Contents lists available at ScienceDirect

Journal of Contextual Behavioral Science

"Using the implicit relational assessment procedure (IRAP) to explore common humanity as a dimension of self-compassion"

Eman Alasiri[a], Diana Bast[b,*], Russell L. Kolts[a]

[a] Eastern Washington University, Cheney, WA, 99004, USA
[b] Trinity Dublin College, College Green, Dublin 2, Ireland

ABSTRACT

Self-compassion encompasses three factors: self-kindness, mindfulness, and common humanity. This study aimed to explore the factor of common humanity (the belief that suffering is a shared human experience) and whether people tend to identify with this factor more in relation to their own experiences or those of others. Previously, it has been suggested that when applying compassion, people tend to be harsher on themselves and more compassionate toward others. Most research in self-compassion has typically used explicit measures, which can be scientifically limiting. The present study employed an implicit measure, the Implicit Relational Assessment Procedure (IRAP) to explore this domain. Participants ($N = 55$) completed the IRAP and its analog scale as well as Neff's self-compassion scale. Results showed that explicitly, participants reported high levels of common humanity and lack of isolation for both themselves as well as others. Implicitly, however, while they showed similar patterns in three of these four trial-types (i.e., common humanity regarding self and others and isolation with respect to others) they were neutral regarding self-isolation. Implications and limitations are discussed.

Compassion involves being moved by suffering, combined with the motivation to alleviate and prevent such suffering (Gilbert, 2010). Self-compassion involves compassion directed at one's own suffering, and is an important concept that has often been related to well-being and mental health. Neff (2003a, 2003b) operationally defined self-compassion in terms of three factors: self-kindness, mindfulness, and common humanity. Self-kindness means nourishing the self with kindness and acceptance rather than judgments and harsh criticism (Neff, 2003a). Mindfulness is a balanced state of mind involving acceptance of thoughts regardless of their directions (Brown & Ryan, 2003; Tirch, 2010). Common humanity is the realization that personal flaws, failures, vulnerabilities, and suffering are shared by all humans rather than being unique and isolated experiences of the individual (Neff, 2003a, 2003b, 2009).

The focus of the present study is the factor of common humanity. The experience of common humanity allows the individual who is suffering to contextualize their painful experience not as something that is isolating or unique to them, but as a universal experience that is shared with other beings – in particular, other humans. A high sense of common humanity can decrease self-judgment and self-blaming, associated with thoughts such as, "I am not alone in this" and "this is not my fault" (Neff, 2003b), allowing experiences of uniqueness and isolation to be replaced with feelings of normalization and connectedness.

Within a common humanity framework, there is a shared understanding and support system when thinking about self-shortcomings and weaknesses (Neff, 2009).

Even though common humanity provides the experience of an important and accessible support system (Neff, 2009), hardship and suffering can easily provoke people to feel isolated as they focus on their own difficulties (Neff, 2008). When people feel their experiences of suffering are unique and isolated, they become prone to psychological problems such as anxiety and depression (Lee, Robbins, & Hill, 1995a, 1995b; Neff, 2009). In sum, previous research suggests that understanding and feeling the common nature of personal difficulties is beneficial, yet challenging (Neff, 2008). However, previous research did not explore the difficulty of perceiving the experiences of others as part of a shared common humanity, and thus this is a key goal of the present study.

Neff (2003b) developed the Self-Compassion Scale (SCS) to measure the various factors involved in of self-compassion. Her operationalization included three factors, each of which was reflected in items tapping both positive and negative dimensions of the constructs: self-kindness vs self-judgment, mindfulness vs over-identification, and common humanity vs isolation. More recent analyses of the SCS have confirmed these factors and their polarities as six distinct dimensions that can be scored separately or combined into a single measure of self-compassion

(Neff et al., 2018).

In characterizing the intrapersonal dynamics of self-compassion, Neff (2003a) suggests that people tend to be harsher on themselves and kinder to others. If that was the case, it is important to investigate this relationship using implicit measures instead of merely focusing on self-reports. Self-reports can be easily influenced by external factors such as social desirability (Tourangeau & Yan, 2007), self-deceptive positivity (Paulhus, 1991), or even the tendency to answer without being aware of the contextual control influencing one's answers (Bast & Barnes-Holmes, 2014; Weber & Cook, 1972). It is especially challenging when using explicit measures for self-compassion factors, not only due to self-biases that might influence the way people respond, but also because people may not be aware of their emotional experiences. Schwarz and Clore (1983) suggest that emotional state temporarily affects how people evaluate their life experiences and thus explicit measures may also be prone to mood-related factors in this context.

