

A Statistical Evaluation and Modeling on the Social Transitivity Behavior

Chiao-Yi Yang^a, Frederick Kin Hing Phoa^{a,*}, Yen-Sheng Chiang^b

^a Institute of Statistical Science
Academia Sinica, Taiwan

^b Department of Sociology
Chinese University of Hong Kong, China

*Corresponding author's email: fredphoa@stat.sinica.edu.tw

ABSTRACT--- *Transitivity is one of the most important mechanisms to form a social network in a human society, but it remains unclear how such behavior is quantified and affected by some key factors, including the social environment and the participants' characters. This study investigates the sharing behavior based on the notion of transitivity and an experiment is conducted on a variety of populations from kindergarteners to teenagers. The key effects that have high impacts to the sharing behavior are identified from the statistical analysis of the experimental results. A mathematical model is built for the experimental results and its performance compared with other models is also illustrated.*

Keywords--- Social transitivity, Social Networks, Sharing Behavior, Design of Experiments

1. INTRODUCTION

Human social networks are characterized by rich variation at the individual level. Some stay in tie-loosen groups with little overlap between friends, whereas others are embedded in tightly-knit groups where everyone knows each other. Fowler et al. (2009) pointed out that most social network analyses have ignored the effect of individual characteristics. This paper focuses on the investigation of one of these individual characteristics: the Social Transitivity.

Newman and Park (2003) officially defined "transitivity" as the tendency between two nodes to be connected if they share a mutual neighbor. In general, transitivity of relations suggests a tendency to establish a new link between nodes b and c if there have been a link between nodes a and b and a link between nodes a and c . If homophily has driven the links between nodes a and b and between nodes a and c , then there would be similar attributes that would drive the connection between nodes b and c in a homophilous relation.

In this paper, we present an analysis of sociality propagation based on the notion of transitivity, which is classified as five possible outcomes, given the assumptions that subjects A and C are in the same social group (acquainted) and subject B is in another group (unacquainted): (1) Subject A will treat subjects B and C well to some extent whether subject B treats subject C well or subject C treats subject B well; (2) If subject C treats subject B badly, then subject A will also treat subject C badly to some extent; (3) If subject C treats subject B badly, then subject A will treat subject B well to some extent; (4) If subject B treats subject C badly, then subject A will treat subject B as the usual; and (5) If subject B treats subject C badly, then subject A will treat subject C well to some extent.

We conduct an experiment on behavior of sharing to a variety of populations from kindergarteners to teenagers. It aims at investigating when a person starts to develop a sense of sociality that would take into account the third party when evaluating his/her relationship with an altar. The study attempts to investigate the development of a person's sociality, or transitivity in specific, in his/her life course. The functioning of transitivity moves one's perception of the social world from a dyad to a triad, which is a significant transition to the maturity of sociality. The discrepancy between a dyad and a triad in social relationships is also seen as a distinction of research focus between psychology and sociology.

2. EXPERIMENTAL SETUP

The experiment is similar to the design by Fehr et al. (2008). Subjects are given items, which would be anything that appears valuable to the subjects in question, to decide whether to share with others. For this study, the items are the stickers and the subjects range from kindergarteners to middle-school students. Subjects are facing two different kinds of altars: those in the same class and in the different class. They decide on how many of the items in their hands to be

shared with altars. The major part of the experiment lies in the second stage, where they will face the same decision-making as before, but this time they are told how many items others, whom they care or not, have given to altars. An illustration is given below. Assume that there are three subjects i, j and k in the experiment. i originally proposed $d_{i,j,0}$ and $d_{i,k,0}$ to j and k respectively. Upon knowing that k gave s stickers to j , i is invited to reconsider his offer to j and k , denoted as $d_{i,j,1}$ and $d_{i,k,1}$. The comparison on the difference between $d_{i,j,0}$ and $d_{i,j,1}$ leads to an understanding if the action of i to j is influenced by how k treats j and how much i cares about the decision of k . i, j and k could be either in the same class or in a different class. The experiments are conducted in a number of different scenarios with and without the information about s .

All experiments are anonymous. Subjects know that they would play the game with an altar randomly chosen from the same class or a different class, but they do not know altar's real identity. Every subject is required to participate in the original scenario (where the subject has no information about the interaction between other participants and altars) and four pre-designed scenarios. In each scenario every subject has ten stickers to share with altars. Subjects are expecting to play the game for real. They will get whatever amount they have after their sharing decision.

