Survey Research Methods (2022)
Vol. 16, No. 3, pp. 329-352
doi:10.18148/srm/2022.v16i3.7908
European Survey Research Association

© 2022Author(s)
CC BY-NC 4.0

Adapting the Robust Effect Size Cliff's Delta to Compare Behaviour Profiles

Frank Bais
Cito Institute for Educational Measurement
Arnhem, The Netherlands

Joost van der Neut
Delft University of Technology
Delft, The Netherlands

Cliff's Delta is a non-parametric effect size that is based on data observations. In this paper, we elaborate on an adaptation of Cliff's Delta in order to compare behaviour profiles. Behaviour profiles are density distributions in which survey answer behaviour is summarized for specific groups of respondents or items. Such profiles are useful, as they take into account the varying number of survey items that is filled out per respondent due to filter questions. By the adapted profile-based Cliff's Delta, two subgroups of respondents (for instance higher and lower educated respondents) can be compared on the occurrence of specific answer behaviour (for instance giving 'don't know'-answers). By means of simulations, we show that the adapted profile-based Cliff's Delta converges towards the original Cliff's Delta as the number of items that is filled out by respondents increases. The uncertainty that comes along with a finite number of items is taken into account by the adapted profile-based Cliff's Delta. As a result, the adapted profile-based Cliff's Delta has a restricted magnitude for a finite number of items. We conclude that the adapted profile-based Cliff's Delta is a solid and conservative statistic that is both useful and advantageous to compare behaviour profiles. We close with two survey data examples and by discussing our findings.

Keywords: Cliff's Delta; effect size; behaviour profiles; answer behaviour; survey statistics

1 Introduction

The relation between survey answer behaviour and measurement error has been studied extensively. According to the literature, the occurrence and size of measurement error and hence response data quality can be influenced by respondent characteristics (see for instance Olson and Smyth, 2015; Pickery and Loosveldt, 1998; Tourangeau, Rips, and Rasinski, 2000) and survey item characteristics (see for instance Campanelli et al., 2011; Saris and Gallhofer, 2007; Tourangeau et al., 2000). Both respondent and item characteristics may lead to undesirable answer behaviour, like answering "won't tell" (Bradburn, Sudman, Blair, & Stocking, 1978; Shoemaker, Eichholz, & Skewes, 2002) or answering "don't know" (Beatty & Herrmann, 2002; Leigh & Martin Jr, 1987).

The relation of undesirable answer behaviour to respondent and item characteristics could be used as a starting point for future survey design. Surveys could be adapted to specific respondent and item characteristics in order to minimize the occurrence of undesirable answer behaviour and measurement error. This means that exposed relations could function as guides in optimizing future data collection. In order to investigate such relations, so-called "behaviour profiles" can be constructed. A behaviour profile is a density distribution that gives a statistical summary of survey answer behaviour for a specified group of respondents or items. Such a profile takes into account the uncertainty that exists around the actual occurrence of answer behaviour. In practice, this means that two behaviour profiles can be compared on the extent to which they show a specific undesirable answer behaviour. For instance, the profiles for men and women or for sensitive and non-sensitive items could be compared on the occurrence of don't know-answers. See section 2 for a brief explanation about behaviour profiles and see Bais (2021) for an extensive statistical elaboration on this topic.

A consequence of using behaviour profiles is that we do not consider individual data observations for the analyses, as each individual respondent estimate is transformed into a density distribution. This means that we cannot use customary effect sizes like Cohen's *d* (Cohen, 1962, 1988) or the Mann-Whitney *U* statistic (Mann & Whitney, 1947) to compare two groups of respondents. Thus, in order to compare two group profiles of respondents or items on the occurrence of answer behaviour, an inventive statistical measure is needed. In this paper, we elaborate on an adaptation of the robust and non-parametric effect size Cliff's Delta (Cliff, 1993, 1996a, 1996b) to compare two behaviour profiles. The main benefit of our adapted Cliff's Delta is that two behaviour pro-

Contact information: Frank Bais, Department of Research and Innovation, Cito Institute for Educational Measurement, Amsterdamseweg 13, 6814 CM Arnhem (E-mail: frank.bais@cito.nl).

files of any type or shape can be compared without difficulty.

Cliff's Delta was originally developed by Cliff (1993) for the use with ordinal data. It is a measure of how frequently the data values in one group are larger than the data values in a second group. In our study, we address our first research question by showing how we transform the original Cliff's Delta for data observations into an adapted Cliff's Delta for density distributions. Second, we address our second research question by illustrating that the adapted Cliff's Delta is a justifiable approximation of a fixed reference Cliff's Delta based on two estimated behaviour profiles by means of simulations. See Bais (2021) and Bais, Schouten, and Toepoel (2022) for an application of the adapted Cliff's Delta to survey data.

In section 2, we briefly explain the concept and interpretation of behaviour profiles. In section 3, we elaborate on the concept of Cliff's Delta and its transformation from the original into the adapted version. In section 4, we show that the adapted Cliff's Delta is a solid and conservative measure that is suitable for comparing two behaviour profiles. In section 5, we give two answer behaviour examples of estimating the adapted Cliff's Delta from two behaviour profiles. We close with a conclusion and discussion of the simulation outcomes in section 6.

2 Behaviour profiles

In order to compare the occurrence of answer behaviour for the two categories of a respondent or item characteristic, so-called "behaviour profiles" can be constructed. We define a behaviour profile as a density distribution that summarizes answer behaviour for a specified group of respondents or items. We may distinguish respondent and item behaviour profiles, which summarize answer behaviour for types of respondents and items respectively. A behaviour profile represents the relative proportions of a group of respondents or items (for instance lower educated respondents or items containing difficult language) in showing a specified answer behaviour (for instance answering "don't know") for all possible probabilities from 0 to 1. From here, we first give reasons to make use of behaviour profiles for comparing specific groups of respondents or items. Second, we elaborate on how to construct behaviour profiles.

2.1 Why using behaviour profiles?

In order to compare types of respondents or items on the occurrence of answer behaviour, a few aspects need to be taken into account regarding the answer behaviour. First, the number of respondents that participates in a survey and the number of items that is applicable to a specific answer behaviour per survey is finite. This means that an extent of uncertainty exists around the actual occurrence of behaviour, since the behaviour is based on a by definition delimited number of respondents or items. In other words, the actual occurrence of behaviour for respondents is surrounded by more uncertainty as they fill out a smaller number of items. And the actual occurrence of behaviour for items is surrounded by more uncertainty as they are filled out by a smaller number of respondents. Second, when a survey contains filter questions that may or may not branch out into follow-up questions, each respondent is likely to fill out a different number of items for that survey. Therefore, the actual occurrence of behaviour is indicated with varying uncertainty across different respondents and items within such a survey. Hence, to compare groups of respondents or items on their answer behaviour, simply using individual behaviour proportions is insufficient. Our method of using behaviour profiles to estimate behaviour occurrences takes into account the aforementioned uncertainties in order to compare types of respondents or items.

Let us briefly discuss why using individual behaviour proportions is not sufficient. Consider the situation in which we want to construct a behaviour profile for a group of respondents for giving "won't tell"-answers in a survey containing filter questions. Suppose that almost all respondents fill out relatively many items and give "won't tell"-answers only occasionally, while one single respondent who fills out only few items gives "won't tell"-answers frequently. This means that we are uncertain about the behaviour of this single respondent compared to the behaviour of the other respondents. Simply averaging individual behaviour proportions would lead to a biased (in this case higher) group mean proportion. More importantly, comparing either weighted or unweighted mean proportions by using a customary measure like Cohen's d is not robust against skewed or heavy-tailed distributions and outliers (Rousselet, 2016). As a difference between two groups may increase by including an outlier, Cohen's d is affected by this inclusion and may be *smaller* than when this outlier would have been removed (Rousselet, 2016). In summary, behaviour profiles can be used to omit parametric assumptions and account for important uncertainties, and a robust statistical measure is needed to compare such profiles.

2.2 Constructing behaviour profiles

Except for the notation and the interchangeability of the terms respondent and item, the concept and construction of respondent and item profiles are identical. For the purpose of this paper, we therefore only discuss the concept and construction of respondent profiles. In constructing a respondent profile, we make use of the binomial distribution to take into account the abovementioned uncertainties. Here, each item that is filled out can be interpreted as an independent Bernoulli trial with probability p.

Let I_r be the number of items for which a specific answer behaviour is possible that has been presented to a spe-

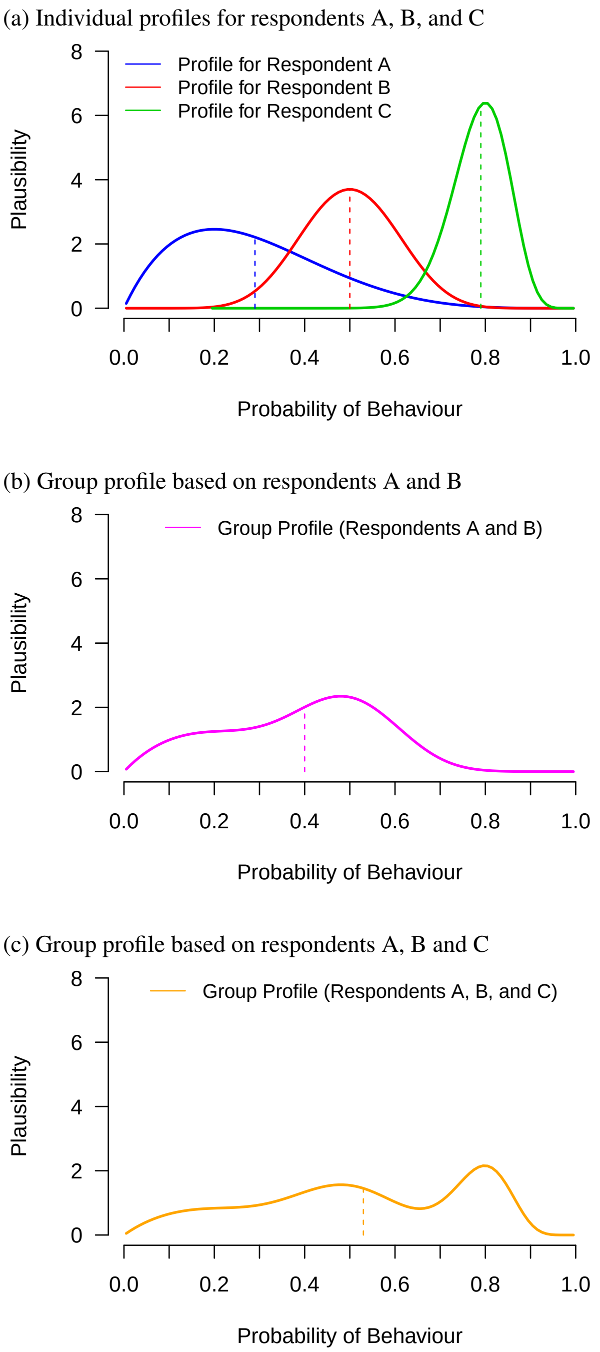


Figure 1. Example profiles

cific respondent r. We interpret each item as an independent Bernoulli experiment, with an unknown probability p that the behaviour is shown (and a probability $1-p$ that the behaviour is not shown). Consequently, the total number of items X_r for which respondent r shows the behaviour is expected to have a binomial distribution (Dekking, Kraaikamp, Lopuhaä, & Meester, 2005)

$$P(X_r = K_r) = \binom{I_r}{K_r} p^{K_r}(1-p)^{I_r-K_r} \quad , \tag{1}$$

where $K_r \in \{0, 1, \ldots, I_r\}$. From our observations, we know the number of times G_r that the behaviour is actually shown by respondent r. The likelihood λ_r is the probability $P(X_r = G_r)$ that this observation is correctly predicted by formula (1) (Dekking et al., 2005; Fisher, 1925), leading to

$$\lambda_r(p) = \binom{I_r}{G_r} p^{G_r}(1-p)^{I_r-G_r} \quad . \tag{2}$$

Unlike probability density functions, $\lambda_r(p)$ does not necessarily integrate to 1. In this study, we assign this favorable property to each distribution by normalization and we average behaviour distributions of multiple respondents to construct behaviour profiles for groups of respondents. From here, we therefore define the normalization of the likelihood of an individual respondent as the plausibility Λ_r that the behaviour is shown by this respondent:

$$\Lambda_r(p) = \frac{\lambda_r(p)}{\int_0^1 \lambda_r(p)\, \mathrm{dp}} \quad . \tag{3}$$

For a single respondent r, the average or expected value E_r for the behaviour occurrence can be estimated on the basis of the respondent's profile and the following integral over p:

$$E_r = \int_0^1 p\Lambda_r(p)\, \mathrm{dp} \quad . \tag{4}$$

The distribution resulting from formula's (2) and (3) is an individual respondent profile. The profile delineates the expected behaviour occurrence across the full potential probability range from 0 to 1 and gives consideration to the amount of occurrence uncertainty.

