Measurement-Theoretic Foundations of Weighted Utilitarianism (Extended Abstract)

Satoru Suzuki

Faculty of Arts and Sciences, Komazawa University,
1-23-1, Komazawa, Setagaya-ku, Tokyo 154-8525, Japan
bxs05253@nifty.com

1 Measurement-Theoretic Considerations of Harsanyi's Aggregation Theorem

Harsanyi [5] attempts to develop expected utility theory of von Neumann and Morgenstern [10] to provide a formalization of *(weighted) utilitarianism*. Weymark [15] refers to this result as Harsanyi's Aggregation Theorem. Here we would like to define such measurement-theoretic concepts as

1. scale types,
2. representation and uniqueness theorems, and
3. measurement types

on which the argument of this paper is based: *First*, we classify *scale types* in terms of the class of *admissible transformations* φ. A scale is a triple $\langle \mathfrak{U}, \mathfrak{B}, f \rangle$ where $\mathfrak{U}$ is an observed relational structure that is qualitative, $\mathfrak{B}$ is a numerical relational structure that is quantitative, and f is a homomorphism from $\mathfrak{U}$ into $\mathfrak{B}$.

– A is the domain of $\mathfrak{U}$ and B is the domain of $\mathfrak{B}$. When the admissible transformations are all the functions $\varphi : f(A) \to B$, where $f(A)$ is the range of f, of the form $\varphi(x) := \alpha x; \alpha > 0$. φ is called a *similarity transformation*, and a scale with the similarity transformations as its class of admissible transformations is called a *ratio scale*. Length is an example of a ratio scale.
– When the admissible transformations are all the functions $\varphi : f(A) \to B$ of the form $\varphi(x) := \alpha x + \beta; \alpha > 0$, φ is called a *positive affine transformation*, and a corresponding scale is called an *interval scale*. Temperature on the Fahrenheit scale and temperature on the Celsius scale are examples of interval scales.
– When a scale is unique up to order, the admissible transformations are *monotone increasing functions* φ satisfying the condition that $x \geq y$ iff $\varphi(x) \geq \varphi(y)$. Such scales are called *ordinal scales*. The Mohs scale is an example of a ordinal scale.
– A scale is called a *log-interval scale* if the admissible transformations are functions φ of the form $\varphi(x) := \alpha x^{\beta}; \alpha, \beta > 0$. Psychophysical functions are examples of log-interval scales.

Second, we state about *representation* and *uniqueness theorems*. There are two main problems in measurement theory:

1. the *representation problem*: Given a numerical relational structure $\mathfrak{B}$, find conditions on an observed relational structure $\mathfrak{A}$ (necessary and) sufficient for the *existence* of a homomorphism f from $\mathfrak{A}$ to $\mathfrak{B}$ that preserves all the relations and operations in $\mathfrak{A}$.
2. the *uniqueness problem*: Find the transformation of the homomorphism f under which all the relations and operations in $\mathfrak{A}$ are preserved.

A solution to the former can be furnished by a *representation theorem* that specifies conditions on $\mathfrak{A}$ are (necessary and) sufficient for the existence of f. A solution to the latter can be furnished by a *uniqueness theorem* that specifies the transformation up to which f is unique. *Third*, we classify *measurement types*.

1. *ordinal measurement*
2. *cardinal measurement*
   (a) extensive measurement
   (b) difference measurement
      i. *algebraic-difference measurement*
      ii. positive-difference measurement
      iii. absolute-difference measurement