The alternative to explicit self-report measurement is the use of measures which seek to target implicit processes. These typically work by requiring participants to respond under time pressure, which can limit the influence of extraneous factors (see e.g., Greenwald & Banaji, 1995). The measure used in the present research was the Implicit Relational Assessment Procedure (IRAP). This is a relatively new, yet well-established, computer-based task that asks participants to respond quickly and accurately to certain relational stimuli (Barnes-Holmes, Barnes-Holmes, Luciano, & McEnteggart, 2017; Barnes-Holmes, Barnes-Holmes, Stewart, & Boles, 2010) so as to measure the valence and directionality of specific attitudinal relationships (e.g., Self-Good, Other-Bad). The logic behind the IRAP is that the latency of responding to the tasks should be shorter across trial blocks that are consistent with participants' pre-existing knowledge/beliefs, and it should be longer across blocks that are inconsistent with participants' pre-existing knowledge/beliefs (Barnes-Holmes et al., 2006).

For example, in a study investigating the implicit relational assessment procedure as a measure of forgiveness of self and others (Bast & Barnes-Holmes, 2014), the IRAP contained statements about faults that pertained to the respondents, as well as similar statements that pertained to the faults of others. For example, a *self*-trial presented the statement "My faults are acceptable" while an *other* trial presented "Other people's faults are acceptable." As is typical in an IRAP, on some (consistent) blocks of trials participants had to respond consistent with one particular attitudinal direction while on other (inconsistent) blocks they had to respond opposite to that direction. In this case, some trial-blocks required participants to confirm self-forgiveness but deny forgiveness of others, while other blocks required them to confirm forgiveness of others and deny self-forgiveness. That is, during a "self-forgiveness" block, participants responded "My faults are acceptable"–TRUE and "Other people's faults are acceptable"–FALSE. In contrast, during an "other-forgiveness" block, participants responded "Other people's faults are acceptable"–TRUE and "My faults are acceptable"–FALSE. Interestingly, the IRAP results yielded a response bias towards forgiving oneself more readily than others, whereas the explicit self-report instrument suggested the opposite (a bias toward forgiving others rather than oneself). In addition, correlational analyses yielded limited evidence that the implicit and explicit measures were related, thus confirming that implicit versus explicit forgiveness of self and others may be under different sources of control, at least in the context of the psychological testing environment of the Bast and Barnes-Holmes study.

Another recent study (Kavanagh et al., 2019) used the Natural Language-IRAP (NL-IRAP) across a series of experiments to investigate deictic relational responding with regard to the mental states of self and others - including *unspecified* and *specified* others. The use of the NL-IRAP allowed for the presentation of relatively complex statements that required participants to infer the thoughts or beliefs of others on a trial-by-trial basis within the IRAP. A 'self-focused IRAP' required participants to respond to both positive and negative statements about themselves, whereas an 'other-focused IRAP' required participants to respond to similar statements about others. The results of the experiments that had sentences regarding *unspecified* others indicated that the other-focused IRAP produced overall bias scores that were significantly stronger than responding to the self-focused IRAP. However, non-significant differences were recorded across the experiments that had sentences when other was *specified*.

In the present study, we used natural language sentences in the IRAP to explore common humanity; that is, the relationship between how people perceive their difficulties versus those of others (e.g., When I have difficulties, I am unlike others; When another person has difficulties, they are unlike others) and the results were correlated with explicit measures (i.e., self-compassion scales and an IRAP analog).

This study represents an innovation in research in two ways. First, common humanity is typically measured in terms of how people perceive their own suffering in relation to that of others; however, in this present study, we also investigated how people perceive other people's suffering in relation to that of others. Second, at the time of writing, no published study had attempted to use an implicit measure to assess common humanity of the self versus others (an important aspect of self-compassion) or self-compassion *per se*. Given the lack of such previous work, the current research should be seen as largely exploratory and constituting only a first step toward a more complete analysis of self-compassion.

1. Methods

1.1. Participants

One hundred and twenty-four college students ranging in age from 18 to 49 with a mean of 21 were recruited from Psychology classes at a public university located in the inland northwest of the United States. Fifty-five of these (14 male, 40 female, 1 gender-fluid) met the requirements of the study and hence, only their data were used. None had been previously exposed to the implicit measure. Participants received course credits as compensation for taking part in the study.

1.2. Setting, apparatus and materials

The study took place in a quiet room free from distractions. Both measures were presented to participants using a standard laptop. The 2012 version of the IRAP was employed. Explicit measures (questionnaires) were completed using an electronic survey format.

Implicit Relational Assessment Procedure (IRAP). The IRAP is a computerized program that allows participants to read the instructions at their own pace and start once they are ready by directing them to press the "spacebar" in the keyboard. The IRAP measures response latencies; how quickly and accurately participants respond to stimuli as they are presented in different ways. The IRAP also provides participants with feedback indicating if they have met the requirements of a given block or not, and prompts them to go fast when they take too long on a given trial. The purpose of the feedback is to constantly remind the participants of the expected criteria to which they need to adhere. For the current IRAP, there were two different sets of instructions, alternating between different blocks. The first set of instructions stated, "For this block, answer in a way that is compassionate toward yourself, but not compassionate toward others," while the opposite set of instructions stated, "For this block, answer in a way that is compassionate toward others, but not compassionate toward yourself." One of the two sets occurred before each block and participants could spend as much time as they needed to read. Then, when ready, they could press the "spacebar" to proceed to the task. It is important to mention here that we chose to use the word "compassionate" to avoid priming participants to the idea of common humanity.

