Working Paper 2021-97

Analyzing relations of necessity in survey research: Incorporating notions of fuzzy-set Qualitative Comparative Analysis and bootstrap

Daisuke Mori

Faculty of Law

Kumamoto University

Abstract: Correlation coefficients often used in quantitative research, such as survey research, cannot adequately measure certain key aspects within the relation of necessity. In this paper, I focus on Qualitative Comparative Analysis (QCA), enabling the analysis of set relations or relations of necessity and sufficiency, and try to incorporate it into the analysis of survey research. I present the parameters of fit in a fuzzy-set QCA to measure such a relation. I also show the methods used to conduct statistical inferences to understand the properties of a population based on a random sample from that population. For example, I analyzed the survey data on the purpose of tort damages in the case of a defective car accident in Japan. In this survey, respondents were asked about the extent to which they considered five factors in deciding the amount of damages: compensation for monetary loss, punishment, compensation for mental suffering, deterrence, and satisfying the feeling of retribution. I found that compensation for monetary loss is not correlated with punishment, deterrence, and satisfying the feeling of retribution. However, compensation for monetary loss is a necessary condition. The analysis of relations of necessity can highlight relations that have not been detected by conventional statistical analysis based on correlations.

Keywords: Necessity, Set Relation, QCA, Fuzzy-Set, Bootstrap, Survey Research

---

COMPASSS Working Paper 2021-97
Available at http://www.compasss.org/wpseries/Mori2021.pdf
COMPASSS Working Paper Series Managing Editor, Claude Rubinson

Analyzing relations of necessity in survey research: Incorporating notions of fuzzy-set Qualitative Comparative Analysis and bootstrap

Abstract

Correlation coefficients often used in quantitative research, such as survey research, cannot adequately measure certain key aspects within the relation of necessity. In this paper, I focus on Qualitative Comparative Analysis (QCA), enabling the analysis of set relations or relations of necessity and sufficiency, and try to incorporate it into the analysis of survey research. I present the parameters of fit in a fuzzy-set QCA to measure such a relation. I also show the methods used to conduct statistical inferences to understand the properties of a population based on a random sample from that population. For example, I analyzed the survey data on the purpose of tort damages in the case of a defective car accident in Japan. In this survey, respondents were asked about the extent to which they considered five factors in deciding the amount of damages: compensation for monetary loss, punishment, compensation for mental suffering, deterrence, and satisfying the feeling of retribution. I found that compensation for monetary loss is not correlated with punishment, deterrence, and satisfying the feeling of retribution. However, compensation for monetary loss is a necessary condition. The analysis of relations of necessity can highlight relations that have not been detected by conventional statistical analysis based on correlations.

Keywords: Necessity, Set Relation, QCA, Fuzzy-Set, Bootstrap, Survey Research

Daisuke Mori
Faculty of Law
Kumamoto University
2-40-1 Kurokami Chuou-ku Kumamoto-shi Kumamoto 8608555 JAPAN
E-mail: mdai@kumamoto-u.ac.jp

Acknowledgements: This study was funded by the Japan Society for the Promotion of Science (grant/award nos. 19K13484). I am thankful for the helpful comments and suggestions provided by the anonymous reviewers. I am also grateful to Atsushi Ishida, Taisuke Fujita, Shuichi Takahashi, Yasuhiro Ikeda, and the participants of the Online Workshop on Qualitative Comparative Analysis (September 2020), and the 66th Conference of the Japanese Association for Mathematical Sociology (August 2018).

1. Introduction

In quantitative research, such as survey research, one frequently derives correlation coefficients between various variables in the first stage of data analysis. In such cases, it is important to avoid evaluating the presence of relations between variables based solely on the correlation coefficients because they can only measure the strength of linear relations. However, there are few simple indicators that measure other kinds of relations.

In this paper, I analyze one such relation - the relation of necessity. I present the parameters of fit in fuzzy-set Qualitative Comparative Analysis (QCA) to measure this relationship and explore ways to apply them to survey research. I also present methods to conduct statistical inferences because one of the purposes of survey research is to infer the properties of a population based on a random sample from that population.

QCA is a technique that strives to deliver the advantages of both the "qualitative" (case-oriented) and "quantitative" (variable-oriented) techniques. It allows the systematic comparison of cases, with the help of formal tools and with a specific conception of cases (Rihoux and Ragin 2008:6). It is also often said that QCA enables analysis of small-N and medium-N datasets that are not suitable for statistical analysis (Ragin et al. 2003:324; Rihoux 2003:353).

However, QCA has another feature that enables us to analyze set relations or relations of necessity and sufficiency. Schneider and Wagemann (2012:12) state that "the use of QCA would be appropriate even if the *N* is large if, and only if, researchers are interested in set relations rather than correlations." In this paper, I focus on this feature of QCA and try to incorporate it into the analysis of survey research.[1]

The remainder of this paper is organized as follows. In Section 2, I explain the difference between the relationship of necessity and correlation. I present the parameters of fit in a fuzzy-set QCA to measure such a relationship. I also present the methods used to conduct statistical inferences. In Section 3, I analyze the relations of necessity in survey research. In Section 4, I present conclusions and suggest future areas of research.

2. Analysis of relation of necessity and its statistical inference

2. 1. Relation of necessity and correlation

First, I explain the difference between set relations and correlations using a hypothetical example.[2] Suppose 14 students in a certain school take two mathematical tests, Test 1 and Test 2. Test 1 is about linear equations, and Test 2 is about quadratic equations. The perfect score for each test is five.

The students' scores on these tests are shown in Table 1 (I put aside Test 3 in Table 1 for now). The Pearson correlation coefficient between the scores of Tests 1 and 2 was 0.074.[3] Therefore, it can be concluded that these two scores are almost uncorrelated.

[1] I do not intend to apply QCA's many important features, such as the systematic comparison of cases, to the analysis of survey research. I solely incorporate this one aspect.
[2] For the explanation about the difference between set relations and correlations, see also Goertz and Starr (2003), Ragin (2006), and Schneider and Wagemann (2012:83-90).
[3] The statistical software I use is R version 4.0.3. I use the "psych" package (version 2.0.12) to calculate the confidence intervals and *p*-values of the Pearson correlation coefficients. The R script of the analysis in Section 2 is shown in Appendix.

Table 1 Students' Scores of Mathematical Tests

ID	Test 1	Test 2	Test 3
1	5	0	5
2	3	0	4
3	4	0	5
4	5	0	4
5	2	1	4
6	3	1	5
7	4	3	5
8	5	4	5
9	3	2	4
10	3	3	5
11	4	4	4
12	3	3	4
13	4	4	5
14	4	5	5

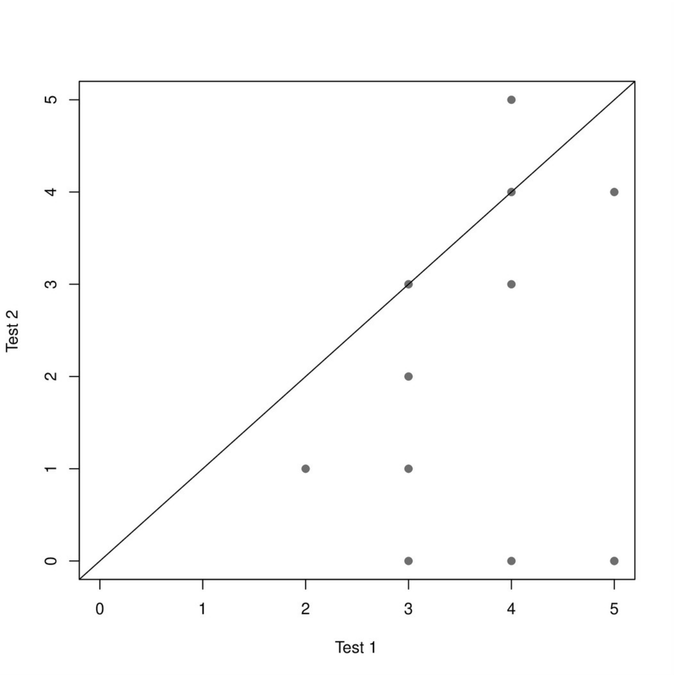

Figure 1 XY Plot of Scores of Test 1 and 2

However, we cannot conclude that there is no relationship between the two scores. Figure 1 shows an XY plot of these two scores. Almost all of the data are below the diagonal line that runs from the bottom-left corner to the top-right corner.

