale on which uses of “better than” reside. For example, Carlson [3–5] defends ratio scales with incompleteness, work by Rabinowicz [20, 21] is based on “favorings”, which seem to imply at least interval scales, and recent work on the semantics of evaluative adjectives involves interval scales [16] and modified ratio scales [30].

This article aims to defend the thesis that “better than” comparisons do not have a uniform scale type. Uses of “better than” and other thin value predicates instead serve as a proxy for comparisons in multiple dimensions with potentially differing underlying scale types. Extralinguistic considerations concerning the underlying dimensions determine which scale can be attributed to a feature of betterness. Thus, the answer to the title’s question is No. The scale type may change according to the context in which “better than” is used, and different comparison dimensions may reside on different scale types.

✉ Erich H. Rast
erast@fcsh.unl.pt; erich@snaflu.de

¹ IFILNOVA Institute of Philosophy, Universidade Nova de Lisboa, IFILNOVA Campus Campolide, Av. de Berna, 26, Sala 119, Lisbon 1069-061, Portugal

2 What Are Scales and Why Are They Relevant?

Scale types go back to work by psychologist Stevens [32], and in what follows the classification from [26: p. 64] is taken as a basis.¹ In measurement theory, a measurement function maps comparisons by an ordering relation to numeric values under a representation condition. For example, suppose a is better than or as good as b (weak betterness) and write this relation as $a \succsim b$. A corresponding strict betterness relation is often written with the symbol ‘ $>$ ’ and is defined as $a > b$ iff. $a \succsim b$ and not $b \succsim a$, and an equivalence relation “equally good” can be defined as $a \sim b$ iff. $a \succsim b$ and $b \succsim a$. The representation condition states that $u(a) \geq u(b) \Leftrightarrow a \succsim b$, where ‘ \geq ’ is “greater than” for numbers and function $u(x)$ is a utility function from comparison objects to real numbers. According to the condition, it represents the ordering expressed by the betterness comparisons. However, utility functions can in principle encode more information than the mere ordering of objects that the right-hand side of the condition expresses. This is where Stevens’s scales come into play.

How can utility functions encode more information than the mere ordering of objects? Intuitively, the result of a utility function may also be interpreted as an intensity, as a measure of how much of a certain quantity or value a comparison object has.² Moreover, utility functions may be used to compare the magnitude of differences between comparisons. Stevens’s scale types make these ideas precise by defining scales in terms of admissible transformations between utility functions.³ The more admissible transformations are restricted, the more informative the utility functions on that scale become. Conversely, less restricted transformations lead to a weaker scale.

An *ordinal scale* is one of the weakest scales. If any increasing transformation of a utility function is admissible, the “better than” comparisons of items in the given domain rest on an ordinal scale. A function $f(x)$ is an increasing function if and only if for all x and y it holds that $f(x) \geq f(y) \Leftrightarrow x \geq y$. If $f(x)$ is an increasing function, then the function $v(x) = f(u(x))$ is an increasing transformation of $u(x)$, and the condition $v(a) \geq v(b) \Leftrightarrow u(a) \geq u(b)$ holds. Allowing arbitrary increasing transformations, as long as they preserve the representation condition, implies that *only* the ordering of the numbers counts; shifting the function on one of the axes or stretching it is irrelevant. Consequently, intensities and value differences are not represented on a scale of this type. On ordinal scales, the utility function is merely an alternative representation of the qualitative ordering relation.

In contrast, cardinal scales allow utility functions to encode more information than the mere ordering secured by the representation condition. Ratio and interval scales are cardinal scales. There are more specific types of these scales such as the logarithm-

¹ Cf. Krantz et al. [15] for details and seminal proofs in standard measurement theory.

² “Intensity” is used informally here to designate how much of something is there, i.e., the magnitude or strength of the measured phenomenon, as it is expressed by the numerical value assigned. The word is used differently, with a more specific meaning, in formal measurement theory.

³ The following exposition of scales rests on Roberts [26: pp. 66-7]. The first to characterize scales by admissible transformations was Stevens [32].

mic Decibel scale, which is a special type of ratio scale, but for current purposes it suffices to only consider ratio and interval scales.

The admissible transformations for *interval scales* are $v(x) = a \cdot u(x) + b$, where a is a positive, non-zero real number and b is a real number. This means that if $u(x)$ is a utility function fulfilling the representation condition $u(a) \geq u(b) \Leftrightarrow a \geq b$, then any linear transformation $v(x) = a \cdot u(x) + b$ represents the same information as $u(x)$ on the interval scale. Difference comparisons are meaningful on this type of scale. For example, $u(a) - u(b) = u(c) - u(d)$ is a sensible comparison, whereas this comparison makes no sense on an ordinal scale. However, the 0-point is not meaningful on an interval scale. Suppose that for some comparison object a , the utility is $u(a) = 0$. Then a function $v(x) = 1 \cdot u(x) + 3$ is allowed as a transformation of $u(x)$, i.e., it is an alternative representation of the same ordering and utilities. However, $v(a)$ is 3 and not 0 in this alternative representation. Hence, the fact that $u(x)$ has its 0-point at a neither makes a nor the value 0 special in this representation. The 0-point on an interval scale is only a convention.