Suppose A is a set, $\succ$ is a binary on A, $\bigcirc$ is a binary operation on A, $\succ'$ is a quaternary on A, and f is a real-valued function. Then we call

the representation $a \succ b$ iff $f(a) > f(b)$

ordinal measurement. When such f exists, then $\langle \mathfrak{A}, \mathfrak{B}, f \rangle$ is an *ordinal scale*. We call

the representation $a \succ b$ iff $f(a) > f(b)$ and $f(a \bigcirc b) = f(a) + f(b)$

extensive measurement. When such f exists, then $\langle \mathfrak{A}, \mathfrak{B}, f \rangle$ is a *ratio scale*. We call

the representation $(a, b) \succ' (c, d)$ iff $f(a) - f(b) > f(c) - f(d)$

, when the direction of differences is taken into consideration, *positive-difference measurement*, when the direction of differences is not taken into consideration, *algebraic-difference measurement*. In the latter case, when such f exists, then $\langle \mathfrak{A}, \mathfrak{B}, f \rangle$ is a *interval scale*. We call

the representation $(a, b) \succ' (c, d)$ iff $|f(a) - f(b)| > |f(c) - f(d)|$

absolute-difference measurement. In terms of these measurement-theoretic concepts, Harsanyi's Aggregation Theorem can be stated in the following way:

Theorem 1 (Harsanyi's Aggregation Theorem). *Suppose that individual and social binary preference relations $\succsim_i$ $(i = 1, \ldots, n)$ and $\succsim$ on the set of lotteries satisfy von Neumann-Morgenstern axioms, and also suppose that $\succsim_i$ and $\succsim$ satisfy the Strong Pareto condition. Furthermore, suppose that $\succsim_i$ and $\succsim$ are represented by individual and* *social expected utility functions $U_i (i = 1, \ldots, n)$ and U respectively. Then, there are real numbers $\alpha_i (> 0)(i = 1, \ldots, n)$ and β such that*

$$U(p) = \sum_{i=1}^{n} \alpha_i U_i(p) + \beta,$$

for any lottery p.

The next corollary directly follows from this theorem:

Corollary 1 (Weighted Utilitarianism on Set of Lotteries). *Lotteries are socially ranked according to a weighted utilitarian rule:*

$$U(p) \geq U(q) \text{ iff } \sum_{i=1}^{n} \alpha_i U_i(p) \geq \sum_{i=1}^{n} \alpha_i U_i(q),$$

for any lotteries $p.q$.

Harsanyi's Aggregation Theorem follows from the next lemmas:

Lemma 1 (Representation). *Suppose that $\succsim_i$ $(i = 1, \ldots, n)$ and $\succsim$ satisfy Weak Order, Continuity, and Independence. Then, there exist individual and social expected utility functions $U_i (i = 1, \ldots, n)$ and U such that*

$$\begin{cases} p \succsim_i q \text{ iff } U_i(p) \geq U_i(q), \\ p \succsim q \text{ iff } U(p) \geq U(q), \end{cases}$$

for any lotteries p, q.

Lemma 2 (Uniqueness). *Suppose that $\succsim_i$ $(i = 1, \ldots, n)$ and $\succsim$ on the set of lotteries satisfy not only the conditions for the representation above but also Nondegeneracy. Then, the individual and social expected utility functions U_i and U are unique up to a positive affine transformation.*

There are *at least two* well-known criticisms on Harsanyi's Aggregation Theorem. The *first* criticism is by Sen [14]: Von Neumann-Morgenstern axioms on individual binary preference relations in Lemma 1 are for *ordinal measurement* and, therefore, any *monotone increasing (even non-affine) transform* of an expected utility function is a satisfactory representation of individual binary preference relations. However, (weighted) utilitarianism requires a theory of *cardinal utility*, and so Harsanyi is not justified in giving his theorems utilitarian interpretations. The *second* criticism is based on the following probability agreement theorem that is provided by Broome [2]:

Theorem 2 (Probability Agreement Theorem). *Suppose that individual and social binary preference relations $\succsim_i$ $(i = 1, \ldots, n)$ and $\succsim$ on the set of lotteries satisfy von Neumann-Morgenstern axioms. Then, $\succsim_i$ and $\succsim$ cannot satisfy the strong Pareto condition unless every individual agrees about the probability of every elementary event.*

In fact, under many circumstances, the members of a society have different beliefs (probabilities).

