Title

RELATIONSHIP BETWEEN LUCES CHOICE AXIOM, THURSTONES THEORY OF COMPARATIVE JUDGMENT, AND DOUBLE EXPONENTIAL-DISTRIBUTION

Journal

JOURNAL OF MATHEMATICAL PSYCHOLOGY, 15(2)

ISSN

0022-2496

Author

YELLOTT, JI

Publication Date

1977

DOI

10.1016/0022-2496(77)90026-8

JOURNAL OF MATHEMATICAL PSYCHOLOGY 15, 109–144 (1977)

The Relationship between Luce's Choice Axiom, Thurstone's Theory of Comparative Judgment, and the Double Exponential Distribution[1]

JOHN I. YELLOTT, JR.

School of Social Sciences, University of California, Irvine 92717

Holman and Marley have shown that Thurstone's Case V model becomes equivalent to the Choice Axiom if its discriminal processes are assumed to be independent double exponential random variables instead of normal ones. It is shown here that for pair comparisons, this representation is not unique; other discriminal process distributions (specifiable only in terms of their characteristic functions) also yield a model equivalent to the Choice Axiom. However, none of these models is equivalent to the Choice Axiom for triple comparisons: There the double exponential representation is unique. It is also shown that within the framework of Thurstone's theory, the double exponential distribution, and hence the Choice Axiom, is implied by a weaker assumption, called "invariance under uniform expansions of the choice set."

1. INTRODUCTION

1.1. *Historical Background*

As premises for a model of choice behavior, Luce's (1959) Choice Axiom and Thurstone's (1927) Theory of Comparative Judgment seem at first glance to be not only different, but quite unrelated. The Axiom simply imposes an intuitively plausible constraint on observable choice probabilities, while the theory postulates an imaginary psychological process, underlying observable behavior, wherein choice objects are represented by random variables ("discriminal processes") which the subject compares in order to arrive at a decision. However, it has been recognized since the earliest days of the Choice Axiom that this apparent dissimilarity is only superficial, and that the two ideas are in fact surprisingly closely related. In his 1959 monograph, Luce included a table showing that for pair-comparison experiments, the predictions of the Choice Axiom are virtually identical to those of Case V of Thurstone's theory. (Recall that in Case V the discriminal processes corresponding to a set of objects $o_1, o_2, \ldots$ take the form $u_1 + \mathbf{X}_1$, $u_2 + \mathbf{X}_2, \ldots$, where $u_1, u_2, \ldots$ are constants (scale values) and $\mathbf{X}_1, \mathbf{X}_2, \ldots$ are independent identically distributed normal random variables.) He then

[1] I thank J. Aczel, D. A. Darling, I. I. Kotlarski, R. D. Luce, A. A. J. Marley, J. E. K. Smith, B. Wandell, and J. L. Zinnes for helpful advice; M. Vernoy for assistance in preparing Fig. 1, and the Institute for Advanced Study, Princeton, for support in the form of a visiting membership during 1970–1971, when this research was begun.

ISSN 0022-2496

pointed out that this near identity is not simply coincidental; instead, it reflects the mathematical fact that for pair-comparison experiments the Choice Axiom is equivalent to a version of Thurstone's theory in which the differences between the discriminal processes (that is, the random variables $(u_i + \mathbf{X}_i) - (u_j + \mathbf{X}_j)$) have a logistic distribution instead of the normal distribution implied by Case V. This fact had already been noted by Adams and Messick (1957), who had also shown that the logistic is unique in this respect: A Thurstone model with independent discriminal processes is equivalent to the Choice Axiom for pair comparisons if and only if the differences between discriminal processes are logistic random variables. (This result is proved here in Section 2.6.)

Luce then raised two questions: (1) What must be the distribution of the discriminal processes themselves, in order for their differences to be logistic (so that the resulting model is equivalent to the Choice Axiom for pair comparisons)? (2) Are there any discriminal process distributions for which the resulting model is equivalent to the Choice Axiom for experiments with larger sets of alternatives? (see Luce, 1959, open problem B-2, p. 144.).

A partial answer to both questions was subsequently supplied by Holman and Marley (cited in Luce & Suppes, 1965), who showed that if Thurstone's discriminal processes are assumed to have the *double exponential distribution*:

$$P[u_i + \mathbf{X}_i \leqslant x] = \exp\{-e^{-[a(x-u_i)+b]}\} \qquad (-\infty < x < \infty)$$

(where a and b are arbitrary constants, $a > 0$; see Fig. 1B), then $\mathbf{X}_i - \mathbf{X}_j$ will be logistic, and the resulting model is equivalent to the Choice Axiom for any choice experiment, not simply for pair comparisons. (This result is proved here in Section 2.6.) However, Holman and Marley did not show that the double exponential is the *only* distribution with this property. In presenting their result, Luce and Suppes (1965) remarked that "it is conjectured that [the double exponential distribution is] the only reasonably well behaved example, but no proof has yet been devised" (p. 339).

The present paper carries this line of development two steps further. First, it provides a solution to the uniqueness problem: For pair-comparison experiments the double exponential is *not* the only distribution that yields a Thurstone model equivalent to the Choice Axiom (Section 3.2), but for experiments involving both pair and triple comparisons, it *is* (Theorem 5, Section 3.4). Second, it provides a kind of explanation for the special status of the double exponential distribution: This distribution (and consequently, the Choice Axiom) can be derived in a straightforward way, starting from Thurstone's original discriminal process notion, by applying Luce's idea that observable choice probabilities ought to satisfy some intuitively plausible constraint. It turns out that there is a constraint ("invariance under uniform expansions of the choice set") which is a good deal weaker than the Choice Axiom itself, but which implies that Axiom when combined with Thurstone's discriminal process assumption.

In order to solve the original uniqueness problem, it was natural to generalize it, and consequently a good deal of the paper is devoted to the uniqueness properties of arbitrary discriminal process distributions.

1.2. *Overview and Summary of Results*

1.2.1. *Uniqueness*

We start with Thurstone's basic idea that choice objects $o_1, \ldots, o_n$ are represented by discriminal processes $u_1 + \mathbf{X}_1, \ldots, u_n + \mathbf{X}_n$, where $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are independent identically distributed random variables with a common distribution function F, and, for any given set of objects, $u_1, \ldots, u_n$ are real constants—the scale values of those objects. When the subject is presented with a set of objects and required to choose one, he picks the object with the largest discriminal process. Now we can imagine a class of *Thurstone models* of this type, each one corresponding to a different probability distribution F: Thurstone's original Case V model corresponds to the assumption that F is a normal distribution, and Holman and Marley's result shows that if F is a double exponential distribution, the resulting model is *equivalent* to the Choice Axiom, in the sense that any set of choice probabilities generated by assigning arbitrary values to the $\{u_i\}$ will satisfy the Choice Axiom, and any set of choice probabilities that satisfy the Choice Axiom can be duplicated by a suitable choice of values for the $\{u_i\}$. Then our specific uniqueness problem is to determine whether any other type of distribution also yields a model equivalent to the Choice Axiom.

One might suppose that this problem could be solved by appealing to Adams and Messick's result, since a model with F arbitrary could be equivalent to the Choice Axiom for pair comparison experiments only if F is a distribution for which the differences $\mathbf{X}_i - \mathbf{X}_j$ are logistic. However, it turns out that F need not be double exponential in order for $\mathbf{X}_i - \mathbf{X}_j$ to be logistic: There are other distributions, technically acceptable as bases for Thurstone models, though not at all appealing, that have the same property. (This is shown in Section 3.2.) Consequently, if we restrict our attention to pair-comparison experiments, the double exponential discriminal process distribution is *not* the only one that yields a Thurstone model equivalent to the Choice Axiom.

Generalizing the problem, let $\mathscr{T}_F$ denote the Thurstone model corresponding to distribution function F, and consider the set $\{\mathscr{T}_F\}$ consisting of all such models, one for every distribution function that satisfies certain minimal constraints designed to preserve the spirit of Thurstone's original idea. (See Definition 3, Section 2.6.) Then the general uniqueness question is whether two models $\mathscr{T}_F$ and $\mathscr{T}_G$ can be equivalent if F and G are different distributions. We know already, from the case of the double exponential distribution, that this can happen if we consider only pair-comparison experiments, and this turns out to be true for other distributions as well; the model with F exponential is another example (Section 3.2. This has a bearing on the "Threshold" model proposed by Dawkins, 1969.) In fact the only important exception seems to be the normal (i.e., Case V) model, whose pair-comparison predictions cannot be entirely duplicated by any other Thurstone model (Section 3.2).

Since pair-comparison predictions alone do not uniquely identify the discriminal process distribution of a Thurstone model, it is natural to wonder whether this is also true of experiments in which the subject is confronted with more than two objects at a time. We define a *complete choice experiment* to be one in which choice probabilities are determined for every subset of objects, the simplest nontrivial case being a complete experiment with three objects. It is not difficult to show that two models $\mathscr{T}_F$ and $\mathscr{T}_G$ cannot be equivalent for complete experiments involving choice sets that are arbitrarily large unless F and G are both distributions of the same type (e.g., both double exponential, but perhaps with different means and variances). This is proved in Section 3.3 (Theorem 3). However, this still leaves open the possibility that models with different types of discriminal process distributions might be equivalent for all practical purposes. For an important subclass of Thurstone models—all those in which the distribution F has a nonvanishing characteristic function—this possibility can be ruled out, because it can be shown that if F has a nonvanishing characteristic function, and G is an arbitrary distribution, the models $\mathscr{T}_F$ and $\mathscr{T}_G$ can be equivalent for complete experiments with three (or more) objects if and only if F and G are distributions of the same type (Theorem 4, Section 3.4). Since double exponential distributions have nonvanishing characteristic functions, and we know that $\mathscr{T}_F$ is equivalent to the Choice Axiom if F is double exponential, it follows that another model $\mathscr{T}_G$ can also be equivalent to the Choice Axiom (and consequently to $\mathscr{T}_F$) for complete experiments with three objects only if G is also double exponential (Theorem 5, Section 3.4).

This result provides the sharpest possible solution to our original uniqueness problem. However, the general version of that problem cannot be disposed of quite so neatly, because F can have a characteristic function with zeros and still yield a perfectly acceptable Thurstone model $\mathscr{T}_F$. Thus the general question, whether $\mathscr{T}_F$ and $\mathscr{T}_G$ can be equivalent for complete experiments with three objects if F and G are distributions of different types, remains open. Arguments given in Section 3.4 show that if there are distributions with this property, they must be fairly exotic, and my conjecture is that they do not exist at all. However, I have not been able to prove this. Technically, the answer turns out to depend on whether a certain functional equation involving characteristic functions (Eq. (23)) has only solutions of a certain form (Eq. (24)). I have not been able to answer this question, but I suspect that a specialist in characteristic functions would find it fairly straightforward. In any event, it is left here as an open problem.

1.2.2. *Deriving the Choice Axiom*

When Thurstone decided to assign normal distributions to his discriminal processes, his choice was tentative, and he clearly envisioned the possibility that some other distribution might do a better job:

> The residuals [discrepancies between observed and predicted results] should be investigated to ascertain whether they are a minimum when the normal or Gaussian distribution of discriminal processes is used as a basis for defining the psychological scale. Triangular and other forms of distribution might be tried. Such an experimental demonstration would constitute perhaps the most fundamental discovery that has yet been made in the field of psychological measurement. Lacking such proof and since the Gaussian distribution of discriminal processes yields scale values that agree very closely with the experimental data, I have defined the psychological continuum that is implied in Weber's Law, in Fechner's Law, and in educational quality scales, as that particular linear spacing of the stimuli which gives a Gaussian distribution of discriminal processes (1927, p. 285).

Now if Thurstone had, for some reason, decided on the double exponential distribution instead of the normal, he would in effect have discovered the Choice Axiom. Since this distribution was in those days quite unknown (it only began to become prominent in statistics in 1928, when Fisher and Tippet showed that it is one of three possibilities for the limit distribution of the maximum of a sequence of random variables. See Section 2.6) Thurstone would not have been likely to think of it offhand. But it is interesting to speculate on how he might have been led to it in a logical way, starting from some intuitively plausible premise about choice behavior. Section 4 describes one such premise: that choice probabilities should be "invariant under uniform expansions of the choice set." This notion is motivated by examples of the following kind: Imagine a subject confronted either with a choice set containing a cup of coffee, a cup of tea, and a glass of milk, or with a set containing k cups of coffee, all indistinguishable from the first one, and similarly k identical cups of tea and k identical glasses of milk. In both cases the subject only gets to choose one object. Under these circumstances it seems fairly reasonable to suppose that the probability of choosing coffee (that is, the sum of the probabilities for each individual cup) should be the same in both cases, no matter what the value of k. Invariance of this sort is implied by the Choice Axiom, but taken alone it is clearly a weaker assumption than the Choice Axiom, and does not imply it. However, Section 4 shows that a Thurstone model $\mathscr{T}_F$ guarantees this sort of invariance only when F is a double exponential distribution (Theorem 6). Thus within the framework of Thurstone's basic discriminal process idea, invariance under uniform expansions of the choice set is an axiomatic filter that singles out the double exponential distribution, which in turn implies the Choice Axiom. It is natural to wonder whether, if Thurstone had thought along these lines, he would still have settled on the normal distribution, or whether the prominent role played by that distribution in the history of psychometrics might not have fallen instead to the double exponential.