To illustrate this feature, consider the Celsius and the Fahrenheit scales for temperatures. We can translate between these scales by applying a linear transformation. The degree in Fahrenheit is $32 + 9/5 \cdot$ the degree in Celsius. So, 0 degrees Celsius is 32 degrees Fahrenheit, and 0 degrees Fahrenheit is -17.78 degrees Celsius. In both cases, when one measures 0 on one scale, one measures a non-zero value on the other scale because the points are shifted just like any other values are shifted by the linear transformation that translates between them. The 0-point behaves like any other point on an interval scale.

In contrast, the 0-point is special on a *ratio scale*. This scale bears the strictest condition on transformations among the three: $v(x) = a \cdot u(x)$ for non-zero positive real number a . It follows from this condition that if $u(x)$ is zero, so will be any alternative representation $v(x)$. Measuring temperature on the Kelvin scale is often given as an example, since the 0 on this scale represents absolute zero. Height is also on a ratio scale, and so are monetary costs. Unless this is prohibited by an additional constraint, ratio scales can be negative. For example, wealth is on a ratio scale and may become negative when a person is in debt.

On a ratio scale, intensities, ratios, and differences are meaningful, although the exact number does not matter when comparing intensities. For example, one may say that John is 1.18 times taller than Mary. If we change the measurements from inches to centimeters, John will still be 1.18 times taller than Mary. In this example, measuring takes place with a ruler, and the possibility of measuring with a ruler is a clear-cut indicator of a ratio scale. Mathematicians express this possibility more abstractly with a concatenation operation.⁴ For example, putting two sticks behind each other amounts to "concatenating" them; the resulting concatenated object's length is the sum of the lengths of its parts. Whenever such a concatenation operation with the corresponding algebraic properties is available, the measurement is called *extensive* and the underlying scale is known to be a ratio scale. Such an operation is not readily available for interval scales because these lack a unique zero value. However,

⁴This approach goes back to Hölder [13]. Using \circ to denote the concatenation of two objects, the condition $u(a \circ b) = u(a) + u(b)$ formally expresses this idea. See Krantz et al. [15: pp. 71-135] for details.

although there are no rulers of negative length, in the abstract sense ratio scales allow values below zero, even though sometimes these are impossible, as the example of temperature in Kelvin illustrates.⁵

There are other scale types. For example, on an *absolute scale* the only admissible transformation is the identity transformation $v(x)=u(x)$. This is the scale used for counting. When counting apples, for instance, a function representing a collection a of three apples should return the number 3 for a ; a function that returns 5 for collection a would not be an adequate count.⁶ As another example, the Decibel scale used to assess loudness is logarithmic. The underlying scale type is important because it defines what can be meaningfully compared with the comparative of a value adjective. The scale type is also crucial for aggregating different dimensions or aspects of a value comparison and for aggregating value with other types of information. For example, the expected utility hypothesis presupposes that it is permissible to multiply the value function ('utility' in this case) with the probability of an outcome. This operation requires at least an interval scale, and a ratio scale is generally more adequate for it.⁷

3 Arguments for Particular Scales

Arguments for or against the idea that "better than" has a certain scale type can be divided into three categories: linguistic concerns, arguments by practicality, and normative considerations. These are outlined in the following paragraphs.

Although linguistic examples suggest that "better than" is multidimensional—see, for example, Weidman Sassoon [36, 37]—, authors focusing on linguistic indicators for scale types have concentrated on monist uses of "better than."

⁵ Stevens [32] did not allow negative values. However, it turned out that there are many applications where negative values make sense, and the prohibition is no longer assumed in more recent works like Krantz et al. [15, 26].

⁶This does not mean that other increasing transformations could not be used for counting apples at all. However, these representations would only be meaningful together with a function or algorithm that allows one to retrieve the number of apples counted. In other words, they would not represent apple counting on their own.