2 Harvey's Aggregation Theorem and Cardinal Utility

In order to escape these two criticisms, we *might* resort to Harvey's Aggregation Theorem ([6]) that has *quaternary preference relations* as primitive that can be represented by *utility differences*, and is concerned only with quaternary preference relations on the set of *outcomes* but is not concerned with binary preference relations on the set of lotteries in Harsanyi's Aggregation Theorem. Lange [8] is the first to connect formally the ranking of *utility differences* with *positive affine transformations* of utility functions. However, he does not use the expression "cardinal utility".[1] Alt [1] is considered to be the first to prove the representation theorem for quaternary preference relations that can be represented by utility differences, and the uniqueness theorem on the uniqueness of the utility functions up to positive affine transformations. However, he also dose not connect utility differences with the expression "cardinal utility". Samuelson [12] is the first to connect utility differences in which utility functions are unique up to positive affine transformations "cardinal utility", though he takes a negative toward cardinal utility. Harvey [6, p.69] defines *difference-worth conditions* as follows: We will use conditions on a quaternary preference relation $\succsim$ as any set of conditions that are satisfied iff there exists a *worth function* w such that

$$(a, b) \succsim (c, d) \text{ iff } w(a) - w(b) \geq w(c) - w(d)$$

for any outcome a, b, c, d, and we will refer to any such conditions as a set of difference-worth conditions. Then Harvey's Aggregation Theorem can be stated in the following way:

Theorem 3 (Harvey's Aggregation Theorem). *Suppose that individual and social quaternary preference relations* $\succsim_i$ $(i = 1, \dots, n)$ *and* $\succsim$ *on the set of outcomes satisfy a certain set of difference-worth conditions. Then,* $\succsim_i$ *and* $\succsim$ *satisfy the strong Pareto condition iff there are real numbers* $\alpha_i(> 0)(i = 1, \dots, n)$ *and* β *such that*

$$w(a) = \sum_{i=1}^{n} \alpha_i w_i(a) + \beta,$$

for any outcome a.

The next corollary directly follows from this theorem:

Corollary 2 (Weighted Utilitarianism on Set of Outcomes). *Outcomes are socially ranked according to a weighted utilitarian rule.*

Harvey's Aggregation theorem follows from the next lemmas:

Lemma 3 (Representation). *Suppose that* $\succsim_i$ $(i = 1, \dots, n)$ *and* $\succsim$ *on the set of outcomes satisfy a certain set of difference-worth conditions. Then, there exist individual and social worth functions* $w_i (i = 1, \dots, n)$ *and* w *such that*

$$(1) \begin{cases} (a, b) \succsim_i (c, d) \textit{ iff } w_i(a) - w_i(b) \geq w_i(c) - w_i(d), \\ (a, b) \succsim (c, d) \textit{ iff } w(a) - w(b) \geq w(c) - w(d), \end{cases}$$

[1] About the history of cardinal utility, refer to [9, pp.95–116].

for any outcome a, b, c, d.

Lemma 4 (Uniqueness). *Suppose that $\succsim_i$ $(i = 1, \ldots, n)$ and $\succsim$ on the set of outcomes satisfy the conditions for the representation above. Then, $w_i (i = 1, \ldots, n)$ and w are unique up to a positive affine transformation.*

Because any set of difference-worth conditions is for *algebraic-difference measurement* that is a kind of *cardinal measurement*, this theorem *might* escape the first criticism. When Hammond [4] attempts to salvage utilitarianism in the way that the (strong) Pareto condition can apply only to *outcomes*. Harvey takes the same position as Hammond that *might* enable this theorem to escape the second criticism.

3 Criticisms of Harvey's Aggregation Theorem from Measurement-Theoretic Point of View

Now we inspect Harvey's Aggregation Theorem from a measurement-theoretic point of view. We offer *two* criticisms on Harvey's Aggregation Theorem: The *first* criticism is as follows: As Roberts [11, p.139] says, the only set of necessary and sufficient difference-worth conditions is due to Scott [13], and requires the assumption that the set of outcomes is *finite*. So when there is no domain-size limitation, the set of necessary and sufficient difference-worth conditions is still unknown. The *second* criticism is as follows: The most essential task of aggregation theorem from a measurement-theoretic point of view is to prove the existence of individual and social worth functions that represent individual and social quaternary preference relations which satisfy not only difference-worth conditions but also the strong Pareto condition. Harvey's Aggregation Theorem is not of such a form. Harvey [6, p.72] himself seems to acknowledge this fact:

> I view the result in Harsanyi [5] and the result presented here as uniqueness results rather than as existence results. ... an expected-utility function or a worth function is unique up to a positive affine function.