Outline of the Balance of the Paper

Section 2 deals with preliminary matters: Sections 2.1 and 2.2 introduce the notation for choice probabilities and systems of such probabilities; Section 2.3 explicates the notion of equivalence between choice models (that is, theories); Sections 2.4 and 2.5 review Luce's Choice Axiom and Thurstone's Theory of Comparative Judgment; Section 2.6 defines the general class of Thurstone models and discusses three cases of special interest: the normal (Case V) model, the double exponential model (earlier work on the relationship between this model and the Choice Axiom is reviewed here), and the exponential model (which turns out to be equivalent to Dawkin's (1969) Threshold Model, but only for pair comparisons). Then Section 3 deals with the uniqueness problem: Section 3.1 provides an overview of the results; Section 3.2 considers the special case of pair comparisons, with particular emphasis on the three specific models just mentioned; Section 3.3 considers the case of complete experiments with choice sets of arbitrary size, and Section 3.4 deals with triple comparisons—this section contains the final solution for the special problem of the Choice Axiom. Section 4 defines the idea of "invariance under uniform expansions of the choice set" and, by appealing back to an earlier result, shows how this principle implies the double exponential distribution. Finally, Section 5 comments briefly on an "irreversibility" theorem proved by Luce in his 1959 monograph, and also outlines the relationship between Thurstone models and the more general class of "independent random utility models" defined by Luce and Suppes (1965). New results of central interest are called theorems; old results and those that are only preparatory are called lemmas. To make the paper self-contained, explicit proofs are given for every important step in the mathematical development, even though in a few places this has meant re-proving results already available in the literature.

2. Notation, Definitions, Conventions

2.1. *Basic Notation*

Suppose $o_1, o_2, \ldots, o_n$ is a set of n *choice objects* (e.g., n tones, to be compared for loudness), and that on each trial of an experiment the subject is presented with a subset of these objects and required to choose one (e.g., "which tone seems loudest?"). In a *pair-comparison experiment* only subsets containing exactly two objects are presented, and here p_{ij} *denotes the probability that* o_i *is chosen when* $\{o_i, o_j\}$ *is presented.* In the general case, when the subsets are not always pairs, we identify each subset of objects with the set of integers that are the indices of these objects, and then $p_S(i)$ *denotes the probability that* o_i *is chosen when* $\{o_j \mid j \in S\}$ *is presented.* C_n denotes the total set of indices of the n choice objects, i.e., $C_n = \{1, 2, \ldots, n\}$. Thus, p_{ij} is shorthand for $p_{\{i,j\}}(i)$; $p_{C_n}(i) = p_{\{1,2,\ldots,n\}}(i)$; the probability of choosing o_i from the set $\{o_i, o_j, o_k\}$ is $p_{\{i,j,k\}}(i)$, etc.

An experiment that determines the probability $p_S(i)$ for every i in every subset S ($S \subseteq C_n$, $|S| \geqslant 2$) will be called a *complete choice experiment with n objects.*

2.2. *Systems of Choice Probabilities*

A pair-comparison experiment with n objects results in $\binom{n}{2}$ binary probability distributions, each of the form $\{p_{ij}, p_{ji}\}$. The entire set $\{\{p_{ij}, p_{ji}\} \mid i, j \in C_n, i < j\}$ will be called a *system of pair-comparison probabilities for n objects,* and be denoted by $\{p_{ij}/C_n\}$.

The result of a *complete* choice experiment with n objects is a set of the form

$$\{\{p_S(i) \mid i \in S\} \mid S \subseteq C_n, |S| \geqslant 2\},$$

containing $2^n - (n + 1)$ discrete probability distributions; one for each subset of two or more objects. A set of this form will be called a *complete system of choice probabilities for n objects,* and be denoted by $\{p_S/C_n\}$.

Convention: Choice probabilities can never be zero or one. The Thurstone models considered below all imply that no choice probability can ever be zero or one. To save repetition, we stipulate at the outset that every system $\{p_S\}$ or $\{p_{ij}\}$ mentioned later contains only probabilities in the open interval (0, 1).

Convention on terminology: "Model" used in the sense of "theory." Some writers use "model" to mean a numerical structure that satisfies a "theory" (e.g., Luce & Suppes, 1965): They would say, for example, that a system $\{p_S/C_n\}$ which satisfies the Choice Axiom is a "model of the Choice Axiom." Here model has the same logical status as theory, as in "the Case V model is a special case of Thurstone's Theory of Comparative Judgment."

2.3. *Equivalence between Choice Theories*

Intuitively, two choice theories are ***equivalent*** (that is, indistinguishable) for a certain kind of experiment if no potential result of such an experiment could allow us to reject one theory in favor of the other. To formalize this idea in the standard way (Burke & Zinnes, 1965; Luce & Suppes, 1965), let R_n denote the set of all possible complete systems of choice probabilities $\{p_S/C_n\}$, and $R^* = \bigcup_2^\infty R_n$. Then R_n is the set of all possible results of a complete choice experiment with n objects, and R is the set of all possible results of complete experiments of any size. A *theory of complete choice experiments* (e.g., Luce's Choice Axiom; any of the Thurstone models defined in Section 2.6) can be thought of as partitioning every R_n into two disjoint subsets: A set of *admissible* systems (those that satisfy the theory) and a set of *inadmissible* systems (those that do not). (In line with this way of thinking, theories will be defined here by specifying the systems they admit.) If two theories partition R_n in exactly the same way, then no complete choice experiment with n objects can distinguish between them. In this case, the two theories will be called *completely equivalent for (choice experiments with) n objects.* If two theories are equivalent for n objects for every n (i.e., partition R^* in exactly the same way), then no complete choice experiment can distinguish between them. In this case the theories will simply be called *completely equivalent.*

It is also useful to define a restricted type of equivalence that reflects only the pair-comparison predictions of two theories. Let B_n denote the set of all possible pair-comparison systems $\{p_{ij}/C_n\}$. Two theories will be called *equivalent for pair-comparison experiments with n objects* if they admit exactly the same systems in B_n. If this is true for every n, the two theories are *equivalent for pair comparisons.*

Relationships between different types of equivalence. Every complete system $\{p_S/C_n\}$ contains a pair comparison system $\{p_{ij}/C_n\}$, and for the theories considered here, a given pair-comparison system is admissible iff it is contained in some admissible complete system. Thus by definition *complete equivalence for n objects implies pair-comparison equivalence for n objects, and complete equivalence implies pair-comparison equivalence.* By the same token, *complete equivalence for n objects implies complete equivalence for n* $- 1$.

The converses of these statements, however, are not logically implied. In particular, pair-comparison equivalence does not imply complete equivalence: The Choice Axiom, for example, is equivalent to several different Thurstone models for pair comparisons (Section 3.2), but it turns out to be completely equivalent to only one—the model with double exponential discriminal processes (Section 3.3 and 3.4). In general, for the class of Thurstone models defined in Section 2.6, pair-comparison equivalence does not imply complete equivalence, except in one special case (Section 3.2). Nor does complete equivalence for $n - 1$ objects imply complete equivalence for n. However, it will be shown that for an important subset of Thurstone models, complete equivalence for three objects implies that both models have the same type of discriminal process distribution, and consequently are completely equivalent (Section 3.4).

The next two sections review the Choice Axiom and Thurstone's Theory of Comparative Judgment. Then Section 2.6 defines a general class of "Thurstone models"—models like Thurstone's original Case V except that their discriminal processes need not be normal.

2.4. *Luce's Choice Axiom*

DEFINITION 1. A complete system of choice probabilities $\{p_S/C_n\}$ satisfies the *Choice Axiom* (Axiom 1 in Luce, 1959) iff, for every i and S, $i \in S \subseteq C_n$:

$$p_S(i) = p_{C_n}(i) \Big/ \sum_{j \in S} p_{C_n}(j). \tag{1}$$

The following fundamental result is proved in Luce, 1959:

LEMMA 1. *A complete system $\{p_S/C_n\}$ satisfies the Choice Axiom iff there exists a set of numbers $v_1, v_2, \ldots, v_n$ ("v scale values") such that*

$$p_S(i) = v_i \Big/ \sum_{j \in S} v_j \tag{2}$$

for every i, S; $i \in S \subseteq C_n$. The scale values $v_1, \ldots, v_n$ are uniquely determined by the system $\{p_S/C_n\}$ up to multiplication by a constant.

Equation (2) implies that the pair-comparison probabilities of any complete system satisfy the following relationship:

LEMMA 2. *For every $i, j, k \in C_n$*

$$p_{ik} = p_{ij}p_{jk}/(p_{ij}p_{jk} + p_{ji}p_{kj}). \tag{3}$$

Proof. Equation (3) follows directly from the fact that

$$p_{ik}/p_{ki} = (p_{ij}/p_{ji})(p_{jk}/p_{kj}),$$

which follows in turn from the relationship $v_i/v_k = (v_i/v_j)(v_j/v_k)$. ∎

2.5. *Thurstone's Theory of Comparative Judgment*

This theory (Thurstone, 1927) is based on the idea that choice objects $o_1, o_2, \ldots, o_n$ are represented in an underlying psychological space by real valued random variables $\mathbf{D}_1, \mathbf{D}_2, \ldots, \mathbf{D}_n$ called *discriminal processes.* When o_i and o_j are presented the subject picks o_i iff $\mathbf{D}_i > \mathbf{D}_j$, so that $p_{ij} = P[\mathbf{D}_i > \mathbf{D}_j] = P[\mathbf{D}_j - \mathbf{D}_i < 0]$. To generate testable predictions, Thurstone arbitrarily (but not thoughtlessly) assumed that the discriminal processes were normal random variables, and considered five special cases corresponding to increasingly severe constraints on their variances and covariances. Of these, only Case V is directly comparable to the Choice Axiom—the others have too many free parameters. In Case V, the discriminal processes are assumed to have identical variances ($\mathrm{Var}\,\mathbf{D}_i \equiv \sigma^2$) and a common covariance ($\mathrm{Cov}(\mathbf{D}_i, \mathbf{D}_j) \equiv r\sigma^2$), so that their marginal distributions differ only in their locations along the axis. Then $\mathbf{D}_i - \mathbf{D}_j$ has a normal distribution with mean $E(\mathbf{D}_i) - E(\mathbf{D}_j)$ and variance $2\sigma^2(1 - r)$. To capture the idea that the discriminal process $\mathbf{D}_i$ of an object o_i represents a "true" scale value that is perturbed by random noise, it is natural to express $\mathbf{D}_i$ in the form $u_i + \mathbf{X}_i$, where u_i is a real number (the Thurstone scale value of o_i) and $\mathbf{X}_i$ is a normal random variable, with $E(\mathbf{X}_i) \equiv \mu$, $\mathrm{Var}(\mathbf{X}_i) \equiv \sigma^2$, $\mathrm{Cov}(\mathbf{X}_i, \mathbf{X}_j) \equiv r\sigma^2$. Then the pair comparison probability p_{ij} is the probability that $u_i + \mathbf{X}_i$ is greater than $u_j + \mathbf{X}_j$, and consequently it can be expressed as

$$p_{ij} = N[(u_i - u_j)/\sigma(2 - 2r)^{1/2}], \tag{4}$$

where N is the normal distribution function

$$N(x) = \int_{-\infty}^{x} (1/(2\pi)^{1/2})\, e^{-(1/2)t^2}\, dt.$$

Equation (4) is Case V of Thurstone's *Law of Comparative Judgment* for pair comparisons. Its testable consequences are entirely captured by the following well-known relationship:

LEMMA 3. *If a system of pair-comparison probabilities* $\{p_{ij}/C_n\}$ *satisfies* (4), *then for every* $i, j, k \in C_n$

$$p_{ik} = N[N^{-1}(p_{ij}) + N^{-1}(p_{jk})]. \tag{5}$$

Proof. (5) follows directly from (4). ∎

Equation (5) is analogous to Eq. (3). Thus for both the Choice Axiom and Case V of Thurstone's law, there is a *triples function* $t(p, p')$ from $(0, 1) \otimes (0, 1) \to (0, 1)$ such that for any p_{ij}, p_{jk}, p_{ik}, $p_{ik} = t(p_{ij}, p_{jk})$. It turns out (Section 3.2) that all of the generalized Thurstone models defined in Section 2.6 also imply triples functions—of the same form as (5), but with N not necessarily a normal distribution function. These functions play a central role in determining whether two models are equivalent for pair comparisons, because the triples function of a model completely determines the pair-comparison systems admitted by that model, and consequently two models can be equivalent for pair comparisons iff they imply the same triples function. (This is shown explicitly in the proof of Theorem 2, Section 3.2.)

Thurstone developed his Case V model explicitly only for the case of pair-comparison experiments, but the following generalization to complete experiments is straightforward:

DEFINITION 2. A complete system of choice probabilities $\{p_S/C_n\}$ satisfies the *Case V* model $T(\mu, \sigma^2, r)$ iff there exist numbers (*scale values*) $u_1, u_2, \ldots, u_n$ such that for every i and S, $i \in S \subseteq C_n$,

$$p_S(i) = P[u_i + \mathbf{X}_i = \mathrm{Max}\{u_j + \mathbf{X}_j \mid j \in S\}], \tag{6}$$

where $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ are normally distributed random variables with common mean μ, common variance σ^2, and $\mathrm{Cov}(\mathbf{X}_i, \mathbf{X}_j) \equiv r\sigma^2$ for all i and j, $i \neq j$. The random variable $u_i + \mathbf{X}_i$ is the *discriminal process* associated with object o_i.