⁷Interval scales are problematic because the lack of a unique zero value can interplay in undesirable ways with Expected Value in decision-making. The expected value of an outcome is calculated by multiplying its utility under certainty with the probability of its occurrence. Suppose we calculate the expected value of an outcome on two different interval scales that represent the same value ordering. In that case, the 0-point on one scale may yield no expected utility with any positive probability. In contrast, the corresponding non-zero value on the other scale may yield different positive expected utilities with these probabilities. This is almost always undesirable, and a ratio scale should be chosen instead. Suppose, for example, that outcomes a and b have probabilities $Pr(a)=0.1$ and $Pr(b)=0.7$, and $u(a)=u(b)=0$ on one scale but $v(a)=v(b)=32$ on the other scale satisfying the representation condition for the comparison $a \sim b$. Then, a and b have expected value 0 on the first scale, whereas $EV(a)=v(a) \cdot Pr(a)=3.2$ and $EV(b)=v(b) \cdot Pr(b)=22.4$ on the second scale and so b is strongly preferable to a from the perspective of expected value on that scale, even though the scales are supposed to be alternative representations of the same value. Such examples indicate that a ratio scale should have been chosen in the first place because 0 is often meaningful, even when that is not immediately obvious.

According to Lassiter [16: pp. 177-8] , the scale of “good” is on an interval scale. This implies, for example, that it is meaningful to say that a certain plan is much better than another plan. Something may also be only slightly better than something else. Modifiers like “slightly” and “much” generally indicate that the comparison by “better than” allows for intensities, and degrees and difference comparisons are meaningful. For example, if $u(x)$ is a faithful representation of the underlying qualitative “better than” comparisons, one may hypothesize that x is slightly better than y iff. $u(x) > u(y)$ and $u(x) - u(y) < k$ for some contextually supplied standard k , and that x is much better than y iff. $u(x) > u(y)$ and $u(x) - u(y) > q$ for some contextually supplied positive standard q . Similar threshold-based definitions are conceivable for related modifiers like “a little bit”, “way”, or “negligibly.”

Soria-Ruiz [30] argues that ratio modifiers like “two times” suggest that “better than” must reside on a ratio scale. Recall that any linear transformation of a utility on an interval scale represents the same information. Assume someone says “It is two times hotter this month than last month” and everyone in the conversation understands that degrees of Celsius are meant. This seems to be perfectly fine, for example it could have been 10° Celsius in April and 20° Celsius in May. However, from a measurement-theoretic standpoint, it is very easy to make incorrect statements by omitting the underlying scale or neglecting the impact of admissible transformations. When the needle climbs from 10° Celsius to 20° Celsius, it seems as if it has gotten twice as hot. Yet, in Fahrenheit the very same temperature change goes from 50 °F to 68 °F. Suddenly, it is not twice as hot. The zero points are particularly problematic, since these are given by convention only on interval scales. Going from 32 °F to 64 °F is twice as hot. However, the equivalent of 32 °F is 0° Celsius and multiplying 0 by 2 remains 0, of course.

People still regularly use ratio modifiers for interval scales on which they agree, but this usage is colloquial and error-prone. A statement like “It is twice as hot as last month” can mean a shift from 10° to 20° Celsius in Europe, whereas in the US going from 50° to 100 °F would indicate a sudden heat wave (100° F \approx 37.8° C). To avoid such confusions and potential sources of error, it is best not to talk about multiples on interval scales at all. So it seems that a ratio scale is needed for a ratio modifier to fully make sense.

Returning to “better than”, Soria-Ruiz [30] argues that uses of ratio modifiers generally indicate a ratio scale instead of an interval scale. We sometimes say things like “This solution is ten times better than the one you proposed yesterday” and “Getting 300 dollars back from the tax authority is three times better than only getting a 100 dollars back.” If statements like these are not just loose ways of talking or hyperbole and can, at least sometimes, be meant literally, then, one may argue, the respective uses of “better than” reside on a ratio scale.

In the end, Soria-Ruiz proposes a modified ratio scale that allows one to explain why saying that something is 1.38 times better than something else is not acceptable. A modifier like “2 times better” is allowed in his proposal because “better than” has a “granularity” of 2, whereas “1.38 times better” is prohibited. Although they are interesting, the details of his proposal are not relevant for this article. What counts is that he suggests a scale stronger than an interval scale based on the idea that *some* ratio modifiers can modify “better than” comparisons and this usage seems to be

incompatible with ordinal and interval scales. His case rests on linguistic uses of ratio modifiers.

The second category of arguments for cardinal scales have been called “arguments from practicality” above for lack of a better term. These are arguments that, in one form or another, try to establish a cardinal scale because of the negative consequences it would have in practice to presume a weaker scale. If “better than” comparisons are exclusively made on an ordinal scale, this has certain undesirable consequences. On an ordinal scale, multiplying a utility with a probability to represent the expected value under risk is not meaningful. Furthermore, because the expected utility hypothesis explains risk attitude through the nonlinearity of a utility function, it cannot be expressed in a purely ordinal framework. A concave utility function represents a risk-averse decision-maker and a convex utility function a risk-prone decision-maker. However, any positive increasing transformation of an ordinal utility function represents the same value, including transformations that drastically change the shape of the utility curve. So it makes no sense to claim that the shape of the utility curve represents risk attitude when the underlying utility function represents only ordinal information.