Experimental Setup. There are 314 subjects in this experiment, comprising 118 kindergarteners, 89 elementary school students (in the fourth and sixth grades) and 107 junior high school students (all are eighth grades). 160 out of 314 participants are female. The number of observations is $N = 3140$ due to five scenarios and two kinds of recipients.

Research Questions. The research question is that how many stickers a subject will share with an altar in five different scenarios. The basic one is no information about the third party provided to the subject. The other four different scenarios considered are described as follows: (1) The subject has been informed at the beginning of the experiment that his/her classmate gave a generous offer (eight stickers) to a student in a different class; (2) The subject has been informed that his/her classmate behaved less generously (two stickers) towards a student in a different class; (3) The subject has been informed that a student in a different class has generously given his/her classmate eight stickers; and (4) The subject has been informed that a student in a different class has less generously given his/her classmate two stickers. Note that in each scenario each subject is required to make two sharing decisions toward each of two kinds of altars, who are in the same class and in the different class.

3. TESTS OF STATISTICAL SIGNIFICANCE ON EXPERIMENTAL EFFECTS

The response variable in this experiment is the number of stickers, ranging from 0 to 10, which a participant (subject, giver) is willing to share with a recipient (altar). A larger value of the response variable represents the more generous the participant is. Five factors to be studied in this investigation are listed as follows: (1) participants' gender (we name the factor "Gender"); (2) participants' school class (we name the factor "School"), ranging from kindergarten to junior high school; (3) recipients' class (we name the factor "Target"): the same class and other class; (4) information categories (we name the factor "Information"), including no information gained, the information about classmates' behavior toward other class students and the opposite information; and (5) the third parties' behavior (we name the factor "Behavior"), including unknown, generous and less generous. The details of them are tabulated as Table 1. For example, the vector $(G, H, T, I, B) = (0, 2, 1, 2, 1)$ stands for the case where a junior high school girl is asked how many stickers she will share with a student in the different class when she knows that a student in the different class treats her classmate well to some extent.

Table 1: Factors and Levels

Factor	Notation	Level		
		0	1	2
Gender	G	Female	Male	
School	H	Kindergarten	Elementary School	Junior High School
Target	T	Same Class	Other Class	
Information	I	No Info	Same to Other	Other to Same
Behavior	B	No Info	Generous	Less Generous

A preliminary study using descriptive and visual statistics. We plot the histograms of the observations at each factor level to view the central tendency and variability (or dispersion) of the experimental data and to roughly check if the data is normally distributed. Figure 1 shows the histograms of the responses at each factor level. These plots reflect that the data follows a (slightly) right-tailed skewed distribution in most of cases and has a single mode at five. All plots have an increase at the end of right tail (ten) except for the cases of $G = 0$ and $H = 2$.

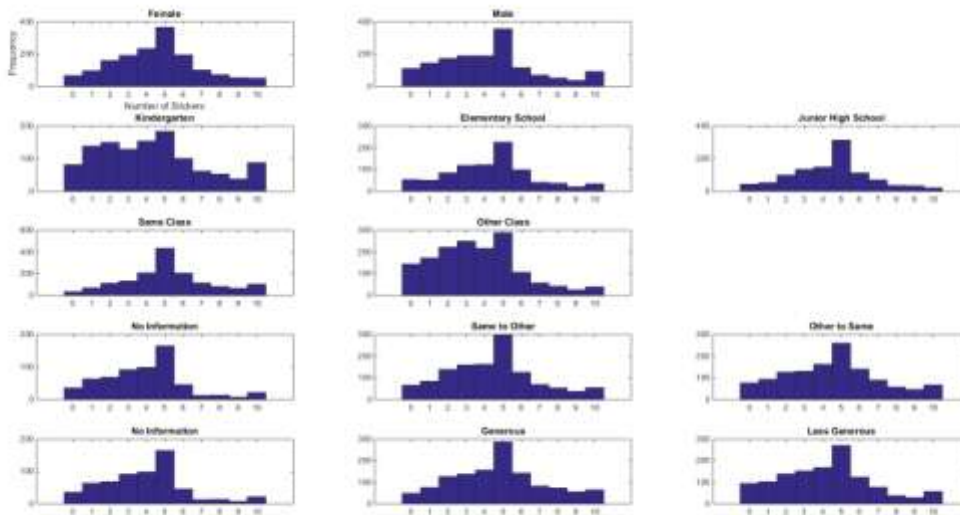


Figure 1. Histograms for all factor levels

The plots in the second row of Figure 1 reveal that preschool children's sharing behavior is remarkably different from the middle school students. The plots in the third row show that children's sharing behavior heavily depends on the relationship between them and the recipients. It implies that the clique structure have been developed on childhood. From the plots in the last two rows of Figure 1, we find that children are more willing to share when the information about third parties is available whether the message is positive or negative. We infer from Figure 1 that these five factors have significant effects on the response.

