Similarity and the Trustworthiness of Distributive Judgments
Appendix 2

Alex Voorhoeve

Arnaldur Smári Stefánsson

Brian Wallace

A2.1 Additional Results

A2.1.1 Preference for the dominant option

Out of 82 subjects, 76 chose the dominant option in all three trials and three subjects chose the dominant option in two out of three trials. Furthermore, one subject only chose the dominant option once out of three trials and two subjects never chose the dominant option. Let π be the probability of a random subject preferring the dominant option in question N v O. If subjects would choose at random we would expect $\pi = 0.5$, that is, subjects would be as likely to choose the dominant option as the dominated option. Therefore, consider the following hypotheses, which can be tested using the exact binomial goodness-of-fit test (equation (A2.1)):

$$H_0 : \pi = 0.5,$$

$$H_1 : \pi \neq 0.5.$$

The null-hypothesis has a p-value of $p_E^{BI} = 0.000$. Therefore, we reject the null hypothesis and conclude that subjects do not choose at random.

A2.1.2 Frequency of choices to aid the better off

Table A2.1 shows the percentage of all *choices* that favour the better off. As mentioned in the main text, the resulting pattern is very similar to the pattern generated when we consider the percentage of all *subjects* who favour the better off (see Table 2 in the paper).

Table A2.1. Choices expressing a preference for aiding the better off, in percent

Choices		Percentage of choices ($n = 79{\cdot}3$) expressing preference for better off
Similar along utility gain dimension; aiding better off lowers total utility	A v B	57.4
	B v C	53.6
	C v D	51.9
	D v E	39.7
Possible dissimilarity along both dimensions; aiding better off lowers total utility	A v C	41.4
	A v D	33.8
	B v D	30.0
	B v E	30.4
	C v E	28.7
Wholly dissimilar; aiding better off lowers total utility	A v E	25.3
	S v R	16.0
	U v T	19.4
	W v V	16.5
Wholly dissimilar; aiding better off raises total utility	G v F	17.3
	Q v P	23.6

Note: $n = 79{\cdot}3$. In the pairwise choices in the second column, the alternative which involves aiding the better off is always listed first.

Using McNemar's exact test (see section A2.2.4 for an explanation of this test), we examine whether the rate of aiding the better off differs between adjacent pairs in the sequence A through E. Table A2.2 reports the results. We cannot reject the hypothesis that the rate of aiding the better off is the same throughout A versus B, B versus C and C versus D. This confirms the hypothesis that, in each of these pairwise comparisons, only the gain dimension is regarded as similar. But we can reject, at the 5% confidence level, the hypothesis that this rate is the same in D versus E as it is for these other paired alternatives. As mentioned in the main text, we conjecture that this is because in this choice, a substantial number of subjects find the alternatives similar along *both* dimensions, which means the similarity heuristic does not make a judgment at Stage 2, but instead moves to Stage 3, at which individuals (we assume) have preferences in line with common theories of distributive justice, and therefore choose E.

Table A2.2. Comparison of the distribution of choices between alternatives similar along the health gain dimension

		B v C Better off		*B v C* Worse off	*C v D* Better off		*C v D* Worse off	*D v E* Better off		*D v E* Worse off
A v B	Better off	40.5		17.7	39.2		19.0	29.1		29.1
			0.69			0.42			0.01^{***}	
	Worse off	13.9		27.8	12.7		29.1	8.9		32.9
B v C	Better off				39.2		15.2	32.9		21.5
						0.83			0.01^{***}	
	Worse off				12.7		32.9	5.1		40.5
C v D	Better off							31.6		20.3
									0.03^{**}	
	Worse off							6.3		41.8

Note: n= 79. Numbers in the comparisons of distributions across (aiding the better off at a cost in total utility, aiding the worse off at a gain in total utility) give percentages of the total population. Numbers in the grey boxes give the probability p of obtaining the observed results, or anything more extreme, if the answers come from the same distribution according to McNemar's exact test. We cannot reject the null hypothesis that A v B, B v C, and C v D are drawn from the same distribution, but can reject the hypothesis that responses to D v E come from the same distribution as other responses.
$^{***}p$>.01
$^{**}p$>.05
$^{*}p$>.10

Again using McNemar's test, Table A2.3 examines the evidence for our prediction that subjects will switch from favouring the better off in choices between alternatives that are similar along the gain dimension to favouring the worse off in choices between wholly dissimilar alternatives. The underlined numbers in the top-right-hand corner of every comparison indicate the predicted switch. Throughout, this switch is substantial; moreover, there is no comparable switch in the opposite direction. For example, 50.6% of all subjects switch from aiding the better off in a choice between A and B to aiding the worse off in a choice between S and R, while only 5.1% switch in the other direction.