This relation in the plot is not a correlation, but a relation of necessity. Since knowledge about linear equations is fundamental to handling quadratic equations, obtaining a good score in Test 1 is necessary to answer questions in Test 2. Therefore, almost all student's Test 1 scores should be higher than their Test 2 scores. However, Test 2 scores might not be strongly correlated with the Test 1 scores. This is because, apart knowledge about linear equations, other knowledge such as concepts of powers and roots, quadratic formula, and factorization are necessary to solve quadratic equations. In other words, the comprehension of linear equations is a necessary condition for solving quadratic equations.

2.2. Methods for analyzing the relation of necessity

As the low correlation coefficient between score of Tests 1 and 2 shows, the relation of necessity cannot be captured by correlation coefficients. The majority of standard statistical techniques are not well suited for detecting subset relations of necessity and sufficiency (Schneider and Wagemann 2012: 89). Although Goertz and Starr (2003) provided seminal work on necessary conditions, it does not offer an integrated framework to systematically analyze the issue.

There are currently two possible options for analyzing the relation of necessity. One is QCA and the other is the Necessary Condition Analysis (NCA) proposed by Dul (2016, 2019). I choose QCA in this study for two reasons. First, it is a more established analytical tool while NCA is relatively new and still developing. NCA is more useful in determining which level of a condition is necessary for a particular level of the outcome, whereas QCA is suitable to analyze only necessity in kind (Vis and Dul 2018: 894). Since I mainly focus on necessity in kind in this study, I adopt the latter. Second, QCA, especially the extensive fuzzy-set version, has multiple indicators such as consistency, coverage, and relevance, enabling one to evaluate various aspects of the relations of necessity.

2. 3. Parameters of fit in relation of necessity

In QCA, one uses membership scores instead of the values of variables. In ordinary (crisp) sets, all elements are either fully in or out of the sets, which correspond to membership scores of 1 and 0, respectively. In fuzzy sets, partial belonging to the sets is also possible, which indicates membership scores between 0 and 1. Membership scores of elements between 0.5 and 1 indicate that they are more in than out of the sets, membership scores of 0.5 indicate that they are neither in nor out of the sets, and membership scores between 0 and 0.5 indicate that they are more out of than in the sets.

When analyzing the relationship between variables in QCA, one needs to transform the values of such variables into fuzzy-set membership scores to sets. This transformation is called "calibration" in QCA. In the present example, I use the sets "comprehension of linear equation" and "comprehension of quadratic equation" (hereinafter, the sets "TEST1" and "TEST2"). I transform Test 1 and 2 scores 0, 1, 2, 3, 4, and 5 into membership scores of 0, 0.2, 0.4, 0.6, 0.8, and 1.[4]

[4] Note that values of variables that have same intervals need not be calibrated into membership scores that have same intervals. When the variation in the variable is conceptually irrelevant, one should consider it to transform values of the variable into membership scores. For example, if we transform GDP per capita into fuzzy-set membership scores in the "rich country" set, several countries with the highest GDP per capita could all be considered "rich" and would be assigned a membership score of 1, despite the large variation in GDP per capita

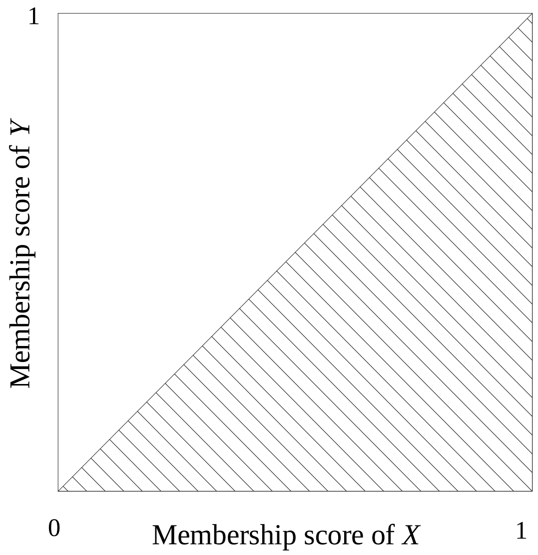

Figure 2　Relation of Necessity in Fuzzy Sets

In fuzzy-set QCA, one can conclude that the set *X* is a superset of *Y*, or *X* is a necessary condition for *Y*, when all of the data are on or below the diagonal of the XY plot, which is shown as the shaded area in Figure 2. In mathematical notation, $X \supseteq Y$ when, for all i, $y_i \leq x_i$, where x_i and y_i are the membership scores of the case of index i ($i = 1, ..., n$) to sets *X* and *Y*. In the present example, x_i and y_i are the ID number *i* students' membership scores to TEST1 and TEST2, respectively.

However, in reality, there are few cases when all of the data are on or below the diagonal. Therefore, we need to decide how much data should be on or below the diagonal for it to be considered a necessary condition. Some indicators called "parameters of fit" are proposed to handle this problem in a fuzzy-set QCA (Schneider and Wagemann 2012:119). "Consistency" of necessity is one such indicator. It is used to measure the degree to which the data are on or below the diagonal. Consistency scores take values from 0 to 1, and one can conclude $X \supseteq Y$ when they are equal to or higher than the threshold value, which is often 0.9.[5]

Consistency of necessity is calculated by the following formula (Ragin 2006:297):

$$\text{Consistency of necessity} = \frac{\sum_{i=1}^{n} \min(x_i, y_i)}{\sum_{i=1}^{n} y_i}$$

Based on this formula, the consistency of the necessity of TEST1 $\supseteq$ TEST2 is calculated to be 0.967. Since this is higher than the threshold value of 0.9, one can conclude TEST1 $\supseteq$ TEST2.

In addition to consistency, which measures the degree of relation of necessity, fuzzy-set QCA has a different kind of indicator. It is called "coverage" of necessity, and it measures the triviality of necessary conditions. Figure 3

among these countries (Schneider and Wagemann 2012:29). For now, I adopt a simple mechanical transformation, since the data I use here is just a hypothetical example for understanding purposes.

[5] For example, Schneider and Wagemann (2012:148, 278) state that 0.9 or an even higher value is advisable for a threshold for consistency of necessity.

illustrates this triviality.[6] In the Venn diagrams in Figure 3, both $X_1 \supseteq Y_1$ and $X_2 \supseteq Y_1$ holds true because both sets X_1 and X_2 are larger than set Y_1. However, set X_2 was larger than set X_1. This means that the number of cases that satisfy condition X_2 but do not satisfy Y_1 are larger than the number of cases that satisfy condition X_1 but do not satisfy Y_1. Therefore, one can say that X_2 is more trivial than X_1 in one sense. Coverage scores take values from 0 to 1, and higher values indicate that the necessary condition is less trivial.

Coverage of necessity is calculated by the following formula (Ragin 2006:303):

$$\text{Coverage of necessity} = \frac{\sum_{i=1}^{n} \min(x_i, y_i)}{\sum_{i=1}^{n} x_i}$$

Based on this formula, the coverage of necessity of TEST1 ⊇ TEST2 is calculated to be 0.558. This indicates that there are a certain number of students who understand linear equations but not quadratic equations.

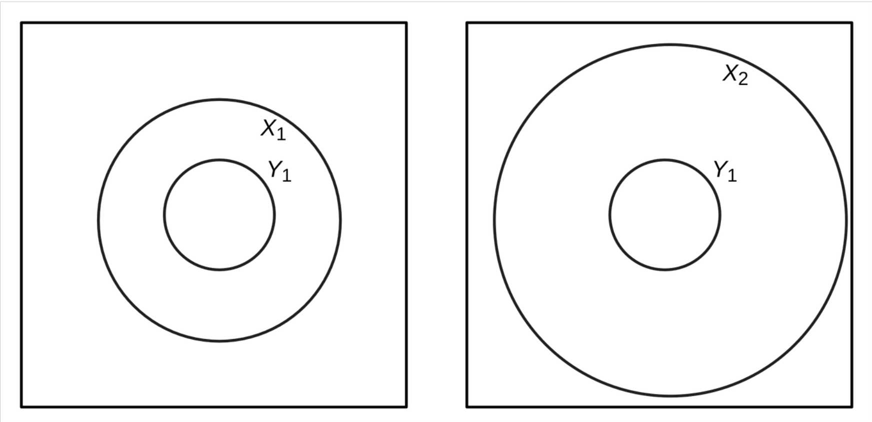

Figure 3 Coverage of Necessity

In addition, another indicator is proposed to measure different types of trivialness. In Figure 4, unlike the right figure in Figure 3, the coverage is high because the size of Y_2 to X_2 is larger than Y_1. However, X_2 might not be an important necessary condition, even in the case shown in Figure 4. This is because the size of X_2 is so large that it is close to the universal set. This means that almost all elements belong to X_2. For example, being able to perform arithmetic calculations is a trivial necessary condition for solving quadratic equations, not only because there are many people who are able to perform arithmetic calculations but unable to solve quadratic equations, but because almost all people can perform arithmetic calculations.[7] The relevance of necessity is an indicator that considers the

[6] Strictly speaking, Venn diagrams can only show crisp set, while I use fuzzy set in the analysis of the data of Table 1. However, I use this diagram in the explanation because this diagram helps with the intuitive understand of the concept of coverage.