See Figure 1a and consider the behaviour profile for respondent A. This respondent filled out five items and a specific undesirable answer behaviour (for instance giving 'don't know'-answers) was selected for one out of these five items by this respondent. The behaviour occurrence for respondent A is relatively uncertain, as only five items were filled out. This uncertainty is evident by the stretched shape of the behaviour profile. This means that a broad range of probabilities for the behaviour occurrence is more or less plausible and that the expected value of 0.29 (see the blue dotted line) is not precisely estimated. Respondent B selected specific behaviour for 10 out of 20 items. The behaviour occurrence is less uncertain for respondent B than for respondent A, considering the higher peak at the expected value of 0.50 and a slightly more squeezed profile. See Figure 1a and consider the behaviour profile for respondent C. This respondent filled out the larger number of 40 items. A specific undesirable answer behaviour was selected for 32 out of these 40 items by respondent C. The behaviour occurrence is more certain for respondent C than for respondent A or B. This can be seen in the behaviour profile for respondent C, which has a relatively squeezed and peaked profile. This means that only a relatively narrow range of probabilities for the behaviour occurrence is plausible and that the expected value of 0.79 is relatively precisely estimated. In summary, our profile method takes into account the degree of uncertainty that comes along with the answer behaviour of each individual respondent. See Appendix A1 for two other examples of individual profiles and to construct individual profiles in R yourself.

By considering all respondents who meet the condition of a specific category for a characteristic (for instance lower educated respondents for educational level), the average respondent group profile can be calculated by averaging their individual plausibilities:

$$\overline{\Lambda}(p) = \frac{1}{R}\sum_{r=1}^{R}\Lambda_r(p) \quad , \tag{5}$$

where $\overline{\Lambda}(p)$ is the respondent profile of the group behaviour occurrence averaged over all respondents, and R is the total number of respondents in the group. By means of this average respondent profile, the averaged expected value $\overline{E}$ for the behaviour occurrence for this group of respondents can be calculated as follows:

$$\overline{E} = \int_0^1 p\overline{\Lambda}(p)\,\mathrm{dp} \quad . \tag{6}$$

The distribution resulting from formula (5) is an averaged group respondent profile. From here, we illustrate how to come to such an averaged group profile step by step. See the group profile in Figure 1b. This group profile consists of the average of the two individual profiles for respondent A and B (see Figure 1a). For each probability, the accompanying plausibilities for respondents A and B are averaged. The result is the group profile for respondents A and B. As can be seen in Figure 1b, the low broad bump and the higher more narrow bump of the individual profiles for respondent A and respondent B respectively are still visible. The dotted line at around 0.40 refers to the expected value of the group profile.

See Figure 1c. This group profile consists of the average of the three individual profiles for respondents A, B, and C (see Figure 1a). Again, the individual profiles for respondents A, B, and C are averaged. The result is the group profile for respondents A, B, and C as can be seen in Figure 1c. The bumps of the individual profiles for respondents A and B are still visible to some extent, but the peak of the individual profile for respondent C is most striking in the group profile. This means that as an individual profile is based on more items, the more influence that individual profile has on the shape of the group profile. The dotted line at around 0.52 refers to the expected value of the group profile. Group profiles can be based on any number of individual respondent profiles and any type of respondent characteristics. See Appendix A2 for two other examples of group profiles and to construct group profiles in R yourself.

In summary, respondent profiles for groups with different characteristics can be constructed. In this way, an idea can be obtained about the difference of the occurrences of specific answer behaviour (for instance answering don't know) between two groups (for instance lower and higher educated respondents). The expected values give an indication of the average behaviour occurrence for the groups as a whole. The next step is to use a solid statistical measure to compare the behaviour occurrences of two groups.

3 Adapting Cliff's Delta

The non-parametric effect size Cliff's Delta δ can be used as a robust alternative to using two independent group means. It measures how often the data observations in one group are larger than the data observations in a second group (Cliff, 1993, 1996a, 1996b). In this section, we first briefly explain the original Cliff's Delta for data observations. Second, we address our first research question by illustrating how we transform the original Cliff's Delta into an adapted version for density distributions.

3.1 The original Cliff's Delta for data observations

Cliff's Delta δ is a robust effect size that indicates to what extent two groups are different. It calculates the probability that a random data observation x_a from a group A is larger than a random data observation x_b from another group B, minus the reverse probability (Hess & Kromrey, 2004; Rousselet, Foxe, & Bolam, 2016; Rousselet, Pernet, & Wilcox, 2017):

$$\delta = p(x_a > x_b) - p(x_a < x_b) \quad . \tag{7}$$

The sample estimate of δ is obtained by comparing each data observation in group A to each data observation in group B:

$$\hat{\delta}_O = \frac{\sum_{a=1}^{R_A}\sum_{b=1}^{R_B} sgn(x_a - x_b)}{R_A R_B} \quad , \tag{8}$$

where $\hat{\delta}_O$ is the *observation-based* sample estimate of δ. The sign function sgn $(x_a - x_b)$ results in 1, 0, or −1 when $x_a > x_b$, $x_a = x_b$, or $x_a < x_b$ respectively. The total number of comparisons is the product of the sample sizes R_A and R_B of group A and B respectively. Calculating $\hat{\delta}_O$ may be considered a dominance analysis, referring to the extent to which the one data distribution overlaps the other (Hess & Kromrey, 2004). The smaller the overlap between the distributions of two groups, the larger the dominance and the more difference between the two groups. A $\hat{\delta}_O$ of −1 or 1 indicates absence of overlap between two groups and a $\hat{\delta}_O$ of 0 refers to group equivalence (Hess & Kromrey, 2004).

3.2 Adapting Cliff's Delta for density distributions

In this subsection, we illustrate how to transform $\hat{\delta}_O$ into an adapted version for density distributions in order to use the statistic for comparing behaviour profiles. Consider $\hat{\delta}_O$ for which each specific observation from a sample A is compared to each specific observation from a sample B exactly once. Therefore, we may regard both observations for each such comparison on its own as having a "frequency" or "weight" of 1. Implementing these frequencies into formula (8) gives

$$\hat{\delta}_O = \frac{\sum_{a=1}^{R_A} \sum_{b=1}^{R_B} sgn(x_a - x_b)(w_a w_b)}{\sum_{a=1}^{R_A} \sum_{b=1}^{R_B} (w_a w_b)} \quad , \qquad (9)$$

where w_a and w_b are the frequencies of the data observations x_a and x_b from groups A and B respectively. By simply considering each possible data observation pair once, these frequencies are all 1 by definition, making formula (9) identical to formula (8).

When we apply this idea to behaviour profiles, we may consider the behaviour probabilities from 0 to 1 (with a specific step size interval) our "observations" and the plausibilities for each probability their accompanying "frequencies" or 'weights'. Implementing the probabilities and their plausibilities into formula (9) gives the adapted Cliff's Delta for density distributions:

$$\hat{\delta}_P = \frac{\sum_{a=1}^{A} \sum_{b=1}^{B} sgn(p_a - p_b)\overline{\Lambda}_A(p_a)\overline{\Lambda}_B(p_b)}{\sum_{a=1}^{A} \sum_{b=1}^{B} \overline{\Lambda}_A(p_a)\overline{\Lambda}_B(p_b)} \quad , \qquad (10)$$

where $\hat{\delta}_P$ is the *profile-based* sample estimate of δ. Here, p_a and p_b are the probabilities from 0 to 1 for group A and group B respectively, and $\overline{\Lambda}_A$ and $\overline{\Lambda}_B$ are the averages of all individual respondent profiles for group A and group B respectively. Note that A and B refer to the same number of step size intervals for both groups and are unrelated to respondent group size. We choose to discretize the probability axis by a step size of 0.01, which means that we have 100 of these intervals for each distribution. The midpoints of the intervals, 0.005, 0.015, 0.025, ..., 0.995, may be considered our "observations" that all have their own accompanying weight in the form of a plausibility. See Figure 2a, where we show the actual 100 plausibility points of the distributions for three groups of respondents. Note that $\hat{\delta}_P$ can also be evaluated as a continuous function. In Appendix B, we discuss several representations for the continuous form of Cliff's Delta.

3.3 Using and interpreting $\hat{\delta}_P$

See Figure 2a for an example of a pair of "observations" in comparing group 1 and group 3. Consider their respective probabilities of 0.305 and 0.655, and their respective

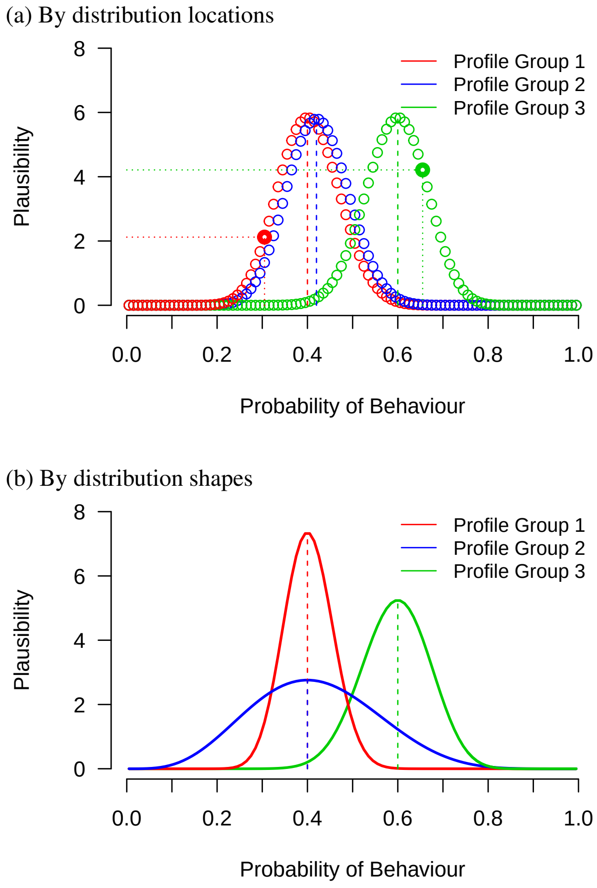


Figure 2. Example profiles for group 1, 2 and 3

plausibilities of about 2.1 and 4.2 (see the dotted lines). Implementing these values into the numerator of formula (10) gives sgn(0.305-0.655)(2.1*4.2), resulting in a negative contribution to the total numerator sum. In the numerator, all positive and negative contributions of all possible pairs of probability observations are summed. The result is divided by the sum of all possible pairs of all (positive) plausibility products. Hence, $\hat{\delta}_P$ will fall between (or at) -1 and 1. In case of complete overlap of two distributions, $\hat{\delta}_P$ will be 0. In case of absence of overlap of two distributions, $\hat{\delta}_P$ will be -1 or 1. We consistently compare the probabilities of group A to the probabilities of group B according to formula (10) throughout this paper. Note that if we would interchange groups A and B, that is using sgn $(p_b - p_a)$ in the numerator of formula (10), the sign of $\hat{\delta}_P$ would flip.

In comparing behaviour profiles, $\hat{\delta}_P$ takes into account both location and shape. See Figure 2a. Considering the *location* of behaviour profiles, when two groups are located close to each other, they *largely* overlap. This means that the expected values for both groups are alike and result in a *smaller* absolute $\hat{\delta}_P$ (as for groups 1 and 2). When two groups are located further from each other, they *hardly* overlap. This means that the expected values for both groups are different and result in a *larger* absolute $\hat{\delta}_P$ (see group 1 and 3). As can be seen in Figure 2a, the pairs of observations for which the probability is larger in group 3 than in group 1, are *frequently* accompanied by substantial plausibilities. However, the pairs of observations for which the probability is larger in group 1 than in group 3, are *rarely* accompanied by substantial plausibilities. The larger the overlapping area of both groups, the smaller the absolute $\hat{\delta}_P$, as positive and negative contributions in the numerator cancel each other out. Hence, in case of substantially different expected group values and little overlap between the two behaviour profiles, the outcome is a relatively *large* absolute $\hat{\delta}_P$. When comparing group 1 to group 3 (see Figure 2a), the result will be a large negative $\hat{\delta}_P$, indicating that the behaviour occurrence is higher for group 3 than for group 1. Calculating $\hat{\delta}_P$ for comparing the profiles for groups 1 and 2 by means of formula (10) yields -0.16, while comparing the profiles for groups 1 and 3 yields a $\hat{\delta}_P$ of -0.95.

