DECISIONS FOR THE SELF VS. DECISIONS FOR OTHERS: NATURE AND SOURCE OF BIASES

By Vedika Sakhardande, Aditi Gupta

Abstract

This research examines the differences in decision-making when made for self versus others, in the domain of career paths. With this study, we focus on studying the choices made by students across three dimensions: a) Choices made for the self b) Choices made for classmates c) Choices made by the average college students. We intend to answer the following questions: 1) Whether there exists any difference in choices made for the self and others 2) Does 'familiarity' play a role in affecting self-other differences 3) How gender affects self-other differences 4) Identify the proximate cause for self-other differences. We found that no self-other difference existed for peers or the average college student. Moreover, gender as a variable did not impact the risk-seeking tendencies of participants both within and across categories. We also see that the attributes considered when making choice for the self coincide with the attributes considered when making choice for the other.

Keywords: decision-making, risk preference, self-other difference

Introduction

As a flourishing domain, behavioural economics has laid the groundwork for revealing the biases and the irrationalities in human behaviour. The work of Thaler, Kahneman and Tversky has revealed how emotions and circumstances impact our presumably 'rational' choices and the real-world consequences this leads to. However, a relatively unexplored dimension of human decision-making lies in the study of how emotions and circumstances affect outcomes when we make choices for others. When the subject facing the consequences changes, are our decisions more or less biased? Given that these decisions made by one for the other are highly prominent not just in our day-to-day interactions but also a key aspect of our liberal democracies, understanding the drivers and biases of these choices holds tantamount importance. Example: Would a financial planner take the same actions they recommend to their client? Would people follow the same dating advice they give their friends? In this study, we attempt to study the differences in choices made for the self and those for others in the domain of career choices. The rationale behind this domain selection is that our sample population—undergraduate students—makes decisions in this domain frequently and are likely to treat the self and other conditions as unique than generalise responses when answering for others.

Beyond this, we also investigate how the "closeness" of the other to the respondent impacts choices by considering two levels of others—classmates and an average college student. Our analysis is complete with the involvement of a gender variable which we use to gain insights into the role of norms and socialization in choice outcomes. Through a self-diagnostic survey, we also attempt to locate the difference (if any) in the factors driving self-other differences.

Literature Review

In contradiction to the classical expected utility theory, Tversky and Kahneman found an alternate model for decision under risk developed which is called the 'prospect theory'. (Kahneman, Daniel and Tversky, Amos, 1979) It states that people don't prefer outcomes that are merely probable in comparison with outcomes that are obtained with certainty. This 'certainty effect' contributes to risk aversion in choices involving sure gains and to risk seeking in choices involving sure losses.

However, a further branch of decision theory seeks to study whether the choices made for the self are qualitatively and quantitatively similar to the decisions made for others. This type of decision-making is highly pervasive in our society and is observed at all levels: from the personal level—involving financial advisors creating investment plans for their clients or doctors taking decisions for their patients, to the social level—involving politicians rolling out policies for the masses.

This has resulted in studies conducted across various domains—the monetary decision domain (Stone et al, 2006), political decision-making (Hibbing and Alford, 2005), medical decision-making (Garcia-Retamero and Galesic, 2012) romantic decision-making (Beisswanger et al., 2010), job selection (Kray and Gonzalez, 1999) etc with each domain producing varying results about a) the existence of a self-other difference b) the extent of the self-other difference.

The above-mentioned studies, however, are further subdivided based on their definition of the 'other' as well as their hypothesis on the potential source of self-other difference. According to Hoch (1987), subjects made to predict the attitudes of one of three target populations: an average married American consumer; their average graduate peer or their spouse made more accurate predictions for their spouses and peers than the average American consumer. This established that the more the subjects are familiar with the 'other' they make assumptions for, the more the assumptions align with their own values and opinions. This is further reiterated in the work of Hsee and Weber (1997).

The self-other difference is also owed to various factors. Kray and Gonzalez (1999) establish that self-other difference occurs due to the difference in weighing of attributes assigned to a decision as well as the consideration of more attributes when deciding for the self. Beisswanger et al. (2010) answer the question—exactly which attributes are ignored when people make decisions for others? And find that 'negative consequences' of choices are considered more stringently when making choices for the self as opposed to choice-making for the other. Stone et al. (2006) consider the role of regret in establishing a self-other difference. Hsee and Weber (1997) study the role of both the hypothesis of 'risk-as-value' which could lead people to believe they are more risk-seeking than others and the 'risk-as-feelings' which leads people to predict other's risk preference based on stereotypes of the group to which the person belongs. Larrick (1993) stated that self-other decision-making discrepancies could emerge due to different strategies for self-image protection.

