EC PSYCHOLOGY AND PSYCHIATRY
Research Article

Noreen's 4I2AFC Area Theorem and its Extension

Jian Bi[1]* and Carla Kuesten[2]

[1]Sensometrics Research and Service, Richmond, VA, United States

[2]Amway, Ada, MI, USA

***Corresponding Author:** Jian Bi, Sensometrics Research and Service, Richmond, VA, United States.

Received: September 24, 2019; **Published:** October 04, 2019

Abstract

This research note proposes a new type of sensory discrimination forced-choice methods, i.e. the paired specified 'M+N' with different (AB and BA) pairs. Noreen's 4I2AFC area theorem in signal detection theory is originally demonstrated and extended. An analytical expression of a general psychometric function for the new type of methods is derived from the extension of the 4I2AFC area theorem. The performances of the new type of methods in both difference testing and similarity/equivalence testing are explored. It shows that the new type of methods is more powerful than the conventional methods.

Keywords: *Signal Detection Theory (SDT); Thurstonian Modeling; Specified 'M+N'; Paired Specified 'M+N'; Psychometric Function; 4I2AFC Area Theorem*

Introduction

There are three main objectives for this paper. One is to propose and explore a new type of sensory discrimination forced-choice methods, i.e., the paired specified 'M+N' with different (AB and BA) pairs. The second is to demonstrate and extend an "Area Theorem" proposed by Noreen [1] in signal detection theory. The third objective is to derive an analytical expression of a general psychometric function for the new type of methods.

This paper can be regarded as a generalization of a recent paper, Bi and Kuesten [2] in which the 4I2AFC as a special situation of the paired specified 'M+N' with different (AB and BA) pairs (M=N=1) is explored.

This paper can also be viewed as a companion to the two recent papers, Bi [3] and Bi, Kuesten, Lee, and O'Mahony [4]. Bi [3] originally derived an analytical expression of a general psychometric function for the specified 'M+N' methods, based on an extension of Green's 2-AFC area theorem in signal detection theory [5,6]. Bi, Kuesten, Lee, and O'Mahony [4] originally derived an analytical expression of a general psychometric function for a paired version of the specified 'M+N' with same (AA or BB) and different (AB or BA) pairs, based on an extension of the same-different area theorem proposed by Noreen [1] and demonstrated independently by Irwin, Hautus, and Butcher [7], Micheyl and Dai [8] and Bi, Kuesten, Lee, and O'Mahony [4]. The current paper will derive originally an analytical expression of a general psychometric function for another paired version of the specified 'M+N', i.e. the paired specified 'M+N' with different (AB and BA) pairs, based on an extension of the 4I2AFC area theorem proposed also by Noreen [1].

The current paper and our recently published papers mentioned above discuss three types of sensory discrimination forced-choice methods, i.e. the specified 'M+N' and two paired versions of the specified 'M+N'; demonstrate and extend three profound area theorems in signal detection theory, i.e. Green's 2-AFC area theorem, Noreen's same-different area theorem, and Noreen's 4I2AFC area theorem; and derive originally three general psychometric functions for the three types of methods. A common motivation of these papers is to establish a theoretical structure to generalize the sensory discrimination forced-choice methods.

Review of the specified 'M+N' and paired versions of the specified 'M+N' methods

As discussed in Bi, Lee, and O'Mahony[9] almost all the sensory discrimination forced-choice methods can be generalized in the 'M+N' framework. There are specified and unspecified versions of the 'M+N' methods. In the specified version, the characteristics of the samples are identified. In the unspecified version, the samples are sorted without being identified. We discuss only the specified 'M+N' in this paper.

There are also paired versions of the specified 'M+N', in which a pair of stimulus presentations is used as a single perceptual event. Bi, Kuesten, Lee, and O'Mahony [4] discussed a paired version of the specified 'M+N' with same (AA or BB) and different (AB or BA) sample pairs. The dual-pair (4IAX) [1,10,11] which is the paired 2-AFC, is a special case of the specified 'M+N' with one same (AA or BB) and one different (AB or BA) pairs (M=N=1). This paper discusses another paired version, i.e., the specified 'M+N' with different (AB and BA) pairs.

In a paired specified 'M+N' with different (AB and BA) pairs, M BA pairs and N AB pairs are presented simultaneously to a subject. The task of the subject is to select the N AB pairs, where A is a signal or a stronger stimulus, B is noise or a weaker stimulus, and AB is a pair with decreasing stimuli change. The 4I2AFC [1,2,10], which is the paired 2-AFC with different (AB and BA) pairs, is a special case of this paired version of the specified 'M+N' with one AB pair and one BA pair (M=N=1).

All the specified 'M+N' and the paired versions of the specified 'M+N' are forced-choice methods without response bias or with less response bias than non-forced-choice methods. Table 1 of this paper lists the names of some specific specified 'M+N' and paired versions of the specified 'M+N' methods with smaller M and N.

M, N	Specified 'M+N'*	Paired specified 'M+N' (with same (AA or BB) and different (AB or BA) pairs)**	Paired specified 'M+N' with different (AB and BA) pairs***
M=N=1	2-AFC	Paired 2-AFC (4IAX)	Paired 2-AFC with different (AB and BA) pairs (4I2AFC)
M=2, N=1	3-AFC	Paired 3-AFC	Paired 3-AFC with different (AB and BA) pairs
M=N=2	Specified tetrad	Paired specified tetrad	Paired specified tetrad with different (AB and BA) pairs
M=3, N=1	4-AFC	Paired 4-AFC	Paired 4-AFC with different (AB and BA) pairs
M=3, N=2	Specified 'two-out-of-five	Paired specified 'two-out-of-five	Paired specified 'two-out-of-five with different (AB and BA) pairs
M=N=3	Specified hexagon	Paired specified hexagon	Paired specified hexagon with different (AB and BA) pairs
M=N=4	Specified octad	Paired specified octad	Paired specified octad with different (AB and BA) pairs