Definition 2 is stated in such a way that to every triple (μ, σ^2, r) there corresponds a distinct Case V model. However, it is well known that in the case of pair comparisons, the parameters μ, σ^2, and r cannot be identified. That is, if a pair-comparison system satisfies any $T(\mu, \sigma^2, r)$, then it satisfies all of them, and in particular it satisfies the model $T(0, 1, 0)$ in which the $\mathbf{X}_i$ have mean zero, variance 1, and are independent. The same is true of the generalization to complete experiments represented by Definition 2, as the following Lemma shows.

LEMMA 4. *Every Case V model $T(\mu, \sigma^2, r)$ is completely equivalent to the model $T(0, 1, 0)$.*

Proof. Suppose $\{p_S/C_n\}$ is a complete system of choice probabilities admitted by $T(\mu, \sigma^2, r)$, with scale values $u_1, u_2, \ldots, u_n$. To show that this system is also admitted by $T(0, 1, 0)$, observe that if $\{u_i + \mathbf{X}_i \mid i = 1, \ldots, n\}$ are the discriminal processes of $T(\mu, \sigma^2, r)$, their joint moment generating function $m(\theta_1, \ldots, \theta_n) = E[\exp(\sum \theta_i(u_i + \mathbf{X}_i))]$ is

$$\exp\left[\sum_{i=1}^{n} \theta_i(u_i + \mu) + \tfrac{1}{2}\sigma^2 \sum_{i=1}^{n} \theta_i^2 + r\sigma^2 \sum_{i \neq j} \theta_i\theta_j\right].$$

However, a straightforward calculation shows that this is also the joint m.g.f. of a set of n normal random variables of the form $\mathbf{X}_i' = u_i + \mu + \mathbf{X}_i^* + \mathbf{Y}$, where X_1^*, $X_2^*, \ldots, \mathbf{X}_n^*$ are n independent identically distributed (i.i.d.) normal R.V.'s with mean zero, variance $\sigma^2(1 - r)$, and $\mathbf{Y}$ is another normal random variable, independent of all the others, with mean zero, and variance $r\sigma^2$. Consequently for any $p_S(i)$ in $\{p_S/C_n\}$

$$\begin{aligned} p_S(i) &= P[u_i + \mathbf{X}_i = \mathrm{Max}\{u_j + \mathbf{X}_j \mid j \in S\}] \\ &= P[u_i + \mu + \mathbf{X}_i^* + \mathbf{Y} = \mathrm{Max}\{u_j + \mu + \mathbf{X}_j^* + \mathbf{Y} \mid j \in S\}] \\ &= P[u_i + \mathbf{X}^* = \mathrm{Max}\{u_j + \mathbf{X}_j^* \mid j \in S\}] \\ &= P[(u_i/c) + (\mathbf{X}_i^*/c) = \mathrm{Max}\{(u_j/c) + (\mathbf{X}_j^*/c) \mid j \in S\}]. \end{aligned}$$

where $c = (\sigma^2(1 - r))^{1/2}$. The random variables $\mathbf{X}_1^*/c, \mathbf{X}_2^*/c, \ldots, \mathbf{X}_n^*/c$ are normal, i.i.d., and have mean zero and variance one. Consequently if $\{p_S/C_n\}$ is admitted by $T(\mu, \sigma^2, r)$ with scale values $u_1, u_2, \ldots, u_n$, it is also admitted by $T(0, 1, 0)$ with scale values $u_1/(\sigma^2(1-r))^{1/2},\ldots, u_n/(\sigma^2(1-r))^{1/2}$. The converse is proved by working backward along the same route. ▮

Remark on nonidentifiable correlations. The proof of the last lemma shows explicitly why correlated and independent discriminal processes in a Case V model yield the same results: If $\{u_i + \mathbf{X}_i \mid i = 1,\ldots, n\}$ is a set of jointly distributed normal discriminal processes with common variances and a common nonzero covariance, then there is another set $\{u_i + \mathbf{X}_i^* + \mathbf{Y} \mid i = 1,\ldots, n\}$, with $\{\mathbf{X}_i^*\}$ i.i.d., and $\mathbf{Y}$ independent of the $\{\mathbf{X}_i^*\}$, which have the same joint distribution and consequently yield the same predictions. In this equivalent set of processes the correlation between the original $\mathbf{X}_i$ is represented entirely by the extra random variable $\mathbf{Y}$. Because this random variable has no effect on the probability of an inequality of the form $u_i + \mathbf{X}_i^* + \mathbf{Y} > u_j + \mathbf{X}_j^* + \mathbf{Y}$, its distribution has no effect on choice probabilities based on such inequalities, and in particular we can assume that $\mathbf{Y}$ is a degenerate R.V. equal to zero. In fact, we could assume that $\mathbf{Y}$ has a nonnormal distribution and the resulting model (in which the discriminal processes would no longer be normal or independent) would yield the same results, since

$$p_S(i) = P[u_i + \mathbf{X}_i^* + \mathbf{Y} = \mathrm{Max}\{u_j + \mathbf{X}_j^* + \mathbf{Y} \mid j \in S\}]$$

is independent of the distribution of $\mathbf{Y}$. This shows that if we start with a Case V model with independent normal discriminal processes, we can then construct an infinite number of distinct models with correlated nonnormal discriminal processes, all of which will yield the same predictions as the original.

The same argument applies also in the general case developed in the next section, and so it can be seen that in general, if the discriminal processes are not assumed to be independent, there is no hope of identifying their distribution from their predictions for choice experiments.

2.6. *Generalization: Thurstone Models with Independent Discriminal Processes*

The essential features of Case V of Thurstone's Theory of Comparative Judgment are (1) the idea that choice objects $o_1, o_2,\ldots, o_n$ are represented on an underlying psychological continuum by random variables $u_1 + \mathbf{X}_1,\ldots, u_n + \mathbf{X}_n$, with

$$p_S(i) = P[u_i + \mathbf{X}_i = \mathrm{Max}\{u_j + \mathbf{X}_j \mid j \in S\}];$$

(2) the $\mathbf{X}_i$ are independent and identically distributed, so that the discriminal processes $\{u_i + \mathbf{X}_i\}$ are themselves independent and identically distributed except for shifts along the axis; (3) the (cumulative) distribution function of the difference $\mathbf{X}_j - \mathbf{X}_i$ is continuous and strictly increasing, so that $p_{ij} = P[\mathbf{X}_j - \mathbf{X}_i \leqslant u_i - u_j]$ is a continuous and strictly increasing function of the scale value difference $u_i - u_j$.

A general class of models that share these properties without presupposing normal distributions can be defined as follows.

Notation. If F is a distribution function, its *difference distribution* D_F is the distribution function

$$D_F(x) = \int_{-\infty}^{\infty} F(x + y)\, dF(y).$$

(Thus if $\mathbf{X}_1, \mathbf{X}_2$ are i.i.d. with common distribution function F, D_F is the distribution function of $\mathbf{X}_1 - \mathbf{X}_2$.)

DEFINITION 3. Suppose F is a distribution for which D_F is continuous and strictly increasing over $(-\infty, \infty)$. A complete system of choice probabilities $\{p_S/C_n\}$ satisfies the *Thurstone model* $\mathscr{T}_F$ if and only if there exist numbers $u_1, u_2, \ldots, u_n$ (*scale values*) such that for every $i \in S \subseteq C_n$

$$p_S(i) = P[u_i + \mathbf{X}_i = \mathrm{Max}\{u_j + \mathbf{X}_j \mid j \in S\}]$$

where $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ are n independent random variables with F as their common distribution function (i.e., $P(\mathbf{X}_i \leqslant x) = F(x)$).[2]

In other words: A given Thurstone model $\mathscr{T}_F$ can be thought of as a specific collection of complete systems of choice probabilities, all those that can be generated by the distribution function F together with every possible set of scale values $u_1, u_2, \ldots$. F here is regarded as completely specified: It has no free parameters of its own. Thus, the model corresponding to the normal distribution with mean zero and variance 1 is conceptually distinct from the model corresponding to the normal distribution with mean 1 and variance 2. Lemma 4 showed that in the case of normal distributions this distinction is artificial, and as might be expected the same is true for any other type of distribution. Theorem 1 in Section 3.1 makes this explicit: Models with distributions of the same type are completely equivalent. Consequently there is usually no harm in talking about, e.g., "the normal model," without specifying precisely which normal distribution we have in mind. Still, it is necessary to have some convention in this regard to keep things logically straight, and the one adopted here has at least the advantage of not taking anything for granted.

For pair-comparison experiments the difference distribution D_F plays a central role, because

$$p_{ij} = P[u_i + \mathbf{X}_i > u_j + \mathbf{X}_j] = D_F(u_i - u_j).$$

The assumption that D_F is continuous and strictly increasing over the whole line guarantees the third essential Thurstone condition, i.e., that p_{ij} be a continuous strictly increasing function of $u_i - u_j$. Note that the scale values of a Thurstone model are uniquely determined by a system $\{p_S/C_n\}$ up to addition of a constant (i.e., if $\mathscr{T}_F$ generates

[2] It should perhaps be emphasized that the class of "Thurstone models" defined here represents a generalization only of Thurstone's Case V, and consequently the theorems proved below for this class need not apply to generalized version of his other cases.

$\{p_S/C_n\}$ with scale values $\{u_i\}$, scale values of the form $\{a + u_i\}$ will work also—but no others, since $u_i - u_j = D_F^{-1}(p_{ij})$).

Definition 3 excludes discriminal process distributions concentrated on finite intervals. Models of that sort are intriguing, because they admit the possibility that choice probabilities can be zero and one, but they are awkward to deal with for the same reason. We can afford to exclude them here without loss of generality, because—for reasons spelled out in Section 4—none of them can be equivalent to the Choice Axiom.

Of the following three examples, only (c) requires special motivation. It is included partly because of its uniqueness properties, which are similar to those of (b), the case of central interest, but easier to demonstrate explicitly. Example (c) also shows that the Choice Axiom is not the only theory of choice behavior that inadvertently turns out to be equivalent to a Thurstone model. However, in (c), this equivalence only extends to pair comparisons.

Examples of Thurstone Models (Fig. 1)

(a) *The Normal Model* (Thurstone's Case V). If F is the normal distribution with mean u, and variance σ^2, $\mathscr{T}_F$ is the Case V model $T(\mu, \sigma^2, 0)$ of Definition 2. Here D_F is again normal, with mean zero and variance $2\sigma^2$.

(b) *The Double Exponential Model (Luce's Choice Axiom).* If $\mathbf{Y}$ is an exponential random variable with density e^{-y}, the random variable $\mathbf{X} = -\log \mathbf{Y}$ has distribution function

$$P(\mathbf{X} \leqslant x) = F(x) = e^{-e^{-x}} \qquad (-\infty < x < \infty). \tag{6}$$

Equation (6) is the *double exponential* distribution function, with mean γ (where γ is Euler's constant 0.5772,...) and variance $\pi^2/6$. (The mean can be derived from formulas found in most handbooks of integrals. The variance takes more work, but it can be obtained using a formula for the derivatives of the γ function given by Whittaker and Watson (1963, p. 241).) This distribution is famous in statistics, because it is one of three possible distributions for the limit (as $n \to \infty$) of the maximum $\mathbf{Z}_n{}^*$ of n i.i.d. random variables $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$. Of course if the $\mathbf{Z}_i$ are not bounded from above, $\mathbf{Z}_n{}^* \to \infty$ with probability 1, but in many cases sequences of normalizing constants $\{a_n\}$, $\{b_n\}$, can be found such that the distribution of $a_n\mathbf{Z}_n{}^* - b_n$ converges to a nondegenerate limit. If this is true when $a_n \equiv 1$ (i.e., if $\mathbf{Z}_n{}^* - b_n$ has a limit distribution), then that distribution must be of the double exponential type. (Gumbel (1958) discusses all this very throughly.) This fact is used in Section 4 to derive the Choice Axiom from a weaker assumption about choice probabilities in experiments with "redundant" objects.