The grey boxes report the probability of finding this pattern if the underlying rate of aiding the better off were the same. We can reject this hypothesis at the 1% significance level for all comparisons but one, and at the 5% level for the remaining comparison.

Table A2.3. Comparison of the distribution of choices between alternatives similar along the health gain dimension

		A v E			*S v R*			*U v T*			*W v V*		
		Better off		Worse off	Better off		Worse off	Better off		Worse off	Better off		Worse off
A v B	Better off	21.5	0.00***	36.7	7.6	0.00***	50.6	7.6	0.00***	50.6	11.4	0.00***	46.8
	Worse off	1.3		40.5	5.1		36.7	6.3		35.4	2.5		39.2
B v C	Better off	17.7	0.00***	36.7	8.9	0.00***	45.6	8.9	0.00***	45.6	12.7	0.00***	41.8
	Worse off	5.7		40.5	3.8		41.8	5.1		40.5	1.3		44.3
C v D	Better off	20.3	0.00***	31.6	11.4	0.00***	40.5	10.1	0.00***	41.8	13.9	0.00***	38.0
	Worse off	2.5		45.6	1.3		46.8	3.8		44.3	0.0		48.1
D v E	Better off	15.2	0.02**	22.8	5.1	0.00***	32.9	6.3	0.00***	31.6	10.1	0.00***	27.8
	Worse off	7.6		54.4	7.6		54.4	7.6		54.4	3.8		58.2

Note: $n=79$. Numbers in the comparisons of distributions across (aiding the better off at a cost in total utility, aiding the worse off at a gain in total utility) are percentages of the total population. Underlined numbers represent the predicted shift from aiding the better off when choosing between partly similar alternatives to aiding the worse off when choosing between wholly dissimilar alternatives. Numbers in the grey boxes give the probability p of obtaining the observed results, or anything more extreme, if the answers come from the same distribution according to McNemar's exact test. We can confidently reject the hypothesis that responses to partly similar alternatives and responses to wholly dissimilar alternatives come from the same distribution.

***p>.01
**p>.05
*p>.10

A2.1.3 In choices between dissimilar alternatives, a large majority consistently choose to aid the worse off

Binomial point estimates and confidence intervals

For questions S v R, U v T and W v V, let π be the probability of aiding the worse off at a gain in total utility, and for questions G v F and Q v P, let it be the probability of aiding the worse off at a cost in total utility. Table A2.4 shows for each question q

the estimate for π using the maximum likelihood estimator expressed in equation (A2.3) and the exact Clopper and Pearson binomial confidence interval expressed in equation (A2.4). Between 83% and 85% aid the worse off when this also improves total utility, in S v R, U v T and W v V, and 76% of subjects do so in all three questions. When aiding the worse off comes at a cost, 83% displaying priority for the worse off in G v F, 73% in Q v P, and 70% in both questions. Finally, 66% of subjects aid the worse off in all five choices. The 99% lower bound is greater than 0.5 for all probabilities.

Correlation

Table A2.5 shows the positive correlations between subjects aiding the worse off in S v R, U v T, W v V, G v F and Q v P.[1] The correlation between subjects aiding the worse off in all three S v R, U v T and W v V, and displaying priority for the worse off in both G v F and Q v P is 0.64.

Table A2.4. Proportion of subjects aiding the worse off

		99% Confidence	
Questions	$\hat{\pi}$	Lower	Upper
S v R	0.85	0.73	0.94
U v T	0.83	0.70	0.92
W v V	0.84	0.71	0.93
S v R, U v T, W v V	0.76	0.62	0.87
G v F	0.83	0.70	0.92
Q v P	0.73	0.59	0.85
G v F, Q v P	0.70	0.55	0.82
S v R, U v T, W v V, G v F, Q v P	0.66	0.51	0.79

Note: $n = 79$. Maximum likelihood estimator (equation (A2.3)), with Clopper and Pearson confidence interval (equation (A2.4)).