[7] The word “trivial” here might be a bit misleading. This does not mean that knowledge on arithmetic calculations is not important for solving quadratic equations. “Triviality” in this sentence means that whether or not people have knowledge about arithmetic calculations is of little use for distinguishing between people that can solve quadratic equations and those that cannot, because almost all people have knowledge related to arithmetic calculations.

trivialness related to the relative size of the universal set. The relevance scores ranged from 0 to 1. When relevance is much smaller than coverage, one should be careful about the trivialness related to the relative size of the universal set.

The relevance of necessity is calculated using the following formula (Schneider and Wagemann 2012:236):

$$\text{Relevance of necessity} = \frac{\sum_{i=1}^{n}(1-x_i)}{\sum_{i=1}^{n}(1-\min(x_i, y_i))}$$

On the basis of this formula, relevance of necessity of TEST1 ⊇ TEST2 is calculated to be 0.439. Since this value is not very different from the coverage score, the problem of trivialness related to the relative size of the universal set is not serious in the present example.

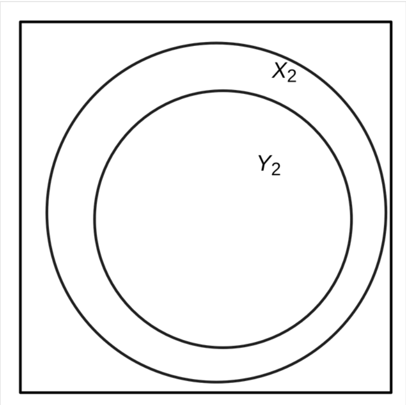

Figure 4 Relevance of Necessity

2. 4. Statistical inference of relation of necessity

2. 4. 1. Confidence interval

Let us assume that 14 students in Table 1 are randomly sampled from the population of students in the same grade. If our purpose is to infer the properties of the population based on the sample in Table 1, the analysis in Sections 2.1 and 2.2 is not enough. For example, the consistency of necessity in the sample calculated in Section 2.2 is unusually high, and might be much lower in the population.

Statistical inferences were used to handle the sampling errors. One of the methods used in inferential statistics is the confidence interval. The correlation coefficient between the scores of Tests 1 and 2 is calculated to be 0.074 in Section 2.1. The 95% confidence interval was [-0.475, 0.582].

Statistical inference has yet to be sufficiently discussed in the QCA. There have been a few previous related studies. Eliason and Stryker (2009) proposed a goodness-of-fit F test to assess the fit between a fuzzy set graph and causal necessity, sufficiency, and necessity and sufficiency hypotheses. However, the disadvantage is that this test does not precisely correspond with parameters of fit, such as consistency, coverage, and relevance.

Braumoeller (2015) and Dul, Jan, Laan, and Kuik (2020) use permutation test. Permutation tests require symmetric null hypotheses requiring, for example, "the data of one variable and the other variable to come from the same distribution." In contrast, the bootstrap method can be applied more generally because it does not require this type of special symmetry (Efron and Tibshirani 1993:178-201). Cooper and Glaesser (2016) and Gibson and Vann (2016) use the bootstrap method in QCA, but their focus is on the stability of QCA's complex form of solutions derived from truth tables. This is the analysis of the relation of sufficiency, while our main focus is on the relationship between necessity and simple indicators.[8]

I explore a way to construct confidence intervals of parameters of fit, such as consistency of necessity. The bootstrap method was adopted.[9] The features of this method are its simplicity and versatility. It can be applied to virtually any statistic, regardless of the complexity of its calculations (Efron and Tibshirani 1993: 43). As statistical inference about parameters of fit in QCA are still under development, this feature is desirable.

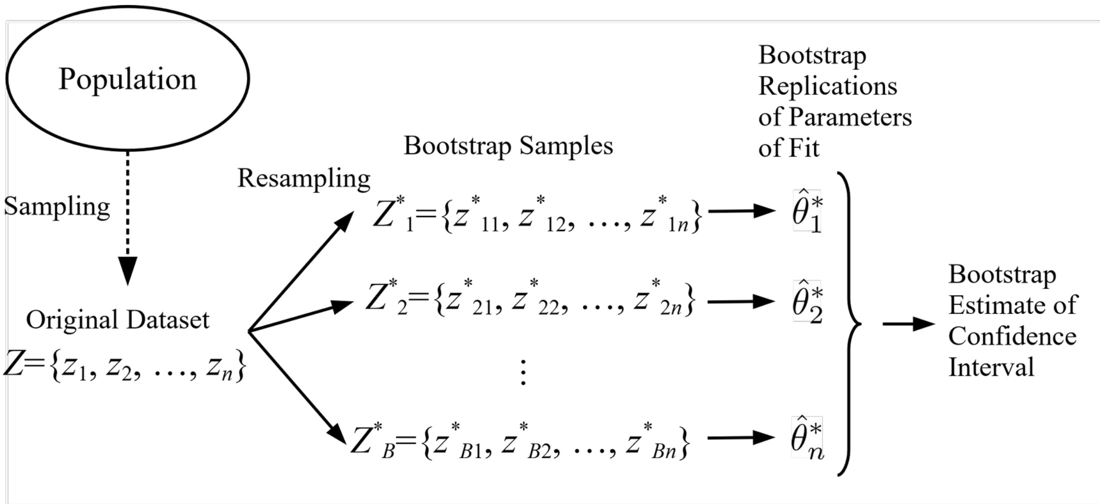

Figure 5 Process of Bootstrap Method

Figure 5 illustrates the bootstrap method. Suppose $Z=\{z_1, z_2, ..., z_n\}$ is a random sample of size n from a population. Each data z_i (i=1, 2, ..., n) in Z takes paired values (x_i, y_i). In the present example, i is student ID number (i=1, 2, ..., 14), x_i is a student with ID number i's membership score to TEST1, and y_i is the same student's membership score with respect to TEST2. A parameter of fit in this sample Z was calculated to be $\hat{\theta}$.

I randomly resample n data with replacement from Z. Repeating this resampling B times gives us bootstrap samples $Z^*_1, Z^*_2, ..., Z^*_B$. Z^*_1 is the first bootstrap sample of which data are $\{z^*_{11}, z^*_{12}, ..., z^*_{1n}\}$, and Z^*_2 is the second bootstrap sample whose data are $\{z^*_{21}, z^*_{22}, ..., z^*_{2n}\}$. I calculate the parameters of fit in bootstrap samples Z^*_1, Z^*_2, ..., Z^*_B, denoting $\hat{\theta}_1^{\text{¿}}, \hat{\theta}_2^{\text{¿}}, \ldots, \hat{\theta}_B^{\text{¿}}$. These are called bootstrap replications.

Based on such bootstrap replications, I derived the percentile confidence interval.[10] I arrange $\hat{\theta}_1^{\text{¿}}, \hat{\theta}_2^{\text{¿}}, \ldots, \hat{\theta}_B^{\text{¿}}$

[8] Schneider and Wagemann (2012:220) point out the "sufficiency bias" in the QCA-based research where most QCA places much more emphasis on the analysis of sufficiency.
[9] For the details of bootstrap method, see Efron and Tibshirani (1993), and Davison and Hinkley (1997).
[10] There are some other methods to construct bootstrap confidence intervals such as BCa method. See, for

from the lowest to the highest. Let $\hat{\theta}^{¿(\alpha B)}$ denote the αBth lowest value among $\hat{\theta}_1^{¿}, \hat{\theta}_2^{¿}, \ldots, \hat{\theta}_B^{¿}$, where $0<\alpha<1$. The $100(1-2\alpha)$% two-sided confidence interval was obtained as follows:

$$\left[\hat{\theta}^{¿(\alpha B)}, \hat{\theta}^{¿((1-\alpha)B)}\right]$$

Since $100(1-2\alpha)=95$ when $\alpha=0.025$, the 95% confidence interval is $\left[\hat{\theta}^{¿(0.025B)}, \hat{\theta}^{¿(0.975B)}\right]$. On the basis of this formula, the 95% confidence interval of consistency of necessity of TEST1 ⊇ TEST2 is calculated to be [0.903, 1.000].[11]

Given the bootstrap method's versatility, I can derive the confidence intervals of coverage and relevance of necessity of TEST1 ⊇ TEST2 the same way. The 95% confidence interval of coverage was calculated to be [0.327, 0.792], and the 95% confidence interval of relevance was calculated to be [0.256, 0.667].