Then, is what Harvey says true? What should be proved is the uniqueness of individual and social worth functions that represent individual and social quaternary preference relations which satisfy *not only* difference-worth conditions *but also* the strong Pareto condition. However, in Lemma 3, individual and social quaternary preference relations satisfy *only* difference-worth conditions. So the uniqueness of individual and social worth functions that represent individual and social quaternary preference relations which satisfy both difference-worth conditions and the strong Pareto condition is not guaranteed. After all, Harvey's Aggregation Theorem can give any answer neither to the *representation problem* nor to the *uniqueness problem*.

4 Our Aggregation Theorems

The *aim* of this paper is that we prove new aggregation theorems, which escape these two criticisms, inspired by Harvey's Aggregation Theorem. Our aggregation representation and uniqueness theorems (main results) can be stated in the following way:

Theorem 4 (Aggregation Representation Theorem (Main Result 1)). *Suppose that individual and social quaternary preference relations $\succsim_i$ $(i = 1, \ldots, n)$ and $\succsim$ on the set of outcomes satisfy Weak Order, Order Reversal, Weak Monotonicity, Soluvability and Archimedean condition in Krantz et al. [7], and also suppose that $\succsim_i$ and $\succsim$ satisfy the strong Pareto condition. Then, there exist individual and social utility functions $u_i (i = 1, \ldots, n)$ and u such that*

$$(1) \quad \begin{cases} (a, b) \succsim_i (c, d) \text{ iff } u_i(a) - u_i(b) \geq u_i(c) - u_i(d), \\ (a, b) \succsim (c, d) \text{ iff } u(a) - u(b) \geq u(c) - u(d), \end{cases}$$

for any outcome a, b, c, d and there are real numbers $\alpha_i (> 0)(i = 1, \ldots, n)$ and β such that

$$u(a) = \sum_{i=1}^{n} \alpha_i u_i(a) + \beta,$$

for any outcome a.

One of key techniques for proving this theorem is a version of *Moment Theorem* in abstract linear spaces in Domotor [3]. The next corollary directly follows from this theorem.

Corollary 3 (Weighted Utilitarianism on Set of Outcomes). *Outcomes are socially ranked according to a weighted utilitarian rule.*

Theorem 5 (Aggregation Uniqueness Theorem (Main Result 2)). *Suppose that $\succsim_i$ $(i = 1, \ldots, n)$ and $\succsim$ on the set of outcomes satisfy the conditions for the representation above. Then, $u_i (i = 1, \ldots, n)$ and u are unique up to a positive affine transformation.*

Because our aggregation theorems do not include any set of necessary and sufficient difference-worth (algebraic difference) conditions but include only some sufficient conditions, it escapes the first criticism. Because our aggregation representation theorem guarantees the existence of individual and social utility functions that represent individual and social quaternary preference relations which satisfy not only difference-worth (algebraic difference) conditions but also the strong Pareto condition, and our aggregation uniqueness theorem guarantees the uniqueness of such functions, they escape the second criticism. Finally, we would like to discuss the following possible criticism, which is similar to the first criticism on Harsanyi's Aggregation Theorem, from a measurement-theoretic point of view to our aggregation representation and uniqueness theorems. We can prove the following propositions similar to Lemma 1 and Lemma 2 of Harsanyi's Aggregation Theorem:

Proposition 1 (Representation). *Suppose that individual and social quaternary preference relations* $\succsim_i$ $(i = 1, \ldots, n)$ *and* $\succsim$ *on the set of outcomes satisfy Weak Order, Order Reversal, Weak Monotonicity, Soluvability and Archimedean condition in Krantz et al. [7]. Then, there exist individual social utility functions* $u_i (i = 1, \ldots, n)$ *and* u *such that*

$$(2) \begin{cases} (a,b) \succsim_i (c,d) \text{ iff } \dfrac{u_i(a)}{u_i(b)} \geq \dfrac{u_i(c)}{u_i(d)}, \\ (a,b) \succsim (c,d) \text{ iff } \dfrac{u(a)}{u(b)} \geq \dfrac{u(c)}{u(d)}, \end{cases}$$

for any outcome a, b, c, d.