The difference distribution corresponding to (6) is quickly found to be the logistic distribution

$$D_F(x) = (1 + e^{-x})^{-1}. \tag{7}$$

Relationship to the Choice Axiom. Adams and Messick (1957) showed that a Thurstone

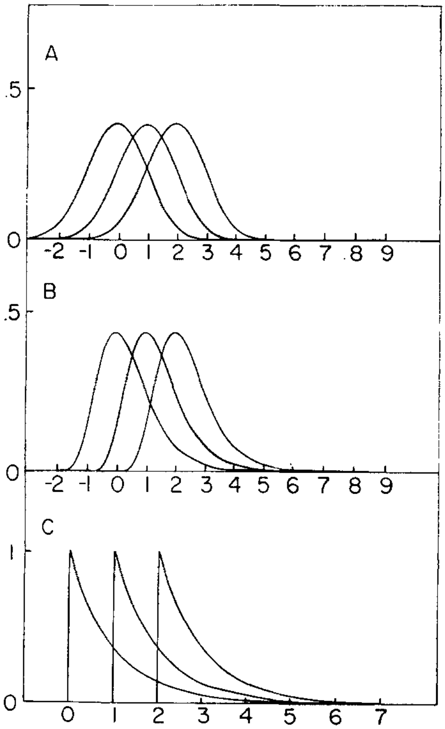

FIG. 1. Discriminal process distributions (shown as probability density functions) for three different Thurstone models. Each panel depicts the density functions $f(x)$ for three discriminal process random variables of the form $u_i + \mathbf{X}_i$, with $u_1 = 0$, $u_2 = 1$, $u_3 = 2$ in each case. Every distribution has variance 1. (A) The Normal (Case V) Model: $f(x) = (2\pi)^{-1/2} \exp[-\frac{1}{2}(x - u_i)^2]$. (B) The Double Exponential (Choice Axiom) Model: $f(x) = a \exp[-a(x - u_i)] \exp[-\exp[-a(x - u_i)]]$, $a = \pi/6^{1/2}$. (The means are $u_i + \gamma a^{-1}$, γ = Euler's constant.) (C) The exponential model: $f(x) = \exp[-(x - u_i)]$. (The means are 1, 2, 3.)

model is equivalent to the Choice Axiom for pair comparisons if and only if its difference distribution is logistic (any logistic with mean zero, not just (7)):

LEMMA 5 (Adams and Messick's theorem). *A Thurstone model $\mathcal{T}_F$ is equivalent to the Choice Axiom for pair-comparison experiments iff $D_F(x) = (1 + e^{-ax})^{-1}$, with $a > 0$.*

Proof. One method (the original) is to show by a functional equation argument that only the logistic implies the triples function (3). Here we can take a shorter route. Suppose $D_G(x)$ is $(1 + e^{-x})^{-1}$, and $\mathcal{T}_G$ admits $\{p_{ij}/C_n\}$. Then

$$\begin{aligned} p_{ij} &= P[u_i + \mathbf{X}_i > u_j + \mathbf{X}_j] \\ &= D_G(u_i - u_j) \\ &= (1 + e^{-(u_i - u_j)})^{-1}. \end{aligned}$$

Setting $u_i = \log v_i$, the last line becomes

$$p_{ij} = v_i/(v_i + v_j).$$

Consequently (via Lemma 1), the system $\{p_{ij}/C_n\}$ satisfies the Choice Axiom. Conversely, if $\{p_{ij}/C_n\}$ satisfies the Choice Axiom, with scale values $v_1 ,..., v_n$, then it satisfies $\mathscr{T}_G$ with $u_i = \log v_i$. Then to establish uniqueness, we appeal to Theorem 2 below, which shows that in general, two Thurstone models $\mathscr{T}_F$ and $\mathscr{T}_G$ can be equivalent for pair comparisons iff $D_F(x) = D_G(ax)$ for some positive constant a. ▌

In view of (7) and Lemma 5, it is clear that the double exponential Thurstone model (6) is equivalent to the Choice Axiom for pair comparisons. Holman and Marley (cited in Luce and Suppes, 1965) showed that this is also true for complete choice experiments for any model of the double exponential type:

LEMMA 6 (Holman and Marley's theorem). *If F is a distribution function of the double exponential type*

$$F(x) = e^{-e^{-(ax+b)}} \qquad (a > 0,\ b \text{ arbitrary})$$

then the Thurstone model $\mathscr{T}_F$ is completely equivalent to the Choice Axiom.

Proof. It is sufficient to prove this only for the particular case $a = 1$, $b = 0$, and then appeal to Theorem 1 below, which shows that two models $\mathscr{T}_F$ and $\mathscr{T}_G$ with distributions of the same type (i.e., $F(x) = G(ax + b)$) are always completely equivalent. Suppose then that a complete system $\{p_S/C_n\}$ is admitted by $\mathscr{T}_G$ ($G(x) = e^{-e^{-x}}$) with scale values $u_1 ,..., u_n$. Set $u_i = \log v_i$, and $\mathbf{X}_i = -\log \mathbf{Y}_i$. Then $\mathbf{Y}_i$ has the exponential distribution function $1 - e^{-y}$, and

$$\begin{aligned} p_S(i) &= P[u_i + \mathbf{X}_i = \mathrm{Max}\{u_j + \mathbf{X}_j \mid j \in S\}] \\ &= P[v_i/\mathbf{Y}_i = \mathrm{Max}\{v_j/\mathbf{Y}_j \mid j \in S\}] \\ &= P[\mathbf{Y}_i/v_i = \mathrm{Min}\{\mathbf{Y}_j/v_j \mid j \in S\}] \\ &= \int_0^\infty v_i e^{-v_i y} \Big(\prod_{j \in S-i} e^{-v_j y} \Big)\, dy \\ &= v_i \Big/ \sum_{j \in S} v_j \,. \end{aligned}$$

Consequently $\{p_S/C_n\}$ also satisfies the Choice Axiom, with v scale values $v_i = e^{u_i}$. The converse is immediate. ▌

Notice that Lemma 6 does not say that the double exponential is the *only* Thurstone model that is completely equivalent to the Choice Axiom. In citing Holman and Marley's result, Luce and Suppes (1965, p. 339) mentioned that the uniqueness question was open. At that time all that was known by way of uniqueness was Adams and Messick's theorem, which shows that any other model that is completely equivalent to the Choice

Axiom must, like the double exponential, have the logistic as its difference distribution. That property, however, does not uniquely characterize the double exponential distribution, for reasons elaborated in Section 3.2. Subsequently, both the author (1971) and McFadden (1974) have independently given proofs (along quite similar lines) that a Thurstone model $\mathscr{T}_F$ can be completely equivalent to the Choice Axiom only if F is a double exponential distribution. Both proofs depend on extra assumptions about F: Mine assumed that F has a density f, with $\lim_{x\to\infty} xf(\log x) \neq 0$;[3] McFadden's assumed that F is translation complete (e.g., has a nonvanishing characteristic function). Theorem 3 below has as a consequence the somewhat better result that without any additional assumptions (beyond those of Definition 3) $\mathscr{T}_F$ is equivalent to the Choice Axiom if and only if F is double exponential. This result, however, is still weak; the strongest one possible is proved in Section 3.4 (Theorem 5).

(c) *The Exponential Model (Dawkins' Threshold Model).* If $F(x)$ is the exponential distribution $1 - e^{-x}$, then D_F is the Laplace distribution

$$D_F(x) = \int_{-\infty}^{x} \tfrac{1}{2} e^{-|y|}\, dy.$$

This model turns out to be equivalent, for pair comparisons, to the *Threshold Model* proposed by Dawkins (1969). That model assumes the existence of a set of "thresholds" $t_1, t_2, \ldots, t_n$, corresponding to the choice objects $o_1, o_2, \ldots, o_n$: The more preferred the object, the lower its threshold—if $t_1 < t_2 < \cdots < t_n$, then o_1 is most preferred, o_n least. $\mathbf{V}$ denotes an "excitation" random variable, with $P(\mathbf{V} \leqslant v) = H(v)$. (We assume $0 < H(v) < 1$ and strictly increasing for all v. H turns out to be unidentifiable.) When a pair o_i, o_j is presented, with $t_i \leqslant t_j$, the subjects' choice depends on where $\mathbf{V}$ is relative to these thresholds: If $\mathbf{V} < t_i \leqslant t_j$, the subject makes no choice, and instead draws another sample of $\mathbf{V}$. If $t_i < \mathbf{V} \leqslant t_j$, he chooses o_i. If $t_i \leqslant t_j < \mathbf{V}$, he picks either o_i or o_j, each with probability .5. Dawkins (1969) showed that the triples function for this model (analogous to (3) and (5)) is given by

$$p_{ik} = 1 - 2p_{kj}p_{ji} \tag{8}$$

in the case $p_{ij} \geqslant 0.5$, $p_{jk} \geqslant 0.5$. The remaining cases can be derived from this expression. (Since the subject might not choose either object—if $\mathbf{V}$ falls below both thresholds—it is necessary to redefine p_{ij} here to mean the probability that i is chosen when $\{o_i, o_j\}$ is presented and the subject actually chooses an object.)

Any Thurstone model with the Laplace distribution as its difference distribution will also imply this same triples function, and consequently be equivalent to Dawkins' model for pair-comparison experiments. The following lemma also shows that this representation is unique: The Laplace stands in the same relationship to Dawkins' model that the logistic does to the Choice Axiom.

[3] This condition makes somewhat more sense in the context $F(x) = P(\mathbf{X} \leqslant x)$; $\mathbf{Y} = e^{-\mathbf{X}}$: Then it implies that the density of $\mathbf{Y}$ is positive at the origin.

LEMMA 7. *A Thurstone model $\mathscr{T}_F$ is equivalent to Dawkins' Threshold model for pair comparisons if and only if its difference distribution D_F is a Laplace distribution*:

$$\begin{aligned} D_F(x) &= \tfrac{1}{2}e^{ax}, \qquad & x \leqslant 0, \\ &= 1 - \tfrac{1}{2}e^{-ax}, \qquad & x > 0. \end{aligned} \tag{9}$$

Proof. Here again we need only show that (9) implies the triples function (8), and then Theorem 2 below establishes uniqueness. Assuming p_{ij} and p_{jk} both $\geqslant 0.5$, $u_i - u_j$ and $u_j - u_k$ are both nonnegative, and

$$\begin{aligned} p_{ik} &= D_F(u_i - u_j + u_j - u_k) \\ &= 1 - \tfrac{1}{2}e^{-a(u_i-u_j)}e^{-a(u_j-u_k)} \\ &= 1 - 2(\tfrac{1}{2}e^{a(u_j-u_i})(\tfrac{1}{2}e^{a(u_k-u_j)}) \\ &= 1 - 2p_{ji}p_{kj} \, . \quad \blacksquare \end{aligned}$$

Surprisingly, this equivalence breaks down as soon as we consider complete experiments. In fact, for complete experiments with three objects, Dawkins' Threshold model is not equivalent to any Thurstone model. To see why, suppose o_1 is preferred to o_2 is preferred to o_3, so that in the Threshold model $t_1 < t_2 < t_3$, and in any corresponding Thurstone model $u_1 > u_2 > u_3$. In this case the Threshold model implies that the following relationship holds regardless of where t_2 is located in the interval (t_1, t_3):

$$p_{\{1,2,3\}}(3) = \tfrac{1}{3}(1 - H(t_3))/(1 - H(t_1)) = \tfrac{2}{3}p_{31} \, . \tag{10}$$

No Thurstone model can duplicate this prediction, because to do so would mean that the probability of the event

$$u_3 + \mathbf{X}_3 = \operatorname{Max}\{u_i + \mathbf{X}_i \mid i = 1, 2, 3\}$$

depended only on the difference between u_1 and u_3, and not on the location of u_2 within the interval (u_3, u_1). For any Thurstone model, however, this probability clearly must always be a strictly decreasing function of u_2.

In his 1969 paper, Dawkins reports a great deal of pair comparison data which is all very well fit by either his model or the Choice Axiom or Thurstone's Case V (the predictions of all three models are nearly identical–not surprisingly, in view of the similarity of their difference distributions). Equation (10), however, seems much too strong to be generally true, and so it seems fair to attribute the success of Dawkins' model for pair-comparison experiments to the fact that in that special case it happens to be a Thurstone model.

3. Uniqueness Properties of Thurstone Models

3.1. *Preliminaries*

The general question here is whether two Thurstone models $\mathscr{T}_F$ and $\mathscr{T}_G$ can be equivalent even though F and G are different distributions. An overview of the results was given earlier in Section 1.2.1, and Fig. 2 shows a graphical summary. Theorem 1 disposes of the trivial case in which F and G are distributions of the same type (e.g., F and G are both double exponential distributions, differing only in their means and variances).

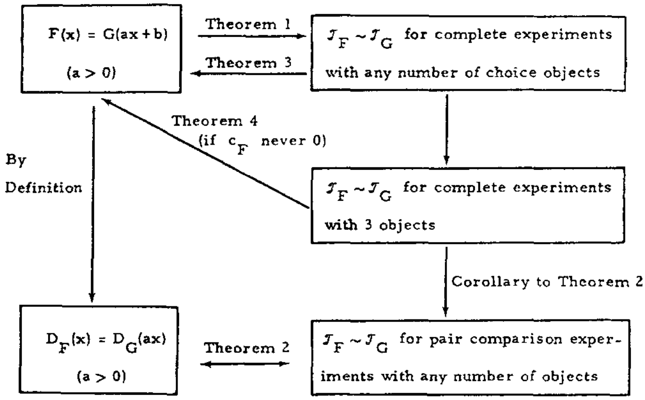

FIG. 2. Logical structure of the relationship between two discriminal process distributions F and G and various types of equivalence ($\sim$) between their corresponding Thurstone models $\mathscr{T}_F$ and $\mathscr{T}_G$. Arrows signify the direction of implication.

THEOREM 1. *If $\mathscr{T}_F$ and $\mathscr{T}_G$ are Thurstone models, and F and G are distributions of the same type (i.e., $F(x) = G(ax + b)$ for all x and some pair of constants a, b, with $a > 0$), then $\mathscr{T}_F$ is completely equivalent to $\mathscr{T}_G$.*

Proof. Suppose $\mathscr{T}_F$ admits the system $\{p_S/C_n\}$ with scale values $u_1, u_2, \dots, u_n$, and that $p_S(i)$ is an arbitrary probability in that system. Then

$$\begin{aligned} p_S(i) &= P[u_i + \mathbf{X}_i = \mathrm{Max}\{u_j + \mathbf{X}_j \mid j \in S\}] \\ &= P[au_i + a\mathbf{X}_i + b = \mathrm{Max}\{au_j + a\mathbf{X}_j + b \mid j \in S\}] \\ &= P[au_i + \mathbf{X}_i^* = \mathrm{Max}\{au_j + \mathbf{X}_j^* \mid j \in S\}] \end{aligned}$$

where

$$\begin{aligned} P[\mathbf{X}_j^* \leqslant x] &= P[a\mathbf{X}_j + b \leqslant x] \\ &= F[(x - b)/a] \\ &= G(x). \end{aligned}$$

Consequently $\{p_S/C_n\}$ is also admitted by $\mathscr{T}_G$, with scale values $au_1, au_2, \dots, au_n$. ∎

In view of Theorem 1 it is unnecessary to maintain a strict distinction between Thurstone models whose discriminal process distributions belong to the same type: There is no ambiguity in talking about, e.g., "the double exponential model" without specifying exactly which double exponential distribution we mean, because all such models are completely equivalent.