Table 1: *Names of Some Specific Specified 'M+N' and Paired Versions of the Specified 'M+N' Methods*

*Note: *M and N are the numbers of noise and signal, respectively;*

*** M and N are the numbers of same pair and different pair, respectively;*

**** M and N are the numbers of BA pair and AB pair, respectively.*

The 4I2AFC area theorem and its extension

The 4I2AFC area theorem

In the literature of mathematical psychology, Noreen [1] proposed, without a strict demonstration, an analogous "Area Theorem" to Green's 2-AFC area theorem. It is referred to as the 4I2AFC area theorem. The theorem suggests that if a pair of stimulus presentations is treated as a single perceptual event, the stimulus pair AB can be used as a 'signal' and pair BA can be used as a 'noise' in a yes-no or A-Not A paradigm. Then, the area under the *ROC* curve (*AUC*) of the yes-no paradigm should equal the probability of correct responses (P_c) of the four-interval, two-alternative forced-choice (4I2AFC) task. This is the so-called Noreen's 4I2AFC area theorem, which is originally demonstrated in Appendix A.1 of this paper as in Equation (1):

$$AUC = \Phi(\delta) = P_c, \quad (1)$$

where Φ() denotes a cumulative distribution function of the standard normal distribution and $P_c=\Phi(\delta)$ is a psychometric function of the 4I2AFC [2].

The importance of the area theorem is that it reveals the inherent relationship between a method with response bias and a forced-choice method without response bias or with less response bias than non-forced-choice methods.

An extension of the 4I2AFC area theorem

Noreen's 4I2AFC area theorem can be extended to link the *AUC* and P_c of all the paired specified 'M+N' with different (AB and BA) pairs. Equation (2) is derived originally in Appendix A.2 of this paper.

$$AUC = N\int_{-\infty}^{\infty}\Phi\left(x+\sqrt{2}\delta\right)^{M}\Phi(-x)^{N-1}\phi(x)dx, \qquad (2)$$

where ϕ() is the density function of the standard normal distribution.

It means that for a *ROC* curve with a *hit* probability $H'=H^N$ and a *false-alarm* probability $F'=1-(1-F)^M$, the area under the *ROC* curve (*AUC*) is in Equation (2), where H and F are the *hit* and *false-alarm* probabilities of the yes-no paradigm with pair AB as a 'signal' and pair BA as a 'noise'. It can be demonstrated that Equation (2) becomes Equation (1) when M=N=1 (See, Appendix A.1). It is $AUC=P_c$ for the 4I2AFC.

However, except for the 4I2AFC, so far, there is no an analytical expression of the psychometric function for the paired specified 'M+N' with different (AB and BA) pairs in the literature. It can be demonstrated by simulation in the next section of this paper that the right-part of Equation (2) equals the P_c of the paired specified 'M+N' with different (AB and BA) pairs.

Simulation-derived psychometric function for the paired specified 'M+N' with different (AB and BA) pairs

We can produce a simulation-derived psychometric function for the paired specified 'M+N' with different (AB and BA) pairs for any specific M and N, using a method similar as that in Bi, Lee, and O'Mahony [9].

For example, for the paired specified two-out-of-five ('M+N' with M=3, N=2) with different (AB and BA) pairs, 10,000 sets (x_1, x_2, x_3, y_1, y_2) of random numbers are drawn from a normal distribution with mean -d' and standard deviation √2, i.e., $X \sim N(-d', \sqrt{2})$, and a normal distribution with mean d' and standard deviation √2, i.e., $Y \sim N(d', \sqrt{2})$, respectively. A correct response will be made if max (x_1, x_2, x_3) < min (y_1, y_2). The probability of correct responses in the 10,000 sets of samples can be observed. For 41 values of *d'* from 0 to 4 with a step of 0.1, we can obtain the 41 corresponding P_c values. With the 41 sets of the P_c and *d'* values, a simulated psychometric function can be established by using the R built-in program 'smooth.spline'. The R code 'sfmnabba(m,n)' for the simulated psychometric function is available from Appendix B in this paper.

For any given *m* (M=*m*) and *n* (N=*n*) as input of the R code '*sfmnabba*(*m,n*)', a simulation-derived psychometric function for the paired specified 'M+N' with different (AB and BA) pairs can be produced as below for 5 specific paired specified 'M+N' with different (AB and BA) pairs methods.

```
> sfmnabba11<-sfmnabba(1,1)
> sfmnabba21<-sfmnabba(2,1)
> sfmnabba31<-sfmnabba(3,1)
> sfmnabba22<-sfmnabba(2,2)
> sfmnabba32<-sfmnabba(3,2)
```

For any given d' value(s), the simulated P_c value(s) can be obtained by using the simulation-derived psychometric function and an R built-in program 'predict'. For example, for the paired 3-AFC with different (AB and BA) pairs (M=2, N=1), for d' = 0, 0.5, 1, 1.5, and 2, the simulated P_c values are 0.3333, 0.5449, 0.7442, 0.8861 and 0.9595, respectively. The *AUC* values based on Equation (2) are 0.3333, 0.5462, 0.7452, 0.8847, and 0.9586, respectively, by using the R code '*pmnabba*' for M=2 and N=1.

```
> predict(sfmnabba21,seq(0,2,0.5))$y
[1] 0.3334645 0.5448578 0.7444223 0.8861241 0.9594556
> pmnabba(2,1,seq(0,2,0.5))
[1] 0.3333 0.5462 0.7452 0.8847 0.9586
```

Figure 1 shows the analytical *AUC* values based on Equation (2) and the simulated P_c values based on the simulation-derived psychometric functions for some paired specified 'M+N' with different (AB and BA) pairs for some specific M and N values. It shows a very good consistency between the AUC values and the P_c values.