A typical economics example illustrates the problem. Consider an offer of (a) getting \$800 in cash for sure, or (b) getting \$2000 with a 50% chance of winning. Let the concave function $u(x)=\sqrt{x}$ represent the utility of a risk-averse decision-maker. The expected utilities of a and b are:

$$EU(a) = Pr(a) \cdot u(a) = 1 \cdot \sqrt{800} \approx 28.28.$$

$$EU(b) = Pr(b) \cdot u(b) = 0.5 \cdot \sqrt{2000} \approx 22.36.$$

Hence, the decision-maker prefers to get \$800 for sure, even though the expected value of b is higher (\$1000), as the expected value of a prospect is just $EV(x)=Pr(x) \cdot x$. The problem is that $v(x)=u(x)^2$ is an admissible transformation of $u(x)$ on an ordinal scale. If the scale really was ordinal, the function $v(x)=x$ would thus be an admissible transformation of $u(x)$, but this is just the expected value function according to which the decision-maker prefers b over a under any positive risk. So, risk attitudes cannot be modeled in this way on ordinal scales.

There are techniques to combine ordinal utility functions with *generalizations* of probabilities. Qualitative counterparts to expected value and expected utility have been studied by Bouyssou D. & Pirlot [1, 7, 18]. However, these strategies are only applicable when information about future events is insufficient or does not yet give rise to full-fledged probabilities for other reasons. In such cases, it may make sense to use formalisms like possibility theory by Dubois & Prade [8, 9] or Haas-Spohn ranking theory by Spohn [31]. However, if there is enough information for assessments of quantified risk, then these formalisms may be too general and full probabilities should be used. If so, the problem remains: Attempting to combine ordinal utility functions with full probabilities leads to a host of technical problems; strictly speaking, multiplication is meaningless in such an approach.

Interval and ratio scale utilities are also customarily used to express a decision maker’s evaluative attitude towards more of something. For instance, according to

the Principle of Diminishing Marginal Utility, many goods, including money, are worth less when someone has more of them than when someone has less of them. For instance, a dollar is subjectively worth less for a billionaire than for someone who has no money at all. Again, this use of nonlinearity is not meaningful with ordinal utilities.

Finally, there are more general normative justifications for ratio scales. As was already apparent in the discussion of expected utility above, it is customary in decision theory to consider values from the perspective of how they enable persons to make rational choices.⁸ For instance, decision-theorists like Ramsey, Savage, and Fishburn [10, 24, 28] develop sets of axioms for preferences and utilities and examine the rationality postulates that make these choice-guiding. The underlying use of the attribute “rational” is normative at least in a weak sense, in the sense that not acting in coherence with the respective postulates would be a mistake. Similarly, the conditions given by Carlson [5] for the extensive measurement of value will, without additional change, make them live on a ratio scale, albeit on a potentially incomplete ratio scale on which some objects remain incomparable.

Arguments for and against such principles are often normative, at least in the weak sense according to which it counts as a mistake if a principle is broken, and occasionally in the stronger sense of prescribing that agents should act by following them. In the decision-making literature the word “rational” frequently conveys this prescriptive aspect, though the relation between normativity and rationality may be left unclear. The often unspoken premise is that being rational is beneficial, yet there is frequently no non-circular description of what it means to be rational. To be rational in this tradition simply is to choose and act in accordance with a set of axioms.

Despite the difficulty in determining precisely what constitutes the normativity of a given theory of “better than” comparisons, it appears reasonable to conclude that many authors who argue from practicality have at least weak normativity in mind. The ought at work in such a theory is not a moral ought but rather an advice to avoid mistakes. This form of recommendation is at play in Dutch book arguments, which can indirectly support a ratio scale (depending on how they are formulated and what they are aiming to show). Broadly conceived, Dutch book-style arguments are hypothetical scenarios in which a bookkeeper or clever salesman exploits flaws in a decision-maker’s evaluative judgments such as cycles and incorrect probability assessments. The bookkeeper gains money or valuable goods from offering bets or options to swap one comparison object for a seemingly better one in such a scenario, and this process can potentially continue ad infinitum, creating a money pump that causes unlimited losses for the decision-maker. For proponents of Dutch book arguments, such losses are a sure sign of lack of rationality.

Suppose an agent uses “better than” in a way that justifies only an ordinal scale. Then, one might argue, there is no way to combine these comparisons with probabilities in a scenario with quantified risk without losing vital information. Not using expected value may lead to value loss. Consider someone who disregards probability

⁸Although many ethicists agree with decision-theorists that values have to have the potential to be choice-guiding, some reject choice-guidance as a *defining* feature of value. See, for example, Chang [6: p.336], Hansson [11: p. 518], [12: p. 546] and [25: p. 4].

and always chooses the option that yields the best outcome in terms of value only. This strategy is terrible if the outcome is unlikely to obtain and the most likely outcome is a huge loss. There are always scenarios where not taking into consideration expected value will result in losses. Since agents *should* avoid losses, the expected value is needed to avoid value loss in some scenarios. Since an expected value principle requires at least an interval scale and makes most sense on a ratio scale, Dutch book arguments involving uncertainty can provide normative support for these scales.