3.2. *Implications of Pair-Comparison Equivalence*

The starting point is the following result:

THEOREM 2. *Two Thurstone models $\mathscr{T}_F$ and $\mathscr{T}_G$ are equivalent for pair comparisons if and only if their difference distributions satisfy*

$$D_F(x) = D_G(ax)$$

for some $a > 0$ and all x.

Proof. Note first that any model $\mathscr{T}_F$ implies a triples function analogous to Eq. 5 for the Case V Model. Since

$$\begin{aligned} p_{ij} &= P[\mathbf{X}_j - \mathbf{X}_i \leqslant u_i - u_j] \\ &= D_F(u_i - u_j), \end{aligned}$$

it follows that for any three pairwise probabilities p_{ij}, p_{jk}, p_{ik} in a pair-comparison system admitted by $\mathscr{T}_F$:

$$p_{ik} = D_F(D_F^{-1}(p_{ij}) + D_F^{-1}(p_{jk})). \tag{11}$$

Equation (11) is the *triples function* for $\mathscr{T}_F$.

Next, observe that the triples function (11) completely determines the set of pair-comparison systems admitted by $\mathscr{T}_F$: That set consists exactly of those systems in which (a) $p_{12}, p_{23}, \ldots, p_{n-1,n}$ are arbitrary probabilities; (b) p_{1i} is the probability obtained by applying the triples function t of $\mathscr{T}_F$ to this chain, i.e.,

$$\begin{aligned} p_{13} &= t(p_{12}, p_{23}), \\ p_{1i} &= t(p_{1,i-1}, p_{i-1,i}), \end{aligned}$$

and (c)

$$p_{ij} = t(p_{i1}, p_{1j}).$$

Consequently two models $\mathscr{T}_F$ and $\mathscr{T}_G$ are equivalent for pair comparisons iff they have the same triples function, i.e., iff for every p, p' in (0, 1)

$$D_F(D_F^{-1}(p) + D_F^{-1}(p')) = D_G(D_G^{-1}(p) + D_G^{-1}(p')). \tag{12}$$

To solve this functional equation, let $x = D_F^{-1}(p)$, $x' = D_F^{-1}(p')$. Then

$$D_G^{-1}D_F(x + x') = D_G^{-1}D_F(x) + D_G^{-1}D_F(x'),$$

so $D_G^{-1}D_F$ satisfies the Cauchy equation $f(x + x') = f(x) + f(x')$ for every x and x'. Since $D_G^{-1}D_F$ is continuous, the only solution is $D_G^{-1}D_F(x) = ax$, and so $D_F(x) = D_G(ax)$. The constant a must be positive because $a = D_G^{-1}D_F(1)$ and $D_F(1) > 0.5$ (since $D_F(0) = 0.5$, and D_F is strictly increasing by assumption), so that $D_G^{-1}D_F(1) > 0$. ▮

The following is included here for use later in Sections 3.3 and 3.4.

COROLLARY 1 (to Theorem 2). *If $\mathscr{T}_F$ and $\mathscr{T}_G$ are equivalent for complete experiments with three (or more) objects, then:* (1) *They are equivalent for pair-comparison experiments with any number of objects; and* (2) *if $\{p_S/C_n\}$ is a system admitted by both models, and the scale values for this system are $u_1, ..., u_n$ according to $\mathscr{T}_F$, and $u_1', ..., u_n'$ according to $\mathscr{T}_G$, then*

$$u_i' - u_j' = a(u_i - u_j) \tag{13}$$

where a is the positive constant for which $D_F(x) = D_G(ax)$.

Proof. (1) Complete equivalence for three objects implies pair-comparison equivalence for three objects, which implies that $\mathscr{T}_F$ and $\mathscr{T}_G$ have the same triples function. The proof of Theorem 2 shows that this implies pair-comparison equivalence for any number of objects. (2) Pair-comparison equivalence implies $D_F(x) = D_G(ax)$. If p_{ij} is in $\{p_S/C_n\}$, $p_{ij} = D_F(u_i - u_j) = D_G(u_i' - u_j')$. Then

$$\begin{aligned} D_G^{-1}(p_{ij}) &= u_i' - u_j' \\ &= D_G^{-1}D_F(u_i - u_j) \\ &= a(u_i - u_j). \quad ▮ \end{aligned}$$

Now suppose that two models $\mathscr{T}_F$ and $\mathscr{T}_G$ are equivalent for pair comparisons, and that one of them has a known distribution—say, F is the double exponential distribution (6). What can we infer about G? From Theorem 2 we know that pair-comparison equivalence implies only that $D_F(x) = D_G(ax)$. Clearly one solution is $F(x) = G(ax + b)$, with $a > 0$: That corresponds to the case in which F is the distribution function of $\mathbf{X}$ and G is the distribution function of $a\mathbf{X} + b$ ($a > 0$), so that F and G are distributions of the same type. The other obvious solution is $F(x) = 1 - G(-ax + b)$ (again, $a > 0$): This corresponds to G being the distribution of $-a\mathbf{X} + b$. In this case G is a distribution of the same type as ^-F, where

$$^-F(x) = 1 - F(-x),$$

i.e., ^-F is the distribution function of $-\mathbf{X}$. (If F is not a symmetrical distribution, ^-F and F are not of the same type. For example, if F is a double exponential, hence not symmetrical, as can be seen from Fig. 1B, ^-F is not.)

If the only solutions to $D_F(x) = D_G(ax)$ were $F(x) = G(ax + b)$ and $F(x) =$

$1 - G(-ax + b)$, Adams and Messick's theorem (Lemma 5) would permit a simple solution to the uniqueness problem for the Choice Axiom, because we would know that $\mathscr{T}_G$ could be equivalent to the Choice Axiom only if G were either a double exponential distribution, or a distribution function of the same type as $1 - e^{-e^x}$. Then a quick calculation would show that the latter possibility does not yield the Choice Axiom for complete experiments, and so the problem would be solved.

However, things are not that simple, because in addition to the obvious solutions, which correspond to linear transformations of random variables, $D_F(x) = D_G(ax)$ also has other solutions, in which F and G are not related to one another in any readily interpretable way. These hidden possibilities emerge when things are considered from the standpoint of characteristic functions. Recall that if $\mathbf{X}$ has distribution function F, its characteristic function (c.f.) $c_F(t)$ is the complex valued function of a real variable t defined by $c_F(t) = E(e^{it\mathbf{X}})$: This function always exists; two distributions F and G are identical iff c_F and c_G are identical; the c.f. of $\alpha X + \beta$ is $e^{i\beta t}c_F(\alpha t)$; and the c.f. of a sum of independent random variables is the product of their individual c.f.'s (e.g., Feller, 1966, Chap. XV). Thus if X_1, X_2 are i.i.d. with distribution function F, their difference $X_1 - X_2$ has the c.f. $c_{D_F}(t) = c_F(t)\, c_F(-t)$.

In terms of characteristic functions, our original equation $D_F(x) = D_G(x)$ (where to simplify appearances we set $a = 1$, without loss of generality) takes the form

$$c_F(t)\, c_F(-t) = c_G(t)\, c_G(-t),$$

and the two obvious solutions are represented by

$$c_G(t) = e^{ibt}c_F(t) \qquad (F(x) = G(x + b)),$$

and

$$c_G(t) = e^{ibt}c_F(-t) \qquad (F(x) = 1 - G(-x + b)).$$

However, in general, these are not the only possible solutions: In the theory of characteristic functions there are many examples in which $c_F(t)\, c_F(-t) = c_G(t)\, c_G(-t)$ even though c_G is not either of the obvious solutions. (Lukacs, 1960, discusses the problem in detail.) In these cases the extra solutions G are not related to F in any probabilistically obvious way,[4] but they can nevertheless often be shown to be well-behaved distributions, with continuous densities and finite moments—distributions, in other words, that would yield plausible Thurstone models in their own right.

The Exponential Case

Suppose, for example, that F is the exponential distribution function $1 - e^{-x}$. Then $c_F(t) = (1 - it)^{-1}$, and $D_F(x)$ is the Laplace distribution, with c.f. $(1 + t^2)^{-1}$. But Lukacs (1960, p. 94) notes that the function

$$c_G(t) = \frac{(1 + it/v)(1 + it/\bar{v})}{(1 - it)(1 - it/v)(1 - it/\bar{v})}$$

[4] In fact these extra solutions are distributions whose Fourier components have the same amplitudes as those of F, but differ in phase.

(where $v = 1 + ir$, $\bar{v} = 1 - ir$, $r \geqslant 2(2^{1/2})$) is also a c.f. (of a continuous density) and it is easy to see that $c_G(t)\, c_G(-t) = (1 + t^2)^{-1} = c_F(t)\, c_F(-t)$ even though $c_G(t)$ is not of the form $e^{ibt}c_F(t)$ or $e^{ibt}c_F(-t)$. Thus $D_G(x) = D_F(x)$, even though G is not an exponential distribution, or the distribution of the negative of an exponential random variable.

This counterexample shows there is a Thurstone model $\mathscr{T}_G$ that is equivalent to the exponential model for pair comparisons even though neither G or ^{-}G is an exponential distribution. Consequently Dawkin's Threshold model is equivalent to both exponential and nonexponential Thurstone models for pair comparisons, and (as noted earlier) not equivalent to any Thurstone model for complete experiments with three or more objects.

The Double Exponential Case

If F is the double exponential distribution $F(x) = e^{-e^{-x}}$, $c_F(t) = \Gamma(1 - it)$ (where Γ is the gamma function), and so the difference distribution $D_F(x) = (1 + e^{-x})^{-1}$ has c.f. $\Gamma(1 - it)\,\Gamma(1 + it)$. However, an example due to Laha (1964) shows that this c.f. can also be factored into $c_G(t)\, c_G(-t)$, where neither factor is the c.f. of a double exponential distribution. Laha studied the general question: If $\mathbf{Y}_1$ and $\mathbf{Y}_2$ are i.i.d. and positive, and $\mathbf{Y}_1/(\mathbf{Y}_1 + \mathbf{Y}_2)$ has the β distribution, must $\mathbf{Y}_1$ and $\mathbf{Y}_2$ be γ distributed? The special case where $\mathbf{Y}_1$ and $\mathbf{Y}_2$ are exponential is relevant here. In this case $\mathbf{Y}_1/(\mathbf{Y}_1 + \mathbf{Y}_2)$ has the uniform distribution on (0, 1) (a special case of the β) which implies that $\log \mathbf{Y}_1 - \log \mathbf{Y}_2$ has the logistic distribution function (7). Thus if there exist positive nonexponential i.i.d. random variables $\mathbf{Y}_1$, $\mathbf{Y}_2$ such that $\mathbf{Y}_1/(\mathbf{Y}_1 + \mathbf{Y}_2)$ is uniform, then $\mathbf{X}_1 = -\log \mathbf{Y}_1$ and $\mathbf{X}_2 = -\log \mathbf{Y}_2$ (and also—but less interestingly, $-\mathbf{X}_1$, $-\mathbf{X}_2$) will be a pair of i.i.d. random variables that are not double exponential, but nevertheless have the logistic as their difference distribution.

Adapted to the problem at hand, Laha's results can be summarized as follows. If $\mathbf{X}_1$ and $\mathbf{X}_2$ are i.i.d., each having distribution function F, and $\mathbf{X}_1 - \mathbf{X}_2$ is logistic, then: (i) F is absolutely continuous and strictly increasing over the whole line; (ii) if F has finite absolute moments of all orders and is infinitely divisible, then ^{-}F or F must be double exponential; (iii) there exist solutions F that have finite absolute moments of all orders, but neither F or ^{-}F is double exponential. The last result is proved by constructing the characteristic function of such a distribution. The expression for this c.f. is very complicated, noninformative to the naked eye, and fills at least an entire page (Laha, 1964, pp. 296–297). So it will not be reproduced here—suffice it to say that no one would ever be tempted to use the corresponding distribution as the basis for a theory of choice behavior, even though from a technical standpoint it makes a perfectly acceptable Thurstone model.

Laha's results show that there exists a Thurstone model in which the discriminal process distribution is not related in any obvious way to the double exponential, but which is equivalent to the double exponential model (hence, to the Choice Axiom) for pair comparisons. This model cannot be excluded unless one is willing to require that "reasonable" Thurstone Models must have infinitely divisible distributions, which does not seem to be a justifiable restriction. (Recall that a distribution is infinitely divisible if, for every n, it is the distribution of the sum of n i.i.d. random variables. Put another way, F is infinitely divisible if $(c_F)^{1/n}$ is a characteristic function for every n.) It is true that the three examples of Thurstone models given in Section 2 all involve discriminal process distributions that are infinitely divisible. (For the normal, infinite divisibility is obvious: this is the prototype of the concept. Feller (1966, Chap. XVII) shows that the exponential distribution is infinitely divisible, and Laha (1966) shows that the double exponential is also.) But Laha's example shows that this is not an essential property, and it is easy to find less exotic distributions that are not infinitely divisible and still lead to perfectly acceptable Thurstone models. For example, if $\mathbf{Y}$ is normal and another, independent, random variable $\mathbf{U}$ is uniform on $(-1, 1)$, the sum $\mathbf{X} = \mathbf{U} + \mathbf{Y}$ has a positive density, and consequently its distribution F satisfies Definition 3. This distribution is not infinitely divisible, because its c.f. vanishes infinitely often (since the c.f. of the uniform term is $\sin t/t$), whereas the c.f. of an infinitely divisible distribution can never be zero (Feller, 1966, p. 532).