Considering the *shape* of behaviour profiles, see Figure 2b. Here, group 1 and 2 are located around the same expected value, but shaped differently. When a group profile is relatively *stretched* across the probability range (see group 2), this means that quite some uncertainty exists around the expected value for this group. In general, a stretched profile is likely to give a relatively *large* area of overlap with another profile (see the overlap between groups 2 and 3). This results in a relatively *small* absolute $\hat{\delta}_P$. However, when a group profile is relatively *squeezed* (see group 1), this means that the expected value for this group is relatively certain. In general, a squeezed profile is likely to give a *small* area of overlap with another profile (see the overlap between group 1 and 3). The outcome is a relatively *large* absolute $\hat{\delta}_P$. Calculating $\hat{\delta}_P$ for groups 2 and 3 by means of formula (10) yields -0.74, while comparing the profiles for groups 1 and 3 yields a $\hat{\delta}_P$ of -0.96. In summary, the further the profiles are located from each other and the more squeezed they are, the more the profiles differ and the more $\hat{\delta}_P$ deviates from zero. Hence, we can use $\hat{\delta}_P$ as a suitable measure to compare the behaviour profiles for two specified groups of respondents or items.

Using $\hat{\delta}_P$ to compare behaviour profiles has many advantages. For instance, $\hat{\delta}_P$ makes no assumption about the shape of the underlying distribution (Cliff, 1993, 1996a, 1996b; Goedhart, 2016; Vargha & Delaney, 2000) and is robust in case of outliers or skewed or otherwise non-normal distributions (Goedhart, 2016). The statistic is easy to calculate, straightforward to interpret, and standardized, which means that different effect size categories can be distinguished (Goedhart, 2016). For our $\hat{\delta}_P$, relatively small or unequal sample sizes are no issue.

4 Simulations for Cliff's adapted Delta

In this section, we simulate data from two fixed respondent group profiles that are accompanied by a specific fixed reference δ. In subsection 4.1, we sample observations from both reference profiles for which we calculate $\hat{\delta}_O$. Here, we consider each observation fixed and certain, without taking into account the uncertainty that comes along with the finite number of items. We illustrate that $\hat{\delta}_O$ estimates δ with more certainty as we sample more observations per group. In subsection 4.2 and 4.3, we take into account the fact that the number of items is finite in practice by constructing behaviour profiles for which we calculate $\hat{\delta}_P$. Here, we consider each probability a latent behaviour occurrence for a single respondent who fills out a specific number of items. We illustrate that $\hat{\delta}_P$ estimates δ with more certainty as the number of respondents per group increases. Subsequently, we answer our second research question by showing that $\hat{\delta}_P$ approaches δ as respondents fill out more items.

The simulation examples are based on random sampling from two fixed reference distributions with pre-specified means and shape. To map the uncertainty regarding the values of the respondent profiles and Cliff's Delta's, we use the method of bootstrapping to construct 99% confidence intervals. All respondent profiles are sampled randomly from the reference profiles 10000 times. As a result, 10000 corresponding Cliff's Delta's can be calculated. For sections 4.2 and 4.3, the occurrence of a specific behaviour per item is sampled on the basis of the sampled behaviour probability for the corresponding respondent. See Table 1 for the R code to construct the reference profiles and for an overview of why and how we set up each simulation. R codes for the simulations can be requested from the authors.

4.1 Increasing the number of observations for $\hat{\delta}_O$

First, we construct two respondent profiles as our reference profiles for this section. See Figure 3a. Comparing group 1 to group 2 results in a fixed reference δ of 0.29. When we sample observations from these respondent profiles randomly with replacement, we can calculate $\hat{\delta}_O$. Let us sample varying numbers of observations with an equal number of observations per group and calculate the accompanying $\hat{\delta}_O$. See Figure 3b. As can be seen in the graph, the average $\hat{\delta}_O$ across all samples is estimated well for any fixed number of observations per group. The certainty of this estimation becomes larger as the number of observations increases. Thus, $\hat{\delta}_O$ based on large numbers of randomly sampled observations from two respondent profiles is an accurate estimation of δ based on those two profiles.

4.2 Increasing the number of respondents for $\hat{\delta}_P$

We showed that increasing the number of sampled observations gives a more certain estimation of $\hat{\delta}_O$. We may con-

Table 1
R Code for the Reference Distributions and an Overview of the Purpose and Execution of the Various Simulations

Construction of the reference profiles G1 and G2 in R

```
step = .01; seq = seq(.5*step, 1 - (.5*step), step)
meanA=.40 and sdA=.04
meanB=.70 and sdB=.10
A = dnorm(seq, meanA, sdA) # construct A
B = dnorm(seq, meanB, sdB) # construct B
A = A/(sum(A)*step) # normalize A
B = B/(sum(B)*step) # normalize B
G1 = (A+B)/2 # construct G1 as the average of A and B
meanG2=.45 and sdG2=.15
G2 = dnorm(seq, meanG2, sdG2) # construct G2
G2 = G2/(sum(G2)*step) # normalize G2
```

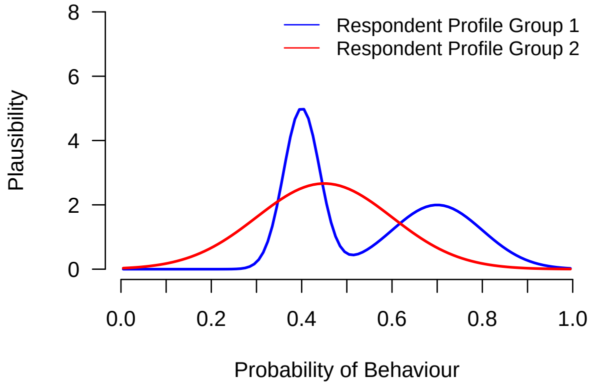
What do we want to illustrate by the simulation?	How do we execute the simulation?
Simulation in section 4.1	
Illustrate that Cliff's Delta $\widehat{\delta_O}$ estimates reference Cliff's Delta δ with more certainty as more observations are sampled from two behaviour profiles.	Sample observations from reference profiles G1 and G2 based on their plausibility to occur.
Simulation in section 4.2	
Illustrate that two constructed behaviour profiles and Cliff's Delta $\widehat{\delta_p}$ that results from comparing them become more certain as they are based on more respondents.	1) Sample probabilities (as latent behaviour occurrences) from reference profiles G1 and G2 based on their plausibility to occur. 2) Construct behaviour profiles based on these probabilities for 50 vs. 500 respondents (for 500 items).
Simulation in section 4.3, part 1	
Illustrate that the estimate of Cliff's Delta $\widehat{\delta_p}$ that results from comparing two constructed behaviour profiles becomes more accurate (converges towards reference Cliff's Delta δ) as the number of items per respondent increases.	1) Sample 500 fixed probabilities (as latent behaviour occurrences) from reference profiles G1 and G2 based on their plausibility to occur. 2) Construct behaviour profiles based on these probabilities for 5 to 500 items per respondent (for 500 respondents).
Simulation in section 4.3, part 2	
Illustrate that constructed behaviour profiles become more accurate as the number of items per respondent increases.	Same as for section 4.3 part 1, but for 5, 20, 50, and 200 items per respondent.

sider these observations the latent behaviour occurrences for individual respondents. In practice, these individual occurrences are based on a finite number of survey items. The uncertainty that comes along with this finite number of items is the reason to construct respondent profiles, as our profile method takes into account this uncertainty. In the following, we sample occurrences or respondents that are based on a finite number of items.

First, let us illustrate the influence of a varying respondent group size on the estimation of δ for the fixed number of 500 items per respondent. For this purpose, we compare groups consisting of 50 and 500 respondents by sampling from the reference profiles.

See Figure 4 for the averaged samples. In Figure 4a, the profile confidence intervals are relatively broad, referring to the present uncertainty of the profiles and hence of

(a) Respondent profiles for group 1 (mean=0.55) and group 2 (mean=0.45) with reference Cliff's δ = 0.29



(b) Original observation-based Cliff's $\hat{\delta}$ = 0.29

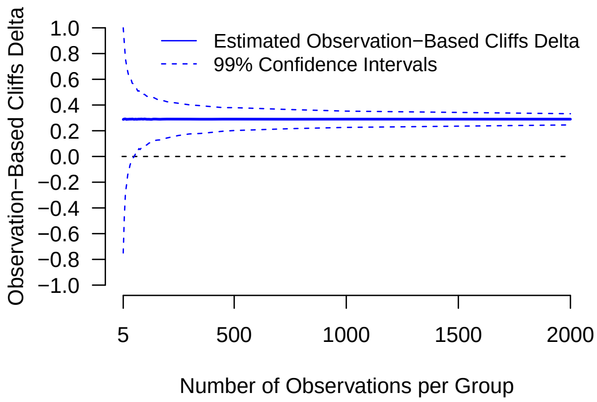
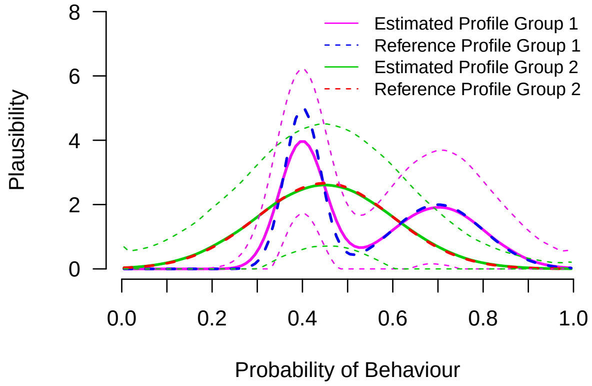


Figure 3. Reference profiles and observation-based Cliff's $\hat{\delta}$

(a) 50 respondents (99% C.I. = [−0.00, 0.56])



(b) 500 respondents (99% C.I. = [−0.20, 0.38])

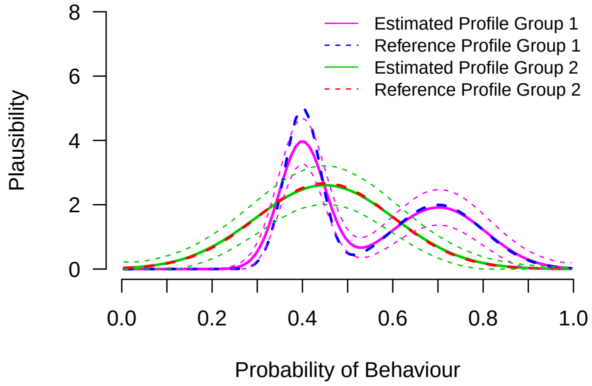


Figure 4. Estimated respondent profiles and their statistical properties for group 1 (mean=0.55) and group 2 (mean=0.45) with estimated profile-based Cliff's $\hat{\delta}$ = 0.29

$\hat{\delta}_P$. Here, the average $\hat{\delta}_P$ across all samples is estimated well considering its value of 0.29, but its confidence interval of (-0.00, 0.56) indicates large uncertainty for the estimation. In Figure 4b, the average $\hat{\delta}_P$ is estimated well and the profile confidence intervals are visibly smaller than in Figure 4a. This refers to a more certain estimation of the profiles and hence of $\hat{\delta}_P$ considering its more narrow confidence interval of (0.20, 0.38). We can conclude that respondent profiles and $\hat{\delta}_P$ become more certain as group size increases. This conclusion is in line with the more certain estimation of $\hat{\delta}_O$ for an increasing number of observations in subsection 4.1, where the observations can be considered respondents who filled out an infinite number of items.

4.3 Increasing the number of items for $\hat{\delta}_P$

In this section, we illustrate the effect of increasing the number of items per respondent. To disentangle this effect from the influence of group size, we keep group size large and fixed at 500 respondents per group. To disentangle this effect from the influence of group behaviour occurrence, we use the same sample of 500 respondent behaviour probabilities for each simulation. We show that $\hat{\delta}_P$ is a conservative statistic and factually an underestimation of δ. As the number of items per respondent increases, the estimation becomes more accurate. We will see that $\hat{\delta}_P$ approaches δ as the number of items per respondent increases. A formal proof on how $\hat{\delta}_{PC}$ converges to δ when the number of items per respondent increases towards infinity is provided in Appendix C.