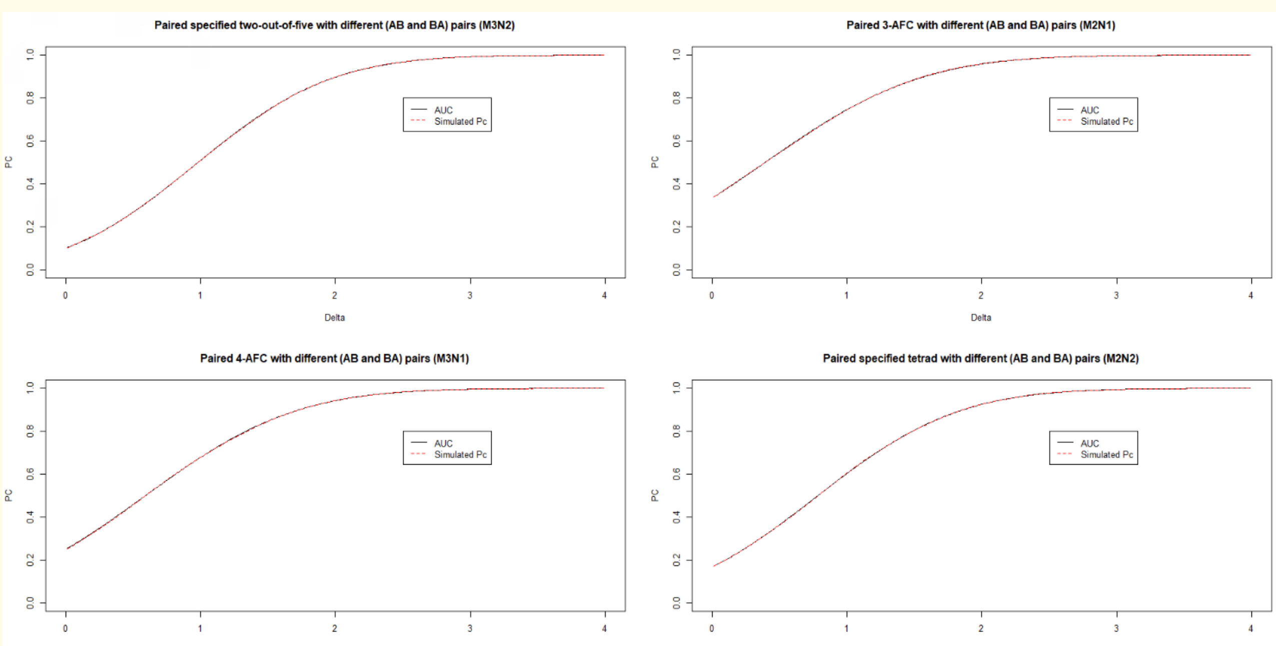

Figure 1: *Area under ROC Curve and Proportion of Correct Responses (P_c) of each of some Methods of the Paired Specified 'M+N' with different (AB and BA) Pairs*

Analytical general psychometric function of the paired specified 'M+N' with different (AB and BA) pairs

The equivalence between the simulated and the AUC values suggests that the analytical expression in Equation (2) for *AUC* can be used as a psychometric function in Equation (3) for the paired specified 'M+N' with different (AB and BA) pairs.

$$P_c = N\int_{-\infty}^{\infty}\Phi\left(x+\sqrt{2}\delta\right)^{M}\Phi(-x)^{N-1}\phi(x)dx. \qquad (3)$$

The psychometric function in Equation (3) has been demonstrated analytically for the 4I2AFC with M=N=1. It has been demonstrated by simulations for some methods of the paired specified 'M+N' with different (AB and BA) pairs with specific M and N values.

With the psychometric function in Equation (3), the P_c value(s) can be estimated for given δ or *d'* value(s) and for any given specific M and N values by using the R code '*pmnabba*(*m,n,d*)'. For example, for the paired 3-AFC with different (AB and BA) pairs (M=2, N=1), for δ =1 and 2, the corresponding P_c values are 0.7452 and 0.9586, respectively. For observed P_c value(s), the corresponding δ or *d'* value(s) can also be estimated from Equation (3) by using the R code '*qmnabba*(*m,n, pp*)'. For example, for the paired 3-AFC with different (AB and BA) pairs (M=2, N=1), for observed P_c values = 0.7452 and 0.9586, the estimated δ or *d'* values are 1 and 2 as below.

```
> pmnabba(2,1,c(1,2))
[1] 0.7452 0.9586
> qmnabba(2,1,c(0.7452,0.9586))
[1] 1 2
```

Estimation of variance of *d'* is often needed to measure precision of the estimated parameter δ and is used in statistical inference for *d'*. As described by Bi, Ennis, and O'Mahony [12] variance of *d'* for a forced-choice method is composed of two components: sample size N and the *B-value*, which is determined solely by the method used. A general form of variance of *d'* for a forced choice method can be expressed as *Var*(*d'*) = *B/N*. The *B*-value can be obtained from $B=p_c(1-p_c)/f'^2(d_0')$, where *f* denotes a psychometric function for a method and $f'(d_0')$ denotes the derivative of the function *f*(*d'*) evaluated at d_0'. For the paired specified 'M+N' with different (AB and BA) pairs, *f* is in Equation (3) and $f'(d_0')$ is in Equation (4).

$$f'(d_0') = \sqrt{2}NM\int_{-\infty}^{\infty}\Phi\left(x+\sqrt{2}d_0'\right)^{M-1}\Phi(-x)^{N-1}\phi(x+\sqrt{2}d_0')\phi(x)dx. \qquad (4)$$

The R code '*bmnabba*(*m,n,dd*)' can be used to estimate the B-values and the variance of *d'*. For example, for the paired 3-AFC with different (AB and BA) pairs (M=2, N=1), if the estimated *d'* values are 1 and 2, the estimated *B*-values are 1.5720 and 4.4364, respectively. If the sample size is 100, then the variances of the estimated *d'* values are 0.0157 and 0.0444, respectively.

```
> bmnabba(2,1,c(1,2))
[1] 1.572020 4.436424
```

Performance of the paired specified 'M+N' with different (AB and BA) pairs in difference testing

Difference testing power

For calculation of difference testing power for a forced-choice method, see, e.g. Ennis [13], Ennis and Jesionka [14] and Bi [15]. For example, for *d'*= 0, the corresponding proportion of correct responses for the paired 3-AFC with different (AB and BA) pairs is P_{c0} = 1/3. For an assumed true difference in terms of δ or *d'*, e.g., d_1' = 0.5, the corresponding proportion of correct responses for the method is p_{c1}= 0.5462, based on the psychometric function in Equation (3) for M=2, N=1.