The IRAP format includes three key stimuli on each trial; a label stimulus, a target stimulus and two responding statements. Label

stimuli for this IRAP were two different statements; one pertaining to the self, ("When I have difficulties, I am …), and one pertaining to others ("When another person has difficulties, they are …"). These two statements were quasi-randomly alternated across trials. In addition, target stimuli were quasi-randomly paired with label stimuli, across trials. For this IRAP, there were 12 target stimuli. Six reflected common humanity, (normal, connected, the same as others, human, similar to others, like others), and six reflected isolation, (abnormal, cut off, different from others, flawed, worse than others, unlike others). These statements were derived from Neff's subscales of common humanity and isolation in Self-Compassion Scale (SCS; 2003b) and indeed were significantly correlated with Neff's (2003b) subscales of common humanity and isolation (reported later). When developing these statements, it was seen as crucially important to make the statements in the common humanity category match the length (number of syllables) of their counterparts in the isolation category, so that reaction time would be an accurate indication of the directionality of the implicit belief rather than the length of the targets. After developing the initial statements, they were tested by gathering pilot data (not included here) and minor changes were made to the statements to better represent the concepts of common humanity and isolation.

The two responding statements appeared at the bottom of the screen prompting participants to choose 'd', which was on the left side of the screen, for "True" and 'k', which was on the right side of the screen, for "False." For the current study, these two options occurred simultaneously and stayed in the same position across trials.

For a given trial, the label and target stimuli represented one of four possible combinations, (a) Self-Common Humanity (e.g. 'When I have difficulties, I am … Connected'); (b) Self-Isolation (e.g. 'When I have difficulties, I am … Cut-off'); (c) Others-Common Humanity (e.g. 'When another person has difficulties, they are … Connected'); (d) Others-Isolation (e.g. 'When another person has difficulties, they are … Cut-off'). Participants could then choose either "True" or "False" for each given combination based on the instructions they were given at the beginning of each block. An example of the four trial-types of the current IRAP is presented in Fig. 1.

Self-Compassion Scale (SCS, Neff, 2003b**).** This scale, one of two 'explicit' measures, contained 26 statements which participants rated from 1 to 5, with 1 being (almost never) and 5 being (almost always). Even though the whole scale was used, only the common humanity and isolation subscales, which are negatively correlated ($r = -.560$, $p < .000$), were included in the following analysis. An example of a common humanity statement is "When things are going badly for me, I see the difficulties as part of life that everyone goes through", and an example of isolation is "When I think about my inadequacies, it tends to make me feel more separate and cut off from the rest of the world". Subscale scores are computed by calculating the mean of subscale item responses and scores of negative subscales items such as isolation are reversed (i.e., 1 = 5, 2 = 4, 3 = 3, 4 = 2, 5 = 1). Research indicates the SCS has an appropriate factor structure, and that a single factor of "self-compassion" can explain the inter-correlations among the six
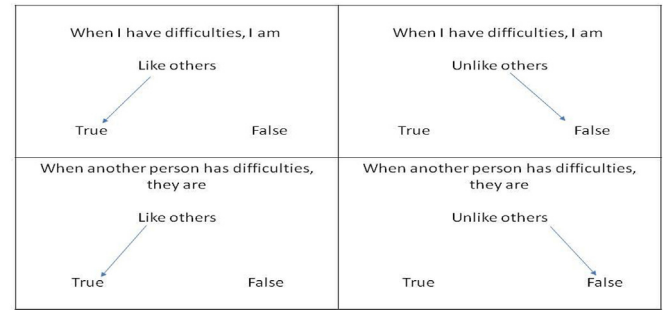
Table 1
Labels, response options and targets deployed in the IRAP in both experiments.

Label 1	Label 2
When I have difficulties, I am	When another person has difficulties, they are
Response Option 1	**Response Option 2**
True	False
Terms deemed consistent with Label 1	**Terms deemed consistent with Label 2**
Normal	Abnormal
Connected	Cut off
The same as others	Different from others
Human	Flawed
Similar to others	Worse than others
Like others	Unlike others

facets (Neff, 2003a). The scale also demonstrated concurrent validity (e.g., correlates with social connectedness), convergent validity (e.g., correlates with therapist ratings), discriminant validity (e.g., no correlation with social desirability), and test-retest reliability ($\alpha = .93$; Neff, 2003a).