For each simulation, our starting points are the reference profiles. Random samples of 500 fixed behaviour probabilities from both reference profiles are used. Based on these probabilities, we construct respondent profiles based on varying numbers of items that run from 5 to 500 items per respondent. For simplicity, we choose an equal number of items per respondent within each respondent profile (which is not a requirement for the methodology). For instance, when a respondent profile is based on 50 items, then 50 items are sampled for each individual respondent based on the corresponding probability. For each specific number of items, re-

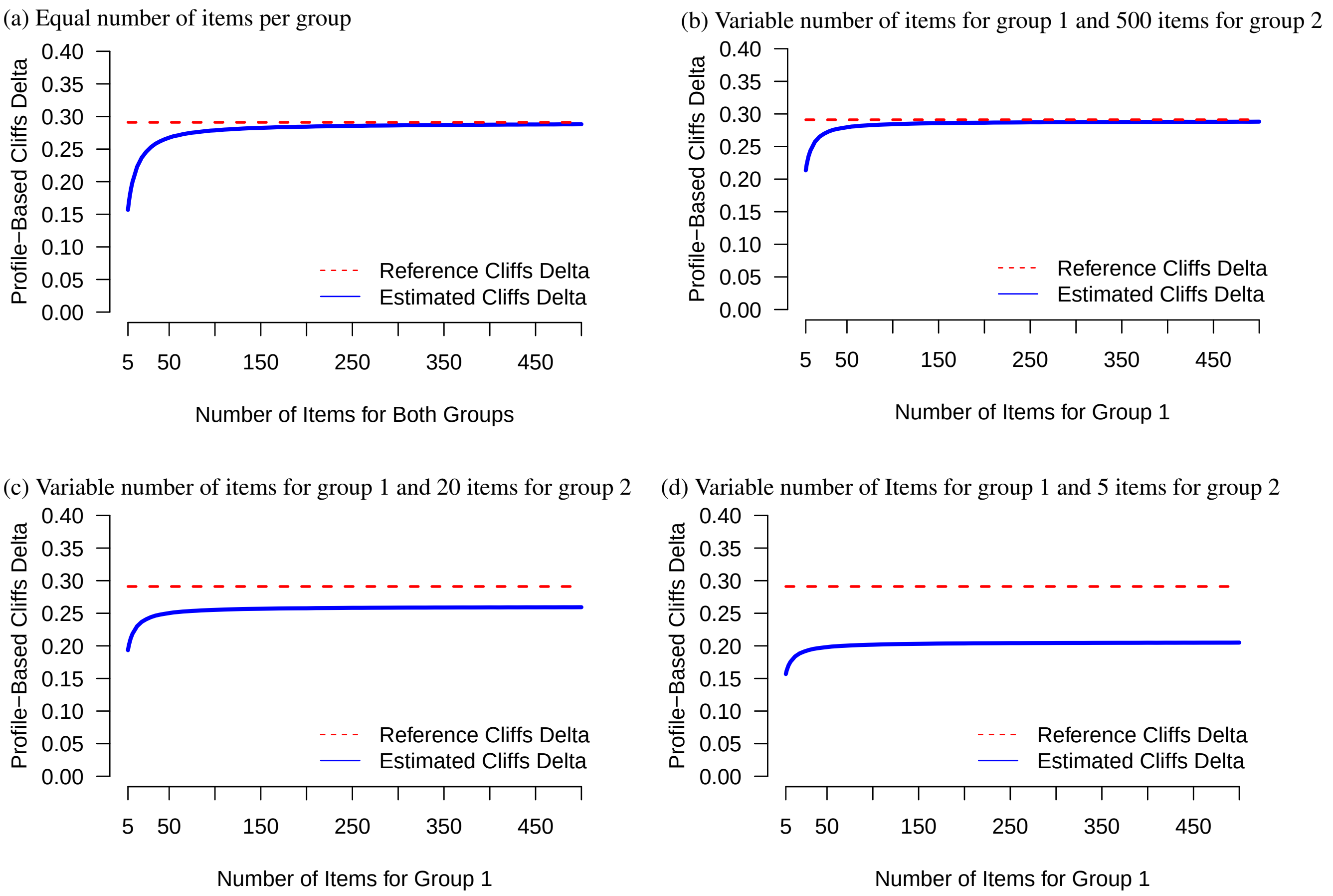


Figure 5. Estimated profile-based Cliff's $\hat{\delta}$ for two distributions from 500 respondents with varying numbers of items per group and $\delta = 0.29$.

spondent profiles are sampled 10000 times in order to obtain 10000 $\hat{\delta}_P$'s for each comparison between two profiles. From these $\hat{\delta}_P$'s, the averaged $\hat{\delta}_P$ is considered.

First, we compare respondent profiles that are based on equal numbers of items. See Figure 5a. For small numbers of items, $\hat{\delta}_P$ is clearly underestimated compared to δ. As the number of items increases, this underestimation decreases and $\hat{\delta}_P$ converges towards δ. The simulation outcome for a respondent profile based on 500 items and a respondent profile based on varying numbers of items is comparable. See Figure 5b. Here, the underestimation of $\hat{\delta}_P$ is more modest for small numbers of items for group 1 and converges more rapidly towards δ than for the situation in Figure 5a. The important similarity between both situations is that $\hat{\delta}_P$ converges towards δ as the number of items increases. The more items that respondent profiles are based on, the more squeezed their shape in general, hence the more they converge towards a precise estimation of the behaviour occurrence for the corresponding group. Theoretically, the profiles would ultimately converge towards precise point estimates. Such point estimates are identical to single data observations for which $\hat{\delta}_O$ can be calculated.

It is interesting to see what happens when we compare a respondent profile based on varying numbers of items to a respondent profile based on a *small* fixed number of items. See Figure 5c for such comparisons for a fixed number of 20 items for group 2. Also here, $\hat{\delta}_P$ rapidly increases as the number of items increases for group 1. However, as can be seen in Figure 5c, this increasement is limited and $\hat{\delta}_P$ does not converge closely towards δ. When the fixed number of items for group 2 is even smaller with only 5 items per respondent, this effect is even stronger. See Figure 5d. Here, $\hat{\delta}_P$ hardly increases with an increasing number of items in group 1. The smaller the fixed number of items that group 2 is based on, the stronger the underestimation of $\hat{\delta}_P$ remains for a large number of items for group 1.

We need to note that this underestimation of $\hat{\delta}_P$ for smaller numbers of items per group is one of the benefits of the statistic. The smaller the number of items that a group is based on, the more uncertainty exists about the true behaviour profile

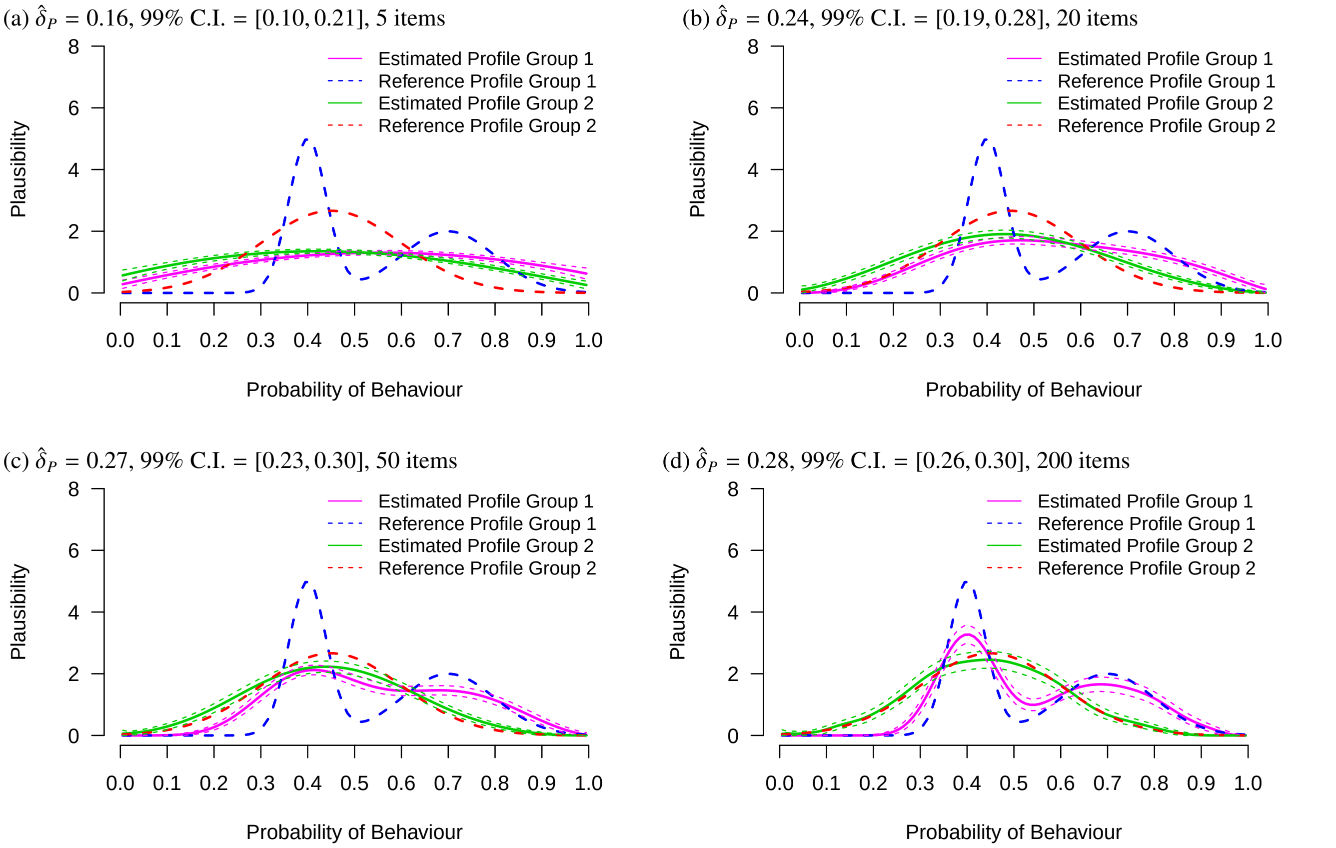


Figure 6. Estimated respondent profiles and their statistical properties for varying numbers of items for both groups of respondents (reference Cliff's δ = 0.29).

for that group. In uncertain circumstances, we want to make a cautious and conservative comparison between groups. In other words, we do not want to find an effect by using $\hat{\delta}_P$ in case two groups with similar distributions are evaluated by too few items. In general, this means that groups need to differ clearly in order to find an effect when one or both groups are based on few items. Thus, the relatively larger underestimation of $\hat{\delta}_P$ for groups that are based on relatively fewer items is a favorable property of the statistic.

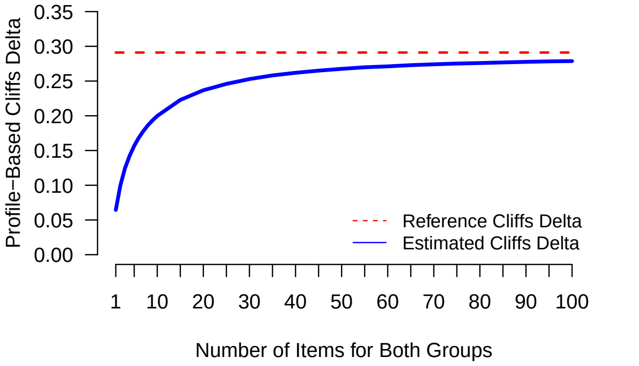
To illustrate the effect of an increasing number of items and a more squeezed shape of the profiles, let us consider some examples of respondent profiles that are sampled from our reference profiles. The sampled respondent profiles are based on equal numbers of items for each separate example. See Figure 6 in which profiles are based on 5, 20, 50, and 200 items. As can be seen in Figure 6a, the estimated profiles are very stretched and contain little detail when based on only 5 items. As the number of items increases in Figure 6b through Figure 6d, the estimated profiles become more squeezed and shaped in accordance with the reference profiles in more detail.