Histograms are used to visually examine the normality of response and the significance of factors. To provide more sufficient evidence of non-normality, Kolmogorov-Smirnov test is conducted under the null hypothesis that the response follows a standard normal distribution. The results of Kolmogorov-Smirnov test at each factor level all indicate that the response does not come from a standard normal distribution with $p\text{-value} < 0.0001$. In our experiment, the response has multiple categories. It's evident that our response is not normally distributed, more like multi-nominally distributed. Moreover, the response is ordinal (or ordered). Hence, we will use a multi-nominal and ordinal logistic regression model to fit the observed data in the next section and further investigate which factors have significant effects on the response.

Next, we employ a graphical tool, the half-normal plot, to judge which factors have significant effects on the response. A half-normal plot is formed by the ordered absolute values of the estimated effects for the factors of interest and their available interactions on the vertical axis and the corresponding coordinates on the half-normal probability scale on the horizontal axis. A half-normal distribution is the distribution of $|X|$, where X has a normal distribution. Note that here we only have an interest in main effects and two-factor interactions. The procedure for assessing which factors are important and which factors are unimportant consists of the following: identifying the line of near-zero factors; then arguing the factors on or close to the near-zero line as unimportant; finally declaring the factors on the remaining off-line factors as important. Figure 2 displays the half-normal plots of the main and two-factor interaction effects from the mean, medians and log variance points of view. The notation " $G:T$ " stands for the interaction of factors G and T , namely $G \times T$. Figure 2 suggests that factor T has significant location effects most likely and that G and H are the most influential factors from the dispersion point of view. To further confirm these findings described here, several formal tests for factorial effects will be introduced in the next subsection.

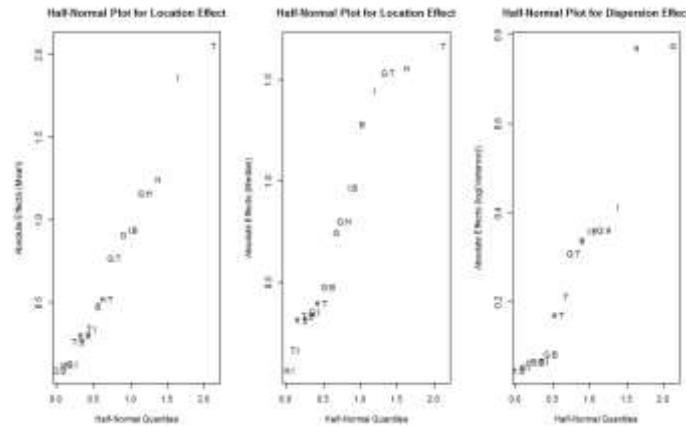


Figure 2. Half-normal plots of location and dispersion effects

Statistical non-parametric tests on factorial effects. The half-normal plot is an informal graphical tool involving visual judgment. We present the Lenth's method that generally performs well for testing the significance of the factors of interest. The null hypothesis of Lenth's test considered all main effects and two-factor interactions as zero. Table 2 summarizes the results of Lenth's test of the location and dispersion models at the significance level 0.05. It reveals that the mean effects of *T* and *I* are significant and *G* and *H* have significant influences on the response in the dispersion model. These results provide a quantitative basis for the informal findings obtained based on the half-normal plots in Section 3.1.

Location/Dispersion Model	Significant Factors (p-value)
Mean	<i>T</i> (0.0276), <i>I</i> (0.0366)
Median	NONE
Log(Variance)	<i>G</i> (0.0339), <i>B</i> (0.0348)

The descriptive and inferential statistical procedures used in Section 3.1 have shown that the response at each factor level cannot be assumed to come from a normal distribution. Hence, we investigate which factors have significant impacts on the response via non-parametric procedures. Table 3 summarizes the results of Kruskal-Wallis one-way analysis of variance (ANOVA), which tests if all treatment levels are equal for each factor by examining whether the distributions of the rank vectors of all treatment levels are the same. The results show that the rank distributions of the responses at different factor levels are significantly different. Moreover, box plots in Figure 3 also visually present the significant difference among all treatment levels for each factor. Figures 3(b) and 3(d) show that the ranks of samples at zero and non-zero levels distribute differently. We transfer all three-level factors "School" (*H*), "Information" (*I*) and "Behavior" (*B*) into two levels by combining their non-zero levels.