IRAP Analog. The second explicit measure used was an analogue of the same stimuli used in the IRAP. The two labels, ('When I have difficulties, I am … ') and ('When another person has difficulties, they are … '), were paired with the six common humanity targets and the six isolation targets. Hence, there were 24 statements, which participants rated on a scale from 1 (Not True at All) to 7 (Very True). Each description involved using one of the twelve label stimuli employed in the IRAP (i.e., six negatively-valenced phrases indicating isolation and six positively-valenced phrases indicating common humanity; see Table 1) and each was applied to both the self and the others, thus giving a total of 24 items. Ratings obtained were used to calculate four separate scores, analogous to the four IRAP trial-type scores. The ratings for the items that targeted negative feelings were reversed (e.g., a score of 7 was rescored as 1) so that all positive scores indicate positive bias while all negative scores indicated negative bias. For purposes of graphical representation, responses on this 7-point scale were recorded to range from -3 (instead of 1) to $+3$ (instead of 7).

1.3. Procedure

Participants were guided to a quiet room equipped with a chair, desk and laptop. After completing consent forms, they were asked to complete the IRAP followed by the explicit measures. Each session took approximately 45 min in total (30 min to complete the IRAP and 15 min to complete the explicit measures).

Participants were first exposed to the IRAP. The researcher read relevant instructions from a script and explained to participants the nature of the task that they would perform. Participants were told to read the instructions at the beginning of each trial and they were shown the keys they needed to press to respond either "True" or "False." It was explained that they needed to answer in the directed way for each trial as fast as they could, and that an "X" mark would appear if they responded in a way that was not consistent with the instructions of that block. They were told that if this "X" appeared, then they couldn't move to the next trial until they chose the correct response.

This study employed the IRAP 2012. The IRAP was composed of six test blocks and two practice blocks. Participants started with the practice blocks and could retake them up to four times. If participants met the requirements of accuracy (at least 80% correct responses) and speed (at or less than 2000 ms), then they could proceed to the actual test blocks. If they did not meet either of these requirements, a "Thank you" screen occurred and they were debriefed and asked to move to the explicit measure (data from such participants was not included in the analysis). After collecting the pilot data, we decided to reduce the required speed to 1500 since most participants were able to respond

Fig. 1. An example of the four different trial -types presented in the current IRAP.

within that time frame. Hussey, Thompson, Mcenteggart, Barnes-Holmes, and Barnes-Holmes (2015) indicated that although most researchers have used the 2000 ms criterion, it is recommended to lean toward a more conservative approach when feasible. As explained earlier, there were two types of blocks alternating in the IRAP. In the first type ('consistent'), participants were instructed to answer in a compassionate way to the self but not to others. In this block, the correct response would be "True" to common humanity targets and "False" to isolation targets that are paired with the *self*, (e.g. responding False to "when I have difficulties, I am … Flawed and True to "when I have difficulties, I am … Human"). They would also respond "False" to common humanity targets and "True" for isolation targets that were paired with *others,* (e.g. responding False to "when another person has difficulties, they are … Human" and True to "when another person has difficulties, they are … Flawed").

For the second type of blocks ('inconsistent'), participants were directed to answer in a compassionate way toward others but not toward the self. In this block, the correct response would be "True" to common humanity targets and "False" to isolation targets that are paired with *others,* (e.g. responding False to "when another person has difficulties, they are … Flawed" and True to "when another person has difficulties, they are … Human"). They would also respond "False" to common humanity targets and "True" for isolation targets that were paired with the *self*, (e.g. responding False to "when I have difficulties, I am … Human and True to "when I have difficulties, I am … Flawed").

The IRAP recorded accuracy rate (percentage of correct answers) and latency (the time taken to make a response) for each participant and generated three Excel files. Once participants completed the practice blocks and the six test blocks, a "Thank you" statement was presented and they were directed to notify the researcher.

Following completion of the IRAP, participants were requested to spend a few minutes answering the explicit measures, which were presented on-screen. They were asked to read each item carefully and to ask for clarification from the researcher if anything seemed unclear.

2. Results

2.1. Data analysis

Two mixed repeated measures (2 × 4) ANOVAs and post-hoc tests were conducted for the IRAP and analog IRAP scale. Single-sample t-tests were conducted for the four trial-types for both the IRAP and the analog for each of the two conditions. Finally, a correlation matrix was created to check for relationships between the implicit and explicit measures.

2.2. IRAP

Data preparation. For the IRAP, response latency, defined as "the time in milliseconds from the onset of trial until the emission of a correct response" (Roddy, Stewart, & Barnes-Holmes, 2009, p. 421) was the primary dependent variable used. As explained previously, participants were required to maintain an accuracy level of ≥80% correct responses and a latency of ≤1500 ms in the test blocks to qualify for inclusion in the study. For each datum to be used, a participant needed to meet these requirements in at least two of the three consecutive pairs of test blocks. For participants who didn't meet the criteria in one of the three pairs of test blocks this pair was discarded and the *D*-IRAP scores were calculated by averaging the remaining two pairs of test blocks (Barnes-Holmes et al., 2010).