[1] The sample Pearson correlation coefficient is

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}}.$$

Table A2.5. Correlation between questions

	U v T	W v V	G v F	Q v P
S v R	0.62	0.51	0.51	0.50
U v T		0.68	0.47	0.46
W v V			0.58	0.29
G v F				0.46

Note: n= 79. Correlations between aiding the worse off in S v R, U v T, W v V, G v F and Q v P.

A2.1.4 Choices between dissimilar alternatives

Table A2.6 shows the frequencies from the choices between dissimilar alternatives and the p-values from the null hypothesis that choices come from the same distribution. Since the p-values are close to one, we cannot reject the hypothesis that choices from these questions come from the same distribution.

Table A2.6. Comparison of the distribution of choices between dissimilar alternatives

		U v T: Better off	U v T	U v T: Worse off	W v V: Better off	W v V	W v V: Worse off
S v R	Better off	8.9		3.8	7.6		5.1
			1.00			1.00	
	Worse off	5.1		82.3	6.3		81.0
U v T	Better off				10.1		3.8
						1.00	
	Worse off				3.8		82.3

Note: n= 79. Numbers in the comparisons of distributions across (aiding the better off at a cost in total utility, aiding the worse off at a gain in total utility) give percentages of the total population. Numbers in the grey boxes give the probability p of obtaining the observed results, or anything more extreme, if the answers come from the same distribution according to McNemar's exact test. We cannot reject any null hypothesis.

A2.1.5 Matching individuals with decision rules

As mentioned in the main text, when matching subjects with the decision rule that best fits their behaviour, the similarity heuristic may benefit unduly from the variability it allows in individuals' perceptions of similarity. In the main text, we attenuate this problem by allowing only two forms of similarity judgments. Here, as a robustness check, we consider a version of the similarity heuristic that imposes the following uniform perceptions of (dis)similarity across all subjects: Only and all adjacent alternatives in the A through E sequence are similar along the gain dimension. This is a very demanding test of this heuristic, since some diversity in individual perceptions of similarity is to be expected, and doesn't imply that individuals do not use the heuristic. Table A2.7 below shows the results. Unsurprisingly, the share of subjects whose behaviour best matches this uniform version of the similarity heuristic is somewhat less than the 41.8% that best fit the version considered in the main text. At 36.7% it is roughly on a par with "always aid the worse off", which is the uniquely best match for 35.4% of the population (and tied for best match for a further 5.1%). As in the version discussed in the main text, the similarity heuristic remains a reasonably good explanation of the choices of individuals who were matched with it: it gets 77.5% of their choices right, which is 10.0% better than the success rate we would achieve if we could not appeal to the similarity heuristic to explain these subjects' choices.

Table A2.7. Matching subjects with decision rules when imposing uniform similarity judgments

Rule	Share (%)	Fit (%)	Fit premium (%)
Similarity heuristic	36.7	77.5	10.0
When no similarity:			
Worst off	35.4	77.9	
Total utility	1.3	66.7	
Worse off	35.4	88.8	8.3
Greater number	17.7	74.8	27.0
Total utility	5.1	79.2	4.8
Worse off/total utility (tie)	5.1	76.8	n.a.

Note: $n=$ 79, with 42 choices per person. "Fit" is the share of choices (in those subjects in whose behaviour it fits best) consistent with the rule in question. The "fit premium" is the difference between the share of these subjects' choices explained by the given rule and the share of these subjects' choices explained by the next-best-fitting rule.