Proposition 2 (Uniqueness). *Suppose that* $\succsim_i$ $(i = 1, \ldots, n)$ *and* $\succsim$ *on the set of outcomes satisfy the conditions for the representation above. Then,* $u_i (i = 1, \ldots, n)$ *and* u *are unique up to a transformation of functions of the form* $\alpha x^{\beta}; \alpha, \beta > 0$.

These propositions imply that Weak Order, Order Reversal, Weak Monotonicity, Soluvability and Archimedean condition in Krantz et al. [7] can satisfy not only (1) but also (2). So our aggregation theorems cannot justify weighted utilitarianism. How can we escape this criticism? Von Neumann-Morgenstern axioms on individual binary preference relations in Lemma 1 are considered, as we have argued earlier, to be for ordinal measurement according to the first criticism by Sen. In this criticism, the fact that any monotone increasing (even non-affine) transform of an expected utility function is a satisfactory representation of individual binary preference relations is used to prove that von Neumann-Morgenstern axioms on individual binary preference relations in Lemma 1 are not for cardinal measurement but for *ordinal measurement*. In Lemma 2, von Neumann-Morgenstern axioms together with the Us being expected utility functions imply the cardinality of Us. So von Neumann-Morgenstern axioms only does not justify the cardinality of Us. On the other hand, because our axioms on individual quaternary preference relations are in nature for *utility-difference measurement (algebraic-difference measurement)* that is a kind of *cardinal measurement*, only our axioms justify the cardinality of ws. Propositions 1 and 2 is not about utility-difference measurement. So we do not have to take these propositions into consideration.

References

1. Alt, F.: Über die Mäßbarkeit des Nutzens. Zeitschrift für Nationalökonomie 7, 161–169 (1936)
2. Broome, J.: Weighing Goods. Blackwell, Oxford (1991)
3. Domotor, Z.: Ordered sum and tensor product of linear utility structures. Theory and Decision 11, 375–399 (1979)
4. Hammond, P.: Utilitarianism, uncertainty and information. In: Sen, A., Williams, B. (eds.) Utilitarianism and Beyond, pp. 85–102. Cambridge University Press, Cambridge (1982)
5. Harsanyi, J.C.: Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. Journal of Political Economy 63, 309–321 (1955)

6. Harvey, C.M.: Aggregation of individuals' preferences intensities into social preference intensity. Social Choice and Welfare 16, 65–79 (1999)
7. Krantz, D.H., et al.: Foundations of Measurement, vol. 1. Academic Press, New York (1971)
8. Lange, O.: The determinateness of the utility function. Review of Economic Studies 1, 218–224 (1934)
9. Moscati, I.: Measuring Utility. Oxford University Press, Oxford (2019)
10. von Neumann, J., Morgenstern, O.: Theory of Games and Economic Behavior. Princeton University Press, Princeton (1944)
11. Roberts, F.S.: Measurement Theory. Addison-Wesley, Reading (1979)
12. Samuelson, P.: The numerical representation of ordered classifications and the concept of utility. Review of Economic Studies 6, 65–70 (1938)
13. Scott, D.: Measurement structures and linear inequalities. Journal of Mathematical Psychology 1, 233–247 (1964)
14. Sen, A.: Welfare inequalities and Rawlsian axiomatics. Theory and Decision 7, 243–262 (1976)
15. Weymark, J.A.: A reconsideration of the harsanyi-sen debate on utilitarianism. In: Elster, J., Roemer, J.E. (eds.) Interpersonal Comparisons of Well-Being, pp. 255–320. Cambridge University Press, Cambridge (1991)