Main Effect	Gender	School	Target	Information	Behavior
p-value	0.0000	0.0020	0.0000	0.0000	0.0000

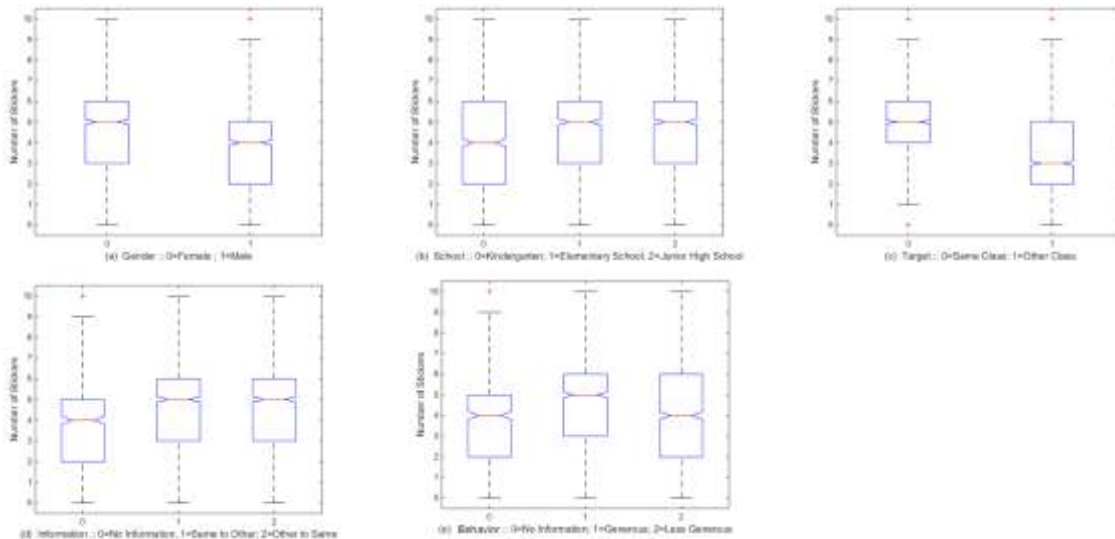


Figure 3: Box Plots for Each Factor

Wilcoxon rank sum test and signed rank test are used to investigate if the two treatment levels of each factor have different effects on the response. The main difference between these two tests is that Wilcoxon signed rank test is used when comparing two related samples, that is, it is used under the assumption that data are paired and come from the same population. This study employs Wilcoxon signed rank test to identify whether the effects of T , I and B are significant and examine the significance of G and H via the Wilcoxon rank sum test. Table 4 shows that all factors are significant. These results are consistent with those in Table 3. Further explanations of the results in Table 4 are given as follows: (1) Female are significantly more generous than male (see Test 1); (2) Preschool children tend to keep more number of stickers than the others. This result is reasonable because stickers are more attractive to preschool children (see Test 2); (3) People treat the person more generously with whom they have a certain connection. It is also an understandable result (see Test 3); (4) A person tends to treat other people more generous in the situation when he/she knows the third parties interaction in advance regardless of whether the interaction is good or not. It reveals that a person's competitive mentality orientates him/her to be more willing to express his/her generosity compared to the situation without information provided (see Tests 4 and 5).

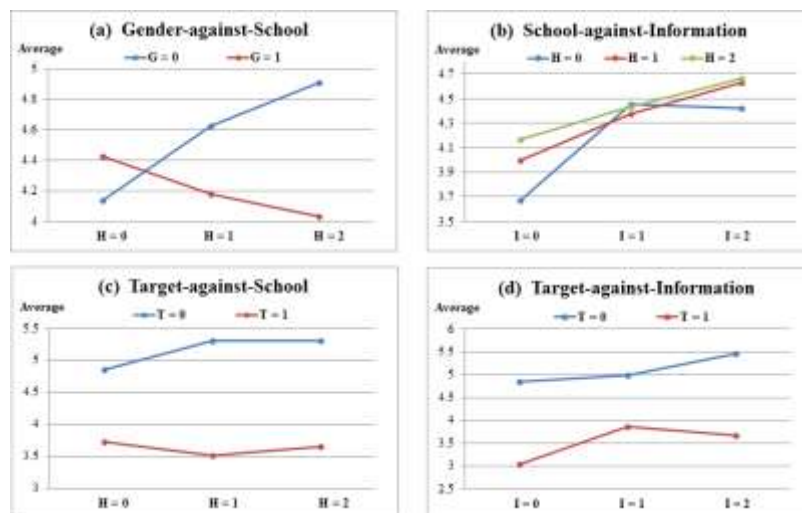


Figure 4: Interaction Plots

Over the past several decades, many non-parametric methods have been proposed for testing whether or not the joint effect of two factors exists (Sawilowsky 1990). We here investigate the interaction effect between two factors, say Y and Z , by comparing the averages of the response over all level combinations of Y and Z . The interaction plots can display the interaction calculations graphically by using the Z factor level as the horizontal axis and the average as the vertical