A key delineation in all of the above-mentioned studies is twofold: a) Whether they can demonstrate a statistically significant self-other difference and b) the degree to which this difference is observed. The respective conclusions are domain-specific and depend on the sample studied. Hsee and Weber (1997) found that participants systematically predicted others to be more risk-seeking than themselves with participants however only predicting abstract others to be more risk-seeking and concrete others to not be so. Beisswanger et al. (2010) found that people were more risk-taking when making decisions for their friends however this difference existed only for circumstances involving a low life impact and vanished for circumstances involving a high life impact. Stone et al. (2006) however found that individual decision-making regarding financial situations can be generalised to decision-making for others and no self-other difference in decision-making was observed.

Study

The primary purpose of our study was to examine whether there exists any difference in self-other risk preference in the domain of career choices among undergraduate students. To accomplish this goal, we surveyed people and asked them to make career choices for either a) themselves, b) their classmates, or c) an average college student. With this design, we wanted to analyse the degree of difference between choices made for these categories. In order to locate the proximate source of disparity we asked the participants to conduct a self-diagnostic test which required them to indicate the factors which led to their choices from the following list of 5: a) Social norms regarding favourable career choices, b) Search for security, c) Search for

higher returns, d) Perceived dangers associated with the option not selected, and e) How options aligned with goals of the subject you were choosing for. These factors summarise the key factors studied in the papers stated in the literature as the major source of self-other difference. Gender as a variable was used to derive intersectional insights about the data.

Methodology

Participants. Participants were 61 females and 35 males aged 18-22 years who were undergraduate students at colleges across India.

Design. The study involved a 3 (Group: self, other-familiar, other-abstract) x 2 (Gender: male vs female) between-subjects design. The dependent variable was the participants' level of risk-seeking. In order to avoid people responding similarly for themselves and others a between-subjects design was employed to eliminate the possibility of a response bias.

Materials. The questionnaire used in this study had 5 scenarios regarding a choice between career paths. This domain was selected so that respondents would give a close approximation of their own opinions and instincts as opposed to them estimating their preferences. Two options followed each scenario with one being risk-seeking (eg: Investing personal funds to establish a venture that has an 8.5% chance of earning Rs. 235 lakhs post the 1 year required to establish the venture) and one being risk-averse (eg: investing in a full-time masters' degree that has a graduate earning a record of a starting salary of Rs. 20 lakhs on average). The expected outcome of both scenarios was the same to enable participants to make choices on the basis of their risk preference/their perception of others' risk preference solely. Participants received a different version of the questionnaire depending on whether they were to make choices for themselves, other classmates or for an average college student. The survey also contained a diagnostic section to enable respondents to indicate the core motivations behind their choices. This section intended to clue us into the main drivers of decisions and identify potential sources of difference.

Procedure. Participants received one of the versions of the survey based on a randomised number selection on Google forms. They were provided with instructions on answering the form and urged to rely on their intuition. The results were analysed using one-way ANOVA using the R platform with subsidiary analysis performed using Excel.

Limitations. The study was a hypothetical at-home study which generated the following drawbacks in analysis a) The lack of consequentialism to the choices could lead to data that doesn't reflect real-world instincts, and b) Participants may have responded similarly to the three categories due to an inability to differentiate the role they played in each circumstance. Lastly, the discrete nature and limited size of the data may have interfered with the prediction power of the tests conducted.

Results

In the questionnaire, each question contained a risk-seeking and risk-averse option with the risk-seeking answer attaining a score of 1 and the risk-averse answer attaining a score of 0. An overall score (participant risk score) was given to each participant by summing up the scores of the 5 scenarios. TA score of 5 indicated perfect risk seeking while a score of 0 indicated perfect risk aversion.

Within the 97 responses overall, each category received > 30 responses. To carry out a fair comparison of the means, we checked our data for normality and homogeneity of variance (see Appendix for results of tests).

The data was analysed for differences across categories: self, peer and stranger and across gender: female and male.

A) Category-wise risk score comparison

To evaluate the differences in risk scores across categories, we carried out different types of tests in R. The gender of the respondents was not a variable under consideration for this analysis.