2. 4. 2. Hypothesis testing

Hypothesis testing is another tool used for statistical inference. In hypothesis testing, I set the null hypothesis and evaluate whether the result is statistically significant based on the *p*-value. For example, hypothesis testing of correlation is often performed under the null hypothesis H_0: $\rho=0$, where ρ denotes Pearson correlation coefficient. In the present example, the *p*-value of the correlation coefficient between scores of Tests 1 and 2 is calculated to be 0.802, which is not statistically significant at the 5% significance level.

Hypothesis testing about the parameters of fit can also be performed using the bootstrap method. Under the null hypothesis H_0: $\hat{\theta}=\theta_o$, the percentile *p*-value is obtained as follows:

$$p=\frac{1}{B}\sum_{b=1}^{B} I\{\hat{\theta}_b^{¿}<\theta_o\backslash\},$$

where $I\backslash\{\cdot\backslash\}$ is called indicator function that takes following values:

$$I\left(\hat{\theta}_b^{¿}<\theta_o\right)=\begin{cases}1, \wedge \hat{\theta}_b^{¿}<\theta_o\\ 0, \wedge \hat{\theta}_b^{¿}\geq\theta_o.\end{cases}$$

Let $\hat{C}_b^{¿}$be a consistency of the necessity of a bootstrap sample and C_o be a consistency of necessity in the null hypothesis. If the threshold value at which one can confirm the relation of necessity is 0.9, the null hypothesis H_0 is $C_o=0.9$. The *p*-value under this null hypothesis is $\sum_{b=1}^{B} I\left\{\hat{C}_b^{¿}<0.9\right\}/B$. This formula calculates what proportion

example, Efron and Tibshirani (1993:223-224). I adopt percentile confidence intervals because of its simplicity.

[11] I set the number of resampling at 10,000 times and the random seed as 1. I use the same settings in subsequent analyses in this paper.

of *B* emerges when the consistency is below the 0.9 threshold.

Based on this formula, the *p*-value of consistency of necessity of TEST1 ⊇ TEST2 is calculated to be 0.018. As this is statistically significant at the 5% level, we can conclude that the consistency of necessity exceeds the threshold value 0.9 and TEST1 ⊇ TEST2 holds in the population.

2. 4. 3. Adjustment in case of multiple testing

Let us assume that 14 randomly sampled students take not only Tests 1 and 2 but also Test 3, which is about arithmetic calculation. Their scores are shown in Table 1. I aimed to analyze the relationships among these three tests. Table 2 shows the correlation coefficients among the scores of the three tests. This table also shows the 95% confidence intervals and *p*-values.

However, when conducting hypothesis testing multiple times to analyze the relationships among multiple variables, one should be careful about using *p*-values in hypothesis testing.[12] In this case, the *p*-value is adjusted, because the familywise error rate, which is the rate at which one or more correct null hypotheses among all hypotheses is incorrectly rejected, becomes high. The adjustment methods frequently used are the Bonferroni method and Holm method. The former is simpler, whereas the latter is more powerful.

Table 2 shows *p*-values adjusted by Holm method. Since all of them are larger than 0.05, none of the correlation coefficients in Table 2 are statistically significant at the 5% level.

Table 2 Correlation of Scores of Three Mathematical Tests

	r	95% CI	*p*	Adj. *p*
Test1-Test2	0.074	[-0.475, 0.582]	0.802	0.821
Test1-Test3	0.375	[-0.195, 0.755]	0.187	0.561
Test2-Test3	0.239	[-0.334, 0.683]	0.411	0.821

Note: *N*=14. *r* : Pearson correlation coefficients, CI: percentile confidence intervals, *p*: percentile *p*-values, Adj. *p*: percentile *p*-values adjusted by the Holm method.

I also analyzed the relation of necessity for comprehension of three mathematical subjects. Table 3 shows the parameters of fit of the relation of necessity, their 95% confidence intervals, and *p*-values. The "TEST3" in this table denote the set "comprehension of arithmetic calculation." Test 3 scores of 0, 1, 2, 3, 4, and 5 were transformed into membership scores of 0, 0.2, 0.4, 0.6, 0.8, and 1.

The family-wise error rate also becomes high when one conducts hypothesis testing about the consistency of the relation of necessity multiple times. Therefore, the adjustment of *p*-values is also necessary in this case. This adjustment of *p*-values is also necessary when hypothesis testing is performed.

[12] For the details about the problems of conducting hypothesis testing multiple times, see, for example, Bretz et al. (2011).

Table 3 shows *p*-values of consistency adjusted by Holm method.[13] The adjusted *p*-value of consistency of relation of necessity TEST1 ⊇ TEST2 is 0.070, which is not statistically significant at the 5% level. This is in contrast to the result in Section 2.3.2 where *p*-value of consistency of relation of necessity TEST1 ⊇ TEST2 is statistically significant.

Table 3 Relation of Necessity about Comprehension of Three Mathematical Subjects

	Consistency	95% CI	*p*	Adj. *p*	Coverage	95% CI	Relevance	95% CI
TEST1⊇ TEST2	0.967	[0.903, 1.000]	0.018	0.070	0.558	[0.327, 0.792]	0.439	[0.256, 0.667]
TEST1⊇ TEST3	0.797	[0.716, 0.877]	0.994	1.000	0.981	[0.942, 1.000]	0.947	[0.800, 1.000]
TEST2⊇ TEST1	0.558	[0.327, 0.792]	0.999	1.000	0.967	[0.903, 1.000]	0.976	[0.909, 1.000]
TEST2⊇ TEST3	0.469	[0.279, 0.661]	1.000	1.000	1.000	[1.000, 1.000]	1.000	[1.000, 1.000]
TEST3⊇ TEST1	0.981	[0.942, 1.000]	0.000	0.001	0.797	[0.716, 0.877]	0.316	[0.150, 0.500]
TEST3⊇ TEST2	1.000	[1.000, 1.000]	0.000	0.000	0.469	[0.279, 0.661]	0.150	[0.065, 0.255]

Note: *N*=14. CI: percentile confidence intervals, *p*: percentile *p*-values, Adj. *p*: percentile *p*-values adjusted by the Holm method.

In addition to these adjusted *p*-values, different methods can be used to handle multiple hypothesis testing. We can derive the percentile probability that multiple relations hold simultaneously. From Table 3, I hypothesize that there is a relation TEST3 ⊇ TEST1 ⊇ TEST2, which means TEST3 is a necessary condition for both TEST1 and TEST2, and TEST1 is a necessary condition for TEST2. I obtain a picture of these relations using the Venn diagram in Figure 6, although it is not accurate because Venn diagrams can only express crisp sets.

Let the consistency of the relation of necessity TEST1⊇TEST2 be *C*(TEST1 ⊇ TEST2). The percentile probability that TEST3 ⊇ TEST1 ⊇ TEST2 holds is obtained by $\sum_{b=1}^{B} I\{C(\text{TEST}3 \supseteq \text{TEST}1)<0.9 \vee C(\text{TEST}3 \supseteq \text{TEST}2)<0.9 \vee C(\text{TEST}1 \supseteq \text{TEST}2)<0.9\}/\text{B}$. This formula calculates what proportion of *B* occurs when any one of the consistencies of TEST3 ⊇ TEST1, TEST3 ⊇ TEST2, or TEST1 ⊇ TEST2 is below the 0.9 threshold.[14] The result of the calculation was 0.018, which was

[13] Confidence intervals can also be adjusted by some methods such as Bonferroni method. I omit this adjustment due to the space limitation.

[14] The percentile *p*-value of consistency calculates what proportion of *B* emerges when the consistency is below the 0.9 threshold and when the relation of necessity is not established. Similarly, the percentile probability of TEST3 ⊇ TEST1 ⊇ TEST2 calculates what proportion of *B* occurs when this relation is not established. If any one of the consistencies of TEST3 ⊇ TEST1, TEST3 ⊇ TEST2, or TEST1 ⊇ TEST2 is below the 0.9 threshold, the TEST3 ⊇ TEST1 ⊇ TEST2 relationship is not established. Therefore, $\sum_{b=1}^{B} I\{C(\text{TEST}3 \supseteq \text{TEST}1)<0.9 \vee C(\text{TEST}3 \supseteq \text{TEST}2)<0.9 \vee C(\text{TEST}1 \supseteq \text{TEST}2)<0.9\}/\text{B}$ is the formula representing the TEST3 ⊇ TEST1 ⊇ TEST2 percentile probability. One might wonder whether $\sum_{b=1}^{B} I\{C(\text{TEST}3 \supseteq \text{TEST}1)<0.9 \vee C(\text{TEST}1 \supseteq \text{TEST}2)<0.9\}/\text{B}$ is enough. However, when we accept that relation of necessity holds under conditions where consistency is less than 1.0, there is a possibility that

smaller than 0.05. Therefore, we can conclude that TEST3 ⊇ TEST1 ⊇ TEST2 holds at the 5% significance level.