A2.2 Methodology

A2.2.1 Notation

Using a probabilistic version of the preference relation (see Tversky (1969)), let $\Pr(0,1)$ be the probability of choosing 0 when choosing between 0 and 1. Hence, $\Pr(0,1)+\Pr(1,0) = 1$, and 0 is said to be preferred to 1 if is chosen more often than 1:

$$0 \succ 1 \quad \text{if and only if} \quad \#(0) > \#(1).$$

This definition for preferences implies that for each question, instead of three, we only have one observation on each subject. Therefore, for the analysis, each subject i is given one observation for each question q:

$$y_{i,q} = \begin{cases} 0 & \text{if } \#(0) > \#(1), \\ 1 & \text{if } \#(0) < \#(1). \end{cases}$$

A2.2.2 Goodness-of-fit

Exact binomial and multinomial tests for goodness-of-fit

Let π be the probability of $y_{i,q} = 1$, and x be the number of times that $y_{i,q} = 1$ in the obtained data:

$$\pi = \Pr(y_{i,q} = 1),$$
$$x = \#(y_{i,q} = 1).$$

Consider the null hypothesis that the true probability is π^0, and the alternative hypothesis that it is not π^0:

$$H_0 : \pi = \pi^0,$$
$$H_1 : \pi \neq \pi^0.$$

The p-value from the exact binomial goodness-of-fit test is the probability of observing the obtained data, or any other data that is less likely to occur under the null hypothesis:[2]

$$p^{BI} \equiv \sum_{z:\Pr(z \mid n,\pi^0) \leq \Pr(x \mid n,\pi^0)} Pr(z \mid n, \pi^0). \tag{A2.1}$$

Exact trinomial tests for goodness-of-fit for $H_0 : \pi_1 = \pi_2 = \pi^0_{12}$

As explained in Section 4.4 of the main text, to assess whether intransitivities are more frequent in the direction explicable by use of the similarity heuristic, we need a trinomial test. Let $y_{i,T}$ be a trinomial random variable taking values $k = 1, 2, 3$. Furthermore, let π_k be the probability of $y_{i,T}$ taking the value k and, and x_k be the number of times that $y_{i,T}$ takes the value k in the obtained data:

$$\boldsymbol{\pi} = \begin{pmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \end{pmatrix}, \quad \pi_k = \Pr(y_{i,T} = k),$$

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \quad x_k = \#(y_{i,T} = k).$$

The null hypothesis we want to test is that $\pi_1 = \pi_2 = \pi^0_{12}$.[3] There is a number of ways this can be tested, since π^0_{12} can be any number between 0 and 0.5. One approach would be to construct a test statistic that integrates over values of π^0_{12}, giving each equal weight. Doing so would, however most likely give very small p-values since the data we have observed will be very unlikely given a number of values for π^0_{12}. The most conservative test statistic is constructed by picking π^0_{12} such that the the probability of observing the

[2]For the binomial distribution, the probability of an outcome that occurs with probability π occurring x times in n independent observations, is (Forbes et al., 2011)

$$f(x; n, \pi) = \Pr(x \mid n, \pi) = \begin{pmatrix} n \\ x \end{pmatrix} \pi^x (1 - \pi)^{n-x}.$$

[3]We will use a two-sided test, which is more conservative than a one-sided test and more straightforward to carry out. In our sample $x_1 > x_2$ and therefore, if we reject $\pi_1 = \pi_2$ we can conclude that $\pi_1 > \pi_2$.

obtained data is maximized. Now, if $\pi_1 = \pi_2 = \pi_{12}^0$, and therefore $\pi_3 = 1 - 2\pi_{12}^0$, the probability of observing the obtained data $\boldsymbol{x'} = (x_1, x_2, x_3)$ is

$$\Pr(\boldsymbol{x} \mid \pi_1 = \pi_2 = \pi_{12}^0) = a(\pi_{12}^0)^b(1 - 2\pi_{12}^0)^c \equiv P(\pi_{12}^0),$$

where $a \equiv \frac{n!}{x_1!x_2!x_3!}$, $b \equiv x_1 + x_2$ and $c \equiv x_3$.[4] The most conservative p-value is therefore found by maximising $P(\pi_{12}^0)$ with respect to π_{12}^0. That is, by setting $\hat{\pi}_{12}^0 = \frac{x_1+x_2}{2n}$ and $1 - 2\hat{\pi}_{12}^0 = \frac{x_3}{n}$.[5] Therefore, our hypothesis formally reads:

$$\begin{aligned} &H_0 : \pi = \boldsymbol{\hat{\pi}_{12}^0} = (\hat{\pi}_{12}^0, \hat{\pi}_{12}^0, 1 - 2\hat{\pi}_{12}^0) \equiv \arg\max_{\pi_{12}^0} P(\pi_{12}^0), \\ &H_1 : \pi \neq \boldsymbol{\hat{\pi}_{12}^0}, \end{aligned}$$