Individual response latencies were transformed into four *D*-IRAP scores using the D_{IRAP} algorithm (Barnes-Holmes et al., 2010) which was developed according to the *D* algorithm that was used to analyze data for the Implicit Association Test (IAT) (Schmukle, Egloff, & Gutenberg, 2005). The reason behind transforming latencies into *D* scores is to minimize the impact of factors such as age, motor skills,

and/or cognitive ability on latency data, allowing researchers to measure differences between groups using a response-latency paradigm with reduced contamination by individual differences associated with extraneous factors (Barnes-Holmes et al., 2010; Greenwald, Poehlman, Uhlmann, & Banaji, 2009).

D-IRAP scores were calculated as follows: (1) only latencies from the test blocks were used (data from practice blocks were discarded); (2) latencies over 10,000 ms were eliminated from the data set; (3) if a participant produced more than 10% of trials with latencies that were less than 300, the whole dataset for that person was removed; (4) 12 standard deviations were calculated for the four trial types, hence there were four for Test Blocks 1 and 2, four for Test Blocks 3 and 4, and four for Test Blocks 5 and 6; (5) 24 latency means were calculated for the four trial types in each test block; (6) difference scores were calculated for the four trial types between each pair of blocks by subtracting the latency means of the consistent block from the latency means of the corresponding inconsistent block; (7) each difference score was divided by its corresponding standard deviation, calculated previously in Step 4, generating 12 *D*-IRAP scores, one for each trial type in each pair of the test blocks; and finally (8) 4 *D*-IRAP scores were generated for each trial type by averaging the 12 *D*-IRAP scores calculated in Step 7. The same steps were followed for participants who only met the criteria in two of the three pairs of test blocks with minor adjustments such as calculating 8 standard deviation for each trial types instead of 12 (Bast, Linares, Gomes, Kovac, & Barnes-Holmes, 2016). In addition, to aid in interpretation, scores for the two 'Others' trial types were inverted (i.e., multiplied by −1; see Hussey et al., 2015). Following this transformation, positive D-IRAP scores indicated a positive bias (i.e., showing a bias more towards common humanity than isolation), while negative scores indicated a negative bias (i.e., showing a bias away from common humanity and more in the direction of isolation).

Mean scores analyses. The four *D*-IRAP scores were then entered into a one-way repeated-measures analysis of variance (ANOVA). Results of this analysis showed that there was a significant effect of trial-type on responding, $F\ (3,\ 162) = 26.42$, $p = .001$, $\eta^2 = .60$. Bonferroni post hoc test results showed that self-common humanity trial-type scores differed significantly from self-isolation trial type scores ($p = .001$). The self-common humanity trial-type also differed significantly from the others-common humanity and others isolation trial-types ($p = .001$). Further, results yielded a significant difference between the self-isolation and others-isolation trial types ($p = .017$), as well as between the self-isolation and others-common humanity trial-types, ($p = .020$). Results showed no significant difference between the others-common humanity and others-isolation trial-types. In addition, the four trial-type scores were tested using one-sample *t*-testing and results indicated significant differences for all of the trial-types ($p \leq .001$) except for the self-isolation trial-type. These results are presented in Fig. 2.

Statistical analyses thus indicated that participants showed an implicit bias toward confirming experiences of common humanity for the self and others with a stronger bias toward the self. In other words, participants agreed that in the case of both themselves and others, psychological struggle is something that we share with humanity in general, though they endorsed this somewhat more readily for themselves than others. Regarding the isolation trial type, participants did not show a specific isolation bias toward the self (i.e., they neither confirmed nor rejected that they themselves are isolated when they struggle) but they did show a (positive) isolation bias with respect to other people (i.e., denying that others are different from most people when they struggle). In other words, participants showed an implicit bias toward rejecting the isolation of others when they struggle but did not show any bias with respect to self-isolation when they struggle.

2.3. IRAP analog

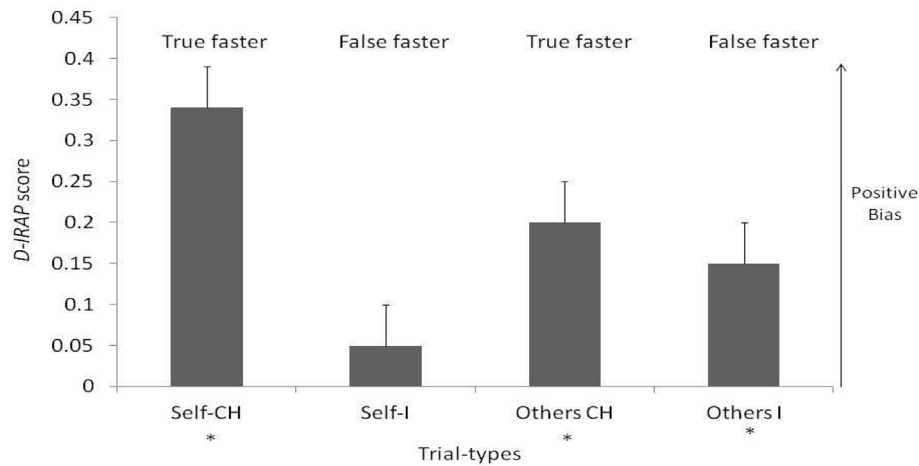**Mean scores analyses.** The overall mean ratings obtained from the