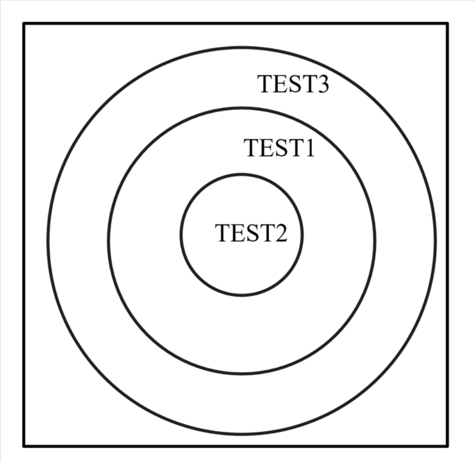

Figure 6 Multiple Relations of Necessity

3. Applying analysis of relations of necessity to survey research

3. 1. Overview of the survey

In this section, I analyze the relation of necessity presented in Section 2 to the survey research. The data I used was derived from the survey of Mori et al. (2017) on the purpose of tort damages in the case of a defective car accident in Japan. I briefly provide an overview of this survey based on the explanation of Mori et al. (2017).

Under Japanese tort law, the tortfeasor will be sued and damages will be imposed on him when he illegally infringes on the rights of other people.[15] There are several explanations for such damages in Japan.

The first explanation is that the purpose of the damages is "compensation," which aims to reimburse the victim an amount of money commensurate with the loss they have suffered. The loss includes not only monetary loss but also mental suffering, such as pain and sorrow.

The second explanation is that the purpose of the damages is "punishment" and "deterrence," which aims to punish the tortfeasor by imposing damages, and preventing similar cases in the future. In addition, satisfying the victims' feelings of retribution is often mentioned as being related to deterrence and punishment.

Japanese legal scholars have long debated which of the above-mentioned purposes is the main one. The current dominant view is that compensation is the main purpose, while the others are merely subordinate (Kihara 2015: 25; Yamamoto 2019: 19).[16] It is important to note that this debate affects the actual legal systems in Japan. For example, many legal scholars oppose the introduction of punitive damage systems because they consider them ill-

TEST3 ⊇ TEST1 and TEST1 ⊇ TEST2 holds while TEST3 ⊇ TEST2 does not hold.

[15] For the details of tort law in Japan, see, for example, Kihara (2015) and Yamamoto (2019).

[16] The Supreme Court of Japan takes a similar view. They stated "[t]he Japanese system of compensatory damages based on torts purports to restore a victim to the state in which it would have been if torts were not performed by the offender....[I]mposing on the offender liability for damages may have the effect of punishment of the offender or of general prevention. However, such is only a subordinate effect". See Northcon I, Oregon Partnership v. Mansei Kogyo Co., Ltd., (Sup. Ct., July 11, 1997) *Minshu* 51(6), 2573 (translated into English in *Japanese Annual of International Law* 41, 104).

suited to the Japanese system of damages, the main purpose of which is compensation.[17]

Views of legal scholars as well as ordinary people can affect the legal system.[18] This is because laws that are not consistent with people's views are hardly complied with by them. However, there has been little research on ordinary people's views on the purpose of damages. Mori et al.'s (2017) survey dealt with this matter.

The online survey was conducted in January 2015. The respondents forming the online panel were between 20 and 60 years of age and lived in the prefectures of the Kanto region of Japan.[19] Although the sampling method adopted was quota sampling, I considered it to be close to random sampling. It collected 546 questionnaires, half of which were used in this study.[20]

The survey first shows a hypothetical scenario concerning respondents' damage. The scenario is as follows:

> Alex is a 30-year-old white-collar worker. He caused an accident while driving a car manufactured by Company B, a domestic auto manufacturer. He lost control of the car and crashed into a guardrail because the front-wheel suddenly detached. It was damaged and cost him five-hundred thousand yen to have it repaired.
>
> Alex suffered broken bones during the accident. It took him 3 months (a 1-month hospital stay and 2-month hospital visit) to recover without after-effects. The treatment cost, including the cost of hospital stay, was three million yen.
>
> Alex had to take a 3-month absence from work. As a result, he did not receive salary for 3 months, which amounted to one million yen.
>
> Alex was not at fault during the accident. The investigation revealed that the cause of the accident was a design flaw in the part that connects the front wheels to the axle. It was also discovered that there had been 40 such accidents caused by cars manufactured by Company B, but no civil lawsuits were filed.
>
> Company B was aware of the design flaw before Alex's accident, but they covered it up and never announced a recall. They also made false reports to the public office that the cause of the accidents was the users' poor maintenance. These facts have recently come to light. As a result, they paid a fine of two hundred thousand yen.
>
> Alex filed a civil lawsuit for damages against Company B. The costs of litigation and attorneys are assumed to be zero. (Mori et al. 2017:617-618).

---

[17] Punitive damages are damages, other than compensatory or nominal damages, awarded against a person to punish him for his outrageous conduct (American Law Institute 1979: §908(1)). The United States is one of the most well-known examples of a country with a punitive damages system.

[18] Vice-chairman of the Justice System Reform Council, which was established under the Cabinet to discuss improvement of legal and judicial system in Japan, stated "if the way of thinking and values of the people in Japan as a whole change in the future, the contents of damages in civil law will also change" (Justice System Reform Council 2001).

[19] The survey was conducted with the aid of NTT Com Online Marketing Solutions Corporation. The online panel is a group of people who had registered to participate in various surveys and earn points from NTT Com. They had 740,000 registered individuals in Japan as of 1 November 2013. See Mori et al. (2017:598).

[20] There are two different versions of questionnaires and respondents were randomly assigned one of them. For the details of this design of the survey, see Mori et al. (2017:597). The reason for using only 271 questionnaires in our paper is that I use only one version of the questionnaire.

After this scenario, respondents were asked how much Company B should pay Alex for damages. They were also asked how much they considered the following five factors in deciding the amount of damages.

(a) Compensation for the monetary loss that Alex suffered.
(b) Punishing Company B.
(c) Compensating Alex for mental suffering, such as pain and sorrow.
(d) Preventing similar kinds of cases in the future by showing that this case results in the imposition of damages.
(e) Satisfying Alex's feelings of retribution

The degree of consideration of various factors by respondents was measured using a five-point scale: 1. disregarded, 2. somewhat disregarded, 3. neutral, 4. somewhat considered, and 5. considered.[21] Hereinafter, I express factor (a) as "Monetary," factor (b) as "Punishment," factor (c) as "Mental," factor (d) as "Deterrence," and factor (e) as "Retribution."

I analyzed the relationships among these factors to reveal the views of ordinary people about the purpose of damages. I particularly examined whether they have the same view as legal scholars, that is to say, whether they considered "Monetary" and "Mental" to be primary and "Punishment," "Deterrence," and "Retribution" to be subordinate.

3. 2. Summary statistics and correlation

Table 4 is a frequency table used to grasp the general features of respondents' answers regarding the five factors. Table 4 shows that sum of percentage of "4. somewhat considered," and "5. considered" exceeds 50% in four factors other than "Retribution". In "Monetary" and "Mental," sum of percentage of "4. somewhat considered," and "5. considered" is particularly high, and in "Monetary," percentage of "5. considered" exceeded 70%.

Table 4 Frequency Table about the Purpose of Damages

	1. Disregarded	2. Somewhat disregarded	3. Neutral	4. Somewhat considered	5. Considered	Total %
Monetary	0.4	1.1	4.4	24.0	70.1	100.0
Punishment	4.8	15.5	17.7	34.3	27.7	100.0
Mental	1.8	3.3	9.6	40.6	44.6	100.0
Deterrence	6.3	18.8	18.8	33.2	22.9	100.0
Retribution	6.6	19.6	25.1	32.8	15.9	100.0

Note: *N*=271; the numbers in the table are percentages and they are rounded to the first decimal place.

[21] I reversed the scale in the original questionnaire to make it easier to interpret the results of the analysis.