In summary, we can conclude that $\hat{\delta}_P$ is a solid and conservative statistic that can be used to compare two behaviour profiles. In principle, $\hat{\delta}_P$ is an underestimation of δ that converges towards δ as the number of items increases. This means that $\hat{\delta}_P$ takes into account the uncertainty regarding the numbers of items that the profiles are based on. Note that $\hat{\delta}_O$ does not take into account this uncertainty. However, the underestimation remains evident and hence the convergence remains limited when one of the profiles is based on a small number of items. This means that the magnitude of $\hat{\delta}_P$ may be substantially restricted by the profile that is based on the least number of items. In fact, mainly this profile determines the degree of underestimation of $\hat{\delta}_P$ relative to δ. Thus, in case of a small number of items for one group, $\hat{\delta}_P$ is underestimated regardless of the number of items for the other group.

4.4 Recommendation on the number of items and respondents

In this section, we provide some guidance on the number of items per respondent and the number of respondents per group for using respondent profiles and $\hat{\delta}_P$. For the number of items per respondent, we zoom in on part of the simulation results for Figure 5a in section 4.3. We consider the numbers

(a) 1 through 100 items per respondent (500 respondents in both groups)



(b) 5 through 500 respondents in both groups (500 items per respondent)

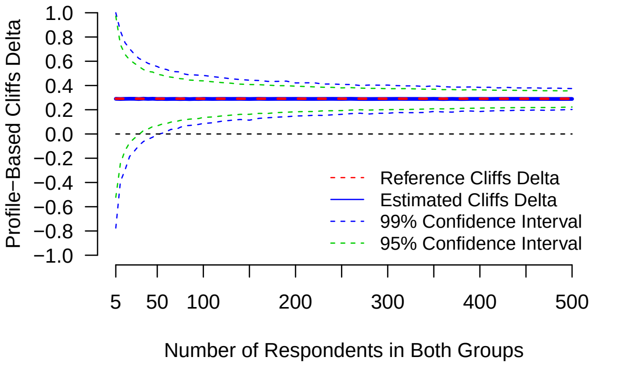
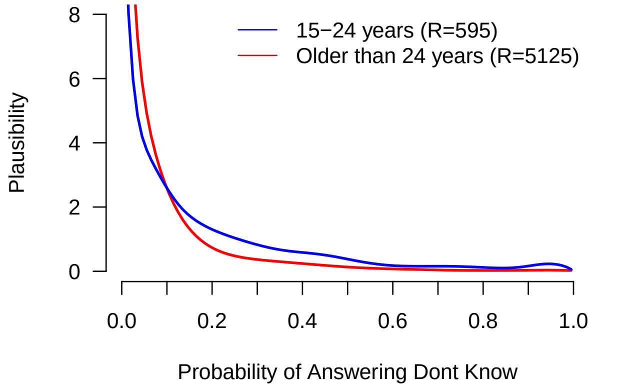


Figure 7. Estimated profile-based Cliff's $\hat{\delta}$

(a) Age subgroups for "don't know" in the survey Politics (Cliff's $\hat{\delta}_P$ = 0.28, 99% C.I. = [0.22, 0.34])



(b) Educational subgroups for "neutral responding" in the survey Personality (Cliff's $\hat{\delta}_P$ = −0.16, 99% C.I. = [−0.23, −0.09])

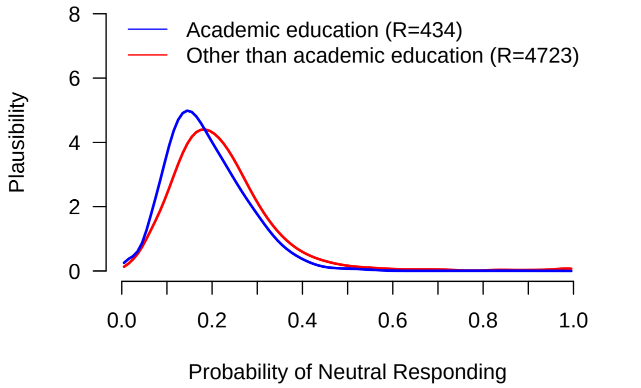


Figure 8. Respondent profiles and their statistical properties

of 1 through 100 items per respondent for a fixed number of 500 respondents per group. See Figure 7a. In order to use $\hat{\delta}_P$ and to omit a clear underestimation of the statistic, we recommend to use at least roughly 30 items per respondent. Again, note that $\hat{\delta}_P$ can be used for any number of items per respondent, but that an existing effect may not be found for lower numbers of items.

To give some guidance on the number of respondents per group, see Figure 7b. We consider the numbers of 5 through 500 respondents for each group for a fixed number of 500 items per respondent. We simulated data by sampling with replacement in the same way as we did in section 4.2 for 50 and 500 respondents per group. In order to obtain a more or less reliable $\hat{\delta}_P$, we recommend to use at least roughly 75 to 100 respondents per group. Note how the results for $\hat{\delta}_P$ in Figure 7b resemble the results for $\hat{\delta}_O$ in the left part of Figure 3b.

5 Survey data examples

In this section, we make use of data of the LISS (Longitudinal Internet studies for the Social Sciences) Panel administered by CenterData at Tilburg University, The Netherlands (CentERdata, 2014a, 2014b). In this section, we show two survey data examples. Each example consists of two respondent profiles for which we calculate $\hat{\delta}_P$. To calculate the 99% confidence intervals for $\hat{\delta}_P$, we bootstrap 10000 profiles for each respondent profile. For each bootstrap profile, respondents are randomly sampled with replacement from the corresponding group data. In sampling, we consider the number of times that a behaviour did and did not occur as given for each respondent. The number of sampled respondents is equal to the group size R. See Figure 8a.

In Figure 8a, we compare respondents between 15 and 24 years old to respondents older than 24 years. They are compared on giving don't know-answers (Beatty & Herrmann, 2002; Leigh & Martin Jr, 1987) for the items of the LISS Panel survey "Politics and Values" (wave 6, 2014b). Only those items are considered for which a don't know-answer was an applicable response option. More than 99% of all respondents in both groups filled out between 64, 66, or 67 applicable items. As can be seen in Figure 8a, the respondents in both age groups give don't know-answers only occasionally; the lower the probability on a don't know-answer, the higher the expected plausibility of the probability. However, respondents between 15 and 24 years old seem to give relatively more don't know-answers considering the plausibility of the probability range of approximately 0.10–0.60 and the small bump at a high probability (0.90–1.00) on the behaviour. This small bump may refer to a subgroup within the respondent group between 15 and 24 years old that shows don't know-answers strikingly often. The statistical result is a $\hat{\delta}_P$ of 0.28 with a confidence interval of (0.22, 0.34). This means that respondents between 15 and 24 years old show *more* don't know-answers than respondents older than 24 years for items about political content.

In Figure 8b, we compare respondents who completed an academic education to respondents who completed another level of education. They are compared on giving neutral responses (see Krosnick and Fabrigar, 1997; O'Muircheartaigh, Krosnick, and Helic, 2000) for the applicable items of the LISS Panel survey "Personality" (wave 6, 2014a). More than 95% of all respondents in both groups filled out 182 or 184 applicable items. As can be seen in Figure 8b, the profile for respondents who completed non-academic education is concentrated at higher probability values compared to the profile for respondents who completed academic education. This means that a higher plausibility is expected for higher probabilities on the behaviour for respondents who completed non-academic education. The statistical result is a $\hat{\delta}_P$ of -0.16 with a confidence interval of (-0.23, -0.09). This means that respondents who completed an academic education show *less* neutral responses than respondents who completed a non-academic education for items about personal and psychological traits.

The resulting effect sizes of 0.28 and -0.16 can be classified as a "medium" and "small" effect respectively. We use the rules that $|\hat{\delta}_P| < 0.11$ indicates no effect, $0.11 \leq |\hat{\delta}_P| < 0.28$ a small effect, $0.28 \leq |\hat{\delta}_P| < 0.43$ a medium effect, and $|\hat{\delta}_P| \geq 0.43$ a large effect, as investigated by Vargha and Delaney (2000). See also Goedhart (2016). To illustrate the broad usefulness of the adapted Cliff's Delta, we refer to Appendix D for two other data examples. For an elaboration on the limitations of the profile method in specific (extreme) situations, see Appendix E.

6 Discussion

In this paper, we introduced an adaptation of the non-parametric effect size measure Cliff's Delta δ. The motivation for this adaptation comes from the use of behaviour profiles for respondent survey data. Behaviour profiles are density distributions in which survey answer behaviour is summarized. Such profiles take into account the varying number of survey items that is filled out per respondent due to filter questions. In order to compare two behaviour profiles, the adaptation of Cliff's Delta $\hat{\delta}_P$ can be used. For instance, the occurrence of answering "don't know" for men and women can be compared. The adapted Cliff's Delta $\hat{\delta}_P$ takes into account both the location and the shape of both profiles. Here, the location refers to the estimated behaviour occurrence and the shape refers to the precision of the estimation. The less profiles overlap each other and the more squeezed their distributions, the larger the difference between two groups. In essence, we showed the relation between the adapted profile-based Cliff's Delta $\hat{\delta}_P$ and the original observation-based Cliff's Delta $\hat{\delta}_O$. In Appendix F, we discuss the relation of $\hat{\delta}_O$ to the Mann-Whitney statistic (Mann & Whitney, 1947; Wilcoxon, 1945).

First, we constructed two fixed behaviour profiles and their accompanying fixed reference δ. We illustrated how $\hat{\delta}_O$ is calculated by sampling random observations from these fixed behaviour profiles. This estimate becomes more certain as the number of sampled observations increases. Second, we sampled random probabilities from the fixed profiles and considered each probability the latent behaviour occurrence of an individual respondent who filled out a specific number of items. We showed that an increase in respondent group size results in more certain behaviour profiles and hence a more certain $\hat{\delta}_P$. Third, we illustrated that $\hat{\delta}_P$ converges towards δ as the number of items per respondent in both groups is sufficiently large. Interestingly, $\hat{\delta}_P$ approaches δ only to a limited degree in case one of the groups is based on a small number of items. In fact, the profile that is based on the least number of items determines the degree of convergence of $\hat{\delta}_P$. The smaller the number of items that this profile is based on, the more this profile needs to differ in behaviour occurrence from another profile in order for $\hat{\delta}_P$ to detect an effect. This means that two profiles based on any number of items can be compared. In summary, $\hat{\delta}_P$ is a solid and conservative statistic that is both useful and advantageous to compare two behaviour profiles.

Although $\hat{\delta}_P$ can be used to compare any two density distributions, our specific purpose to transform $\hat{\delta}_O$ was to compare groups of respondents or items on survey answer behaviour. For instance, higher and lower educated individuals can be compared on their frequency of answering 'don't know'. Or items that do and do not contain sensitive content can be compared on how frequently respondents give "won't tell"-answers. By such comparisons, relations between respondent and item characteristics on the one hand, and undesirable answer behaviour on the other, may be exposed. These relations can be investigated per survey and topic, but also across surveys to indicate their stability and consistency for multiple survey topics and characteristics. The relations may function as a guide in designing surveys that minimize the occurrence of undesirable answer behaviour and hence optimize data collection.

Acknowledgement

For this paper, we made use of respondent data of the core surveys "Personality" (wave 6, 2014a) and "Politics and Values" (wave 6, 2014b) from the LISS (Longitudinal Internet studies for the Social Sciences) Panel administered by CentERdata (Tilburg University, The Netherlands). We would like to thank CentERdata for the availability of these data. We would also like to thank Guillaume Rousselet for providing us the inspiration to choose and adapt Cliff's Delta based on his online document "Robust effect sizes for 2 independent groups".

References

Bais, F. (2012). *Intolerance of uncertainty and its effect on future-oriented decision making* (Master's thesis, University of Amsterdam).

Bais, F. (2021). *Constructing behaviour profiles for answer behaviour across surveys* (Doctoral dissertation, Utrecht University). doi:10.33540/538

Bais, F., Schouten, B., & Toepoel, V. (2022). Is undesirable answer behaviour consistent across surveys? An investigation into respondent characteristics. *Survey Methodology*, *48*(1), 191–224.

Beatty, P., & Herrmann, D. (2002). To answer or not to answer: Decision processes related to survey item nonresponse. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), *Survey nonresponse* (pp. 71–86). New York: Wiley.

Bradburn, N. M., Sudman, S., Blair, E., & Stocking, C. (1978). Question threat and response bias. *Public Opinion Quarterly*, *42*(2), 221–234. doi:10.1086/268444

Campanelli, P., Nicolaas, G., Jäckle, A., Lynn, P., Hope, S., Blake, M., & Gray, M. (2011). *A classification of question characteristics relevant to measurement (error) and consequently important for mixed mode questionnaire design*. Paper presented at the Royal Statistical Society.