The Normal Case

The two examples just considered illustrate the general rule that a difference distribution D_F does not imply the type of the discriminal process distribution F, because F and G can be distributions of nontrivially different types (i.e., $G \neq$ either F or ${}^{-}F$) and still yield $D_F = D_G$. However, there is one important exception. A famous theorem due to Cramér (Feller, 1966, p. 408) shows that if $\mathbf{X}_1$ and $\mathbf{X}_2$ are independent and $\mathbf{X}_1 + \mathbf{X}_2$ is normal, then $\mathbf{X}_1$ and $\mathbf{X}_2$ must both be normal also. It follows that if D_F is normal, then F itself must be normal:

COROLLARY 2 (to Theorem 2). *A Thurstone model $\mathscr{T}_F$ is equivalent to the normal (i.e., Case* V) *model for pair comparisons iff F is a normal distribution.*

In other words, the pair-comparison predictions of the normal model, taken altogether, uniquely characterize that model; they cannot be entirely duplicated by any other model.

As a historical note, it is interesting that in his original 1927 paper, Thurstone seems to have taken for granted that experimental confirmation of the Law of Comparative Judgment (i.e., confirmation of the triples function (5)) would imply that the discriminal processes must be normal. Cramér's theorem, however, was not proved until 1936. In the interim, no one could have known whether confirmation of the normal model for pair comparisons actually implied normal discriminal processes, or whether there might not also be nonnormal discriminal process distributions that could yield the same set of pair-comparison predictions. As things turned out, of course, Thurstone was right to ignore the problem for his Case V model. But if he had started with virtually any other distribution, he would have either been wrong, or else the answer would still be unknown, because Cramér's theorem seems to be the only one of its kind (except for an analogous result for Poisson distributions, Raikov's theorem (Lukacs, 1960), which is not helpful here).

3.3. *Uniqueness Implications of Complete Equivalence*

The last section showed that two models $\mathscr{T}_F$ and $\mathscr{T}_G$ can be equivalent for pair comparisons even though F and G are distributions of completely different types—distributions for which the corresponding random variables are not related by any linear transformation. A natural question is whether $\mathscr{T}_F$ and $\mathscr{T}_G$ can be equivalent for all complete experiments when F and G are not the same type of distribution. The next theorem shows that this cannot happen: The set of all complete systems admitted by $\mathscr{T}_F$ determines F, up to its type.

THEOREM 3. *Two Thurstone models $\mathscr{T}_F$ and $\mathscr{T}_G$ are completely equivalent if, and only if, $F(x) = G(ax + b)$, with $a > 0$.*

Proof. The proof of this theorem depends on the characteristic function analysis developed in Section 3.4, and so is deferred to the end of that section.

As a solution to our uniqueness problem, Theorem 3 is incomplete, because it leaves open the question of whether two models $\mathscr{T}_F$ and $\mathscr{T}_G$ can be equivalent for complete experiments with a *finite* number of objects when F and G are distributions of different types. If this were true for a sufficiently large number of objects, then $\mathscr{T}_F$ and $\mathscr{T}_G$ could be equivalent for all practical purposes even though they corresponded to different discriminal process distributions. The next section shows that for an important subset of cases, this possibility can be ruled out.

3.4. *Uniqueness Implications of Complete Equivalence for Experiments with Three Objects*

The sharpest possible result would be to show that $\mathscr{T}_F$ and $\mathscr{T}_G$ can be equivalent for complete experiments with three objects iff F and G are distributions of the same type. The following lemma provides the basis for obtaining such a result by way of a functional equation involving the characteristic functions of F and G:

LEMMA 8. *If $\mathscr{T}_F$ and $\mathscr{T}_G$ are equivalent for complete experiments with n objects ($n \geqslant 3$), the characteristic functions of F and G satisfy the equation*

$$c_G(t_1)\, c_G(t_2) \cdots c_G(t_{n-1})\, c_G\left(-\sum_{i=1}^{n-1} t_i\right)$$

$$= c_F(at_1)\, c_F(at_2) \cdots c_F(at_{n-1})\, c_F\left(-\sum_{i=1}^{n-1} at_i\right), \tag{16}$$

where a is the constant for which $D_F(x) = D_G(ax)$.

Proof. To avoid cumbersome expressions we prove the special case $n = 3$; the reasoning in the general case is exactly the same. Suppose $\mathscr{T}_F$ and $\mathscr{T}_G$ are equivalent for complete experiments with three objects, that both admit an arbitrary system $\{p_S/C_3\}$, and that the scale values for this system according to $\mathscr{T}_F$ are u_1, u_2, u_3. Then from Corollary 2 to Theorem 2 we know that the corresponding scale values for $\mathscr{T}_G$,

u_1', u_2', u_3', satisfy the relationship $u_i' - u_j' = a(u_i - u_j)$ where a is the positive constant that makes $D_F(x) = D_G(ax)$. Let $\{u_i + \mathbf{X}_i \mid i = 1, 2, 3\}$ denote the discriminal processes of $\mathscr{T}_F$, and $\{au_i + \mathbf{X}_i' \mid i = 1, 2, 3\}$ the discriminal processes of $\mathscr{T}_G$. Then

$$\begin{aligned} p_{C_3}(1) &= P[\mathbf{X}_2 - \mathbf{X}_1 \leqslant u_1 - u_2, \mathbf{X}_3 - \mathbf{X}_1 \leqslant u_1 - u_3], \\ &= P[\mathbf{X}_2' - \mathbf{X}_1' \leqslant a(u_1 - u_2), \mathbf{X}_3' - \mathbf{X}_1' \leqslant a(u_1 - u_3)]. \end{aligned} \tag{17}$$

Let

$$J_F(x, y) = P(\mathbf{X}_2 - \mathbf{X}_1 \leqslant x, \mathbf{X}_3 - \mathbf{X}_1 \leqslant y),$$

and

$$J_G(x, y) = P(\mathbf{X}_2' - \mathbf{X}_1' \leqslant x, \mathbf{X}_3' - \mathbf{X}_1' \leqslant y).$$

Then (17) implies

$$J_F(x, y) = J_G(ax, ay). \tag{18}$$

Now consider the joint characteristic function of $(\mathbf{X}_2' - \mathbf{X}_1', \mathbf{X}_3' - \mathbf{X}_1')$:

$$\begin{aligned} E[\exp(it_1(\mathbf{X}_2' - \mathbf{X}_1') + it_2(\mathbf{X}_3' - \mathbf{X}_1')] &= E[e^{it_1\mathbf{X}_2'} e^{it_2\mathbf{X}_3'} e^{-i(t_1+t_2)\mathbf{X}_1'}] \\ &= c_G(t_1)\, c_G(t_2)\, c_G(-t_1 - t_2) \end{aligned} \tag{19}$$

since the $\mathbf{X}_i$ are i.i.d. The same c.f. can also be expressed as the integral

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{it_1x + it_2y}\, dJ_G(x, y). \tag{20}$$

Making the change of variable $x = ax'$, $y = ay'$, and using (18), (20) becomes

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{it_1ax' + it_2ay'}\, dJ_F(x', y'). \tag{21}$$

However, this last integral is also the joint c.f. of $(\mathbf{X}_2 - \mathbf{X}_1, \mathbf{X}_3 - \mathbf{X}_1)$ evaluated at (at_1, at_2), so (21) equals

$$E[\exp[iat_1(\mathbf{X}_2 - \mathbf{X}_1) + iat_2(\mathbf{X}_3 - \mathbf{X}_1)]] = c_F(at_1)\, c_F(at_2)\, c_F(-at_1 - at_2). \tag{22}$$

Consequently (19) and (22) are equal, as claimed.[5] ▌

Specializing to the case $n = 3$, Lemma 8 says that if $\mathscr{T}_F$ and $\mathscr{T}_G$ are completely equivalent for experiments with three objects, the characteristic functions of F and G satisfy the functional equation

$$c_G(t_1)\, c_G(t_2)\, c_G(-t_1 - t_2) = c_F(at_1)\, c_F(at_2)\, c_F(-at_1 - at_2). \tag{23}$$

[5] It can be seen from this proof that the uniqueness problem for Thurstone models boils down to the question of whether the distribution of a set of independent R.V.'s. $\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_n$ is determined by the joint distribution of the differences $\mathbf{X}_2 - \mathbf{X}_1, \mathbf{X}_3 - \mathbf{X}_1, ..., \mathbf{X}_n - \mathbf{X}_1$. I. I. Kotlarski has pointed out to me that there is a rather extensive literature on this question, beginning with his 1966 paper which shows that if the joint c.f. of the differences is nonvanishing, the answer is yes. Although arrived at independently, Theorem 4 is essentially a corollary of Kotlarski's result.

One acceptable solution to (23) is

$$c_G(t) = e^{ibt} c_F(at), \tag{24}$$

which implies $F(x) = G(ax + b)$. (The solution $c_G(t) = e^{wt} e^{ibt} c_F(at)$ (w, b real) is not acceptable because it cannot be a characteristic function. This follows from the fact that if $c(t) = c_1(t) + ic_2(t)$ is a c.f., c_1 must be even and c_2 must be odd (Feller, 1966, p. 474). Consequently, $e^{wt}c(t)$ (w real) can never be a c.f., because $e^{wt}c_1(t)$ will not be even, and $e^{wt}c_2(t)$ will not be odd. Since we know that $e^{ibt}c_F(at)$ is a c.f., $e^{wt}e^{ibt}c_F(at)$ is not). If (24) is the only solution to (23), then complete equivalence for experiments with three objects implies that F and G are distributions of the same type. The next theorem shows that this is always true if F (or G) has a nonvanishing characteristic function.

THEOREM 4. *If $\mathscr{T}_F$ and $\mathscr{T}_G$ are equivalent for complete choice experiments with three objects and the characteristic function of F (or G) is never zero, then $F(x) = G(ax + b)$.*

Proof. To simplify the notation, let $f(t) = c_F(t)$, $g(t) = c_G(t)$. Then (23) becomes

$$g(s)\,g(t)\,g(-s-t) = f(as)\,f(at)\,f(-as-at), \tag{25}$$

which holds for all real s and t. The problem is to show that the only acceptable solutions to (25) are of the form

$$g(t) = e^{ibt} f(at), \tag{26}$$

where b is an arbitrary constant. Observe first that when $s = 0$, (25) becomes

$$g(t)\,g(-t) = |g(t)|^2 = f(at)\,f(-at) = |f(at)|^2 \tag{27}$$

(since $g(0) = f(0) = 1$). Clearly if f never vanishes, neither does g, and vice versa, so the factors in (25) can be divided freely. Using (27) to rewrite $f(-as - at)/g(-s - t)$, (25) becomes

$$\frac{g(s)}{f(as)}\frac{g(t)}{f(at)} = \frac{g(s+t)}{f(as+at)}. \tag{28}$$

Now let $h(t) = g(t)/f(at)$. Then (28) becomes

$$h(s)\,h(t) = h(s+t). \tag{29}$$

Functional equation (29) is a complex version of the well-known Cauchy equation—if h were a real valued function, the only continuous nonzero solution would be $h(t) = e^{bt}$ (Aczél, 1966, p. 38). Since h is a complex function of a real variable, the solution here takes a bit more work, but in the end the result (when h is a ratio of c.f.'s) is a complex analog of the real solution: $h(t) = e^{ibt}$. To show this, let h_1 and h_2 denote the real and imaginary parts of h, i.e., $h(t) = h_1(t) + ih_2(t)$, and similarly, $f(t) = f_1(t) + if_2(t)$ and $g(t) = g_1(t) + ig_2(t)$. Then using the fact that f_1 and g_1 are even, and f_2 and g_2 are odd, it is straightforward to show that h_1 is even and h_2 is odd. Writing (29) in terms of h_1 and h_2, expanding, and equating the real and imaginary parts on both sides, we obtain a pair of real functional equations:

$$h_1(s)\, h_1(t) - h_2(s)\, h_2(t) = h_1(s + t), \tag{30}$$

$$h_1(s)\, h_2(t) + h_2(s)\, h_1(t) = h_2(s + t). \tag{31}$$

Putting $-s$ in place of s in (30), and using the fact that h_1 is even, and h_2 odd, (30) becomes

$$h_1(s)\, h_1(t) + h_2(s)\, h_2(t) = h_1(t - s). \tag{32}$$

Aczél (1966, pp. 176–180) shows that the only continuous nonconstant solution to (32) is $h_1(t) = \cos bt$, $h_2(t) = \sin bt$. Since h is the ratio of two continuous functions (every c.f. is continuous) h_1 and h_2 are continuous, and h cannot be any constant except one (i.e., $b = 0$) because $h(0) = 1$. Consequently the only acceptable solution to (29) is

$$h(t) = \cos bt + i \sin bt = e^{ibt}.$$

Thus (26) is the only solution to (25) under the assumption that f and g are c.f.'s, one of which is known to be nonvanishing. ∎

Applying Theorem 4 to the special case of the double exponential distribution, we obtain finally a complete solution to the uniqueness problem for the Choice Axiom:

THEOREM 5. *A Thurstone model $\mathcal{T}_F$ is equivalent to the Choice Axiom for complete experiments with three (or more) objects if, and only if, F is a double exponential distribution, i.e., iff $F(x) = e^{-e^{-(ax+b)}}$, with $a > 0$.*

Proof. Holman and Marley's theorem (Lemma 6) shows that any double exponential model implies the Choice Axiom, in particular, the model $\mathcal{T}_G$, with $G(x) = e^{-e^{-x}}$. This distribution has the nonvanishing characteristic function $\Gamma(1 - it)$, and consequently any model $\mathcal{T}_F$ that is equivalent to $\mathcal{T}_G$ (hence, to the Choice Axiom) for complete experiments with three objects must have $F(x) = G(ax + b)$. (To show explicitly that $\Gamma(1 - it)$ is never zero we can use the fact that for any complex z, $\Gamma(z)\, \Gamma(1 - z) = \pi/(\sin \pi z)$ (e.g., Ahlfors, 1966, p. 198). Recall also that $\Gamma(1 + z) = z\Gamma(z)$, and that by definition $\sin z = (e^{iz} - e^{-iz})/2i$. Then if $g(t) = \Gamma(1 - it)$:

$$\begin{aligned} |\, g(t)|^2 &= \Gamma(1 - it)\, \Gamma(1 + it) \\ &= it\pi/\sin it\pi \\ &= 2\pi t/(e^{\pi t} - e^{-\pi t}). \end{aligned}$$

Clearly this last expression could only vanish at $t = 0$, and applying L'Hospital's rule shows that $|\, g(0)|^2 = 1$. ∎

(Notice that Theorem 5 does not simply state that the double exponential is the only distribution with a nonvanishing c.f. that is equivalent to the Choice Axiom; that would be a weaker result than the one actually proved, which is completely general.)