Fig. 2. Mean *D*-IRAP trial-type scores obtained for the Self and Other Common Humanity and Isolation IRAP. The letters *T* and *F* indicate the direction of the IRAP effect in terms of responding True or False, respectively. Asterisk (*) represents p < .000.
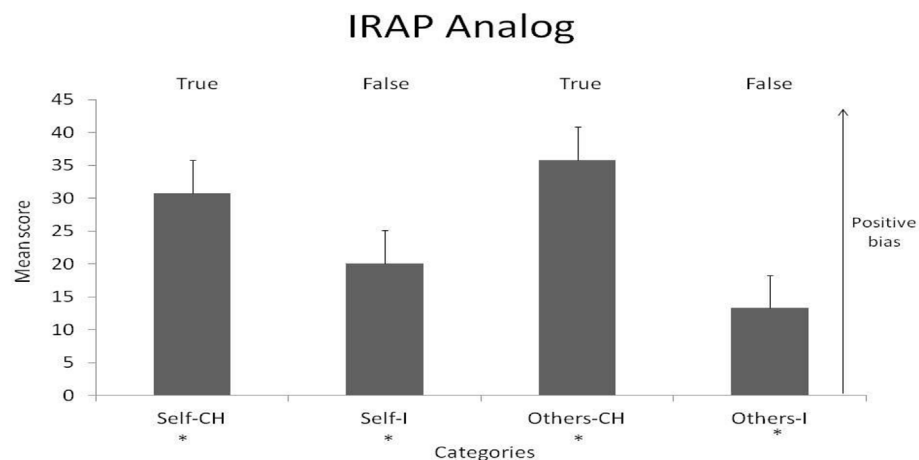

Fig. 3. Mean Scores of the IRAP-analog obtained through the sub-scales of common humanity and isolation.

IRAP Analog are presented in Fig. 3. Mean scores support the idea that participants tended to confirm common humanity both towards Self and Others, with the scores being slightly higher for Others-common humanity, indicating more positive bias towards Others. Participants also tended to reject isolation for both the self and for others, albeit with a higher score on isolation towards the self, indicating that explicitly, participants showed higher denial that they are unlike others when they struggle, indicating more positive bias towards the self in comparison to others.

Means and standard errors for the ratings of the four sets of statements were the following; self-common humanity, $M = 30.80$, $SE = 1.00$; self-isolation, $M = 20.13$, $SE = 1.19$; others-common humanity, $M = 35.87$, $SE = 0.87$; others-isolation, $M = 13.33$, $SE = 0.85$. The four mean scores were entered into a one-way repeated-measures analysis of variance (ANOVA). Mauchly's test indicated that the assumption of sphericity had been violated, x^2 (5) = 83.53, $p = .001$, and so the Greenhouse-Geisser correction ($\varepsilon = .55$) was utilized in the *F*-ratio calculation. Results indicated that there was a significant difference between ratings across the four sets of statements, F (1.64, 88.79) = 92.51, $p = .001$, $\eta^2 = .99$. Bonferroni post hoc test results concluded that all of the ratings differed significantly from one another, ($p = .001$). One sample *t*-testing was conducted on each of the four trial-types yielding significant values ($p < .001$).

These analyses thus suggest that, on the explicit measure, participants reported a significantly higher level of common humanity toward the experiences of others than experiences of the self while reporting a significantly higher level of isolation toward experiences of others than experiences of self. In other words, participants reported lower ratings of common humanity towards the self than towards others, while reporting less isolation regarding others' difficulties than their own.

2.4. Self-Compassion Scale

For the Self-Compassion Scale (SCS), the average was obtained for common humanity ($M = 3.10$, $SD = .80$) and isolation ($M = 3.15$, $SD = .93$) subscales. According to Neff (2003b), on a scale from 1 to 5, a score of 1–2.5 represents low self-compassion, 2.5 to 3.5 represents moderate self-compassion, and 3.5 to 5.0 represent high self-compassion when computing the overall mean of self-compassion. These results indicated that participants fell in the moderate range for these two subscales. As in Neff (2003a), the subscales of common humanity and isolation were found to be negatively correlated ($r = -.560$, $p < .000$).

2.5. Correlation matrix

The four IRAP trial-types, the IRAP analog and the SCS were entered into a correlation matrix. Results showed no significant correlations between the IRAP and the explicit measures. Meantime, all the explicit measures were intercorrelated. Self-common humanity on the IRAP analog was positively correlated with Neff's (2003a) subscale of common humanity ($r = .38$, $p = .004$), and negatively correlated with

Table 2
IRAP analog and SCS scores correlations.