The null and alternative hypotheses for the following tests were:

H_0*: No difference among the means of the categories.*

H_1*: At least one category has a different mean.*

i) One-way ANOVA test

```
               Df Sum Sq Mean Sq F value Pr(>F)
study$Category  2   7.16   3.580   2.529 0.0852 .
Residuals      94 133.09   1.416
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The p-value = 0.0852 > 0.05 implying that we fail to reject the null hypothesis at 5% significance.

∴ There is no significant difference between the categories according to this test.

We can conclude that there exist no self-other differences both when decisions are made for "familiar" peers and "abstract" others.

ii) Tukey HSD Post hoc test

Next we conducted the Tukey test to run a pairwise comparison of the means to find any difference among each of the categories.

```
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = study$Risk.score ~ study$Category, data =
 study)

$`study$Category`
                    diff        lwr       upr     p adj
Self-Peer     -0.5151515 -1.2127326 0.1824295 0.1892398
Stranger-Peer  0.1055718 -0.6031712 0.8143149 0.9330432
Stranger-Self  0.6207234 -0.0880197 1.3294664 0.0982939
```

The p-values > 0.05 across all category comparisons and thus we fail to reject the null hypothesis.

∴ There is no significant difference between the categories: self-peer, stranger-peer and stranger-self

Note 1: At a 10% significance level, the difference between the stranger and self categories becomes statistically significant.

Note 2: The high p-value between Stranger and Peer implies that the responses for these categories exhibited the greatest homogeneity implying people had a nearly congruent perception of risk for both peers and the average college student.

iii) Kruskal-Wallis Rank Sum test

Taking into consideration the uncertainty of the normality of our data we conducted the Kruskal-Wallis test which is a non-parametric alternative to the ANOVA test.

```
        Kruskal-Wallis rank sum test

data:  study$Risk.score by study$Category
Kruskal-Wallis chi-squared = 3.3929, df = 2,
p-value = 0.1833
```

The p-value = 0.1833 > 0.05 implying that we fail to reject the null hypothesis.

∴ There is no significant difference between the categories.

iv) Wilcoxon Rank Sum test

We lastly ran a pairwise comparison between the categories for any discrepancy.

```
        Pairwise comparisons using Wilcoxon rank sum test wi
th continuity correction

data:  study$Risk.score and study$Category

         Peer Self
Self     0.17 -
Stranger 0.97 0.17

P value adjustment method: BH
```

The p-values > 0.05 across all category comparisons and thus we fail to reject the null hypothesis.

∴ There is no significant difference between the categories.

All four tests failed to reject the null hypothesis and gave the same result of no statistically significant difference between any category.
∴ *There was no difference found in decision-making for self vs. others.*

B) Gender-wise risk score comparison

Secondly we wanted to evaluate if gender was a factor in any discrepancy in decision-making between and within the categories. Gender as a variable was used to enable us to study the potential impact of norms and social values on risk-taking. Gender as a potential source of difference in risk behaviour was studied both within and across categories.

i) Comparing within a category

Comparison of means was done within categories for the two genders: females and males.

The null and alternative hypotheses for the following tests were:

H_0*: No difference among mean risk scores of males and females within the category.*

H_1*: There is a difference in the mean risk score of males and females within the category.*

a) Self

The data from female and male samples were found to be normal and variances to be equal, fulfilling the assumptions to fairly conduct an unpaired T-test to compare means.

```
        Two Sample t-test

data:  self$Male and self$ï..Female
t = -0.17734, df = 31, p-value = 0.8604
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.9135023  0.7673484
sample estimates:
mean of x mean of y
 2.076923  2.150000
```

The p-value = 0.8604 > 0.05, thus we fail to reject the null hypothesis.

∴ *There is no significant difference in the means of gender in the self category.*

b) Peer

The independent samples of female and male responses of the peer category were not found to be normally distributed. A Mann-Whitney U test was performed as a non-parametric alternative to the T-test.

```
        Wilcoxon rank sum test with
        continuity correction

data:  peer$Male and peer$ï..Female
W = 139, p-value = 0.475
alternative hypothesis: true location shift is not equal to 0
```

The p-value = 0.475 > 0.05, thus we fail to reject the null hypothesis.

∴ *There is no significant difference in the means of gender in the peer category.*

c) Stranger

The independent samples of female and male responses of the peer category were not found to be normally distributed. A Mann-Whitney U test was performed as a non-parametric alternative to the T-test.