Table 5 shows Pearson correlation coefficients among the five factors.[22] It also shows 95% confidence intervals, *p*-values non-adjusted and adjusted by Holm method. On the one hand, we can see that "Monetary" has statistically significant correlation with "Mental." This is plausible because "Monetary" and "Mental" fall into the same category of the purpose of damages, that is to say, compensation. "Monetary" is compensation for monetary loss and "Mental" is compensation for mental suffering.

On the other hand, "Monetary" does not have significant correlations with "Punishment," "Deterrence," and "Retribution." Respondents consider the former factor and the latter three factors as falling into the different categories of the purpose of damages, in a manner similar to legal scholars. Moreover, "Punishment," "Deterrence," and "Retribution" are significantly correlated. The respondents considered these three as part of the same category.[23]

In addition, "Mental" has a significant correlation with "Punishment," "Deterrence," and "Retribution." Some legal scholars find some punitive nature in damages to compensate for mental suffering.[24] Ordinary people may have similar views.

Table 5 Correlation Coefficients about the Purpose of Damages

	r	95% C.I.	*p*	Adj. *p*
Monetary-Punishment	0.104	[-0.015, 0.221]	0.087	0.218
Monetary-Mental	0.393	[0.287, 0.489]	0.000	0.000
Monetary-Deterrence	0.109	[-0.010, 0.225]	0.073	0.218
Monetary-Retribution	0.031	[-0.088, 0.150]	0.611	0.611
Punishment-Mental	0.488	[0.391, 0.574]	0.000	0.000
Punishment-Deterrence	0.579	[0.494, 0.653]	0.000	0.000
Punishment-Retribution	0.468	[0.370, 0.556]	0.000	0.000
Mental-Deterrence	0.418	[0.314, 0.511]	0.000	0.000
Mental-Retribution	0.395	[0.289, 0.491]	0.000	0.000
Deterrence-Retribution	0.473	[0.375, 0.560]	0.000	0.000

Note: *N*=271.

3. 3. Analysis of relations of necessity

As we see in Section 3.2, "Monetary" does not have statistically significant correlation with "Punishment," "Deterrence," and "Retribution." However, if we depict XY plots, we can find certain relationships between the former and the latter three.

[22] I regard five-point scales as interval scales when calculating Pearson correlation coefficients.
[23] An exploratory factor analysis also shows this fact. When conducting a maximum likelihood factor analysis (in which the number of factors is set to 2 based on the parallel analysis and rotation method being promax with Kaiser normalization), "Deterrence," "Punishment," and "Retribution" are integrated into one factor, and "Monetary" and "Mental" are integrated into the other factor.
[24] For the details about damages to compensate mental suffering, see Mori et al. (2017:594, 603).

Figure 7 is an XY plot between "Monetary" and "Punishment".[25] In this Figure, almost all of the data are on or below the diagonal. This means that most respondents consider "Monetary" and consider "Punishment" to a lesser degree. This relationship can be expressed by the analysis of the relations presented in Section 2.

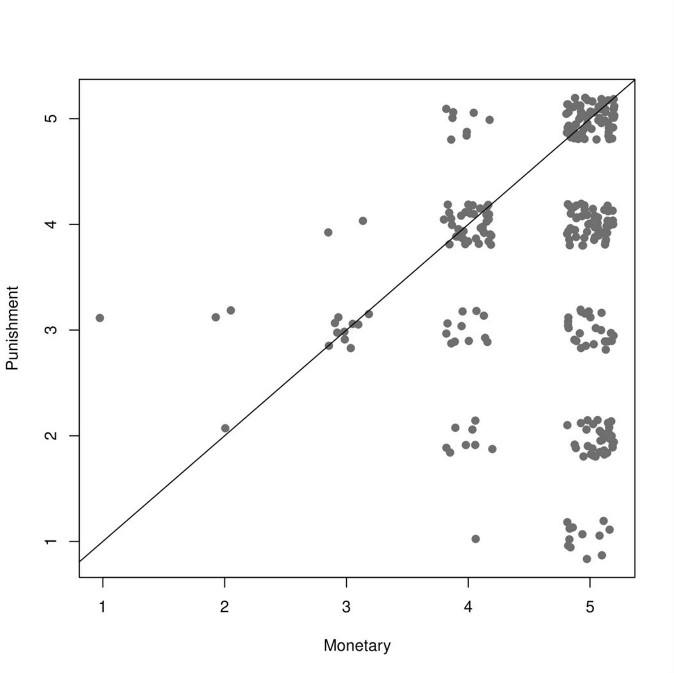

Figure 7 XY Plot between Monetary and Punishment

First, I undertook calibration to transform the values of the five factors into membership scores to sets. I use the sets "consideration of compensation for the monetary loss" (hereinafter, "MONETARY"), "consideration of punishment" (hereinafter, "PUNISHMENT"), "consideration of compensation for mental suffering" (hereinafter, "MENTAL"), "consideration of deterrence" (hereinafter, "DETERRENCE"), and "consideration of satisfying feeling of retribution" (hereinafter, "RETRIBUTION").

I transform "5. considered" into a membership score of 1. Since "4. somewhat considered" is slightly lower than "5. considered," I transform it into a membership score of 0.8. Although "3. neutral" is the middle of the scale, I avoid transforming it into 0.5. One should be careful about assigning the membership score of 0.5 because "doing so represents the weakest possible conceptual statement about that case." (Schneider and Wagemann 2012: 101). I transform "3. neutral" to a membership score of 0.4, which indicates more out of the set than in the set. This is because I focus on the fact that they decided not to choose 4 and 5 when many of the respondents chose them. I transform "1. disregarded" into membership score 0. Since "2. somewhat disregarded" is slightly higher than "1. disregarded," I transformed it into a membership score of 0.2.[26]

[25] I assigned jitters to the values of Monetary and Punishment to clarify the volume of data gathered in the plot.
[26] Emmenegger et al. (2014:8) conducted calibration using a five-point Likert scale. They transformed the scale values 1, 2, 3, 4, and 5 into membership scores 0, 0, 0.2, 0.8, and 1. In the present case, I decided to transform the scale value 2 into 0.2 instead of 0 because I wanted to incorporate the degree in difference between "1. disregarded" and "2. somewhat disregarded". I transform the scale value 3 into 0.4 instead of 0.2 because I assign

After calibration, I calculated the consistency of necessity. Table 6 presents the results of this study. This indicates that MONETARY is a necessary condition for the other four, because their consistency scores exceed the threshold value of 0.9, and they are statistically significant at the 5% level. MENTAL is a necessary condition for PUNISHMENT, DETERRENCE, and RETRIBUTION. If we regard legal scholars' argument that compensation is primary while punishment, deterrence, and retribution are subordinate, as compensation is a necessary condition for (or a superset of) punishment, deterrence, and retribution, we can say that ordinary people adopt similar views as legal scholars.

Table 6 Relation of Necessity about the Purpose of Damages

	Consistency	95% CI	*p*	Adj. *p*	Coverage	95% CI	Relevance	95% CI
MONETARY⊇ PUNISHMENT	0.982	[0.971, 0.991]	0.000	0.000	0.703	[0.660, 0.744]	0.243	[0.188, 0.302]
MONETARY⊇ MENTAL	0.980	[0.968, 0.990]	0.000	0.000	0.876	[0.848, 0.903]	0.435	[0.351, 0.523]
MONETARY⊇ DETERRENCE	0.981	[0.968, 0.991]	0.000	0.000	0.652	[0.610, 0.695]	0.215	[0.167, 0.268]
MONETARY⊇ RETRIBUTION	0.979	[0.965, 0.991]	0.000	0.000	0.601	[0.560, 0.643]	0.193	[0.150, 0.241]
PUNISHMENT⊇ MONETARY	0.703	[0.660, 0.744]	1.000	1.000	0.982	[0.971, 0.991]	0.967	[0.947, 0.984]
PUNISHMENT⊇ MENTAL	0.777	[0.737, 0.815]	1.000	1.000	0.971	[0.954, 0.985]	0.948	[0.918, 0.973]
PUNISHMENT⊇ DETERRENCE	0.883	[0.848, 0.915]	0.834	1.000	0.821	[0.782, 0.859]	0.748	[0.696, 0.801]
PUNISHMENT⊇ RETRIBUTION	0.884	[0.847, 0.918]	0.804	1.000	0.759	[0.717, 0.800]	0.688	[0.633, 0.741]
MENTAL⊇ MONETARY	0.876	[0.848, 0.903]	0.958	1.000	0.980	[0.968, 0.990]	0.919	[0.874, 0.958]
MENTAL⊇ PUNISHMENT	0.971	[0.954, 0.985]	0.000	0.000	0.777	[0.737, 0.815]	0.502	[0.438, 0.567]
MENTAL⊇ DETERRENCE	0.967	[0.948, 0.984]	0.000	0.000	0.720	[0.678, 0.761]	0.445	[0.388, 0.505]
MENTAL⊇ RETRIBUTION	0.984	[0.971, 0.995]	0.000	0.000	0.676	[0.634, 0.717]	0.410	[0.353, 0.467]
DETERRENCE⊇ MONETARY	0.652	[0.610, 0.695]	1.000	1.000	0.981	[0.968, 0.991]	0.971	[0.951, 0.987]
DETERRENCE⊇ PUNISHMENT	0.821	[0.782, 0.859]	1.000	1.000	0.883	[0.848, 0.915]	0.847	[0.801, 0.889]
DETERRENCE⊇ MENTAL	0.720	[0.678, 0.761]	1.000	1.000	0.967	[0.948, 0.984]	0.952	[0.923, 0.976]
DETERRENCE⊇ RETRIBUTION	0.846	[0.807, 0.885]	0.997	1.000	0.781	[0.738, 0.823]	0.747	[0.694, 0.796]
RETRIBUTION⊇ MONETARY	0.601	[0.560, 0.643]	1.000	1.000	0.979	[0.965, 0.991]	0.974	[0.957, 0.989]
RETRIBUTION⊇ PUNISHMENT	0.759	[0.717, 0.800]	1.000	1.000	0.884	[0.847, 0.918]	0.871	[0.831, 0.910]
RETRIBUTION⊇ MENTAL	0.676	[0.634, 0.717]	1.000	1.000	0.984	[0.971, 0.995]	0.980	[0.964, 0.994]
RETRIBUTION⊇ DETERRENCE	0.781	[0.738, 0.823]	1.000	1.000	0.846	[0.807, 0.885]	0.836	[0.795, 0.876]