SCS (Neff)	IRAP analog	r	p
Common Humanity	Self-Common Humanity	.381**	< 0.004
Common Humanity	Self-Isolation	.560**	< 0.000
Isolation	Self-Isolation	-.494**	< 0.000
Isolation	Self-Common Humanity	-.482**	< 0.000

**0.01 level (2-tailed).

the isolation subscale ($r = -.48$, $p = .001$). In other words, participants' ratings of statements pertaining to common humanity on the IRAP analog were positively correlated with their responses to the common humanity subscale and negatively correlated with their responses to the isolation subscale of the SCS. Similarly, self-isolation on the IRAP analog was positively correlated with the isolation subscale ($r = .56$, $p = .001$), and negatively correlated with the common humanity subscale ($r = -.49$, $p = .001$). These results support the construct validity of the statements derived from the SCS in the IRAP (see Table 2).

3. Discussion

Self-compassion, according to Neff (2003a; 2003b), is comprised of three, dimensional factors: self-kindness vs. self-self-judgment, self-common humanity vs. self-isolation, and mindfulness vs. over-identification. The present research focused on the factor of common humanity *vs.* isolation. Common humanity can be defined as the awareness that individual flaws, vulnerabilities, difficulties, and challenges are part of the human experience, whereas isolation can be defined as the notion that individual difficulties and suffering are unique and isolated experiences of the individual. This study aimed to explicitly and implicitly investigate common humanity and whether people identify with this factor more in relation to their own experiences or toward the experiences of others. The inclusion of an investigation at the implicit level was deemed important as most of the empirical studies on self-compassion have used explicit measures (questionnaires) (Gilbert, Mcewan, Matos, & Rivis, 2011; Neff, Pisitsungkagarn, & Hseih, 2008; Neff, Rude, & Kirkpatrick, 2007), which can be easily influenced by extraneous variables (Greenwald, McGhee, & Schwartz, 1998).

The current study found that on the questionnaire measure, participants reported that they experience others' suffering as a common human experience - that these (other) people are going through what everyone goes through in life - whereas they reported experiencing less of that sense of common humanity with regard to their own experiences of struggle. Thus, others' experiences of struggle were perceived as more "normal" than those experienced by the self. This finding was expected, and it is aligned with what Neff (2003a) explained as the common experience that people relate to when it comes to how they perceive their own struggle versus that of others. Regarding isolation, participants responded that they consider neither themselves nor others to be isolated when experiencing psychological struggles; further they reported feeling less isolated when facing a difficult situation in comparison to when other people go through a similar experience.

Comparing these findings with those from the IRAP as the implicit measure, we can see certain commonalities along with some interesting divergence. The broad thrust of the IRAP findings were that participants (i) agreed that both they themselves and others shared their struggles with humanity more broadly; though (ii) they endorsed this more strongly with respect to Self than Others; (iii) indicated that others are not isolated when they struggle; while (iv) not endorsing an attitude one way or the other as to whether they themselves are isolated or not during their own struggles (i.e., thus not rejecting the possibility that they might be isolated).

The first of these effects is obviously the same as that found for the

explicit measure. The second one though is divergent and thus interesting. This effect might suggest that participants are showing an implicit pattern of being easier on themselves than they are showing explicitly (see also Bast & Barnes-Holmes, 2014). This would accord with the idea that people are responding explicitly more on the basis of social desirability and what they think others such as the researcher might want to hear while implicitly having a divergent perhaps less socially acceptable attitude.

At the same time however, it is also possible that this particular pattern might have been influenced to at least some extent by a methodological issue, namely the so-called 'single trial type dominance effect'. In this effect the IRAP trial type that involves the highest levels of response coherence (because it involves endorsing agreement between the stimuli with the highest levels of valence) produces the strongest IRAP effect. In the present IRAP this trial type was the self-common humanity trial type which of course involves responding to Self rather than Others and Common Humanity rather than Isolation and thus (because these are both relatively more positively valenced factors) involves stronger coherence than any of the other trial types and thus is likely to produce a stronger *D*-IRAP effect than the others (Finn, Barnes-Holmes, & McEnteggart, 2018). While this methodological effect is not the only variable involved in producing the trial type effect seen, it is likely to be at least a partial response and thus it is harder in the case of this trial type to isolate the effect of relational responding alone.

While the single trial type dominance effect may have played a role in the strength of the effect seen for the Self-Common Humanity trial type, the suggestion that people might associate higher levels of common humanity with their own experiences than with the experiences of others is also consistent with previous research done in the field of social psychology. According Forsyth (2008), self-serving bias postulates that people tend to attribute positive qualities to themselves even if that means reaching a distorted conclusion about the self. The reason behind this possible distortion is to maintain an enhanced self-esteem by associating self with positive outcomes (Leary, 2004). Words used in the current IRAP for the common humanity factor can be described as positive and desired characteristics (normal, connected, human … etc.), so participants were faster in associating them to the self than to others. Researchers found similar results in previous research that examined self-esteem by using both implicit and explicit measures. They found that, when comparing self to others, participants associated more positive attributions with the self than with other people (Karpinski, 2004).