```
        Wilcoxon rank sum test with continuity correction

data:  stranger$Female and stranger$Male
W = 103, p-value = 0.9647
alternative hypothesis: true location shift is not equal to 0
```

The p-value = 0.9647 > 0.05, thus we fail to reject the null hypothesis.

∴ *There is no significant difference in the means of gender in the stranger category.*

There was no gender-based difference found in all categories.

ii) Comparing between categories within a gender

Comparison of means was done across categories within the two genders: females and males.

The null and alternative hypotheses for the following tests were:

H_0*: No difference among means of the categories across the gender.*

H_1*: At least one category has a different mean across the gender.*

a) Female

The independent samples of females of the peer and stranger categories were not found to be normally distributed. The Kruskal-Wallis test was performed to compare the means between the categories.

```
        Kruskal-Wallis rank sum test

data:  my_data_fem$Score by my_data_fem$Group
Kruskal-Wallis chi-squared = 2.1344, df = 2, p-value = 0.344
```

The p-value = 0.344 > 0.05, thus we fail to reject the null hypothesis.

∴ *There is no significant difference in the means of the categories within females.*

b) Male

The independent samples of males of the peer and stranger categories were found to be not normally distributed. The Kruskal-Wallis test was performed to compare the means between the categories.

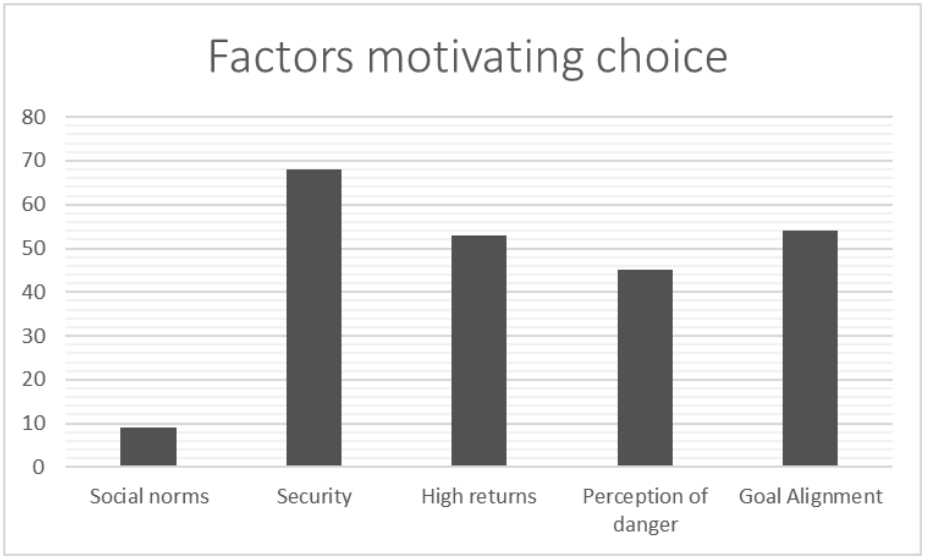
```
        Kruskal-Wallis rank sum test