From a general standpoint, there remains the open question of whether Theorem 4 can be strengthened to cover all Thurstone models, rather than just those that are equivalent to models with nonvanishing characteristic functions. The answer hinges on whether the functional equation (23) can have solutions other than (24) if we drop the requirement that c_F never vanishes. Ignoring for the moment the fact that such solutions are relevant here only if they actually represent characteristic functions, it is not difficult to construct alternative solutions to (23) when c_F is allowed to vanish over entire intervals. For example, suppose

$$\begin{aligned} c_F(t) &= 1 - |t|, && t \in [-1, 1], \\ &= 1 - ||t| - 5|, && |t| \in [4, 6], \\ &= 0, && \text{elsewhere} \end{aligned}$$

(so that the graph of c_F consists of three triangles, centered at -5, 0, and $+5$, each two units wide), and (for some real r)

$$\begin{aligned} c_G(t) &= c_F(t), && |t| \notin [4, 6], \\ &= e^{ir} c_F(t), && t \in [4, 6], \\ &= e^{-ir} c_F(t), && t \in [-4, -6]. \end{aligned}$$

Then clearly $c_G(t)$ does not have the form $e^{ibt} c_F(t)$, but by considering the various possible values of s and t, one can see that

$$c_G(s)\, c_G(t)\, c_G(-s - t) = c_F(s)\, c_F(t)\, c_F(-s - t),$$

so c_G and c_F satisfy (23), but not (24). (That is, if s and t both lie in $[-1, 1]$, then either $-s - t$ does also, in which case $c_G = c_F$, so the equation is satisfied, or else $c_F(-s - t) = c_G(-s - t) = 0$, in which case the same is true. And so on.) However, I have not been able to determine whether examples constructed along these lines can also be the characteristic functions of distributions satisfying Definition 3. This is left as an open problem, for someone with a better understanding of characteristic functions.

One incomplete, but still useful, way to broaden the scope of Theorem 4 is to notice that even if c_F has zeros, there will always be a neighborhood of the origin in which $|c_F(t)|^2$ is strictly positive (since $c_F(0) = 1$ and c_F is continuous), and consequently if $\mathcal{T}_F$ and $\mathcal{T}_G$ are equivalent for three objects, the ratio $h = c_F/c_G$ will always satisfy (29) for all s and t in some neighborhood of zero. It follows (applying Theorem 1 in Aczél, 1966, p. 46, to the real and imaginary parts of log h) that the solution (24) is always valid for $|t|$ sufficiently close to zero, and consequently the c.f. of $F(x)$ always agrees with that of $G(ax + b)$ in some neighborhood of the origin. This means that all the moments of F agree with the moments of $G(ax + b)$. Consequently *if F is a distribu-*

tion that is determined by its moments, $\mathcal{T}_G$ can be equivalent to $\mathcal{T}_F$ for complete experiments with three or more objects iff *G is another distribution of the same type.* This argument applies to many special cases that are not covered by Theorem 4. But F need not have finite moments or a nonvanishing c.f. in order to make an acceptable discriminal process distribution: The convolution of a Cauchy distribution with any symmetric distribution having one of the periodic c.f.'s described by Feller (1966 Example "a," p. 594) will have a positive density (so that Definition 3 is satisfied), but its c.f. will vanish over intervals, and its moment sequence will be uninformative (since only the odd moments will exist, and these will all be zero because of symmetry).

To prove Theorem 3 using c.f.'s, notice that (16) implies

$$c_G{}^n(t)\, c_G(-nt) = c_F{}^n(at)\, c_F(-ant). \tag{33}$$

As noted in the last paragraph, (24) holds in some neighborhood N of the origin. If $\mathcal{T}_F$ and $\mathcal{T}_G$ are completely equivalent, (16) holds for every n, and for any s, there will be a t in N such that $s = -nt$ for some n. Then $c_G{}^n(-t) = e^{-ibs}c_F{}^n(-at)$ and substituting in (33), we get $c_G(s) = e^{ibs}c_F(as)$ for all s.

4. Invariance under Uniform Expansions of the Choice Set: An Assumption That Implies the Choice Axiom

It is evident from his 1927 paper that Thurstone's grounds for assigning normal distributions to his discriminal processes were entirely heuristic, and had nothing to do with considerations specific to choice behavior: The normal distribution was chosen because of its ubiquitous role in statistics, and the suitability of this choice was left as an open problem, to be decided empirically. Another strategy, more in keeping with modern trends in choice theory, would have been to start with some intuitively plausible assumption about choice behavior, and then derive from this the type of the discriminal process distribution. For example, Thurstone might have begun (anachronistically) by postulating the Choice Axiom, which (in the general case, as shown in the previous section) would have led him to the double exponential distribution instead of the normal. Of course in that case the discriminal process mechanism would have been redundant excess baggage, because the Choice Axiom alone leads to the same predictions as the double exponential Thurstone model, and does so in a much more direct fashion.

There is, however, another assumption, weaker than the Choice Axiom itself (and therefore perhaps intuitively more attractive), which also implies that the discriminal process distribution must be double exponential, and which consequently implies the Choice Axiom—under the assumption that some Thurstone model is true.

To motivate this assumption, suppose that $T_{n,1} = \{o_{11}, o_{21}, \dots, o_{n1}\}$ is a set of n choice objects (the reason for the double subscript will become clear in a moment), and let $p_{C_{n,1}}(i)$ denote the probability of choosing object o_{i1} from $T_{n,1}$. Then suppose we expand this choice set by adding $k-1$ objects that are identical to o_{11} (call these $o_{12}, o_{13}, \dots, o_{1k}$), and $k-1$ objects identical to o_{21} ($o_{22}, o_{23}, \dots, o_{2k}$), and so on for each of the original objects, so that the new choice set is $T_{n,k} = \{o_{11}, ..., o_{1k}; o_{21}, ..., o_{2k}; ...; o_{n1}, ..., o_{nk}\}$. This new set $T_{n,k}$ will be called the (kth-order) *uniform expansion* of the original set $T_{n,1}$.

Now let $p^*_{C_{n,k}}(i)$ denote the probability that in choosing from the expanded set $T_{n,k}$ (where again the subject chooses only one object), the object chosen is one of the type i objects, i.e., one of the objects $o_{i1}, o_{i2}, ..., o_{ik}$. In some contexts it seems intuitively plausible that $p^*_{C_{n,k}}(i)$ should be constant for all values of k, i.e., the probability of choosing a type i object should not be affected by a uniform expansion of the choice set.

For example, suppose the original set consists of three objects: a cup of coffee, a cup of tea, and a glass of milk. Then the kth-order uniform expansion of this set consists of k identical cups of coffee, k cups of tea, and k glasses of milk. Assuming that the desirability of the choice objects themselves is the only consideration, it seems natural to suppose that the probability of choosing a cup of coffee from the expanded set (as opposed to tea or milk) will be the same as it was originally, regardless of k.

Clearly this is what the Choice Axiom predicts, since the redundant objects in the expanded choice set will have the same v scale values as those in the original set: If o_{i1} has scale value v_i, then $o_{i2}, o_{i3}, ..., o_{ik}$ must also have scale value v_i, since they are all identical to o_{i1}, and consequently

$$p^*_{C_{n,k}}(i) = kv_i/k \sum_{j=1}^{n} v_j = p_{C_{n,1}}(i).$$

To state our assumption concisely, we will say that *the predictions of a choice model are invariant under uniform expansions of the choice set iff $p^*_{C_{n,k}}(i) \equiv p_{C_{n,1}}(i)$ for all k.*

Now consider what it would mean for a Thurstone model $\mathscr{T}_F$ to imply this sort of invariance. Suppose the original objects $o_{11}, o_{21}, ..., o_{n1}$ have discriminal processes $u_1 + \mathbf{X}_{11}, u_2 + \mathbf{X}_{21}, ..., u_n + \mathbf{X}_{n1}$ (where the $\mathbf{X}_{i1}$ are all i.i.d.), so that

$$\begin{aligned} p_{C_{n,1}}(i) &= P[u_i + \mathbf{X}_{i1} = \mathrm{Max}\{u_j + \mathbf{X}_{j1} \mid j = 1, ..., n\}] \\ &= \int_{-\infty}^{\infty} \left[\prod_{\substack{j=1 \\ j \neq i}}^{n} F(x + u_i - u_j)\right] dF(x). \end{aligned} \tag{34}$$

In the kth-order uniform expansion of this set the discriminal processes of the type i objects will be $u_i + \mathbf{X}_{i1}, u_i + \mathbf{X}_{i2}, ..., u_i + \mathbf{X}_{ik}$, and the probability of choosing a type i object (as opposed to a type j object, $j \neq i$) will be the probability that one of these random variables is the maximum of the entire set of discriminal processes

$$\{u_j + \mathbf{X}_{jm} \mid j = 1, ..., n; m = 1, ..., k\}.$$

Clearly it does not matter whether the subject ranks all of these discriminal processes without regard for object type, and then selects the object corresponding to the largest, or first ranks the discriminal processes within each type (i.e., determines $\mathrm{Max}\{u_j + \mathbf{X}_{jm} \mid m = 1, ..., k\}$ for each j) and then rank orders these maximums. Consequently the probability of selecting a type i object is the probability that the largest of the type i

discriminal processes is greater than the largest discriminal process for every other object type:

$$p^*_{C_{n,k}}(i) = P[\mathrm{Max}\{u_i + \mathbf{X}_{im} \mid m = 1,..., k\} = \mathrm{Max}\{\mathrm{Max}\{u_1 + \mathbf{X}_{1m} \mid m = 1,..., k\},$$

$$\mathrm{Max}\{u_2 + \mathbf{X}_{2m} \mid m = 1,..., k\},..., \mathrm{Max}\{u_n + \mathbf{X}_{nm} \mid m = 1,..., k\}\}]$$

$$= P[u_i + \mathrm{Max}\{\mathbf{X}_{im} \mid m = 1,..., k\} \tag{35}$$

$$= \mathrm{Max}\{u_1 + \mathrm{Max}\{\mathbf{X}_{1m} \mid m = 1,..., k\},..., u_n + \mathrm{Max}\{\mathbf{X}_{nm} \mid m = 1,..., k\}\}].$$

Since the random variable $\mathrm{Max}\{\mathbf{X}_{jm} \mid m = 1,..., k\}$ has distribution function $F^k(x)$, (35) can be written as

$$p^*_{C_{n,k}}(i) = \int_{-\infty}^{\infty} \left[\prod_{\substack{j=1 \\ j\neq i}}^{n} F^k(x + u_i - u_j)\right] dF^k(x). \tag{36}$$

Now the model $\mathscr{T}_F$ is invariant under uniform expansions iff the integrals in (34) and (36) are equal for every choice of scale values $\{u_j\}$ and every integer k. However this is the same as saying that the Thurstone model $\mathscr{T}_{F^k}$ is completely equivalent to the model $\mathscr{T}_F$, since a given set of scale values $u_1, u_2,..., u_n$ will yield exactly the same choice probabilities in either model (that is, when inserted into either (36) or (34).) Consequently it follows from Theorem 3 that

$$F^k(x) = F(a_k x + b_k). \tag{37}$$

Moreover the constant a_k here must always be one, since if $u_1 = 0$, $u_2 = -u$, we have $p_{C_{2,1}}(1) = D_F(u) = p^*_{C_{2,k}}(1) = D_{F^k}(u)$, and (37) implies $D_{F^k}(u) = D_F(a_k u)$. Thus, the Thurstone model $\mathscr{T}_F$ can be invariant under uniform expansions of the choice set if and only if, for every k,

$$F^k(x) = F(x + b_k), \tag{38}$$

where b_k is a constant that depends on k.