axis, where the averages having the same level of Y are joined by a line. In this experiment, ten kinds of interactions are derived from the five factors. Four possibly significant interaction effects are shown in Figure \ref{Interaction plots}. The (a) plot of Figure 4 shows that the difference in the degree of generosity between gender increases with children's school grades. Girls are willing to show generosity as they get order, but boys have the opposite sign. The (b) plot presents that all participants are willing to share when more information about other people's interactions is provided. The degrees of generosity of preschool children and middle school students are related to the source of information. One finding from the (c) plot in Figure 4 is that children share less with the one from other class than with their classmate, especially for the middle school students, that is, middle school students show much more partiality toward those who study in the same class as them than those in other class, compared to kindergarten children. The final plot, (d), shows that providing information about the third parties' interaction increases students' willingness to share whether the recipient is their classmate or not. It implies that children's sharing behavior are related to the combination of the information source and the receiving target.

4. A MATHEMATICAL MODEL FORMULATION, ESTIMATION AND MODEL COMPARISON

In order to find out under what sort of conditions the students are most willing to share, we build a connection between our factors of interest (G, H, T, I and B) and the response based on generalized linear models.

Main and interaction effects re-parameterization. All experimental factors are qualitative. The results in the previous analysis methods bring us to being interested in the comparison of sharing behavior between female ($G = 0$) and male ($G = 1$), preschool children ($H = 0$) and middle school students ($H \neq 0$), acquainted recipient and unacquainted one, the cases with and without information provided, and the comparison between the cases with and without generosity. To this end, the linear contrasts $(-1, 1)$ and $(-1, 0, 1)$ are naturally used to represent the comparisons between two levels of the two-level factor and levels 0 and 2 of the three-level factor, respectively. We choose the following standardized contrasts to represent the degrees of freedom for the main effects of each qualitative factor: $X_G = (-1, 1)/\sqrt{2}$ for level 0 and 1 of factor G ; $X_{H,01} = (-1, 1, 0)/\sqrt{2}$ for level 0 and 1 of factor H and $X_{H,02} = (-1, 0, 1)/\sqrt{2}$ for level 0 and 2 of factor H ; $X_T = (-1, 1)/\sqrt{2}$ for level 0 and 1 of factor T ; $X_{I,01} = (-1, 1, 0)/\sqrt{2}$ for level 0 and 1 of factor I and $X_{I,02} = (-1, 0, 1)/\sqrt{2}$ for level 0 and 2 of factor I ; $X_{B,01} = (-1, 1, 0)/\sqrt{2}$ for level 0 and 1 of factor B and $X_{B,02} = (0, -1, 1)/\sqrt{2}$ for level 1 and 2 of factor B . The interaction between two three-level factors is decomposed into four effects. Due to the difficulty of interpreting high-order interactions, we only consider the main effects of the five factors and the $\$25\$$ products between the five groups ($X_G, \{X_{H,01}, X_{H,02}\}, X_T, \{X_{I,01}, X_{I,02}\}, \{X_{B,01}, X_{B,02}\}$) of contrasts for the two-factor interactions.

Multinomial logistic regression model. The response of this experiment is the number of stickers, ranging from 0 to 10. A larger value of the response implies the subject is more willing to share. It is exactly an ordinal variable. We use the multinomial and ordinal logistic regression model (MOLR) to fit the response and expect the predictor model to be used to forecast children's sharing behavior when the social environment and the subject's characters are provided. The model M_i is suggested as

$$\begin{aligned} \ln\left(\frac{P(s < i)}{P(s > i)}\right) &= \beta_0^{(i)} + \beta_1^{(i)} X_G + \beta_2^{(i)} X_{H,01} + \beta_3^{(i)} X_{H,02} + \beta_4^{(i)} X_T + \beta_5^{(i)} X_{I,01} + \beta_6^{(i)} X_{I,02} \\ &+ \beta_7^{(i)} X_{B,01} + \beta_8^{(i)} X_{B,02} + \beta_9^{(i)} X_G X_{H,01} + \dots + \beta_{15}^{(i)} X_G X_{B,02} + \beta_{16}^{(i)} X_{H,01} X_T + \dots \\ &+ \beta_{20}^{(i)} X_{H,01} X_{B,12} + \beta_{21}^{(i)} X_{H,02} X_T + \dots + \beta_{33}^{(i)} X_{I,02} X_{B,12}, \text{ where } i = 0, 1, \dots, 9, \end{aligned} \quad (1)$$