The power of a difference test is the probability to correctly reject the null hypothesis H_0 when H_0 is false and an alternative hypothesis H_1 is assumed. The power of the difference test with the null hypothesis H_0: d_0' = 0 against the alternative hypothesis H_1 : d_1' = 0.5 is in fact the power of a one-sample binomial test with the null hypothesis H_0: $P_c=P_{c0}$ = 1/3 against the alternative hypothesis H_1: $P_c=p_{c1}$ = 0.5462. For a given sample size, e.g. n = 50, Type I error α = 0.05, the power of difference testing using the paired 3-AFC with different (AB and BA) pairs can be calculated by using the built-in S-Plus program *binomial.sample.size* [16] or an R code '*powf*' in Appendix B of this paper for a one-sample test of binomial proportion as below. The power is about 0.93 in this case.

```
> binomial.sample.size(p = pmnabba(2, 1, 0), p.alt = pmnabba(2, 1, 0.5), alternative = "great", n1 = 50, correct = F)$power
[1] 0.9287289
> powf(p0=pmnabba(2,1,0),p1=pmnabba(2,1,0.5),alpha=0.05,n=50)
[1] 0.93
> pmnabba(2, 1, 0)
[1] 0.3333
> pmnabba(2, 1, 0.5)
[1] 0.5462
```

Power comparisons

Difference testing powers of the following 5 specific methods of the paired specified 'M+N' with different (AB and BA) pairs are compared.

M1N1: Paired 2-AFC with different (AB and BA) pairs (4I2AFC)

M2N1: Paired 3-AFC with different (AB and BA) pairs

M3N1: Paired 4-AFC with different (AB and BA) pairs

M2N2: Paired specified tetrad with different (AB and BA) pairs

M3N2: Paired specified two-out-of-five with different (AB and BA) pairs

Figure 2 presents the five power curves plotting the powers against *d'* from 0.1 to 1 with sample size n = 30. It shows that the paired specified tetrad with different (AB and BA) pairs (M2N2) and the paired specified two-out-of-five with different (AB and BA) pairs (M3N2) are the most powerful among the five methods while the 4I2AFC (M1N1) has the lowest difference testing power in difference testing.

It should be mentioned that the conclusion that the new type of methods with larger M and N have larger testing powers is based solely on theoretical derivation and computer simulations. Considering that a larger number of samples in a method may involve more adaptation and fatigue, the actual operational power may be lower than that theoretical or simulated results suggest.

It is noted that power comparisons have been conducted between the 4I2AFC and the conventional methods in Bi and Kuesten [2]. The results show that the 4I2AFC is more powerful than the conventional ones. Hence, Figure 2 suggests that the paired specified 'M+N' with different (AB and BA) pairs are more powerful than the conventional methods in difference testing.

Performance of the paired specified 'M+N' with different (AB and BA) pairs in similarity/equivalence testing

Similarity/equivalence testing power

For similarity/equivalence testing power using a forced-choice method, see, e.g. Bi [15,17]. The power of the similarity test in this situation is in fact the power of a one-sample binomial test with the null hypothesis H_0: $P_c=p_{c0}$ against the alternative hypothesis H_1: $P_c=p_{c1}<p_{c0}$, where p_{c0} and p_{c1} correspond to d'_0 and d'_1, respectively.

For example, consider the paired 3-AFC with different (AB and BA) pairs. The power of a similarity test is about 0.76 if d'_0= 1 is a given similarity limit and the true difference is assumed to be d'_1= 0.8, i.e., p_{c0} = 0.745 and p_{c1}= 0.671, and is based on a given sample size, e.g., n

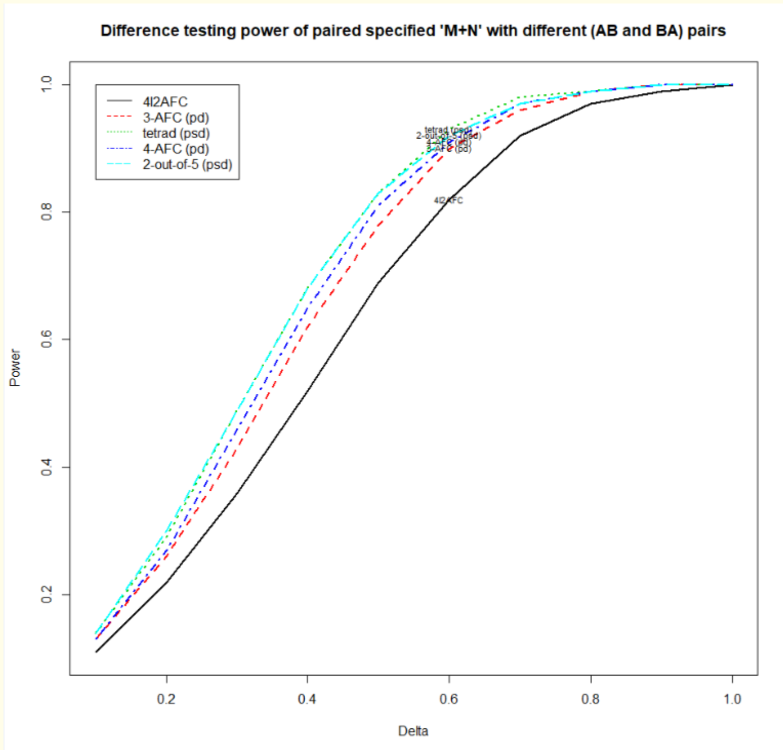

Figure 2: *Difference Testing Powers for the Paired Specified 'M+N' with different (AB and BA) Pairs.*

Note:

tetrad (psd): Paired specified tetrad with different (AB and BA) pairs (M2N2).