Now functional Eq. (38) is precisely the one that Fisher and Tippet used in 1928 to prove that the limit distribution of the (normalized) maximum Z_k^* of k i.i.d. random variables $Z_1,..., Z_k$ (that is, the limit distribution of $Z_k^* - b_k$ as $k \to \infty$) is the double exponential. Their original argument is explained by Gumbel (1958). However, to obtain a proof that is sufficiently general for the present context we use here a different argument, based on Feller's version of a theorem due to Gnedenko:

LEMMA 9. *F is a distribution function that satisfies functional Eq.* (38) (*i.e., for every k there exists a b_k such that $F^k(x) = F(x + b_k)$ for all x*) *iff F is a distribution of the double exponential type, i.e.,*

$$F(x) = e^{-e^{-(ax+b)}}.$$

Proof. If F is double exponential, then

$$F^k(x) = e^{-ke^{-(ax+b)}}$$
$$= \exp\{-e^{-[a(x-(1/a)\log k)+b]}\}$$
$$= F(x + b_k)$$

where $b_k = -(1/a) \log k$.

To show the converse, suppose $\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_k, ...,$ are i.i.d., with F as their common distribution function. Let $\mathbf{X}_k^* = \text{Max}\{\mathbf{X}_1, ..., \mathbf{X}_k\}$, and $\mathbf{Y}_k = e^{\mathbf{X}_k}$. Then $\mathbf{Y}_k^* = e^{\mathbf{X}^*}$. Feller (1966, pp. 270–271) shows that if a sequence of constants $\{a_k\}$ exists such that $G(y) = \lim_{k\to\infty} P[\mathbf{Y}_k^*/a_k \leqslant y]$ exists and is not concentrated at the origin, then

$$G(y) = e^{-cy^{-a}} \qquad (c > 0,\ a > 0,\ y > 0).$$

Now

$$P[\mathbf{Y}_k^*/a_k \leqslant y] = P[e^{\mathbf{X}_k^*} \leqslant a_k y]$$
$$= P[\mathbf{X}_k^* \leqslant \log y + \log a_k]$$
$$= F^k[\log y + \log a_k]$$
$$= F(\log y + \log a_k + b_k).$$

So if F satisfies (38) for some sequence b_k (in fact we can readily show that $b_k = D_F^{-1}(1/(k+1))$), then $G(y)$ exists (for the sequence $a_k = e^{-b_k}$) and equals $F(\log y)$, which clearly is not concentrated at the origin. Thus

$$F(\log y) = e^{-cy^{-a}}$$

i.e.,

$$F(x) = e^{-ce^{-ax}} = e^{-e^{-(ax+b)}},$$

as claimed. ∎

Combining Lemma 9 with the argument that preceded it, we obtain the following result:

THEOREM 6. *The predictions of a Thurstone model $\mathscr{T}_F$ are invariant under uniform expansions of the choice set if and only if F is a double exponential distribution, i.e., iff $F(x) = e^{-e^{-(cx+d)}}$, where $c > 0$.*

In other words, within the class of Thurstone models, invariance under uniform expansions implies the Choice Axiom.[6]

Theorem 6 could stand to be sharpened a bit, so that it requires only invariance under uniform expansions of choice sets with three objects, instead of an arbitrary number as in the present version. This would be straightforward if Theorem 4 could

[6] For a different approach to motivating the double exponential distribution, also based on the statistics of extremes, see Thompson and Singh, 1967.

be strengthened to cover arbitrary discriminal process distributions instead of just those with nonvanishing characteristic functions. (If the invariance condition is stated only for initial choice sets with three objects, it implies the functional equation $c_{F^k}(x)\, c_{F^k}(y)\, c_{F^k}(-x-y) = c_F(x)\, c_F(y)\, c_F(-x-y)$, by the same argument used to prove Lemma 8 in Section 3.4. Then if Theorem 4 held generally, it would follow that $F^k(x) = F(x+b_k)$, and applying Lemma 9 would complete the proof.)

Thurstone Models with Discriminal Processes Concentrated on Finite Intervals

It was noted earlier (in Section 2.6) that no Thurstone model $\mathscr{T}_F$ with F concentrated on a finite interval (say (α, β)) can be equivalent to the Choice Axiom. This is readily seen by considering the predictions of such a model under uniform expansions of the choice set. Suppose the discriminal processes of the original objects $o_{11}, o_{21}, \ldots, o_{n1}$ are $u_1 + \mathbf{X}_{11}, \ldots, u_n + \mathbf{X}_{n1}$, where the $\mathbf{X}_{i1}$ are bounded between α and β. Then for the kth-order uniform expansion of this set, the probability $p^*_{C_{n,k}}(i)$ of choosing a type i object is the probability that

$$u_i + \mathbf{X}^*_{i,k} = \operatorname{Max}\{u_j + \mathbf{X}^*_{j,k} \mid j = 1, \ldots, n\},$$

where

$$\mathbf{X}^*_{j,k} = \operatorname{Max}\{\mathbf{X}_{j1}, X_{j2}, \ldots, \mathbf{X}_{jk}\}.$$

Now as $k \to \infty$, the random variables $\mathbf{X}^*_{j,k}$ will all converge to β, and consequently $u_j + \mathbf{X}^*_{j,k}$ will converge to $u_j + \beta$, so that the probabilities $p^*_{C_{n,k}}(i)$ will all converge to zero, except for the one corresponding to the largest scale value. In other words, if the discriminal processes are concentrated on a finite interval, then under uniform expansions the subject will always wind up deterministically choosing one of the type i objects, where u_i is the largest scale value.

5. Remarks on Irreversibility and on Independent Random Utility Models

5.1. *The Irreversibility Paradox*

In his 1959 monograph Luce proved a very surprising and counterintuitive result: No Thurstone model can simultaneously satisfy the Choice Axiom both for choices based on the principle "pick the object with the *largest* discriminal process" and for choices based on the principle "pick the object with the *smallest* discriminal process." To motivate this result, consider two complete choice experiments, with the same set of three objects o_1, o_2, o_3: In Experiment B (for best) the subject is instructed to always choose the object he prefers most, and in Experiment W (for worst) he is instructed to always choose the object he prefers least. Let $\{b_S(i)/C_3\}$ denote the system of choice probabilities resulting from Experiment B, and $\{w_S(i)/C_3\}$ the system resulting from Experiment W. It seems almost inescapable that these systems should satisfy the condition

$$b_{\{i,j\}}(i) = w_{\{i,j\}}(j). \tag{39}$$

It is easily shown that (39) will hold and both $\{b_S/C_3\}$ and $\{w_S/C_3\}$ will satisfy the Choice Axiom if there exist v scale values v_1, v_2, v_3 such that

$$b_S(i) = v_i \Big/ \sum_{j \in S} v_j\,, \tag{40}$$

and

$$w_S(i) = (1/v_i) \Big/ \sum_{j \in S} (1/v_j). \tag{41}$$

Now suppose we apply a Thurstone model $\mathscr{T}_F$ to the same situation. It seems natural to assume that the discriminal processes of the objects will be the same in both experiments, i.e., o_i corresponds to $u_i + \mathbf{X}_i$ in both cases, and then to suppose that in Experiment B, the subject picks the object with the largest discriminal process, while in Experiment W he picks the object with the smallest. These assumptions seem quite innocuous, but Luce (1959, p. 57) showed that if they hold there is no distribution F for which the resulting systems $\{b_S/C_3\}$ and $\{w_S/C_3\}$ will both satisfy the Choice Axiom and also condition (39), except in the trivial special case in which all objects are chosen with equal probability.

This impossibility theorem was proved without reference to the double exponential representation of the Choice Axiom. In light of the results described here in Section 3 we can now see why it is true: In order for a Thurstone model to satisfy the Choice Axiom for Experiment B, the $\mathbf{X}_i$ must be double exponential, but when the subject in Experiment W picks the object with the smallest discriminal process (i.e., "pick o_i from S iff $u_i + \mathbf{X}_i = \operatorname{Min}\{u_j + \mathbf{X}_j \mid j \in S\}$") he is actually following the rule "pick o_i iff $-u_i - \mathbf{X}_i = \operatorname{Max}\{-u_j - \mathbf{X}_j \mid j \in S\}$". Consequently in Experiment W the subject is behaving according to a Thurstone model in which the discriminal processes are not double exponential (since that distribution is not symmetrical, $u_i + \mathbf{X}_i$ and $-u_i - \mathbf{X}_i$ cannot both be double exponential), and which therefore is not equivalent to the Choice Axiom.

There is, however, a simple way around this difficulty. Suppose that in Experiment W the subject reverses only the scale values of each object (instead of the entire discriminal process) so that his decision rule is "pick o_j from subset S iff $-u_i + \mathbf{X}_i = \operatorname{Max}\{-u_j + \mathbf{X}_j \mid j \in S\}$". Then if the random variables $\mathbf{X}_i$ are double exponential in both Experiments B and W, the resulting choice probabilities will satisfy the Choice Axiom in both directions, i.e., Eqs. 39, 40, and 41 will all hold simultaneously. (To see this explicitly, set $\mathbf{X}_j = -\log \mathbf{Y}_j$, where $\mathbf{Y}_j$ is exponential with density e^{-y}, and $u_j = \log v_j$. Then all three equations fall out immediately.)

5.2. *Relationship between Thurstone Models and Independent Random Utility* (IRU) *Models*

Luce and Suppes (1965) have defined a class of IRU models which represent the ultimate generalization of Thurstone's Case V notion: A system of choice probabilities $\{p_S/C_n\}$ satisfies an IRU model iff there exist independent random variables $\mathbf{U}_1, \ldots, \mathbf{U}_n$ such that for every $i \in S \subseteq C_n$

$$p_S(i) = P[\mathbf{U}_i = \operatorname{Max}\{\mathbf{U}_j \mid j \in S\}].$$

The $\{\mathbf{U}_i\}$ here need not have any systematic relationship between their distributions, and in particular, these distributions need not be identical except for shifts along the axis—the Thurstone models of Section 2.6 represent that special case, where $\mathbf{U}_i = u_i + \mathbf{X}_i$, with the $\mathbf{X}_i$ all i.i.d. Clearly if the discriminal processes of any Thurstone model $\mathscr{T}_F$ are transformed monotonically (i.e., from $\mathbf{U}_i = u_i + \mathbf{X}_i$ to $\mathbf{U}_i' = m(\mathbf{U}_i)$, where m is strictly increasing), the result will be a non-Thurstonian IRU model that is completely equivalent to $\mathscr{T}_F$. It is natural to define equivalence classes of IRU models of the form M_F, where M_F represents the set of all IRU models obtained by applying monotonic transformations to the Thurstone model $\mathscr{T}_F$: Then the application of the basic uniqueness theorems of Section 3 to these classes of IRU models is straightforward.

Starting from the other direction, Levine (1970) has considered the question, When can an arbitrary family of independent random variables be transformed by a common monotonic transformation in such a way that the resulting distribution functions are identical except for shifts along the axis? The essential condition turns out to be that none of the original distribution functions can intersect (i.e., if $F \neq G$, then either $F(x) < G(x)$ for all x, or vice versa). If Levine's conditions hold for an arbitrary set of random variables that constitute an IRU model, then that model is equivalent to some Thurstone model, and its uniqueness status can be determined accordingly, using the results of Section 3.

References

Aczél, J. *Lectures on functional equations.* New York: Academic Press, 1966.

Adams, E., & Messick, S. An axiomatization of Thurstone's successive intervals and paired comparisons scaling models. Technical Report 12, Applied Mathematics and Statistics Laboratory, Stanford University, 1957.

Ahlfors, L. V. *Complex analysis*, 2nd ed. New York: McGraw–Hill, 1966.

Burke, C. J., & Zinnes, J. L. A paired comparison of pair comparisons. *Journal of Mathematical Psychology*, 1965, **2**, 53–76.

Dawkins, R. A threshold model of choice behavior. *Animal Behavior*, 1969, **17**, 120–133.

Feller, W. *An introduction to probability theory and its applications, Vol. II.* New York: Wiley, 1966.

Gumbel, E. G. *Statistics of extremes.* New York: Columbia University Press, 1958.

Kotlarski, I. I. On some characterizations of probability distributions in Hilbert spaces. *Annali di Mathematica pura ed applicata*, 1966, **74**, 129–134.

Laha, R. G. On a problem connected with beta and gamma distributions. *Transactions of the American Mathematical Society*, 1964, **113**, 287–298.

Levine, M. V. Transformations that render curves parallel. *Journal of Mathematical Psychology*, 1970, **7**, 410–443.

Luce, R. D. *Individual choice behavior.* New York: Wiley, 1959.

Luce, R. D., & Suppes, P. Preference, utility, and subjective probability. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.) *Handbook of Mathematical Psychology, Vol. III.* New York: Wiley, 1965.

Lukacs, E. *Characteristic functions.* London: Griffen, 1960.

McFadden, D. Conditional logit analyses of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in econometrics.* New York: Academic Press, 1974.

Thompson, W. A., & Singh, J. The use of limit theorems in paired comparison model building. *Psychometrika*, 1967, **32**, 255–264.

THURSTONE, L. L. A law of comparative judgment. *Psychological Review*, 1927, **34**, 273–286.
TVERSKY, A. Choice by elimination. *Journal of Mathematical Psychology*, 1972, **9**, 340–367.
WHITTAKER, E. T., & WATSON, G. N. *A course of modern analysis*, 4th ed. Cambridge: Cambridge University Press, 1963.
YELLOTT, J. Generalized Thurstone representations for three choice theories: uniqueness theorems. Paper given at annual Mathematical Psychology Meeting, Princeton, September 1971.

RECEIVED: November 14, 1975