CentERdata. (2014a). Personality, wave 6. LISS-Panel Data file cp13f, published by CentERdata, Tilburg University, The Netherlands. doi:10.17026/dans-x5h-4cxd

CentERdata. (2014b). Politics and values, wave 6. LISS-Panel Data file cv13f, published by CentERdata, Tilburg University, The Netherlands. doi:10.17026/dans-zms-r5rz

Cliff, N. (1993). Dominance statistics: Ordinal analyses to answer ordinal questions. *Psychological Bulletin*, *114*(3), 494–509. doi:10.1037/0303-2909.114.3.494

Cliff, N. (1996a). Answering ordinal questions with ordinal data using ordinal statistics. *Multivariate Behavioral Research*, *31*(3), 331–350. doi:10.1207/s15327906mbr3103_4

Cliff, N. (1996b). *Ordinal methods for behavioral data analysis*. doi:10.4324/9781315806730

Cohen, J. (1962). The statistical power of abnormal-social psychological research: A review. *The Journal of Abnormal and Social Psychology*, *65*(3), 145–153. doi:10.1037/h0045186

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Lawrence Erlbaum Associates.

De Bruin, G. O., Rassin, E., van der Heiden, C., & Muris, P. (2006). Psychometric properties of a Dutch version of the Intolerance of Uncertainty Scale. *Netherlands Journal of Psychology*, *62*(2), 87–92.

Dekking, F. M., Kraaikamp, C., Lopuhaä, H. P., & Meester, L. E. (2005). *A modern introduction to probability and statistics: Understanding why and how*. London: Springer.

Fisher, R. A. (1925). *Statistical methods for research workers*. Edinburgh: Oliver & Boyd.

Goedhart, J. (2016). *Calculation of a distribution free estimate of effect size and confidence intervals using VBA/excel*. doi:10.1101/073999

Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment*, *4*(1), 26–42.

Guenole, N., & Chernyshenko, O. S. (2005). The Suitability of Goldberg's Big Five IPIP Personality Markers in New Zealand: A Dimensionality, Bias, and Criterion Validity Evaluation. *New Zealand Journal of Psychology*, *34*(2), 86–96.

Hess, M. R., & Kromrey, J. D. (2004). *Robust confidence intervals for effect sizes: A comparative study of Cohen's d and Cliff's Delta under non-normality and heterogeneous variances*. Annual meeting of the American Educational Research Association.

Krosnick, J. A., & Fabrigar, L. R. (1997). Designing rating scales for effective measurement in surveys. In L. Lyberg, P. Biemer, M. Collins, E. De Leeuw, C. Dippo, N. Schwarz, & D. Trewin (Eds.), *Survey measurement and process quality* (pp. 141–164). New York: Wiley.

Leigh, J. H., & Martin Jr, C. R. (1987). "Don't Know" Item Nonresponse in a Telephone Survey: Effects of Question Form and Respondent Characteristics. *Journal of Marketing Research*, *24*(4), 418–424. Retrieved from https://www.jstor.org/stable/3151390

Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, 50–60. doi:10.1214/aoms/1177730491

O'Muircheartaigh, C. A., Krosnick, J. A., & Helic, A. (2000). *Middle alternatives, acquiescence, and the quality of questionnaire data*. Working Papers 0103, Harris School of Public Policy Studies, University of Chicago.

Olson, K., & Smyth, J. D. (2015). The effect of CATI questions, respondents, and interviewers on response time. *Journal of Survey Statistics and Methodology*, *3*(3), 361–396.

Pickery, J., & Loosveldt, G. (1998). The impact of respondent and interviewer characteristics on the number of "no opinion" answers. *Quality and Quantity*, *32*(1), 31–45.

Rousselet, G. A. (2016). *Robust effect sizes for 2 independent groups*. Retrieved from https://garstats.wordpress.com/2016/05/02/robust-effect-sizes-for-2-independent-groups/

Rousselet, G. A., Foxe, J. J., & Bolam, J. P. (2016). A few simple steps to improve the description of group results in neuroscience. *European Journal of Neuroscience*, *44*(9), 2647–2651.

Rousselet, G. A., Pernet, C. R., & Wilcox, R. R. (2017). Beyond differences in means: Robust graphical methods to compare two groups in neuroscience. *European Journal of Neuroscience*, *46*(2), 1738–1748. doi:10.1111/ejn.13610

Saris, W. E., & Gallhofer, I. (2007). Estimation of the effects of measurement characteristics on the quality of survey questions. *Survey Research Methods*, *1*(1), 29–43. doi:10.18148/srm/2007.v1i1.49

Shoemaker, P. J., Eichholz, M., & Skewes, E. A. (2002). Item nonresponse: Distinguishing between don't know and refuse. *International Journal of Public Opinion Research*, *14*(2), 193–201.

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge University Press.

Vargha, A., & Delaney, H. D. (2000). A critique and improvement of the CL common language effect size statistics of McGraw and Wong. *Journal of Educational and Behavioral Statistics*, *25*(2), 101–132.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin*, *1*(6), 80–83. doi:10.2307/3001968

Zar, J. H. (2010). *Biostatistical analysis*. Prentice-Hall: Pearson.

Appendix A
R-Code examples

A1. Individual profiles

This appendix is provided with an R code in order to construct *individual* profiles yourself. You can simply fill out numbers of items for respondents A and B, and then run the complete code.

```
# Fill out numbers for the construction of individual profiles A and B yourself;
# set GA for the number of items for which a behaviour has been shown by respondent A;
# set IA for the number of items that has been filled out by respondent A;
# set GB for the number of items for which a behaviour has been shown by respondent B;
# set IB for the number of items that has been filled out by respondent B.

GA = 14
IA = 30
GB = 16
IB = 30

# After filling out GA, IA, GB, and IB, just run the rest of the code:
step = 0.01; seq = seq(0.5 * step, 1 - (0.5 * step), step)

# Construct the individual profiles:
respA = dbinom(GA, IA, seq)
respB = dbinom(GB, IB, seq)

# Normalize the individual profiles:
respA = respA / (sum(respA) * step)
respB = respB / (sum(respB) * step)

# Plot the individual profiles:
plot(seq, respA, type='l', col='blue', lwd=2, ylim=c(0,8), axes=FALSE,
main = c(paste('Constructing Behaviour Profiles Yourself'),
paste('for Respondent A and B')),
xlab = 'Probability of Behaviour',
ylab = 'Plausibility')
axis(side=1, at=seq(0, 1, 0.1))
axis(side=2, at=seq(0, 8, 2), las=1)
lines(seq, respB, type='l', lwd=2, col='red')
legend('top', bty='n', c('Profile for Respondent A',
'Profile for Respondent B'), cex=0.9, fill=c('blue', 'red'))

# Calculate expected values for the individual profiles:
sum(seq * respA) * step
sum(seq * respB) * step

# Calculate adapted Cliff's Delta for the individual profiles:
zA = respA
zB = respB
DIF = matrix(0, length(zA), length(zB))
for(x in 1:length(zA))
{
    for(y in 1:length(zB))
}
dif = (x-1) * step - (y-1) * step
DIF[x,y] = dif
DIF = sign(DIF)
DEN = matrix(0, length(zA), length(zB))
for(x in 1:length(zA))
{
    for(y in 1:length(zB))
}
den = zA[x] * zB[y] * step^{}2
DEN[x,y] = den
CD = sum(DIF * DEN) / sum(DEN)
CD
```

A2. Group profiles

This appendix is provided with an R code in order to construct *group* profiles yourself. You can simply fill out numbers of items for respondents in groups A and B, and run the complete code.

```
# Fill out numbers for the construction of group profiles A and B yourself;
# set GA and GB for the number of items for which a behaviour has been shown
# by the three respondents in group A and the three respondents in group B respectively;
# set IA and IB for the number of items that has been filled out
```

```
# by the three respondents in group A and the three respondents in group B respectively.
# Note that each group can be enlarged by any number of respondents by simply
# adding new respondents under 'Input for group A' and/or 'Input for group B'.

# Input for group A:
GA = c(10, 15, 20)
IA = c(40, 40, 40)

# Input for group B:
GB = c(15, 20, 25)
IB = c(40, 40, 40)

# After filling out GA, IA, GB, and IB, just run the rest of the code:
step = 0.01; seq = seq(0.5 * step, 1 - (0.5 * step), step)

# Construct profile A based on the average of all group members:
inputA = matrix(c(GA, IA), ncol=2)
nrespA = nrow(inputA)
groupA = matrix(0, 1/step, 1)
nA = 0
for (i in 1:nrespA)
{
    db = dbinom(inputA[i,1], inputA[i,2], seq)
    dbA = matrix(db, 1/step, 1) / (sum(db) * step)
    groupA = groupA + dbA
    nA = nA + 1
}
groupA = groupA / nA

# Construct profile B based on the average of all group members:
inputB = matrix(c(GB, IB), ncol=2)
nrespB = nrow(inputB)
groupB = matrix(0, 1/step, 1)
nB = 0
for (i in 1:nrespB)
{
    db = dbinom(inputB[i,1], inputB[i,2], seq)
    dbB = matrix(db, 1/step, 1) / (sum(db) * step)
    groupB = groupB + dbB
    nB = nB + 1
}
groupB = groupB / nB

# Plot the group profiles:
plot(seq, groupA, type='l', col='blue', lwd=2, ylim=c(0,8), axes=FALSE,
main = c(paste('Constructing Behaviour Profiles Yourself'),
paste('for Respondent Groups A and B')),
xlab = 'Probability of Behaviour',
ylab = 'Plausibility')
axis(side=1 , at=seq(0, 1, 0.1))
axis(side=2 , at=seq(0, 8, 2), las=1)
lines(seq, groupB, type='l', lwd=2, col='red')
legend('top', bty='n', c('Profile for Group A',
{}'Profile for Group B'), cex=0.9, fill=c('blue', 'red'))

# Calculate expected values for the group profiles:
sum(seq * groupA) * step
sum(seq * groupB) * step

# Calculate adapted Cliff's Delta for the group profiles:
zA = groupA
zB = groupB
DIF = matrix(0, length(zA), length(zB))
for(x in 1:length(zA))
{
   for(y in 1:length(zB))
   {
       dif = (x-1) * step - (y-1) * step
       DIF[x,y] = dif
   }
}
DIF = sign(DIF)
DEN = matrix(0, length(zA), length(zB))
for(x in 1:length(zA))
{
   for(y in 1:length(zB))
   {
       den = zA[x] * zB[y] * step^{}2
       DEN[x,y] = den
   }
```

```
}
CD = sum(DIF * DEN) / sum(DEN)
CD
```

Appendix B
The continuous form of $\hat{\delta}_P$

For reasons of practical utility, we computed $\hat{\delta}_P$ for discrete behaviour profiles. However, we can adjust our formulation for continuous profiles. In order to do this, we multiply both the numerator and denominator in formula (10) by step sizes Δp_a and Δp_b of both respondent profiles:

$$\hat{\delta}_P = \frac{\sum_{a=1}^{A}\sum_{b=1}^{B} \mathrm{sgn}(p_a - p_b)\overline{\Lambda}_A(p_a)\overline{\Lambda}_B(p_b)\Delta p_a \Delta p_b}{\sum_{a=1}^{A}\sum_{b=1}^{B} \overline{\Lambda}_A(p_a)\overline{\Lambda}_B(p_b)\Delta p_a \Delta p_b} \quad . \tag{11}$$

As the profiles are normalized to have an area of 1, the outcome of the denominator in formula (11) is 1. Removing the denominator in formula (11) yields

$$\hat{\delta}_P = \sum_{a=1}^{A}\sum_{b=1}^{B} \mathrm{sgn}(p_a - p_b)\overline{\Lambda}_A(p_a)\overline{\Lambda}_B(p_b)\Delta p_a \Delta p_b \quad . \tag{12}$$

By taking the limit $\Delta p_a \to 0$ and $\Delta p_b \to 0$, we can write the summations in formula (12) as integrals, transforming $\hat{\delta}_P$ into its continuous form $\hat{\delta}_{\mathrm{PC}}$:

$$\hat{\delta}_{\mathrm{PC}} = \int_0^1 \int_0^1 \mathrm{sgn}(p_A - p_B)\overline{\Lambda}_A(p_A)\overline{\Lambda}_B(p_B)\,\mathrm{dp}_A\,\mathrm{dp}_B \quad . \tag{13}$$