4 Why “Better Than” Has No Uniform Scale

As seen in the preceding section, there are numerous grounds to believe that “better than” comparisons take place on a ratio scale. This section’s goal is to make it plausible that not all value comparisons take place on ratio scales. Two distinct types of arguments show this. First, the above linguistic arguments for interval and ratio scales are not conclusive because intensifiers and other modifiers can be given an alternative semantics that does not require such a scale. Second, “better than” assessments are multidimensional, and a deeper examination reveals that these dimensions can differ greatly from one another. Although some of them may be on ratio scales, others are just based on qualitative comparisons, with no evidence to suggest a stronger scale type.

Before proceeding, a general issue with linguistic reasoning should be addressed. Making normative claims based on the way we talk is putting the cart before the horse. Using a ratio modifier with “better than” is *only* meaningful *if* the corresponding aspect of betterness resides on a ratio scale. There must be a substantial explanation for why we *should* use the term “better than” in a way that implicates a particular scale type; it is not sufficient to suggest that we *should* take it that way because we already *do*.

That being said, there is also a modeling argument against the alleged linguistic evidence. Suppose the scale of an aspect of “better than” comparisons is *prima facie* ordinal. Then it is still possible to give a rank-based definition of “two times better than.” Rank-based definitions only include relative ranking and ignore cardinal information. Consider the following example: $a > b > c > d$. The relative rank of a comparison object can be defined by assigning 1 to the worst objects, 2 to the ones immediately better than the worst, and so forth. Going back to work on voting by mathematician Jean-Charles de Borda, these numeric scores are also called Borda Scores. For the example, the scores are $B(a)=4$, $B(b)=3$, $B(c)=2$, and $B(d)=1$. In this context, saying “a is twice as good as c” can make sense, as long as the scores for the relative ranks are assigned according to a uniform method. For example, in a ranking of mobile phones according to multiple criteria, some of which may be very subjective, it may make sense to say that phone *a* is, overall, all things considered, twice as good as phone *c*. This view is also implicit to [25], in which I use the term *canonical utility functions* for a uniform method to assign numbers to ranks. In the present example, the Borda Score of *a* is two times the Borda Score of *c*. Only the relative rank between comparison objects is required for this definition, not cardinal utility. Rank-based definitions can be adjusted to deal with ties, and they can also

be normalized to become independent of the cardinality of the sets of comparison objects. To deal with ties, we have to average ranks. For example, if $a \sim b > c > d$, then a and b get score $(4+3)/2=3.5$. To make scores independent from set sizes, one may either divide by the number of objects n in the comparison set or by the sum of the Borda Scores, which is $n(n+1)/2$.

It is also possible to develop compelling rank-based intensifier definitions. For instance, “ a is much better than b ” may be expressed formally as $B(a)/n - B(b)/n \geq k$, where n is the size of a non-empty finite set of comparison objects and k serves as a threshold for “much.” For example, for comparisons $a > b > c > d > e > f > g$ and threshold $k=0.3$, a is much better than e because $7/7 - 3/7 \geq 0.3$, whereas c is not much better than e (though still better) because $5/7 - 3/7 < 0.3$. Technically speaking, rank-based methods define absolute scales by prescribing uniform ways to assign numbers to ranks, and then these scales are relaxed to allow for arithmetic operations that only concern ranks. For instance, it can be postulated that any multiplication of a rank is only an admissible scale transformation if the result is the possible rank of an item.

The fact that rank-based definitions can be given shows that the use of “ratio modifiers” and intensifiers do not require scales as strong as interval or ratio scales. Conclusions to the contrary by Lassiter [16: pp. 177-8] and Soria-Ruiz [30: pp. 603-5] are premature, although very natural to draw, since ranked-based definitions are unusual and not commonly discussed in texts on measurement theory.⁹ Linguistic evidence suggests ratio, or at least interval scales for uses of “better than”, but does not establish their existence. On a side note, Soria-Ruiz’s intuition that only integer ratio modifiers like “two times” and “ten times” are admissible is confirmed by the rank-based approach when technical complications from ties are ignored; although this depends on the chosen representational conventions for canonical utilities, it is very natural to think of ranks as integers from lowest integer l to highest integer n for a set of n comparison objects.