2-out-of-5 (psd): Paired specified Two-out-of-five with different (AB and BA) pairs (M3N2).

4-AFC (pd): Paired 4-AFC with different (AB and BA) pairs (M3N1).

3-AFC (pd): Paired 3-AFC with different (AB and BA) pairs (M2N1).

4I2AFC: Paired 2-AFC with different (AB and BA) pairs (M1N1).

= 200, Type I error α = 0.05. The power may be determined using the built-in S-Plus program *binomial.sample.size* or an R code '*powf*' in Appendix B as below.

```
> binomial.sample.size(p = pmnabba(2,1,1), p.alt = pmnabba(2,1,0.8), alternative = "less", n1 = 200, correct = F)$power
[1] 0.7604942
> powf(p0=pmnabba(2,1,1),p1=pmnabba(2,1,0.8),alpha=0.05,n=200)
[1] 0.76
> pmnabba(2,1,1)
[1] 0.7452
> pmnabba(2,1,0.8)
[1] 0.671
```

Power comparisons

Similarity testing powers of the following 5 specific methods of the paired specified 'M+N' with different (AB and BA) pairs are compared: the paired 2-AFC with different (AB and BA) pairs (4I2AFC) (M1N1), the paired 3-AFC with different (AB and BA) pairs (M2N1), the paired 4-AFC with different (AB and BA) pairs (M3N1), the paired specified tetrad with different (AB and BA) pairs (M2N2), and the paired specified two-out-of-five with different (AB and BA) pairs (M3N2).

Figure 3 presents the five power curves plotting the powers against δ or *d'* from 0.5 to 1 with sample size *n*=100 and a similarity limit δ or *d'*= 1. It shows that the paired specified two-out-of-five with different (AB and BA) pairs (M3N2) is the most powerful while the 4I2AFC (M1N1) is the least powerful among the 5 methods in similarity testing.

It is noted that similarity testing power comparisons have been conducted between the 4I2AFC and the conventional methods in Bi and Kuesten [2]. The results show that the 4I2AFC is more powerful than the conventional ones. Hence, Figure 3 suggests that the paired specified 'M+N' with different (AB and BA) pairs are more powerful than the conventional methods in similarity testing.

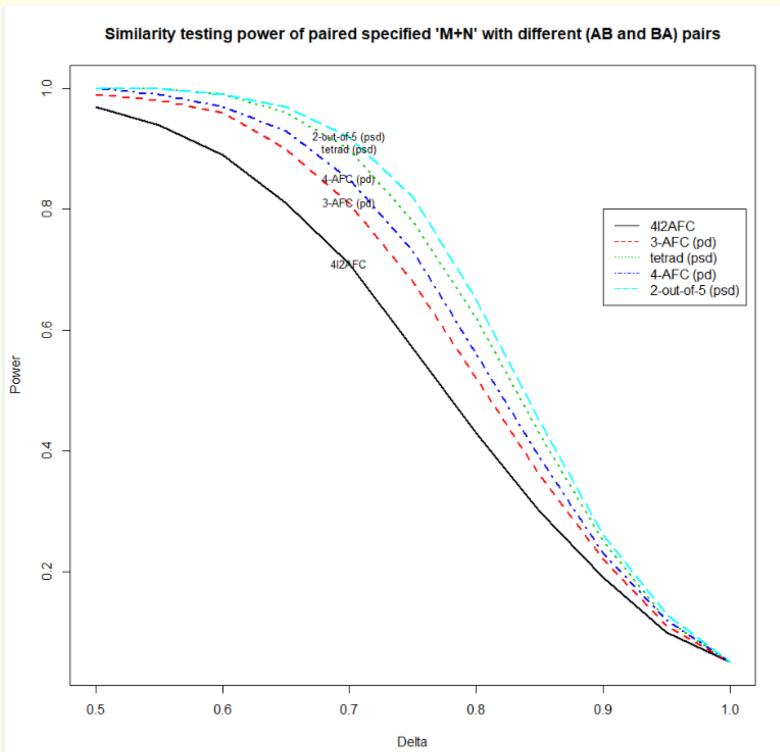

Figure 3: *Similarity/Equivalence Testing Powers for the Paired Specified 'M+N' with Different (AB and BA) Pairs.*

Note:

2-out-of-5 (psd): Paired specified two-out-of-five with different (AB and BA) pairs (M3N2).

tetrad (psd): Paired specified tetrad with different (AB and BA) pairs (M2N2).

4-AFC (pd): Paired 4-AFC with different (AB and BA) pairs (M3N1).

3-AFC (pd): Paired 3-AFC with different (AB and BA) pairs (M2N1).

4I2AFC: Paired 2-AFC with different (AB and BA) pairs (M1N1).

Concluding Remarks

The new type of methods, i.e. the paired specified 'M+N' with different (AB and BA) pairs proposed and explored in this paper, have potential wide application perspective. It is particularly applicable to assessment of before-after treatment effects for visual or manual inspection of food or non-food products. A comprehensive case study using the 4I2AFC method for the facial images was conducted in a recent paper [2]. Another industrial application of the new type of methods is in progress and will appear in a separate applied paper.

This paper makes two novel theoretical contributions to signal detection theory (SDT) and Thurstonian modeling. One is to demonstrate originally Noreen's 4I2AFC area theorem [1] and extend it to link the area under *ROC* curve (*AUC*) with the probability of correct responses (P_c) of each of the methods of the paired specified 'M+N' with different (AB and BA) pairs. Another contribution is to derive originally an analytical expression of a general psychometric function for the paired specified 'M+N' with different (AB and BA) pairs.