In Appendices C and F, it appears helpful to write $\hat{\delta}_{\mathrm{PC}}$ in an alternative form. To derive this form, we may first rewrite formula (13) as

$$\hat{\delta}_{\mathrm{PC}} = \int_0^1 \left(\int_0^1 \mathrm{sgn}(p_A - p_B)\overline{\Lambda}_A(p_A)\,\mathrm{dp}_A \right) \overline{\Lambda}_B(p_B)\,\mathrm{dp}_B \quad . \tag{14}$$

Next, we can split up the integral between parentheses into two parts:

$$\int_0^1 \mathrm{sgn}(p_A - p_B)\overline{\Lambda}_A(p_A)\,\mathrm{dp}_A = \int_0^{p_B} \mathrm{sgn}(p_A - p_B)\overline{\Lambda}_A(p_A)\,\mathrm{dp}_A + \int_{p_B}^1 \mathrm{sgn}(p_A - p_B)\overline{\Lambda}_A(p_A)\,\mathrm{dp}_A \quad . \tag{15}$$

Since $p_A < p_B$ in the first integral and $p_A > p_B$ in the second integral in the right hand side of the formula, we obtain $\mathrm{sgn}\,(p_A - p_B) = -1$ and $\mathrm{sgn}\,(p_A - p_B) = 1$ respectively. Therefore, we may write formula (15) as

$$\int_0^1 \mathrm{sgn}(p_A - p_B)\overline{\Lambda}_A(p_A)\,\mathrm{dp}_A = -\int_0^{p_B} \overline{\Lambda}_A(p_A)\,\mathrm{dp}_A + \int_{p_B}^1 \overline{\Lambda}_A(p_A)\,\mathrm{dp}_A \quad . \tag{16}$$

We can substitute this result into formula (14), leading to

$$\hat{\delta}_{\mathrm{PC}} = \int_0^1 \left(-\int_0^{p_B} \overline{\Lambda}_A(p_A)\,\mathrm{dp}_A + \int_{p_B}^1 \overline{\Lambda}_A(p_A)\,\mathrm{dp}_A \right) \overline{\Lambda}_B(p_B)\,\mathrm{dp}_B \quad . \tag{17}$$

As both behaviour profiles are normalized to have an area of 1, we can derive that

$$\int_0^{p_B} \overline{\Lambda}_A(p_A)\,\mathrm{dp}_A + \int_{p_B}^1 \overline{\Lambda}_A(p_A)\,\mathrm{dp}_A = \int_0^1 \overline{\Lambda}_A(p_A)\,\mathrm{dp}_A = 1 \quad , \tag{18}$$

which means that

$$-\int_0^{p_B} \overline{\Lambda}_A\,(p_A)\;\mathrm{dp}_A = \int_{p_B}^1 \overline{\Lambda}_A\,(p_A)\;\mathrm{dp}_A - 1 \quad . \tag{19}$$

Substituting this outcome into formula (17) gives

$$\hat{\delta}_{\mathrm{PC}} = \int_0^1 \left(2\int_{p_B}^1 \overline{\Lambda}_A(p_A)\,\mathrm{dp}_A - 1 \right) \overline{\Lambda}_B(p_B)\,\mathrm{dp}_B \quad , \tag{20}$$

resulting in the alternative way of writing $\hat{\delta}_{\mathrm{PC}}$:

$$\hat{\delta}_{\mathrm{PC}} = 2\int_0^1 \left(\int_{p_B}^1 \overline{\Lambda}_A(p_A)\,\mathrm{dp}_A \right) \overline{\Lambda}_B\,(p_B)\;\mathrm{dp}_B - 1 \quad . \tag{21}$$

Note that we may also rewrite formula (13) as follows:

$$\hat{\delta}_{\mathrm{PC}} = -\int_0^1 \left(\int_0^1 \mathrm{sgn}(p_B - p_A)\overline{\Lambda}_B(p_B)\,\mathrm{dp}_B \right) \overline{\Lambda}_A(p_A)\,\mathrm{dp}_A \quad . \tag{22}$$

After using similar steps of derivation, we then obtain

$$\hat{\delta}_{\mathrm{PC}} = -2\int_0^1 \left(\int_{p_A}^1 \overline{\Lambda}_B(p_B)\,\mathrm{dp}_B \right) \overline{\Lambda}_A\,(p_A)\;\mathrm{dp}_A + 1 \quad . \tag{23}$$

Formula's (13), (21), and (23) pose three different forms to compute $\hat{\delta}_{\mathrm{PC}}$ for continuous behaviour profiles. These forms may be utilized for specific purposes, as we do in Appendices C and F.

Appendix C
Profiles with an infinite number of items

In section 4.3, our numerical simulation suggests that $\hat{\delta}_P$ converges to δ as the number of items tends to infinity. In this appendix, we provide a mathematical proof for this claim. For this purpose, we use the alternative continuous form of Cliff's Delta, as derived in Appendix B. See formula (21). With help of formula (5), this form can be written as

$$\hat{\delta}_{\mathrm{PC}} = \frac{1}{R_A R_B} \sum_{r_A=1}^{R_A} \sum_{r_B=1}^{R_B} \hat{\delta}_{\mathrm{PC}_{r_A r_B}} \quad , \tag{24}$$

with

$$\hat{\delta}_{\mathrm{PC}_{r_A r_B}} = 2 \int_0^1 \left(\int_{p_B}^1 \Lambda_{r_A}(p_A)\, \mathrm{d}p_A \right) \Lambda_{r_B}(p_B)\, \mathrm{d}p_B - 1 \quad . \tag{25}$$

Here, Λ_{r_A} is the profile of an individual respondent r_A in group A, which can be expressed as

$$\Lambda_{r_A}(p_A) = \frac{p_A^{x_{r_A} I_{r_A}} (1-p_A)^{(1-x_{r_A}) I_{r_A}}}{\int_0^1 p_A^{x_{r_A} I_{r_A}} (1-p_A)^{(1-x_{r_A}) I_{r_A}}\, \mathrm{d}p_A} \quad , \tag{26}$$

with $x_r = \frac{G_r}{I_r}$ being the proportion of items for which the behaviour is shown. Similarly,

$$\Lambda_{r_B}(p_B) = \frac{p_B^{x_{r_B} I_{r_B}} (1-p_B)^{(1-x_{r_B}) I_{r_B}}}{\int_0^1 p_B^{x_{r_B} I_{r_B}} (1-p_B)^{(1-x_{r_B}) I_{r_B}}\, \mathrm{d}p_B} \quad . \tag{27}$$

When formula's (26) and (27) are substituted into (25), we obtain

$$\hat{\delta}_{\mathrm{PC}_{r_A r_B}} = 2 \frac{\int_0^1 \left(\int_{p_B}^1 p_A^{x_{r_A} I_{r_A}} (1-p_A)^{(1-x_{r_A}) I_{r_A}}\, \mathrm{d}p_A \right) p_B^{x_{r_B} I_{r_B}} (1-p_B)^{(1-x_{r_B}) I_{r_B}}\, \mathrm{d}p_B}{\left(\int_0^1 p_A^{x_{r_A} I_{r_A}} (1-p_A)^{(1-x_{r_A}) I_{r_A}}\, \mathrm{d}p_A \right) \left(\int_0^1 p_B^{x_{r_B} I_{r_B}} (1-p_B)^{(1-x_{r_B}) I_{r_B}}\, \mathrm{d}p_B \right)} - 1 \quad . \tag{28}$$

In the limit of $I_{r_A} \to \infty$, we can rewrite this result as

$$\hat{\delta}_{\mathrm{PC}_{r_A r_B}} = 2 \frac{\int_0^1 \left(\left(1_{x_{r_A} > p_B} \right) \int_0^1 p_A^{x_{r_A} I_{r_A}} (1-p_A)^{(1-x_{r_A}) I_{r_A}}\, \mathrm{d}p_A \right) p_B^{x_{r_B} I_{r_B}} (1-p_B)^{(1-x_{r_B}) I_{r_B}}\, \mathrm{d}p_B}{\left(\int_0^1 p_A^{x_{r_A} I_{r_A}} (1-p_A)^{(1-x_{r_A}) I_{r_A}}\, \mathrm{d}p_A \right) \left(\int_0^1 p_B^{x_{r_B} I_{r_B}} (1-p_B)^{(1-x_{r_B}) I_{r_B}}\, \mathrm{d}p_B \right)} - 1 \quad , \tag{29}$$

where $1_{x>y}$ is the indicator function

$$1_{x>y} = \begin{cases} 1 & \text{if } x > y \\ \frac{1}{2} & \text{if } x = y \\ 0 & \text{if } x < y \end{cases} \quad . \tag{30}$$

When we substitute formula (30) into (29), it follows that

$$\hat{\delta}_{\mathrm{PC}_{r_A r_B}} = 2 \frac{\int_0^{x_{r_A}} p_B^{x_{r_B} I_{r_B}} (1-p_B)^{(1-x_{r_B}) I_{r_B}}\, \mathrm{d}p_B}{\int_0^1 p_B^{x_{r_B} I_{r_B}} (1-p_B)^{(1-x_{r_B}) I_{r_B}}\, \mathrm{d}p_B} - 1 \quad . \tag{31}$$

As $I_{r_B} \to \infty$, we can rewrite this result with help of the indicator function as

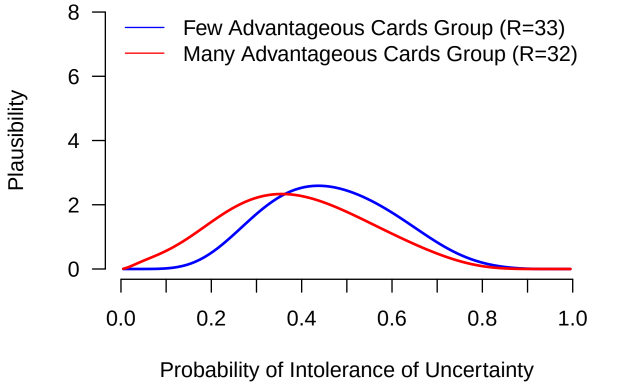
$$\hat{\delta}_{\mathrm{PC}_{r_A r_B}} = 2 \left(1_{x_{r_A} > x_{r_B}} \right) - 1 = \mathrm{sgn}\left(x_{r_A} - x_{r_B} \right) \quad . \tag{32}$$

Substitution into formula (24) yields

$$\hat{\delta}_{\mathrm{PC}} = \frac{1}{R_A R_B} \sum_{r_A=1}^{R_A} \sum_{r_B=1}^{R_B} \mathrm{sgn}\left(x_{r_A} - x_{r_B} \right) \quad , \tag{33}$$

which is identical to formula (8) in section 3. Hence, we have proven that the profile-based Cliff's Delta converges to the observation-based Cliff's Delta as the number of items approaches infinity.

(a) Subgroups that Chose Few versus Many Advantageous Cards and Their Intolerance of Uncertainty Level. 99% C.I. of $[-0.02, 0.54]$ and estimated profile-based Cliff's Delta of 0.28



(b) PVV Voters versus GroenLinks Voters and Their Openness to Experience Level. 99% C.I. of $[0.10, 0.32]$ and estimated profile-based Cliff's Delta of 0.21

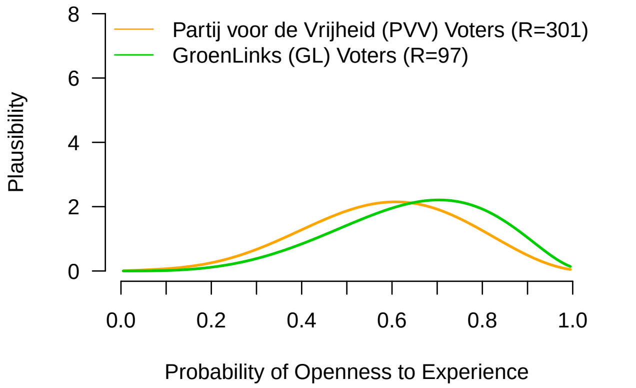


Figure D1. Respondent Profiles and Their Statistical Properties

Appendix D
Other data examples

In this appendix, we illustrate the broad usefulness of the adapted Cliff's Delta. For this purpose, we show two different types of examples than in section 5. See Figure D1a for the first example. The outcomes resulted from research in which individuals participated in playing the IOWA Gambling Task in order to clarify the effect of intolerance of uncertainty on future-oriented decision making (see Bais, 2012). Intolerance of uncertainty was measured by the Dutch version of the Intolerance of Uncertainty Scale (IUS) consisting of 27 items on a 5-points Likert scale (see De Bruin, Rassin, van der Heiden, & Muris, 2006). The IOWA Gambling Task consisted of choosing one card during 100 subsequent trials from four piles of cards that either yielded or cost fictitious money for every ten cards. From these piles, two piles consistently resulted in long term gain and two piles consistently resulted in long term loss of money. The goal for the participants was to prevent long term future loss of money. Hence, they had to discover that they should choose the (advantageous) cards from the two piles that resulted in long term money gain. Findings from a simple regression analysis showed a negative association between number of advantageous cards and intolerance of uncertainty. In other words, a higher level of intolerance of uncertainty was related to choosing fewer advantageous cards (Bais, 2012).