where s is the number of stickers, $X_G, X_{H,01}, X_{H,02}, X_T, X_{I,01}, X_{I,02}, X_{B,01}$ and $X_{B,12}$ are defined before. Based on the model coefficient estimates $\hat{\beta}^{(i)} = [\hat{\beta}_0^{(i)}, \hat{\beta}_1^{(i)}, \dots, \hat{\beta}_{33}^{(i)}]^T$ for $i = 0, 1, \dots, 9$, the estimates of the probability of the number of stickers, s , are calculated via

$$\hat{P}(s = 0) = \frac{e^{X^T \hat{\beta}^{(0)}}}{1 + e^{X^T \hat{\beta}^{(0)}}}, \hat{P}(s = j) = \frac{e^{X^T \hat{\beta}^{(j)}}}{1 + e^{X^T \hat{\beta}^{(j)}}} - \frac{e^{X^T \hat{\beta}^{(j-1)}}}{1 + e^{X^T \hat{\beta}^{(j-1)}}}, \hat{P}(s = 10) = 1 - \frac{e^{X^T \hat{\beta}^{(9)}}}{1 + e^{X^T \hat{\beta}^{(9)}}} \quad (2)$$

where $X^T = (1, X_G, X_{H,01}, \dots, X_{B,12})$ is the collection of the contrasts constructed based on the degrees of freedom of the main effects and their paired interactions, and $\hat{\beta}^{(j)}$ is the vector of the coefficient estimates corresponding to the model M_i . Table 5 tabulates the significant coefficient estimates at the 0.05 level and their corresponding t statistics for each

model shown in Equation (1). It indicates that the factors G, H and T have significant impacts on the response, and the two-factor interactions $G \times H$ and $T \times H$ are also significant.

Table 5: Multinomial Logistic Regression Model

	M_0	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8	M_9
Intercept	-3.40 (-22.64)	-2.24 (-29.81)	-1.37 (-26.70)	-0.67 (-16.14)		1.10 (24.73)	1.72 (31.83)	2.23 (33.06)	2.78 (31.68)	3.40 (27.98)
X_G	0.54 (4.27)	0.55 (6.32)	0.40 (6.09)	0.29 (5.21)	0.18 (3.35)	0.24 (3.98)				
$X_{H,01}$	-0.89 (-2.37)	-0.40 (-2.43)	-0.26 (-2.41)							
$X_{H,02}$		-0.61 (-3.78)	-0.53 (-4.89)	-0.40 (-4.65)	-0.28 (-3.59)	0.22 (2.36)	0.30 (2.67)	0.55 (3.75)	0.58 (3.05)	1.01 (3.61)
X_T	1.40 (6.89)	1.07 (10.77)	0.96 (13.60)	0.97 (16.56)	0.89 (16.32)	0.76 (12.32)	0.70 (9.47)	0.73 (7.98)	0.79 (6.75)	0.80 (4.88)
$X_G X_{H,01}$	0.94 (3.36)	0.69 (3.60)	0.38 (2.77)							
$X_G X_{H,02}$				0.41 (3.59)	0.39 (3.64)	0.53 (4.43)	0.76 (5.12)	0.92 (4.90)	1.10 (94.58)	0.90 (2.68)
$X_T X_{H,01}$	1.41 (2.81)				0.39 (3.42)					
$X_T X_{H,02}$			0.37 (2.48)	0.40 (3.29)	0.25 (2.29)	0.34 (2.73)				

Number in parentheses represent the t-statistics.

We further discuss about the results shown in Table 5. First, the "Gender" factor has a significant effect on the probability of $s = 1, \dots, 6$, but becomes insignificant in estimating the probability of s larger than and equal to seven. It also shows that behaving generously toward others is harder for boys than for girls. Second, the sign and magnitude of the coefficient estimates of the contrasts of the "School" factor present that the preschool children tend to give the number of stickers less than 5, but are more willing to show most generous (10 stickers) toward others, compared to the middle school students. Thirdly, the "Target" is a very important factor in all models, M_0, \dots, M_9 and the coefficient estimates show that children behave with greater generosity toward their classmates than other children from other class.