So what about the above arguments from practicality? These are based on a fallacious appeal to consequences. Although it is true that not assuming interval or ratio scales leads to difficult technical challenges, which may be why cardinal utilities are frequently assumed in the literature on multicriteria decision-making, the negative consequences of not assuming these scale types do not demonstrate that “better than” resides on them. As previously stated, perhaps many authors rather take arguments by practicality as shortcuts for non-fallacious normative arguments for these scale types such as Dutch book-style arguments.

However, Dutch book-style arguments may concern two different issues. If they concern expected value, they can only be formulated under the assumption that the value already resides on an interval or ratio scale. Hence, these types of arguments are not *for* such scales but instead for using an expected value principle for dealing with uncertainty whenever those scales have already been established.

The second type of Dutch book arguments concerns the transitivity of “better than” judgments. Larry Temkin and Stuart Rachels [22, 23, 33, 34] have argued that “better than” comparisons are not necessarily transitive, whereas others such as Broome and

⁹ Rank-based methods are used in “order statistics” for making ordinal comparisons and aggregations of statistical data.

Nebel [2, 19] have maintained that they are. Suppose transitivity may indeed fail, for example that John considers a better than b and b better than c , but also c better than a . Temkin and Rachels discuss examples of this sort with larger ranges of comparison objects (“Spectrum Cases”). When John has a with these cyclic value judgments, a smart salesman can offer to sell b to him for some money (or, any other good), then offer c to John for some money, and later offer John to swap c for a again. In each case, the transaction is validated by John’s value judgments. This is an example of a diachronic Dutch book arguments under certainty, which is also called a money pump. These arguments are supposed to show that it is irrational to violate the transitivity of value judgments under certainty.

Entering the complicated discussion of the arguments for and against the transitivity of value judgments would lead us far astray. However, from the way in which money pumps are set up, the following is apparent. If one does not accept Dutch book arguments under certainty and follows Temkin that value judgments may fail to be transitive, then the value relation also does not reside on an ordinal scale or the canonical absolute scale type laid out above, since the transitivity of comparisons under certainty is a requirement for any of the above scales.¹⁰ In contrast to this, if money pumps are accepted and establish the transitivity of value judgments, then they establish this transitivity for all scales. Dutch book arguments under certainty therefore cannot be used to support or discredit a particular scale type among those discussed so far.

As a result, the normative reasons for cardinal scales are inconclusive and require further scrutiny. However, what *is* the scale type of “better than” if there appear to be no decisive arguments for a specific scale type? As is often the case in ethics, we have to resort to some form of “intuitions” and more general normative considerations. Taking a closer look at different instances of “better than,” it becomes clear that the underlying scale type may vary with the underlying dimension of the comparisons, and that overall betterness is the result of a mixture of potentially diverse dimensions. At the very least, there are always costs in terms of money, efforts, and resources to consider. For example, Scanlon [29] is right in emphasizing that friendship has intrinsic value, but this is a critique of consequentialism. Someone who gives up a good friend in order gain five other friends is not a good friend and does not understand the intrinsic value of friendship. Nevertheless, value comparisons involving friendships will remain multidimensional in practice and involve considerations about duties, costs, and resources in addition to friendship. Having a night out with a good friend means that time is not spent with family and that bills have to be paid, for instance. Every concrete use of an intrinsic value must still be balanced against other values. There are no “better than” comparisons in practice with less than two dimensions, and there are typically many more to consider.

There is no evidence that these various dimensions have the same scale type. Consider monetary worth once more to demonstrate this. Money clearly resides on a ratio

¹⁰ However, it is not a requirement for representations based on semiorders and interval orders with non-transitive equal goodness; see Luce [17, 27, 35] on these kind of representations. An anonymous reviewer to an earlier draft of this article also mentions the possibility of using sets of ordering relations to model different types of incommensurability.

scale: The 0-point is meaningful, and there is evidence that money has diminishing marginal utility. Calculating monetary differences and comparing them is the basis of established trade. Phrases like “we obtained 1.38 times the return on investment than previous year” are meaningful. Consequently, it may also be possible to use corresponding ratio modifiers in an appropriate context when the value of money is considered. For example, “John is 1.38 times better than Bob because he ran 1.38 times faster” can make sense in the context of a running contest. It may sound marked or contrived to talk this way but it makes sense because it can be justified from the properties of the dimension of what is valued.

In contrast, consider the following statement: “Democracy is 1.38 times better than oligarchy.” Anyone would (and should) be skeptical of this claim, not because there is something wrong with democracy but because there is something wrong with the ratio modifier. As a remedy, Soria-Ruiz advises that scales for “better than” need to be “round ratio scales” and gives a technical definition for these scales that only allow values that are rounded to (usually) integral values. However, this bandaid does not solve the general problem. Sentences like “democracy is 2 times/3 times/10 times/100 times better” do not fare better. They could be hyperbole, or the modifier could be a disguised intensifier, but, as Lassiter [16] points out, it is hard to see how they could be taken literally. Why is it so difficult to interpret the modifier literally in these examples? Lassiter’s explanation is that “better than” only resides on an interval scale, and the evidence for this is in his opinion that we do accept intensifiers but not ratio modifiers. However, this argument is inconclusive since it was shown above that a purely ordinal semantics can also be given for intensifiers. The competing explanation that the type of overall betterness at work, in this case, is on an ordinal scale remains feasible and is supported by the observation that many dimensions entering the overall assessment also reside on ordinal scales.