data:  my_data_male$Score by my_data_male$Group
Kruskal-Wallis chi-squared = 1.8421, df = 2, p-value = 0.3981
```

The p-value = 0.3981 > 0.05, thus we fail to reject the null hypothesis.

∴ *There is no significant difference in the means of the categories within males.*

There was no categorical difference found within the genders.

C) Qualitative participant evaluation



The above graph shows the distribution of responses when asked to diagnose the factors guiding their choices. We can see that the primary motivator appears to be security which is followed by goal attainment. Social norms appear to not be a dominant factor guiding choice outcomes among undergraduate students.

The following table gives us an understanding of the tally of factors considered when making a decision within each category:

	Social norms	**Security**	**Higher returns**	**Perception of danger**	**Goal alignment**
Self	9.09%	69.69%	57.57%	36.36%	100%
Peer	3.03%	30.30%	15.15%	12.12%	57.57%
Stranger	16.12%	64.51%	54.84%	41.44%	64.51%

For all three categories, security and goal alignment appear as the primary motivators. The centrality of all factors falls for the peer category indicating that a factor not summarised by the five stated may be a key motivator of choice for peers.

General Discussion

In general, we find that decisions were not influenced by who the decisions were made for- self, peers or average college students. The decisions made for the self, for peers and for strangers had the same mean as per ANOVA testing indicating that a self-other difference does not exist in our domain of career choices. While our findings are consistent with the study conducted by Stone, Yates and Caruthers in the monetary decision-making domain, it goes against the findings of parallel studies conducted by Beisswanger et al., Hsee and Weber, Kray and Gonzalez etc. Further, we found that gender as a variable produced no difference in the results. The risk-taking behaviour of males and females was similar across samples and no gender-wise self-other difference existed. This is consistent with the findings of Stone et al. observed no difference in the level of risk-taking based on who the decision is being made for amongst both males and females.

Proximate reasons:

A key factor for self-other difference as found in the Hsee and Weber (1997) study is the 'risk-as-feelings' notion which states that people make decisions for others based on their perception of the risk level of the other. This perception is socially and normatively determined. For example, in Hsee and Weber (1995), both American and Chinese participants predicted that Americans were more risk-seeking than Chinese people even though in reality the Chinese were found to be significantly more risk-seeking. Homogenous responses across the self and the other could then be elicited if undergraduate students perceive others to have similar risk tastes as themselves by virtue of experiences and stereotypes they have contendedwith. Kray (2000) found that people choose the alternative highest on the most important dimension to reduce the degree of responsibility for unfortunate consequences when advising others. If undergraduate students consider a particular dimension to be important not just for themselves but also for their classmates and college students on average, then there is a possibility of selecting the alternative that scores the highest on this dimension across the categories. Career choices with higher salaries may have been chosen albeit they were riskier as higher income gains were perceived as something others would want.