Based on Equation (2), we obtain the estimates of the probability of the number of stickers in each of the 60 combinations of five factors considered. To further demonstrate the prediction performance of the proposed model, we compare its performance with two models commonly used in Sociology. One is the linear regression model only consisting of the main effects. We call it LR₁. The other is the linear regression model constructed by the five factors and their paired interactions. We call it LR₂. We compare the prediction performance of the three models by the following two measures. One is the correct proportion of the estimation in the number of stickers over all treatment combinations, defined as $M_1 = \frac{1}{b} \sum_{j=1}^b I[\hat{s}_j - s_{j,obs} = 0]$, where $b = 60$ is the number of level combinations of the five factors, I is an indicator function, and $s_{j,obs}$ is the observed number of stickers with the highest frequency in the j th treatment. In MOLR, the estimate \hat{s}_j of the number of stickers in the j th treatment satisfies $\hat{s}_j = \arg \max_{i \in [0,10]} \hat{P}(s = i)$. In LR₁ and LR₂, we round off the forecast of the response after the decimal into an integer as \hat{s}_j . The other measure is the mean squared error (MSE) of the estimation in the number of stickers, defined as $M_2 = \frac{1}{b} \sum_{j=1}^b (\hat{s}_j - \bar{s}_{j,obs})^2$, where $\bar{s}_{j,obs}$ is the average of the number of stickers over all replications in the j th treatment. Here \hat{s}_j in MOLR is redefined as $\hat{s}_j = \sum_{i=0}^{10} i \hat{P}(s = i)$ with the X in (2)-(4) is the vector of contrast corresponding to the j th treatment.

Table 6: Model Comparison

Model	Significant Variables	M_1 (%)	M_2
MOLR	$X_G, X_{H,01}, X_{H,02}, X_T, X_G X_{H,01}, X_G X_{H,02}, X_T X_{H,01}, X_T X_{H,02}$	70.00	0.3013
LR ₁	$X_G, X_T, X_{I,01}, X_{I,02}, X_{B,12}$	16.67	1.3809
LR ₂	$X_G, X_T, X_{I,02}, X_{B,12}, X_G X_{H,02}, X_G X_T, X_T X_{H,01}, X_T X_{I,01}$	16.67	1.3444

The comparison of three models shown in Table 6 strongly indicates that the MOLR model is more appropriate to explain the response of this experiment. The M_1 and M_2 in MOLR are respectively noticeably larger and smaller than LR₁ and LR₂.

5. PRACTICAL APPLICATION

Previous studies show that the children's gender, school class and their relationship with the recipient greatly influence their sharing behavior. We then discuss how the children's characters impact on their sharing behavior. A practical study is conducted to demonstrate the utility of our model in (1). We suggest that what we are interested in are to find the level combination of the three significant factors under which the largest average number of stickers is given and to discover the difference in sharing behavior between children of different ages.

Fortunately, children in the same grade are at the same age in this experiment. The kindergarten children are at the age of about five. The fourth- (sixth-) grade elementary school students are at the age of nine (eleven). The junior high school students are at the age of thirteen. Hence, the factor "School" is renamed to "Age", denoted by A . The A factor has four levels. Due to the quantitiveness of A , we use its original scale to investigate its effect. The value of the A factor is denoted by X_A , where $X_A = 5, 9, 11, 13$ for the four levels. The X_A is used to represent the linear effect of A and X_A^2 to represent its quadratic effect. Then, the model M_i in (1) is modified as follows:

$$\ln\left(\frac{P(s < i)}{P(s > i)}\right) = \beta_0^{(i)} + \beta_1^{(i)}X_G + \beta_2^{(i)}X_A + \beta_3^{(i)}X_A^2 + \beta_4^{(i)}X_T + \beta_5^{(i)}X_{I,01} + \beta_6^{(i)}X_{I,12} + \beta_7^{(i)}X_{B,01} + \beta_8^{(i)}X_{B,02} + \beta_9^{(i)}X_GX_A + \beta_{10}^{(i)}X_GX_A^2 + \dots + \beta_{33}^{(i)}X_{I,02}X_{B,12}, \quad (5)$$

where $i = 0, 1, \dots, 9$. Under the level of 0.05, the coefficient estimators related to the variables $X_G, X_A, X_A^2, X_T, X_GX_A, X_GX_A^2, X_TX_A, X_TX_A^2$ are significant. We construct the model for children's behavior based on these significant variables. Then the predictive probabilities of s are calculated for children aged five to thirteen years, which are the values of level 0-8. For simplicity, we examine the relation between children's characters and the degree of their generosity, categorized as Less Generous ($0 \leq s \leq 5$), Generous ($s = 6$ or 7) and More Generous ($8 \leq s \leq 10$). The probabilities of these three categories in each level combination of factors are displayed diagrammatically in four different forms in Figure 5. The level combinations corresponding to the plots in Figure 5 are listed in the appendix.