There is, perhaps, a battling of intuitions about such examples. A defender of an interval scale could argue in this particular example that the use of intensifiers *indicates* an interval scale without showing it conclusively and supplement this argument by the semantic intuition that such uses of “better than” express intensities of betterness that a rank-based account cannot explain. However, although a rank-based semantics for the “ratio modifier” *can* be provided, as previously stated, some value comparisons may be so qualitative in character that this makes no sense. Moral usage of “better than” is typically of this sort because moral reasoning often involves comparison factors that are hard or impossible to quantify.¹¹ Any value comparison based on following or not following a deontic rule is qualitative in this strict way, for example. Paying off one’s debt is preferable to not paying off one’s debt, but if someone says that it is two, three, or ten times better, this cannot be taken literally.

¹¹A reviewer remarks there might still be quantifiable moral uses like the following example: Giving a blind person directions and helping them across the street is twice as good as only helping them across the street. This judgment seems plausible when both conjuncts fall under a specific value. However, it seems implausible when the values are of a different quality. For example, it is implausible to claim that giving a child a lollipop and preventing a murder is twice as good as preventing a murder. For this article, the question may be left open if and under which circumstances combining value judgments of conjunctions is possible, as long as *some* moral value judgments are qualitative and hard to quantify.

There are also many examples of prudential value comparisons with different scale types. Suppose someone decides on tiles for a bathroom and considers how much they cost, how easy they are to get and replace, and their looks. The costs are clearly on a ratio scale. However, it is difficult to rationalize the second dimension being on a ratio scale. Even if we assign 0 utility when some tiles are completely out of stock and otherwise assign numbers, it will be difficult to justify making difference comparisons in this dimension because the assessments are based on highly subjective and hard to quantify estimates. As for the looks (aesthetic value), a ratio scale appears even more doubtful.

To be clear, judgments can be pressed into a form that makes comparisons reside on a ratio scale using preference elicitation methods like those described by Keeney & Raiffa [14]. We can *posit* a ratio scale or claim that one *should* be used. But does this justify the thesis that utility intensities are always involved? A simple ranking with ties is often the only available information. Unlike monetary costs and other easily quantifiable variables, there is no evidence that aesthetic value comparisons *generally* involve more than a ranking, for instance.

5 Conclusion

The mixed scale type view suggested above has a significant flaw. Ordinal value dimensions cannot be easily aggregated with ratio and interval scale dimensions without losing crucial information. On the one hand, ordinal value aggregation can be based on methods such as Borda Scores and distance measure minimization, but they cannot combine rank-based information like Borda Scores with utilities on a ratio scale in a meaningful way. On the other hand, despite the fact that this is often done in practice, simply converting an ordinal utility function into a ratio scale utility function lacks a measurement-theoretic justification. This quandary leads us back to questions of practicality. To do something interesting with utilities, such as combining them with quantifiable risk, a ratio scale or, at the very least, an interval scale is required. However, stipulating such a scale for merely practical, technical reasons amounts to wishful thinking. Comparisons based on rules (better to X than not to X) do not support cardinal scales, and there are many other examples where assuming a ratio scale is implausible and not supported by the specific justifications that accompany the underlying “better than” comparisons.

Acknowledgements Work on this article was funded under individual grant DL 57/2016/CP1453/CT0002 (norma transitória) by the Portuguese Foundation for Science and Technology and the project UIDB/00183/2020.

Funding Open access funding provided by FCT|FCCN (b-on).

Declarations

Conflict of Interest No conflicts of interests were involved in this work.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License,