and the p-value can be formally stated as:

$$\begin{aligned} p_{\boldsymbol{\hat{\pi}_{12}^0}}^{TRI} &\equiv \sum_{\boldsymbol{z}:\Pr(\boldsymbol{z} \mid n, \boldsymbol{\hat{\pi}_{12}^0}) \leq \Pr(\boldsymbol{x} \mid n, \boldsymbol{\hat{\pi}_{12}^0})} Pr(\boldsymbol{z} \mid n, \boldsymbol{\hat{\pi}_{12}^0}), \\ \boldsymbol{\hat{\pi}_{12}^0} &= (\hat{\pi}_{12}^0, \hat{\pi}_{12}^0, 1 - 2\hat{\pi}_{12}^0) \equiv \arg\max_{\pi_{12}^0} P(\pi_{12}^0). \end{aligned} \tag{A2.2}$$

[4] For the trinomial distribution, the probably of an outcome $\boldsymbol{x}$ under the probability vector $\boldsymbol{\pi}$, is (Forbes et al., 2011)

$$f(\boldsymbol{x}; n, \boldsymbol{\pi}) = \Pr(\boldsymbol{x} \mid n, \boldsymbol{\pi}) = n! \prod_{k=1}^{3} \frac{\pi_k^{x_k}}{x_k!}.$$

[5] Find the value $\hat{\pi}_{12}^0$ such that the probability of obtaining the observed data, x_1, x_2 and x_3, is maximised, constrained on $\pi_1 = \pi_2 = \pi_{12}^0$. That is

$$0 = \frac{\partial P(\pi_{12}^0)}{\partial \pi_{12}^0}\Big|_{\pi_{12}^0 = \hat{\pi}_{12}^0} = ab(\hat{\pi}_{12}^0)^{b-1}(1 - 2\hat{\pi}_{12}^0)^c - 2ac(\hat{\pi}_{12}^0)^b(1 - 2\hat{\pi}_{12}^0)^{c-1}.$$

Rearranging, we see that

$$\begin{aligned} \hat{\pi}_{12}^0 &= \frac{b}{2(b+c)} = \frac{x_1 + x_2}{2(x_1 + x_2 + x_3)} = \frac{x_1 + x_2}{2n}, \\ 1 - 2\hat{\pi}_{12}^0 &= 1 - 2\frac{x_1 + x_2}{2n} = 1 - 2\frac{n - x_3}{2n} = \frac{x_3}{n}. \end{aligned}$$

A2.2.3 Exact binomial estimation

Maximum likelihood estimation, the $\hat{\pi}$ that maximizes the probability of observing the n independent observations, can be used to estimate the π:[6]

$$\hat{\pi} = \frac{x}{n}. \tag{A2.3}$$

The Clopper and Pearson (Clopper and Pearson, 1934) exact confidence interval is

$$\left\{\pi \mid \sum_{j=0}^{x} f(j;n,\pi) > \frac{\alpha}{2}\right\} \cap \left\{\pi \mid \sum_{j=x}^{n} f(j;n,\pi) > \frac{\alpha}{2}\right\}, \tag{A2.4}$$

where $\sum_{j=0}^{x} f(j;n,\pi)$ is the probability that a binomial random variable with probability of success π has x or less successes in n trials and $\sum_{j=x}^{n} f(j;n,\pi)$ is the probability that the same random variable has x or more successes in n trials. Therefore, the interval includes all values between the lowest π such that the probability of obtaining x or more successes in n trials is $\alpha/2$ and the largest π such that the probability of obtaining x or less successes in n trials is at least $\alpha/2$. It is conservative in the sense that the coverage probability is at least $1-\alpha$, and it is sometimes said to be unnecessarily conservative (Newcombe, 1998).

A2.2.4 Choices drawn from same distribution

Consider a sample of n randomly selected subjects. Each subject i is examined under two different scenarios, that is, two different questions, $q = q_a, q_b$. We want to know if the question affects the choice. The McNemar test is appropriate, since each subjects is

[6]The log-likelihood function is

$$\ln \mathcal{L}(\pi|x) = \ln \binom{n}{x} + x \ln \pi + (n-x)\ln(1-\pi),$$

which is maximized at

$$0 = \frac{\partial}{\partial \pi} \log \mathcal{L}(\pi|x)\Big|_{\pi=\hat{\pi}} = \frac{x}{\hat{\pi}} - \frac{n-x}{1-\hat{\pi}}.$$

Rearranging we get (A2.3).

observed twice.[7] Let x_{lk} be the number of subjects choosing l in q_a and k in q_b, see Table A2.8.