We provide a summary of the prediction results. A noticeably conclusion is that the thirteen-aged girl are most willing to share with others in this experiment. All plots of Figure 5 indicate that female children are generally more generous than male at each of ages from five to thirteen whether the recipient of their stickers is their classmate or not. The upper two plots in Figure 5 show that the degree of female generosity have a concave (curved) shape in age and the degree of male generosity decreases with age whether toward their classmates or others in the different class. The upper-right plot also shows that the difference between the degrees of generosity toward classmates and others increases with age whether the subject is female or male. It presents from the bottom-left plot that at least 23.8% of children are willing to share

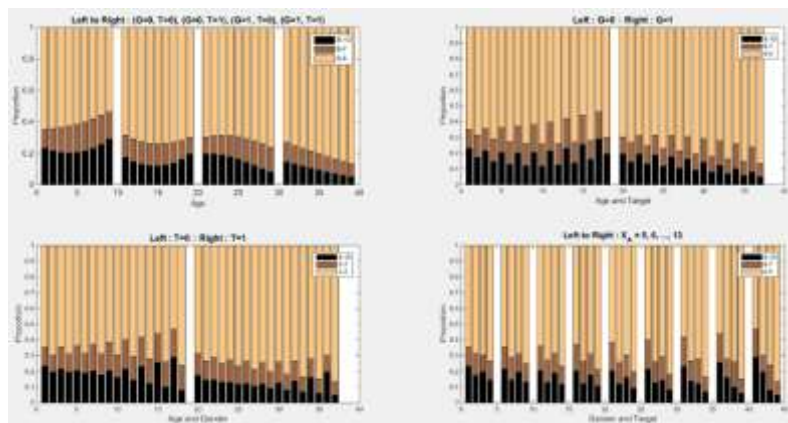


Figure 5: Prediction in the Degree of Children's Generosity.

over five stickers with their classmates in each combination of G and A , but at most 31.3% of children give over five stickers to others in all combinations. The bottom-left plot also shows that the difference between the degrees of female generosity and male generosity increases with age whether toward classmates or others. It indicates from the bottom-right plot in Figure 5 that the variation within four combinations of G and T becomes greater with age. It indirectly shows that children are easy to express their liking and disliking as they get older.

6. SUMMARY

The transitivity behavior is an interesting topic in sociology, behavioral study and social network analysis. In this paper, several statistical testing procedures are used to examine the distribution of data, investigate in the importance of factors and detect the effects between different levels of each factor. The MOLR model is proposed to show that children's gender, school class (or age) and their relationship with the recipients are the important factors in the study of

children's sharing and transitivity behavior. The accuracy of the estimation in the level of children's generosity is extremely higher than two commonly-used models in sociology and the estimation error is greatly less than the other twos. The great significance of factor "Target" is comprehensible, like the proverb, Love me, love my dog. Although this study only analyzes the dataset discussing children's sharing and transitivity behavior, the proposed model can be used in the data with the categorical response. Moreover, the model estimation and prediction will be promising than most of common methods.

7. ACKNOWLEDGEMENT

The authors would like to thank Dr. Yuan-Lung Lin for his constructive suggestions and comments to the improvement of this paper. This work was supported by (a) Career Development Award of Academia Sinica (Taiwan) grant number 103-CDA-M04; (b) Ministry of Science and Technology (Taiwan) grant numbers 102-2628-M-001-002-MY3 and 105-2118-M-001-007-MY2, and (c) Thematic Research Program of Academia Sinica (Taiwan) grant number AS-103-TP-C03.

8. REFERENCES

- Fehr, E., Bernhard, H. and Rockenbach, B. (2008). Egalitarianism in young children. *Nature*, 454, 1079-1083.
- Fowler, J.H., Dawes, C.T. and Christakis, N.A. (2009). Model of genetic variation in human social networks, *Proceedings of the National Academy of Sciences*, 106, 1720-1724.
- Newman, M.E. and Park, J. (2003). Why social networks are different from other types of networks, *Physical Review E*, 68, 036122.
- Sawilowsky, S. S. (1990). Nonparametric tests of interaction in experimental design. *Review of Educational Research*, 60(1), 91-126.