which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- 1 Bouyssou, D., and M. Pirlot. 2003. *Ordinal aggregation and strict preferences for multi-attributed alternatives* (Issue Cahier du LAMSADE 212).
- 2 Broome, J. 1999. *Ethics out of Economics*. Cambridge University Press.
- 3 Carlson, E. 2006. Incomparability and Measurement of Value. In *The Good, the right, life and death*, eds. K. McDaniel, J. R. Raibley, R. Feldman, and M. J. Zimmerman. 19–43. Routledge.
- 4 Carlson, E. 2008. Extensive measurement with incomparability. *Journal of Mathematical Psychology* 52: 250–259.
- 5 Carlson, E. 2018. Value Theory (Axiology). In *Introduction to formal philosophy*, eds. S. O. Hansson, and V. F. Hendricks. 523–534. Springer.
- 6 Chang, R. 2005. Parity, interval value, and choice. *Ethics* 115: 331–350.
- 7 Dubois, D., J.-L. Marichal, H. Prade, M. Roubens, and R. Sabbadin. 2001. The Use of the Discrete Sugeno Integral in Decision-Making: a Survey. *International Journal of Uncertainty Fuzziness and Knowledge-Based Systems* 9(5): 539–561.
- 8 Dubois, D., and H. Prade. 1988. *Possibility theory*. Springer.
- 9 Dubois, D., and H. Prade. 2009. Possibility theory. In *Computational complexity: theory, techniques, and applications*, ed. R. A. Meyers. 2240–2252. Springer.
- 10 Fishburn, P. C. 1970. *Utility theory for decision making*. Wiley.
- 11 Hansson, S. O. 2018a. Formal investigations of value. In *Introduction to formal philosophy*, eds. S. O. Hansson, and V. F. Hendricks. 499–522. Springer.
- 12 Hansson, S. O. 2018b. Preference and choice. In *Introduction to formal philosophy*, eds. S. O. Hansson, and V. F. Hendricks. 535–548. Springer.
- 13 Hölder, O. 1901. Die Axiome Der Quantität Und die Lehre Vom Maß. *Berichte über die Verhandlungen Der Königlich-Sächsischen Gesellschaft Der Wissenschaften zu Leipzig Mathematisch-Physische Klasse* 53: 1–64.
- 14 Keeney, R. L., and H. Raiffa. 1993. *Decisions with multiple objectives: preferences and value tradeoffs*. Cambridge University Press.
- 15 Krantz, D. H., R. D. Luce, P. Suppes, and Tversky. 1971. *Foundations of measurement, volumes I-III*. New York: Academic.
- 16 Lassiter, D. 2017. *Graded modality: qualitative and quantitative perspectives*. Oxford University Press.
- 17 Luce, D. R. 1956. Semiordeers and a theory of utility discrimination. *Econometrica* 24(2): 178–191.
- 18 Marichal, J.-L. 2009. Aggregation functions for decision making. In *Decision-making process: concepts and methods*, eds. D. Bouyssou, D. Dubois, M. Pirlot, and H. Prade. 673–721. Wiley.
- 19 Nebel, J. M. 2017. The Good, the bad, and the transitivity of Better Than. *Nous* 52(4): 874–899.
- 20 Rabinowicz, W. 2008. Value relations. *Theoria* 74: 18–49.
- 21 Rabinowicz, W. 2010. Value relations: Old Wine in New Barrels. In *Philosophical papers dedicated to Kevin Mulligan*, ed. A. Reboul. Université de Genève.
- 22 Rachels, S. 1998. Counterexamples to the transitivity of better than. *Australasian Journal of Philosophy* 76(1): 71–83.
- 23 Rachels, S. 2001. A set of solutions to Parfit’s problems. *Nous* 35(2): 214–238.
- 24 Ramsey, F. P. 1931. Truth and Probability. In *The foundations of Mathematics and other logical essays*, 156–198. Kegan, Paul, Trench, Trubner & Co.
- 25 Rast, E. 2022. *Theory of Value structure*. Lexington.
- 26 Roberts, F. S. 1979. *Measurement theory*. Adison Wesley.

- 27 Roubens, M., and P. Vincke. 1985. *Preference modelling*. Springer.
- 28 Savage, L. J. 1954. *The foundations of statistics*. Wiley.
- 29 Scanlon, T. M. 1998. *What we owe to each other*. Belknap Press of Harvard University.
- 30 Soria-Ruiz, A. 2021. Value and Scale. *Organon F* 28(3): 596–625.
- 31 Spohn, W. 1999. *Ranking Functions, AGM Style*. Technical report. <http://www.lucs.lu.se/spinning/categories/dynamics/Spohn/>.
- 32 Stevens, S. S. 1946. On the theory of scales of Measurement. *Science* 103(2684): 677–680.
- 33 Temkin, L. S. 1987. Intransitivity and the Mere Addition Paradox. *Philosophy & Public Affairs* 16(2): 138–187.
- 34 Temkin, L. S. 2012. *Rethinking the Good: Moral ideals and the nature of practical reasoning*. Oxford University Press.
- 35 Vincke, P., and M. Pirlot. 1997. *Semiororders: Properties, representations, applications*. Springer.
- 36 Weidman Sassoon, G. 2013. A typology of Multidimensional Adjectives. *Journal of Semantics* 30: 335–380.
- 37 Weidman Sassoon, G., and J. Fadlon. 2017. The role of dimensions in classification under predicates predicts their status in degree constructions. *Glossa* 2(1): 1–40.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.