Table A2.8. McNemar contingency table

		q_b 0	q_b 1
q_a	0	x_{00}	x_{01}
	1	x_{10}	x_{11}

What is of interest are x_{01} and x_{10} since they represent subjects who responded differently under the two experimental conditions. Let x^p_{01} and x^p_{10} denote the frequencies of x_{01} and x_{10} in the underlying population, and

$$\pi_{01} \equiv \frac{x^p_{01}}{x^p_{01} + x^p_{10}},$$
$$\pi_{10} \equiv \frac{x^p_{10}}{x^p_{01} + x^p_{10}}.$$

If there is no difference in choices between the experimental conditions, these two probabilities should be the same, $\pi_{01} = \pi_{10} = 0.5$. Therefore, the null hypothesis and the alternative hypothesis are the following:

$$H_0 : \pi_{01} = \pi_{10} = 0.5,$$
$$H_1 : \pi_{01} \neq \pi_{10}.$$

That is, the null hypothesis is that the probability of *giving response 1 in the first question and 0 in the second question* is the same as the probability of *giving response 0 in the first question and 1 in the second question.*

Let $\overline{x} \equiv \max(x_{01}, x_{10})$, $\underline{x} \equiv \min(x_{01}, x_{10})$ and $m \equiv x_{01} + x_{10}$. If the true probabilities are $\pi_{01} = \pi_{10} = 0.5$, then the likelihood of obtaining a frequency of $\overline{x}$ or greater in either x_{01} or x_{10}, which is the same as the likelihood of obtaining a frequency of $\underline{x}$ or less in

[7]See for example Hoffman (1976) and Sheskin (2003, p. 507). Amongst authors using McNemar test are Rutström and Williams (2000),Faravelli (2007) and Manzini and Mariotti (2010).

either x_{01} or x_{10}, is

$$\begin{aligned}\Pr(x_{01} \text{ or } x_{10} \geq \overline{x} \mid \pi_{01} = \pi_{10} = 0.5) &= \Pr(x_{01} \text{ or } x_{10} \leq \underline{x} \mid \pi_{01} = \pi_{10} = 0.5) \\ &= \sum_{j=\overline{x}}^{m} \begin{pmatrix} j \\ m \end{pmatrix} (0.5)^j (0.5)^{m-j}.\end{aligned}$$

Using the null and alternative hypothesis described above, the p-value (the probability of obtaining the observed data, or anything more extreme, given the null) is

$$p^M = 2 \cdot \sum_{j=\overline{x}}^{m} \begin{pmatrix} j \\ m \end{pmatrix} (0.5)^j (0.5)^{m-j}. \tag{A2.5}$$

References

Clopper, C. and Pearson, E. S. (1934). The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika*, 26(4):404–413.

Faravelli, M. (2007). How context matters: A survey based experiment on distributive justice. *Journal of Public Economics*, 91(7):1399–1422.

Forbes, C., Evans, M., Hastings, N., and Peacock, B. (2011). *Statistical distributions.* John Wiley & Sons.

Hoffman, J. (1976). The incorrect use of chi-square analysis for paired data. *Clinical and experimental immunology*, 24(1):227–229.

Manzini, P. and Mariotti, M. (2010). Revealed preferences and boundedly rational choice procedures: an experiment. Unpublished.

Newcombe, R. G. (1998). Two-sided confidence intervals for the single proportion: comparison of seven methods. *Statistics in medicine*, 17(8):857–872.

Rutström, E. E. and Williams, M. B. (2000). Entitlements and fairness: an experimental study of distributive preferences. *Journal of Economic Behavior and Organization*, 43(1):75–89.

Sheskin, D. J. (2003). *Handbook of parametric and nonparametric statistical procedures.* CRC Press.

Tversky, A. (1969). Intransitivity of preferences. *Psychological review*, 76(1):31–48.