So far, we have generalized three types of sensory discrimination forced-choice methods, i.e. the specified 'M+N' and two versions of the specified 'M+N'. The three types of the methods and their theoretical structures for generalization are listed in Table 2. We believe that the key findings are fundamental to sensory analysis methodology and sensory science.

It is noted that this paper and our recently published papers do not touch the unspecified version of the 'M+N'. The conventional unspecified 'M+N' contains the triangle (M=2, N=1), the *m*-alternative oddity [18] where $m > 3$, M=m-1 and N=1, the unspecified tetrads [19] where M=N=2, the unspecified hexagon (M=N=3), the unspecified octad (M=N=4), etc. as its specific situations.

We are curious whether the unspecified 'M+N' can also be extended to the paired versions and whether the unspecified 'M+N' and its paired versions can also be generalized in some ways. Extending the unspecified 'M+N' to the paired versions, exploring the paired versions of the unspecified 'M+N', and generalizing the unspecified 'M+N' and its paired versions might be interesting topics of further research.

Type of methods	General psychometric functions	Associated area theorems
The specified 'M+N'*	$P_c = N\int_{-\infty}^{\infty}\Phi(x)^M\Phi(\delta - x)^{N-1}\phi(x-\delta)dx$	Extension of Green's 2-AFC area theorem
The paired specified 'M+N' (with same and different pairs)**	$P_c = 2M\int_{-\infty}^{0}[\Phi\left(x+\frac{\delta}{\sqrt{2}}\right)+\Phi\left(x-\frac{\delta}{\sqrt{2}}\right)]^N[1 - 2\Phi(x)]^{M-1}\phi(x)dx$	Extension of Noreen's same-different area theorem
The paired specified 'M+N' with different (AB and BA) pairs***	$P_c = N\int_{-\infty}^{\infty}\Phi\left(x+\sqrt{2}\delta\right)^M\Phi(-x)^{N-1}\phi(x)dx$	Extension of Noreen's 4I2AFC area theorem

Table 2: *Theoretical Structure of the Specified 'M+N' and Paired Versions of the Specified 'M+N' Methods*
*Note: *Bi [3]. **Bi, Kuesten, Lee, & O'Mahony [4]. ***This paper.*

Appendix A: Demonstrations of Noreen's 4I2AFC area theorem (1981) and its extension

A.1. Proof of Noreen's 4I2AFC area theorem

The *ROC* curve is a plot of *hit* (*H*) proportions versus *false-alarm* (*F*) proportions. It shows the relationship between the two probabilities as the decision criterion varies. The *ROC* curve can be expressed as *ROC=H(F)*, i.e. *H* is a function of *F*.

A.1.1. *ROC* curve of YN with pair AB as a signal and pair BA as a noise

It is noted that the yes-no (YN) paradigm may represent the A-Not A method in the sensory field. The *hit* probability (*H*) in the YN paradigm is the probability of response "A" when sample A (signal) is presented, i.e. P("A"|A). The *false-alarm* probability (*F*) in the YN paradigm is the probability of response "A" when sample Not A (noise) is presented, i.e., P("A"|N). *F* is also a complement of the probability of correct rejection, i.e., *F*= 1-*Probability of correct rejection*, i.e. P("N"|N). *H* is also a complement of the probability of *miss*, i.e., *H*=1-*Probability of miss*, where the *probability of miss* is the probability of response "*N*" when sample A (signal) is presented, i.e., P("N"|A). The probability of correct responses in the YN paradigm includes both the *hit* probability and the *probability of correct rejection*.

The *hit* probability in the YN paradigm is the probability of Y> *c* for pair AB (signal), i.e., P(Y>c|s), where *c* is a criterion and Y denotes a sensation. The *false-alarm* probability in the YN paradigm is the probability of Y >*c* for pair BA (noise), i.e. P(Y>c|n) . It is usually assumed that the sensation for the AB pair and the sensation for the BA pair follow normal distributions with standard deviation $\sqrt{2}$ and mean δ and -δ, respectively, i.e., ., $Y \sim N(\delta,\sqrt{2})$ for the AB pair and $Y \sim N(-\delta,\sqrt{2})$ for the BA pair. Hence the *hit* and *false-alarm* probabilities are Equations (A1) and (A2), respectively.

$$H = P(Y > c|s) = 1 - \Phi\left(\frac{c-\delta}{\sqrt{2}}\right) = \Phi\left(\frac{\delta-c}{\sqrt{2}}\right), \quad \text{(A1)}$$

$$F = 1 - Probability\ of\ correct\ rejection = 1 - P(Y < c|n) = 1 - \Phi\left(\frac{c+\delta}{\sqrt{2}}\right), \quad \text{(A2)}$$

where Φ(.) denotes a cumulative distribution function of the standard normal distribution.

From Equation (A2), $c=\sqrt{2}\,\Phi^{-1}(1-F)-\delta$, where $\Phi^{-1}(.)$ denotes the inverse standard normal transform. Hence, the *ROC* curve function for the YN with pair AB as a signal and pair BA as a noise should be Equation (A3).

$$ROC = \Phi\left(\frac{\delta-c}{\sqrt{2}}\right) = \Phi(\sqrt{2}\delta - \Phi^{-1}(1-F)). \quad \text{(A3)}$$

A.1.2. The area under the *ROC* curve (*AUC*)

The area under the *ROC* curve (*AUC*) can be calculated in Eq. (A4) ([20], p47).

$$AUC = \int_0^1 (1-F)dH = \int_0^1 \Phi\left(\frac{c+\delta}{\sqrt{2}}\right) d\Phi\left(\frac{\delta-c}{\sqrt{2}}\right). \quad \text{(A4)}$$

Note that $\frac{c}{\sqrt{2}} \to -\infty$, when $\Phi\left(\frac{\delta-c}{\sqrt{2}}\right) \to 1$, hence Eq. (A4) becomes Eq. (A5).