As the experiment was based on hypothetical circumstances and the choices led to no consequential benefits for respondents, the true preferences of the respondents may not have been revealed. However, since Lichtenstein and Slovic (1973) found that there generally is no difference between responses when tasks are hypothetical or real we are motivated to consider other methodological dimensions. Curley et al. found that people's tendency to choose certain outcomes over uncertain outcomes was reduced when decisions were made in private as opposed to those made in public. Hence, at-home survey methods may lead to increased risk-seeking across all categories of responses. Our study also had a poor feedback procedure for respondents. Participants were not told of the impact of their decisions which could've reduced the role of regret and guilt found to have produced self-other differences in other studies.

Arnett's (Arnett, 1992) theory on socialization makes a distinction between narrow and broad specialisation. Narrow socialization refers to cultural influences that leave little room for departure from norms while broad socialization allows for a wide range of behaviour. Stone et al. considers environments such as college as allowing for broad specialisation which leads to respondents being unaffected by other dominant cultural norms. Narrow socialisation could be a source of consistent risk-taking between males and females and also a source of consistent choices between those made for the self and others due to similar normative beliefs of ideal career frameworks and choices.

When making decisions for others, people are more likely to choose what they believe will make other people happy while they rely on their own idiosyncratic preferences when making decisions themselves (Kray, 2000). However, if there is a consensus between what is considered to provide utility to the general populace and the individual, one can hypothesise a bridging of the gap in a choice made for the self and that for the other. In the career domain for example, since the sample population—college students—opinion of the utility of options is derived based on inferences from similar social discourse and mass media rather than from their own experience or insight there could be a convergence in other advice and personal decision making.

Our results may also differ from the previous studies due to the absence of any indication of personal satisfaction in our choice scenarios. The scenario used in Kray and Gonzalez's study contained two jobs to choose from, one high in salary and the other high in personal satisfaction.

In contrast, our study made the respondents choose from two career choices withdifferences in incomes and risk level, with no mention of the self-fulfilment one may gain from any of the choices. Their satisfaction may have differed based on their personal goals and was not pre-specified in the alternatives. The difference in the lack of this particular attribute might have led to a decision strategy where the respondents chose between the alternatives solely on the security of income and made a generalised choice for others too.

A final contributing factor to the lack of difference may be attributed to the nature of the scenarios. Beisswanger et.al found that self-other difference disappeared for scenarios that had a high life impact while it persisted for scenarios that had a low life impact. Since a majority of the scenarios included in our study had a life impact by virtue of dealing with career choices that determined the main sources of income of people a self-other difference was not discernible.

Lastly, it has to be noted that the type of advising that we explored was limited and general in nature, as compared to the advice given to a specific person. The respondents had no idea of the preferences of the others they were choosing and the estimation of the same may have generalised the choice for others to their choice.

Conclusion

Our work has demonstrated that in the career choice domain, there exists no self-other difference both for ‘familiar’ and ‘abstract’ others. We have also shown the insignificance of gender in producing a difference in risk attitudes. Given the nature of our sample and the domain under consideration, this result appears consistent. However, there is a need for further research in the domain with a larger sample and a focus on in-person surveys that provide feedback on choices in order to prove against any methodological limitations foundin this study. Surveys considering both high and low-impact scenarios across this domain would provide broad-scoped insight into the discourse of self-other difference. A better examination of the qualitative motivators with a better understanding of the sample being analysed will deepen our understanding of the field.