In order to use these data to construct two respondent profiles, we first split up the participants into a group that chose few and a group that chose many advantageous cards by using the overall median number of advantageous cards. Second, we transform the participant IUS scores of 1 through 5 on each item into 0, 0.25, 0.50, 0.75, and 1 respectively. Then we sum the transformed scores of all 27 items and round the total to the nearest integer for each participant. The outcome of this process is used as G_r in formula (2) with I_r=27. Finally, the respondent profiles for both groups are computed by formula's (3) and (5); see Figure D1a. When we compare the few to the many advantageous cards profile, the adapted Cliff's Delta is 0.28, which is a medium effect. Interestingly, this value is similar in magnitude to the effect size as found by Bais (2012) and thus confirms the earlier outcome. Note that the confidence interval for Cliff's Delta is broad, as we use 99% confidence intervals and both groups are relatively small.

See Figure D1b for the second example. Central in this example is the Big Five personality trait "openness to experience". Individuals showing a high degree of openness to experience have an intellectual curiosity, an active imagination, and a sensitive, analytical, and self-reflective mindset (see Goldberg, 1992; Guenole and Chernyshenko, 2005). We use respondent data from the LISS Panel surveys "Personality" and "Politics and Values"(both wave 6, 2014a, 2014b) of CentERdata to examine our idea that individuals who vote for a progressive-green party ("GroenLinks" or GL) might show more openness to experience than individuals who vote for a conservative-populistic party ("Partij voor de Vrijheid" or PVV). From the survey "Politics and Values", we consider the item "Which political party did you vote for during the most recent national Dutch elections on September 12th 2012?". All respondents who answered "PVV (Partij voor de Vrijheid)' or "GL (GroenLinks)" are taken into account for the construction of both respondent profiles.

From the survey "Personality", we consider the ten openness to experience items on a 5-points Likert scale (see Guenole and Chernyshenko, 2005). The transformation and summation of the scores is executed by the same procedure as in the former example. Subsequently, the respondent profiles for both political parties are constructed. See Figure D1b. When we compare the GroenLinks to the Partij voor de Vrijheid profile, the adapted Cliff's Delta is 0.21, which is a small but clearly present effect. To some extent, this confirms our idea that individuals who vote for GroenLinks are more opened to experience than individuals who vote for Partij voor de Vrijheid. Finally, note that the confidence interval for Cliff's Delta in this example is more narrow than in the former example, as both groups are relatively large.

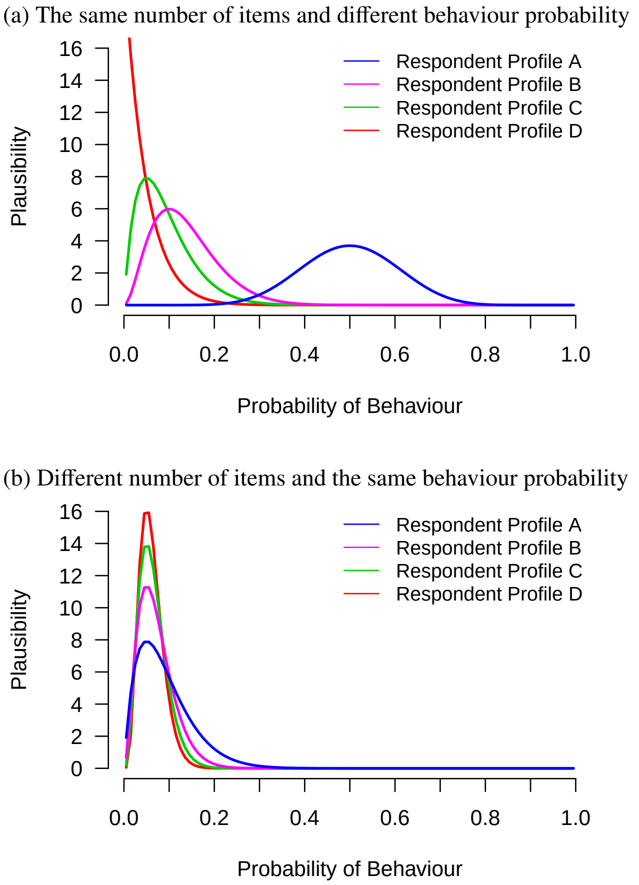


Figure E1. Examples of respondent profiles

Appendix E
Limitations of the behaviour profile method

In this appendix, we discuss and illustrate potentially difficult situations when constructing behaviour profiles. Answer behaviour may occur very rarely or very frequently. This may pose difficulty in comparing behaviour profiles using a statistical measure like the adapted Cliff's Delta. Particularly when one or both behaviour proportions are close to 0 or 1 *and* the one subgroup is based on a relatively small and lower number of items than the other subgroup, three undesirable situations may occur: 1) An effect occurs that does not actually exist; 2) an effect that actually exists is not detected; 3) an effect occurs that may contrast reality. Essentially, the undesirability of these three situations is related to the *different* number of items that two respondent profiles are based on. In other words, the one profile contains more uncertainty than the other profile. In case one or both behaviour occurrences are close to 0 or 1, this difference in uncertainty may be problematic in comparing two respondent profiles. From here, we give examples that refer to behaviour occurrences close to 0.

Before elaborating on the three situations, let us consider respondent profiles that are based on the *same* number of 20 items for some clarification first. See Figure E1a. Respondent profile A (blue) has a behaviour proportion of 0.50 and contains a degree of uncertainty that is equally spread around this value. This means that the expected value equals the behaviour proportion of 0.50. However, when considering respondent profile B (magenta) that has a behaviour proportion of 0.10, the uncertainty cannot be spread well towards the left of the profile center. As a result, the profile loses a bit of its symmetric shape and the expected profile value (0.14) is projected slightly upwards relative to the behaviour proportion. This effect is even a bit stronger for respondent profile C (green) that has a behaviour proportion of 0.05. The profile clearly has a skewed shape towards the right of the profile center and an expected value (0.09) that is projected upwards relative to the proportion. For respondent profile D (red) with a behaviour proportion of 0.00, this effect is striking, as its full area is forced to fall within the probability axis from 0 to 1. The result is a positively skewed profile and an increased expected value (0.05) relative to the proportion.

Now let us consider respondent profiles that have the same behaviour proportion of 0.05, but *differ* in the number of items that they are based on. See Figure E1b. Respondent profile D (red) that is based on 80 items has a nearly symmetric shape and an expected value (0.06) that is only marginally higher than 0.05. This means that a behaviour proportion close to 0 is not of concern in case the number of items is large enough to maintain the profile's nearly symmetric shape and initial centered value. The same effect applies to respondent profile C (green) that is based on 60 items. Its shape is still more or less symmetric and the expected value (0.06) is only slightly higher than 0.05. However, when we consider respondent profile B (magenta) that is based on 40 items, we notice that the profile's shape is a bit skewed towards the right of the profile's center and the expected value (0.07) is projected slightly upwards. The effect is the strongest for respondent profile A (blue) that is based on only 20 items. This profile is clearly positively skewed and its expected value (0.09) clearly deviates from the proportion of 0.05. In summary, the more the behaviour proportion is close to 0 and the lower the number of items that a respondent profile is based on, the more the profile is positively skewed and its expected value is projected upwards relative to the proportion. The same idea can be applied to behaviour proportions close to 1. From here, we elaborate on the undesirability of the three situations as delineated above.

6.1 The diverging case

In situation 1), the behaviour proportion for a group A based on the *larger* number of items is relatively close to 0; the proportion for a group B based on the *smaller* number of items is slightly less close to 0. Due to the smaller number of items and hence larger uncertainty of group B, the expected value for this group is projected more towards the center of the probability axis than for group A. Therefore, the expected value for group B may "diverge" from the expected value for group A. In other words, the expected values may differ *more* than is presumed on the basis of the difference between the proportions. Thus, such a "divergent case" may indicate a difference between two groups, while there may actually not exist a difference between them.

6.2 The converging case

Situation 2) is essentially the opposite of situation 1); the behaviour proportion in a group A based on the *smaller* number of items is relatively close to 0; the proportion in a group B based on the *larger* number of items is slightly less close to 0. Because of the smaller number of items and hence larger uncertainty of group A, the expected value for this group is projected more towards the center of the probability axis than for group B. Therefore, the expected value for group A may "converge" towards the expected value for group B. In other words, the expected values differ *less* than is presumed on the basis of the difference between the proportions. Thus, such a "convergent case" may not detect a difference between two groups, while there may actually exist a difference between them.

6.3 The contrasting case

Situation 3) is essentially a variant of situation 2). The only difference compared to situation 2) is that the expected value for group A (based on the *smaller* number of items) does not only 'converge" towards, but also *surpasses* the expected value for group B (based on the *larger* number of items). This means that based on the proportions, we may presume a more frequent behaviour for group B than for group A. However, based on the expected values, we may presume more frequent behaviour for group A than for group B. Thus, such a "contrasting case" may refer to a specific difference, while the actual difference may be the opposite one.

6.4 A few final notes

We need to note that an undesirable situation may occasionally occur when the behaviour proportions of both groups are not necessarily close to 0 or 1. This may happen when two groups are based on a very large and a very small number of items respectively. The influence of profiles based on very many items on projecting the expected value towards the center of the probability axis is much smaller than this influence of profiles based on very few items. Thus, comparing two respective groups based on very many and very few items may be of concern, regardless of their behaviour proportions.

Theoretically, the most optimal situation is the comparison of two large groups of respondents who all fill out the same large number of items. Practically, this is difficult to accomplish, as most surveys contain filter questions and as both characteristic's categories should then contain the same number of items. This means that in using respondent profiles, the profiles should be scanned on their behaviour proportions and expected values. In this way, distortions can be detected and the abovementioned situations can be omitted.

Appendix F

Relation between Cliff's Delta and Mann-Whitney U

In this appendix, we briefly illustrate the relation between the Mann-Whitney U statistic and Cliff's Delta. This statistic consists of two parts:

$$U_A = R_A R_B + \frac{R_A(R_A + 1)}{2} - T_A \quad (34)$$

and

$$U_B = R_A R_B + \frac{R_B(R_B + 1)}{2} - T_B \quad , \quad (35)$$

where U_A and U_B are the Mann-Whitney U statistics for a group A and B respectively, R_A and R_B are the numbers of observations in group A and B respectively, and T_A and T_B are the sums of the ranks of the observations in group A and B, respectively (see Mann and Whitney, 1947; Wilcoxon, 1945; Zar, 2010). The $\hat{\delta}_O$ is linearly related to U:

$$\hat{\delta}_O = \frac{2U_A}{R_A R_B} - 1 = \frac{-2U_B}{R_A R_B} + 1. \quad (36)$$

Considering formula's (21) and (23) in Appendix B, note that $\hat{\delta}_{\mathrm{PC}}$ is directly related to the form of $\hat{\delta}_O$ in formula (36):

$$\hat{\delta}_{\mathrm{PC}} = 2Z_A - 1 = -2Z_B + 1 \quad , \quad (37)$$

where

$$Z_A = \int_0^1 \left(\int_{p_B}^1 \overline{\Lambda}_A(p_A)\, \mathrm{dp}_A \right) \overline{\Lambda}_B(p_B)\, \mathrm{dp}_B \quad . \quad (38)$$

and

$$Z_B = \int_0^1 \left(\int_{p_A}^1 \overline{\Lambda}_B(p_B)\, \mathrm{dp}_B \right) \overline{\Lambda}_A(p_A)\, \mathrm{dp}_A \quad . \quad (39)$$

Finally, note that formula (36) differs from formula (37) in that $\hat{\delta}_O$ takes into account group size, whereas $\hat{\delta}_{\mathrm{PC}}$ does not. The reason is that $\hat{\delta}_{\mathrm{PC}}$ compares two full behaviour profiles by definition, which can be based on any number of respondents.