$$AUC = -\int_{\infty}^{-\infty} \Phi\left(\frac{c+\delta}{\sqrt{2}}\right) \phi\left(\frac{\delta-c}{\sqrt{2}}\right) d\left(\frac{c}{\sqrt{2}}\right) = \int_{-\infty}^{\infty} \Phi\left(\frac{c+\delta}{\sqrt{2}}\right) \phi\left(\frac{\delta-c}{\sqrt{2}}\right) d\left(\frac{c}{\sqrt{2}}\right). \quad \text{(A5)}$$

Let $x = \frac{c}{\sqrt{2}}$, Eq. (A5) becomes Eq. (A6).

$$AUC = \int_{-\infty}^{\infty} \Phi\left(x + \frac{\delta}{\sqrt{2}}\right) \phi\left(\frac{\delta}{\sqrt{2}} - x\right) dx = \int_{-\infty}^{\infty} \Phi\left(x + \frac{\delta}{\sqrt{2}}\right) \phi\left(x - \frac{\delta}{\sqrt{2}}\right) dx. \quad \text{(A6)}$$

Let $y = x - \frac{\delta}{\sqrt{2}}$, Eq. (A6) becomes Eq. (A7).

$$AUC = \int_{-\infty}^{\infty} \Phi(y + \sqrt{2}\delta)\phi(y)dy, \quad \text{(A7)}$$

where $\phi(.)$ denotes a density function of the standard normal distribution. According to Owen [21], p.403,

10,010.8), for $a = \sqrt{2}\delta$, b=1, Eq. (A7) becomes Eq. (A8).

$$AUC = \int_{-\infty}^{\infty} \Phi(y + \sqrt{2}\delta)\phi(y)dy = \Phi\left(\frac{a}{\sqrt{1+b^2}}\right) = \Phi(\delta). \quad \text{(A8)}$$

Note that the right-hand side of Eq. (A8) is just a psychometric function for the 4I2AFC [2], hence

$$AUC = \Phi(\delta) = P_c. \quad \text{(A9)}$$

It is demonstrated originally and analytically that the area under the *ROC* curve (*AUC*) of a YN paradigm with AB pair as a signal and BA pair as a noise equals the proportion of the correct responses (P_c) in the 4I2AFC for the same stimuli as suggested by Noreen [1].

A.2. An extension of the 4I2AFC area theorem

Let $H'=H^N=\Phi(\delta-c/\sqrt{2})^N$ and , $F'=1-(1-F)^M=1-\Phi(c+\delta/\sqrt{2})^M$, then the *AUC*, i.e. the area under the *ROC* curve with the *hit* probability H' and the *false-alarm* probability F' is in Equation (A10).

$$AUC = \int_0^1 (1-F')dH' = N\int_{-\infty}^{\infty} \Phi\left(x+\frac{\delta}{\sqrt{2}}\right)^M \Phi\left(\frac{\delta}{\sqrt{2}}-x\right)^{N-1} \phi\left(\frac{\delta}{\sqrt{2}}-x\right)dx, \qquad (A10)$$

where $x=c/\sqrt{2}$. Let $y= x-\delta/\sqrt{2}$, Eq. (A10) can be expressed as (A11).

$$AUC = N\int_{-\infty}^{\infty} \Phi\left(y+\sqrt{2}\delta\right)^M \Phi(-y)^{N-1}\phi(-y)dy. \qquad (A11)$$

Equation (A11) can also be expressed as Equation (A12).

$$AUC = N\int_{-\infty}^{\infty} \Phi\left(x+\sqrt{2}\delta\right)^M \Phi(-x)^{N-1}\phi(x)dx. \qquad (A12)$$

Appendix B: R codes used in this paper

No.	Code
1	'sfmnabba(m,n)'
2	'pmnabba(m,n,d)'
3	'qmnabba(m,n,p)'
4	'bmnabba(m,n,d)
5	'powf(p0,p1,alpha,n)'

```
#1
sfmnabba<-function(m, n)
{
#simulated psychometric functions for paired specified 'M+N' with AB and BA
###################
MNpair<-function(m,n,dd){
k<-length(dd)
#########
mnpair<-function(m,n,d)
{b <- 10000
dd1 <- rep(0, b)
for(i in 1:b) {
```

```
s <- (rnorm(n, d, sqrt(2)))
w <-(rnorm(m, -d, sqrt(2)))
if(min(s) > max(w)) {
dd1[i] <- 1
}
}
dd2 <- rep(0, b)
for(i in 1:b) {
s <- (rnorm(n, d, sqrt(2)))
w <-(rnorm(m, -d, sqrt(2)))
if(min(s) > max(w)) {
dd2[i] <- 1
}
}
dd<-c(dd1,dd2)
pc <- sum(dd)/(2*b)
pc
}
##############
pp<-dd
for(i in 1:k){pp[i]<-mnpair(m,n,dd[i])}
pp<-round(pp,4)
pp
}
################
mnptab <- function(m, n)
{
#pcdp tab
dd <- seq(0, 4, 0.1)
k <- length(dd)
```

```
pp<-MNpair(m,n,dd)
dp <- cbind(dd, pp)
dimnames(dp)[[2]] <- c("d'", "pc")
dp
}
################
tab <- mnptab(m, n)
dpf <- smooth.spline(tab[, 1], tab[, 2])
dpf
}
#2
pmnabba<-function(m,n,d){
#Pc of paired specified 'M+N' with AB and BA
pcf<-function(m,n,d){
ff<-function(x,n,m,d){mm<-pnorm(x+sqrt(2)*d)^m*(1-pnorm(x))^(n-1)*dnorm(x)
mm}
pc<-integrate(ff, -99, 99,n=n,m=m,d=d)
pc<-n*pc[[1]]
pc}
k<-length(d)
pp<-rep(0,k)
for(i in 1:k){pp[i]<-pcf(m,n,d[i])}
pp<-round(pp,4)
pp
}
#3
qmnabba<-function(m,n,pp){
#d' paired specified 'M+N' with AB and BA
pdf<-function(m,n,p){
pcdf<-function(d,m,n,p){
```

```
ff<-function(x,n,m,d){mm<-pnorm(x+sqrt(2)*d)^m*(1-pnorm(x))^(n-1)*dnorm(x)
mm}
pc<-integrate(ff, -99, 99,n=n,m=m,d=d)
pc<-n*pc[[1]]
pc-p}
delta<-uniroot(pcdf,interval=c(0,10),m=m,n=n,p=p)
del<-round(delta[[1]],2)
del}
k<-length(pp)
dd<-rep(0,k)
for(i in 1:k){dd[i]<-pdf(m,n,pp[i])}
dd<-round(dd,4)
dd
}
#4
bmnabba<-function(m,n,dd){
#for B values for paired specified 'M+N' with ab and BA
##########
pc4i2afc0<-function(m,n,d){
#Pc of 4I2AFC
pcf<-function(m,n,d){
ff<-function(x,n,m,d){mm<-pnorm(x+sqrt(2)*d)^m*(1-pnorm(x))^(n-1)*dnorm(x)
mm}
pc<-integrate(ff, -99, 99,n=n,m=m,d=d)[[1]]
pc<-n*pc
pc}
k<-length(d)
pp<-rep(0,k)
for(i in 1:k){pp[i]<-pcf(m,n,d[i])}
pp<-round(pp,4)
```

```
pp
}
#########
pp<-pc4i2afc0(m,n,dd)
bb<-dd
k<-length(dd)
############
bf<-function(m,n,d,p){
dpf<-function(x,m,n,d){
dp<-sqrt(2)*m*n*pnorm(x+sqrt(2)*d)^(m-1)*pnorm(-x)^(n-1)*dnorm(x+sqrt(2)*d)*dnorm(x)
dp
}
dpc<-integrate(dpf, -99,99,m=m,n=n,d=d)[[1]]
b<-p*(1-p)/(dpc^2)
b
}
#############
for(i in 1:k){
bb[i]<-bf(m=m,n=n,dd[i],pp[i])}
bb
}
#5
powf<-function(p0,p1,alpha,n){
#power for forced-choice test
a1<-sqrt(p0*(1-p0))*qnorm(1-alpha)
a2<-sqrt(p1*(1-p1))
a3<-abs(p1-p0)
zz<-(sqrt(n)*a3-a1)/a2
pow<-pnorm(zz)
pow<-round(pow,2)
```