Note: N=271.

We also consider the percentile probability that multiple relations of necessity MONETARY ⊇ PUNISHMENT, MONETARY ⊇ MENTAL, MONETARY ⊇ DETERRENCE, MONETARY ⊇ RETRIBUTION, MENTAL ⊇ PUNISHMENT, MENTAL ⊇ DETERRENCE, and MENTAL ⊇ RETRIBUTION simultaneously hold. These relationships are approximately depicted by Venn diagrams in Figure 8. This probability was calculated as 0.000. Since it is statistically significant at the 5% level, we conclude that multiple relations in Figure 8 hold.

---

a larger score than that of "2. somewhat disregarded". However, I confirm that the result of our analysis in this paper hardly changes even if we use the same calibration as Emmenegger et al. (2014).

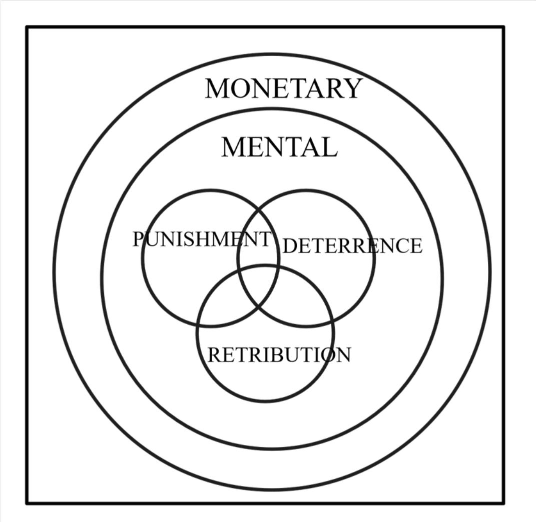

Figure 8 Multiple Relations of Necessity about the Purpose of Damages

Table 6 also shows the coverage of the necessity. This indicates that all of the relations have relatively high coverage scores. The coverage of MENTAL ⊇ MONETARY is high because they are common in terms of compensation.

Table 6 also shows relevance of necessity. This indicates that relevance scores where MONETARY is a necessary condition, such as MONETARY ⊇ PUNISHMENT, are substantially lower than their coverage scores. The frequency table in Table 2 indicates the reason for this. Over 70% choose "5. considered" in MONETARY in Table 2. This makes the size of the set MONETARY close to the universal set, as shown in Figure 4. The fact that MONETARY is a large-sized necessary condition for the other sets might also suggest that ordinary people regard compensation as the main purpose of damages, as we see in the analysis of consistency scores.

4. Conclusion

In this paper, we considered the analysis of the relations of necessity in survey research. The relations of necessity (or subset relations) are different from the correlations. They cannot be measured using correlation coefficients, which have often been used in the first stage of data analysis. We adopted parameters of fit such as consistency, coverage, and relevance in QCA as indicators to measure the relations of necessity. We also proposed bootstrap methods to construct confidence intervals and conduct hypothesis testing. We also dealt with the problem of multiple-hypothesis testing. The analysis of relations of necessity can highlight relations that have not been found by conventional statistical analysis based on correlations.

As an example of an application, we analyzed the survey data on the purpose of tort damages in the case of a defective car accident in Japan. In this survey, respondents were asked about the extent to which they considered five factors in deciding the amount of damages: compensation for monetary loss, punishment, compensation for mental suffering, deterrence, and satisfying the feeling of retribution. We found that although compensation for monetary loss is not correlated with punishment, deterrence, and satisfying the feeling of retribution, compensation for monetary loss is a necessary condition for them. We also found that compensation for monetary loss is a necessary condition for compensation for mental suffering, and compensation for mental suffering is a necessary condition for punishment, deterrence, and retribution.

These views of ordinary people are consistent with that of legal scholars, that compensation is the main purpose while the others are merely subordinate. This might have an impact on the tort law system in Japan. For example, the introduction of punitive damage might be premature as legal scholars as well as ordinary people consider punishment and deterrence to be a subordinate purpose of damages.

Although the views of legal scholars and ordinary people can affect the tort law system in Japan, there has been little research on the latter's views on the purpose of damages. This study's analysis might help improve this situation.

There are some other points to be noted. Although bootstrap methods are useful tools for statistical inference in the analysis of relations of necessity, they are meaningful only when the data are randomly sampled. Statistical inference is a method used to infer the properties of a population based on a random sample from that population. One should be careful when applying the method in this study to data that are not randomly sampled. One should also be careful that statistical inference by the bootstrap method might fail when the data has some problems, such as incomplete data that is censored and dirty data with outliers (Davison and Hinkley 1997:43-44).

Some important issues were not considered in this study. As for the analytical method, we did not deal with analysis of sufficiency in this paper.[27] Even in the randomly sampled survey data, if one is interested in the relationship of sufficiency, they can incorporate some aspects of QCA into the analysis. The bootstrap methods adopted in this study can be useful in the analysis of sufficiency because of its simplicity and versatility.[28] As for the analysis of survey data used in this study, it is important to compare the results of different countries. For example, we should compare the data of Japan and the United States because it is said that the United States, which has a punitive damages system, places more stress on punishment and deterrence than Japan.[29] Extending our method of dealing with these issues is an important future research step.

References

Abraham, Kenneth S. (2017) *The Forms and Functions of Tort Law*. 5th ed. Foundation Press.

American Law Institute (1979) *Restatement (Second) of Torts*. American Law Institute Publishers.

Braumoeller, Bear F., 2015, "Guarding Against False Positives in Qualitative Comparative Analysis." *Political Analysis* 23, 471–487.

Bretz, Frank, Torsten Hothorn, and Peter Westfall (2011) *Multiple Comparisons Using R*. CRC Press.

[27] However, it was implicitly addressed when discussing the coverage of necessity. The formula for the coverage of necessity in Section 2.3 is mathematically identical to the formula for the consistency of sufficiency (Schneider and Wagemann 2012: 139).

[28] There are some previous researchers who have applied bootstrap methods to the analysis of sufficiency in QCA-for example, Cooper and Glaesser (2016) and Gibson and Vann (2016).

[29] Legal scholars in the United States do not necessarily consider compensation to be the main purpose of damages. For example, Abraham (2017:21), who wrote the textbook of tort law in the United States, states, "[i]t is sometimes said that a function of tort law is to promote the compensation of those who have suffered injury. For most analysts of tot law, this is only true in a very limited sense."

Cooper, Barry and Judith Glaesser (2016) “Exploring the Robustness of Set Theoretic Findings from a Large N fsQCA: an Illustration from the Sociology of Education.” *International Journal of Social Research Methodology* 19(4), 445–459.

Davison, Anthony C. and David V. Hinkley (1997) *Bootstrap Methods and Their Application*. Cambridge University Press.

Dul, Jan (2016) “Necessary Condition Analysis (NCA): Logic and Methodology of ‘Necessary but not Sufficient’ Causality.” *Organizational Research Methods* 19(1), 10–52.