References

1. Hsee, C. K., and Weber, E. U. (1997). A fundamental prediction error: Self–others discrepancies in risk preference. *Journal of Experimental Psychology: General, 126*(1), 45–53. https://doi.org/10.1037/0096-3445.126.1.45
2. Stone, E. R., Yates, A. J., and Caruthers, A. S. (2002). Risk taking in decision making for others versus the self. *Journal of Applied Social Psychology, 32*(9), 1797–1824. https://doi.org/10.1111/j.1559-1816.2002.tb00260.x
3. Beisswanger, A. H., Stone, E. R., Hupp, J. M., and Allgaier, L. (2003). Risk taking in relationships: Differences in deciding for oneself versus for a friend. *Basic and Applied Social Psychology, 25*(2), 121–135. https://doi.org/10.1207/S15324834BASP2502_3
4. Kray, L., and Gonzalez, R. (1999). Differential weighting in choice versus advice: I'll do this, you do that. *Journal of Behavioral Decision Making, 12*(3), 207–217. https://doi.org/10.1002/(SICI)1099-0771(199909)12:3<207::AID-BDM322>3.0.CO;2-P
5. Kahneman, D., and Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80*(4), 237–251. https://doi.org/10.1037/h0034747
6. Hibbing, J. R., and Alford, J. R. (2005). Decision making on behalf of others. In *annual meeting of the American Political Science Association, Washington, DC.*
7. Garcia-Retamero, R., and Galesic, M. (2012). Doc, what would you do if you were me? On self–other discrepancies in medical decision making. *Journal of Experimental Psychology: Applied, 18*(1), 38–51. https://doi.org/10.1037/a0026018
8. Hoch, S. J. (1988). Who Do We Know: Predicting the Interests and Opinions of the American Consumer. *Journal of Consumer Research, 15*(3), 315. https://doi.org/10.1086/209169
9. Kray, L. J. (2000). Contingent Weighting in Self-Other Decision Making. *Organizational Behavior and Human Decision Processes, 83*(1), 82–106. https://doi. org/10.1006/obhd.2000.2903
10. Lichtenstein, S., and Slovic, P. (1973). Response-induced reversals of preference in gambling: An extended replication in Las Vegas. *Journal of Experimental Psychology, 101*(1), 16–20. https://doi.org/10.1037/h0035472
11. Arnett, J. (1992). Socialization and adolescent reckless behaviour: A reply to Jessor. *Developmental Review, 12*(4), 391–409. https://doi.org/10.1016/0273-2297(92)90015-t
12. Larrick, R. P. (1993). Motivational factors in decision theories: The role of self-protection. *Psychological Bulletin, 113*(3), 440–450. https://doi.org/10.1037/0033-2909.113.3.440

Appendix

To analyse the data through the parametric one-way ANOVA, we have to check our data for the following assumptions:

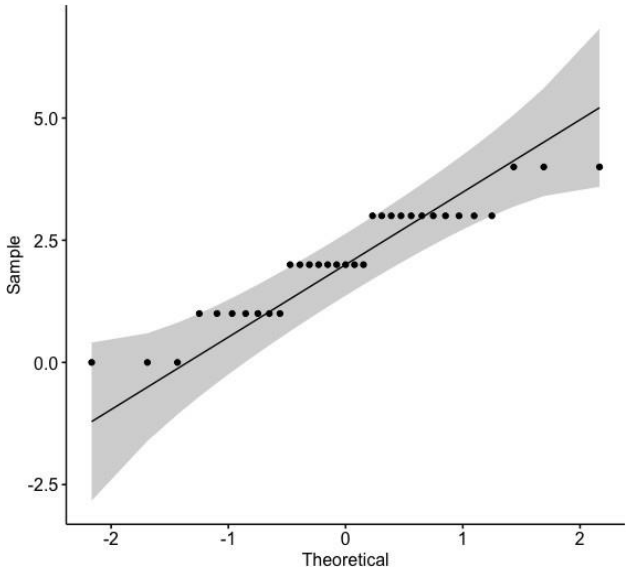
1. Independence of observations: The data were randomly collected from a randomly selected population and there is no association of the populations within or between the groups.
2. Normally-distributed response variable: The data in each group needs to be normally distributed.
3. Homogeneity of variance: The variation within each group being compared should be similar for every group.

Normality:

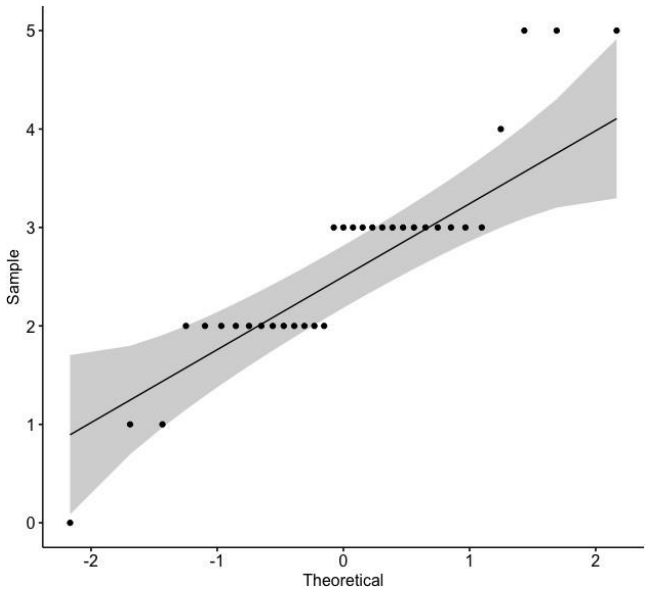
Our sample sizes for each category are ≥ 30. We still test for normality within each category, visually and statistically.

1) Visual normality test by QQ-plot:

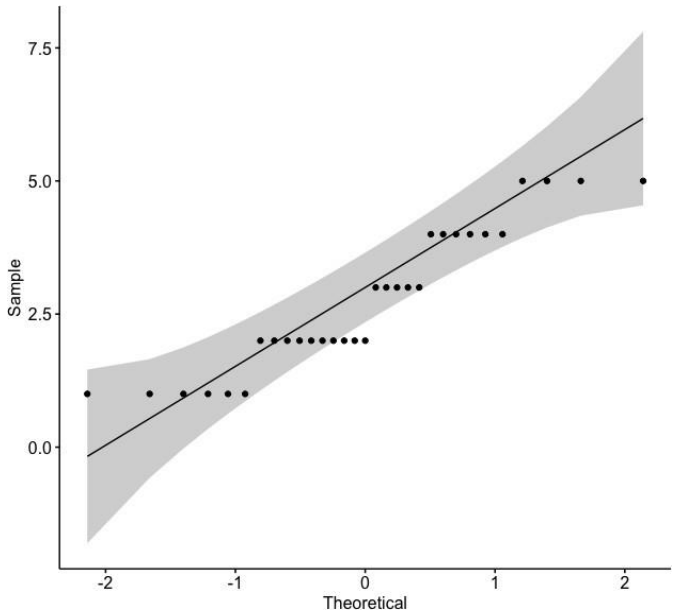
i) SELF:



ii) PEER:



iii) STRANGER:



Visually, all three groups seem to be normally distributed.

2) Statistical tests

We used two tests, the Shapiro-Wilk normality test R and Jarque–Bera test to check for goodness-of-fit.

The null and alternative hypotheses for the tests were:

H_0: *Data is normally distributed.*

H_1: *Data is not normally distributed.*

i) Shapiro-Wilk normality test:

a) SELF:

```
        Shapiro-Wilk normality test

data:  self$Score
W = 0.91194, p-value = 0.01103
```

The p-value = 0.01103 < 0.05, so we reject the null hypothesis.

∴ *Data is not normally distributed.*

b) PEER

```
        Shapiro-Wilk normality test

data:  classmate$Score
W = 0.87439, p-value = 0.001231
```

The p-value = 0.001231 < 0.05 so we reject the null hypothesis.

∴ *Data is not normally distributed.*

c) STRANGER

```
        Shapiro-Wilk normality test

data:  stranger$Score
W = 0.89055, p-value = 0.004222
```

The p-value = 0.004222 < 0.05 so we reject the null hypothesis.

∴ *Data is not normally distributed.*

The Shapiro-Wilk normality test seems to be failing for all three groups.

Another alternative test was the Jarque–Bera test which was conducted in excel.

ii) Jarque–Bera test:

a) SELF

SELF	
observations	33
sample skewness	-0.2511275
sample kurtosis	-0.7106197
JB test statistic	1.04120544
p-value	0.59416233

The p-value = 0.5942 > 0.05 so we fail to reject the null hypothesis.

∴ *Data is normally distributed.*

b) PEER:

CLASSMATE	
observations	33
sample skewness	0.3297781
sample kurtosis	1.1108392
JB test statistic	2.2948449
p-value	0.317454

The p-value = 0.3175 > 0.05 so we fail to reject the null hypothesis.

∴ *Data is normally distributed.*

c) STRANGER:

STRANGER	
observations	31
sample skewness	0.32946443
sample kurtosis	-1.0911232
JB test statistic	2.09861887
p-value	0.35017949

The p-value = 0.3502 > 0.05 so we fail to reject the null hypothesis.

∴ *Data is normally distributed.*

The results of the J-B test correspond to the visual approach. The normality of the data is shown visually and formally.

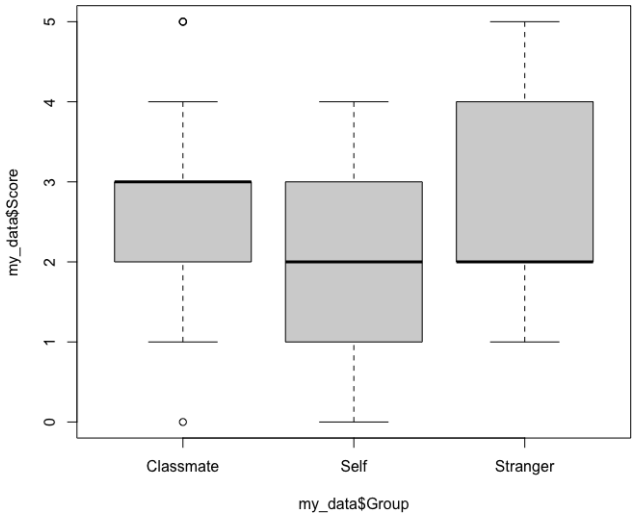
Uncertainty of normality

Even though each category had a minimum sample of 30, there is still ambiguity about the normality of the data because of its discrete nature. Visually and via the Jarque-Bera test, the data was shown normally distributed, but the Shapiro-Wilk test does not work in the individual category samples. Discrete data is usually difficult to prove as normally distributed, so we took into account the uncertainty around the normal nature of data, and it being one of the core assumptions of one-way ANOVA, we additionally used non-parametrictests like the K-W test which doesn't take normality of data as an assumption.

VARIANCE TEST:

First we will determine if the variance between all three categories: self, peer and stranger is equal or not. We can do this visually or by using a statistical test.

1) Visual variance test by using BOXPLOT:



The boxplot here shows a similar variance across the categories as the boxes and the whiskers have a comparable size for all categories.

2) Statistical F test:

Statistically, we can observe homogeneity by comparing the variances between two groups at a time using the F test.

The null and alternative hypotheses for the test were:

H_0: All variances are equal.

H_1: At least one variance is different.

a) Between Self and Peer:

```
        F test to compare two variances

data:  self$Score and classmate$Score
F = 1.1031, num df = 32, denom df = 32, p-value = 0.7831
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.5447884 2.2334185
sample estimates:
ratio of variances
           1.10306
```

The p-value = 0.7831 > 0.05, thus we fail to reject the null hypothesis.

∴ *The variances are equal.*

b) Between Peer and Stranger:

```
        F test to compare two variances

data:  classmate$Score and stranger$Score
F = 0.65419, num df = 32, denom df = 30, p-value = 0.2406
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.3178781 1.3350961
sample estimates:
ratio of variances
         0.6541907
```

The p-value = 0.2406 > 0.05, thus we fail to reject the null hypothesis.

∴ *The variances are equal.*

c) Between Stranger and Self

```
        F test to compare two variances

data:  self$Score and stranger$Score
F = 0.72161, num df = 32, denom df = 30, p-value = 0.366
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.3506385 1.4726905
sample estimates:
ratio of variances
         0.7216113
```

The p-value = 0.366 > 0.05, thus we fail to reject the null hypothesis.

∴ *The variances are equal.*

This result corresponds with the visual test, and so homogeneity is shown visually and formally.