pow

}

Acknowledgement

This research was supported partially by Amway.

Bibliography

1. Noreen DL. "Optimal decision rules for some common psychophysical paradigms". In S Grossberg (Ed.), Mathematical psychology and psychophysiology (Proceedings of the Symposium in Applied Mathematics of the American Mathematical Society and the Society for Industrial and Applied Mathematics, Volume 13). Providence, RI: American Mathematical Society (1981): 237-279.

2. Bi J and Kuesten C. "The four-interval, two-alternative forced-choice (4I2AFC): A powerful sensory discrimination method to detect small, directional changes particularly suitable for visual or manual evaluations". *Food Quality and Preference* 73 (2019): 202-209.

3. Bi J. "An extension of Green's area theorem". *Journal of Mathematical Psychology* 82 (2018): 131-137.

4. Bi J., *et al*. "Paired versions of various sensory discrimination forced-choice methods and the same-different area theorem". *Food Quality and Preference* 63 (2018): 97-106.

5. Green DM. "General prediction relating yes-no and forced-choice results". *Acoustical Society of America* 36.5 (1964): 1042.

6. Green DM and Swets JA. "Signal Detection Theory and Psychophysics". Wiley, New York (1966).

7. Irwin RJ., *et al*. "An area theorem for the same-different experiment". *Perception and Psychophysics* 61.4 (1999): 766-769.

8. Micheyl C and Dai H. "A general area theorem for the same-different paradigm". *Perception and Psychophysics* 70.5 (2008): 761-764.

9. Bi J., *et al*. "Estimations of Thurstonian models for various forced-choice sensory discrimination methods as a form of the 'M+N' test". *Journal of Sensory Studies* 29.5 (2014): 325-338.

10. Creelman CD and Macmillan NA. "Auditory phase and frequency discrimination: A comparison of nine procedures". *Journal of Experimental Psychology: Human Perception and Performance* 5.1 (1979): 146-156.

11. Rousseau B and Ennis DN. "A Thurstonian model for the dual pair (4IAX) discrimination method". *Perception and Psychophysics* 63.6 (2001): 1083-1090.

12. Bi J., *et al*. "How to estimate and use the variance of d' from difference tests". *Journal of Sensory Studies* 12.2 (1997): 87-104.

13. Ennis D. "The power of sensory discrimination methods". *Journal of Sensory Studies* 8.4 (1993): 353-370.

14. Ennis J and Jesionka V. "The power of sensory discrimination methods revisited". *Journal of Sensory Studies* 26.5 (2011): 371-382.

15. Bi J. "Sensory Discrimination Tests and Measurements: Sensometrics in Sensory Evaluation". 2nd Edition, Oxford, UK: Wiley/Blackwell Publishing (2015).

16. Insightful. "S-Plus 6 for Windows: Guide to Statistics, Volume 1". Insightful Corporation, Seattle, Washington (2001).

17. Bi J. “Similarity tests using forced-choice methods in terms of Thurstonian discriminal distance, d’”. *Journal of Sensory Studies* 26.2 (2011): 151-157.

18. Craven BJ. “A table of d’ for M-alternative odd-man-out forced-choice procedures”. *Perception and Psychophysics* 51.4 (1992): 379-385.

19. Ennis J., *et al*. “Thurstonian models for variants of the method of tetrads”. *British Journal of Mathematical and Statistical Psychology* 51.2 (1998): 205-215.

20. Egan JP. “Signal detection theory and ROC analysis”. New York: Academic Press (1975).

21. Owen DB. “A table of normal integrals”. *Communications in Statistics-Simulation and Computation* 9.4 (1980): 389-419.

Volume 8 Issue 11 November 2019