Dul, Jan (2019) *Conducting Necessary Condition Analysis for Business and Management Students*. SAGE Publications.

Dul, Jan, Erwin van der Laan, and Roelof Kuik (2020) “A Statistical Significance Test for Necessary Condition Analysis.” *Organizational Research Methods* 23(2) 385–395.

Efron, Bradley and R.J. Tibshirani (1993) *An Introduction to the Bootstrap*. Chapman and Hall/CRC.

Eliason, Scott R. and Robin Stryker (2009) “Goodness-of-Fit Tests and Descriptive Measures in Fuzzy-Set Analysis.” *Sociological Methods & Research* 38(1), 102–146.

Emmenegger, Patrick, Dominik Schraff, and Andre Walter (2014) “QCA, the Truth Table Analysis and Large-N Survey Data: The Benefits of Calibration and the Importance of Robustness Tests.” COMPASSS Working Paper 2014-79. <http://www.compasss.org/wpseries/EmmeneggerSchraffWalter2014.pdf> (accessed on March 31, 2021).

Gibson, C. Ben and Burrel Vann Jr. (2016) “The Bootstrapped Robustness Assessment for Qualitative Comparative Analysis.” <https://arxiv.org/abs/1606.05000> (accessed on March 31, 2021).

Goertz, Gary and Harvey Starr ed. (2003) *Necessary Conditions: Theory, Methodology, and Applications*. Rowman & Littlefield Publishers.

Justice System Reform Council (2001) “Proceedings of the 54th Justice System Reform Council.” <http://www.kantei.go.jp/jp/sihouseido/dai54/54gijiroku.html> (in Japanese) (accessed on March 31, 2021).

Kihara, Hiroyuki (2015) *Tort Law in Japan*. Wolters Kluwer Law and Business.

Mori, Daisuke, Shuichi Takahashi, and Yasuhiro Ikeda (2017) “Compensation, Punishment, and Deterrence: Survey on the Purpose of Tort Damages in the Case of a Defective Car Accident in Japan.” *Asia-Pacific Journal of Regional Science* 1(2), 589–624.

Ragin, Charles C. (2006) “Set Relations in Social Research: Evaluating Their Consistency and Coverage.” *Political Analysis* 14(3), 291–310.

Ragin, C. Charles, David Shulman, Adam Weinberg, and Brian Gran (2003) “Complexity, Generality, and Qualitative Comparative Analysis.” *Field Methods* 15, 323–340.

Rihoux, Benoîx (2003) “Bridging the Gap between the Qualitative and Quantitative Worlds? A Retrospective and Prospective View on Qualitative Comparative Analysis.” *Field Methods* 15, 351–365.

Rihoux, Benoîx and Charles C. Ragin (2008) *Configurational Comparative Methods: Qualitative Comparative Analysis (QCA) and Related Techniques.* Sage Publications, Inc.

Schneider, Carsten Q. and Claudius Wagemann (2012) *Set-Theoretic Methods for the Social Sciences.* Cambridge

University Press.
Vis, Barbara and Jan Dul (2018) "Analyzing Relationships of Necessity Not Just in Kind but Also in Degree: Complementing fsQCA with NCA." *Sociological Methods & Research* 47(4), 872–899.
Yamamoto, Keizo (2019) *Basic Features of Japanese Tort Law*. Jan Sramek Verlag KG.

Appendix. R script of the analysis in Section 2

```
# Preparing data
test1 <- c(5,3,4,5,2,3,4,5,3,3,4,3,4,4)
test2 <- c(0,0,0,0,1,1,3,4,2,3,4,3,4,5)
test3 <- c(5,4,5,4,4,5,5,5,4,5,4,4,5,5)
test <- data.frame(test1,test2,test3)

# Section 2.1
## Pearson correlation coefficient
round(cor(test$test1,test$test2),3)

# Section 2.3
## Calibration
test$test1f <- test$test1/5
test$test2f <- test$test2/5
test$test3f <- test$test3/5

## Consistency of necessity
consistency <- function(data,x,y) sum(apply(data[,c(x,y)],1,min))/sum(data[,y])
round(consistency(data=test,x="test1f",y="test2f"),3)

## Coverage of necessity
coverage <- function(data,x,y)sum(apply(data[,c(x,y)],1,min))/sum(data[,x])
round(coverage(data=test,x="test1f",y="test2f"),3)

## Relevance of necessity
relevance <- function(data,x,y)sum(1-data[,x])/sum(1-apply(data[,c(x,y)],1,min))
round(relevance(data=test,x="test1f",y="test2f"),3)

# Section 2.4
```

```
## Section 2.4.1
### Confidence interval of correlation coefficient
library(psych)
round(corr.test(test$test1,test$test2,adjust="none",alpha=.05)$ci,3)

### Bootstrap
set.seed(1)
B <- 10000
no_resample <- sample(1:nrow(test), nrow(test)*B, replace=TRUE)
test_resample <- test[no_resample,]
no_group <- gl(B, nrow(test))

### Confidence interval of consistency of necessity
bootcon12                                                                    <-
by(test_resample[,c("test1f","test2f")],no_group,consistency,"test1f","test2f")
round(quantile(bootcon12,c(0.025,0.975)),3)

### Confidence intervals of coverage and relevance
bootcov12                                                                    <-
by(test_resample[,c("test1f","test2f")],no_group,coverage,"test1f","test2f")
round(quantile(bootcov12,c(0.025,0.975)),3)
bootrel12                                                                    <-
by(test_resample[,c("test1f","test2f")],no_group,relevance,"test1f","test2f")
round(quantile(bootrel12,c(0.025,0.975)),3)

## Section 2.4.2
### The p-value of consistency of necessity
round(sum(bootcon12<0.9)/B,3)

## Section 2.4.3
### Analysis of correlation in case of multiple testings
round(corr.test(test[,1:3],adjust="none",alpha=.05)$ci,3)
round(corr.test(test[,1:3],adjust="holm",alpha=.05)$p,3)

### Consistency of necessity in case of multiple testings
#### Consistency scores
```

```
round(consistency(data=test,x="test1f",y="test2f"),3)
round(consistency(data=test,x="test1f",y="test3f"),3)
round(consistency(data=test,x="test2f",y="test1f"),3)
round(consistency(data=test,x="test2f",y="test3f"),3)
round(consistency(data=test,x="test3f",y="test1f"),3)
round(consistency(data=test,x="test3f",y="test2f"),3)

#### Confidence intervals and p-values of consistency
bootcon12_ci <- quantile(bootcon12,c(0.025,0.975))
bootcon12_p <- sum(bootcon12<0.9)/B
bootcon13                                                                              <-
by(test_resample[,c("test1f","test3f")],no_group,consistency,"test1f","test3f")
bootcon13_ci <- quantile(bootcon13,c(0.025,0.975))
bootcon13_p <- round(sum(bootcon13<0.9)/B,3)
bootcon21                                                                              <-
by(test_resample[,c("test2f","test1f")],no_group,consistency,"test2f","test1f")
bootcon21_ci <- quantile(bootcon21,c(0.025,0.975))
bootcon21_p <- sum(bootcon21<0.9)/B
bootcon23                                                                              <-
by(test_resample[,c("test2f","test3f")],no_group,consistency,"test2f","test3f")
bootcon23_ci <- quantile(bootcon23,c(0.025,0.975))
bootcon23_p <- sum(bootcon23<0.9)/B
bootcon31                                                                              <-
by(test_resample[,c("test3f","test1f")],no_group,consistency,"test3f","test1f")
bootcon31_ci <- quantile(bootcon31,c(0.025,0.975))
bootcon31_p <- sum(bootcon31<0.9)/B
bootcon32                                                                              <-
by(test_resample[,c("test3f","test2f")],no_group,consistency,"test3f","test2f")
bootcon32_ci <- quantile(bootcon32,c(0.025,0.975))
bootcon32_p <- sum(bootcon32<0.9)/B
round(rbind(bootcon12_ci,bootcon13_ci,bootcon21_ci,bootcon23_ci,bootcon31_ci,bootcon32_
ci),3)
bootcon_p <- c(bootcon12_p,bootcon13_p,bootcon21_p,bootcon23_p,bootcon31_p,bootcon32_p)
round(bootcon_p,3)

#### The p-value adjusted by Holm method
```

```
round(p.adjust(bootcon_p,method="holm"),3)


#### Percentile probability that multiple relations simultaneously hold
round(sum(bootcon31<0.9 | bootcon12<0.9 | bootcon32<0.9)/B,3)


### [Calculations of coverage and relevance are omitted]
```