The third and fourth points made in respect of the IRAP above pertain not to common humanity but to isolation, which is the opposite concept in important respects. According to Neff (2003a; 2003b; Neff, 2016), common humanity and isolation represents one example of a set of opposite facets. One of the patterns just mentioned was that participants denied that other people are unlike others (i.e., isolated) when they suffer. In this respect the IRAP outcome is similar to that of the explicit measure. The more interesting finding is the fourth one, wherein they did not endorse an attitude one way or the other as to whether they themselves are isolated or not during their own struggles. This pattern is particularly interesting as it shows a divergence from the explicit measure. In the latter, people responded definitively that they were not isolated in their struggles whereas the latter finding shows more ambivalence in this respect. Once again, it might be argued that participants are showing a different attitude towards themselves than towards others. In this case, they might be seen to be describing their own suffering as more profound than that of other people. Once again this might be seen as responding towards their own situation as 'special' and different from that of others at the implicit level while denying this explicitly based on social desirability and what they think the researcher might want to hear. In addition to this furthermore, neither of these patterns of responding with respect to isolation would have been predicted in "the single trial type dominance effect" as according to the

model, neither of the trial types concerned should be affected by coherence to the extent that the self-common humanity trial type was.

This study has contributed by showing people's responses of common humanity and isolation towards the self and the others and indicating that the IRAP is potentially useful as an additional measure of these constructs (i.e., additional to self-report questionnaires). These findings are also relevant as regards Neff's conception of common humanity and isolation. According to this conception, these are polar opposites and therefore, increased common humanity should correspond with decreased isolation, and vice versa. It seems, however, that the two concepts were producing slightly different results using the IRAP; participants showed a high level of self-common humanity, but unexpectedly didn't have any bias, either toward or away, for self-isolation. If these were opposite sides of the same construct, we would expect that self-isolation would be decreasing as self-common humanity was increasing. These findings are consistent with critiques of the bi-dimensionality (i.e., positive versus negative facets) of the three factors. Muris and Petrocchi (2017), argue that the Self-Compassion scale (SCS) should not include both the positive and the negative factors, arguing that they represent very different processes. They found that the three positive factors (self-kindness, common humanity, and mindfulness) of the SCS were positively correlated with overall wellbeing, whereas the negative three (self-judgments, isolation, and over-identification) were positively correlated with psychopathology.

As discussed previously, no correlations were found between the explicit and implicit measures. This finding is consistent with previous research utilizing the IRAP (e.g. Bast & Barnes-Holmes, 2014; Bast et al., 2016). According to relational frame theory (RFT), the fact that the two measures were targeting the same concept but generated different results suggests the two might be mapping into different sets of cognitive and behavioral associations that are contextually controlled (see Hughes, Barnes-Holmes, & Vahey, 2012, for further explanation). Therefore, findings of the current study emphasize the importance of using the IRAP as an implicit relational measure to examine common humanity and common humanity-related behaviors for self and others from an angle that captures the behavioral complexity of human cognition.

This study has a few limitations. First, the rate of participants who did not pass the practice blocks in the IRAP was high. This might be the result of the complexity of the current IRAP (e.g. the wording of the stimuli). It might also be explained by the fact that the required latency, 1500 ms, was not enough time for participants to read and immediately respond to the sentences that were used in the current IRAP. As indicated earlier though, the researchers decided to choose a rigorous approach to limit the influence of individual differences (for further information see: Barnes-Holmes et al., 2010; Hussey et al., 2015). Moreover, previous research has shown that anxiety and depression reduce cognitive functioning, so participants with either condition or both possibly had poor performance and high dropout rate (Bast et al., 2016). Future research could attempt to control for this by implementing depression and anxiety scales. Another possible explanation is that almost half of the data was collected toward the end of the academic quarter at the sponsoring university. This increased the likelihood of having participants who were involved only to get the credit as a compensation, and who may not have paid careful attention to the details of the task.

Second, the correlation between self-common humanity in the explicit IRAP-analogue and the common humanity sub-scales of SCS (Neff, 2003b), although significant, was modest ($r = .38$) given that they were ostensibly measuring the same construct. Since the IRAP requires stimuli to be very brief and direct, it was crucial to create label and target stimuli that would represent the concepts measured with the least number of words. This might have decreased the likelihood that the current IRAP could capture exactly what is measured by the SCS. It is possible that the current IRAP is not the best representation of Neff's subscales of common humanity and isolation. In any event, overall, this

study provides the first IRAP designed to measure the factor of common humanity, and hence, it should be considered an exploratory step toward developing a valid and reliable measure of self-compassion.

Acnowledgement

References

Barnes-Holmes, D., Barnes-Holmes, Y., Luciano, C., & McEnteggart, C. (2017). From the IRAP and REC model to a multi-dimensional multi-level framework for analyzing the dynamics of arbitrarily applicable relational responding. *Journal of contextual behavioral science, 6*(4), 434–445.

Barnes-Holmes, D., Barnes-Holmes, Y., Power, P., Hayden, E., Milne, R., & Stewart, I. (2006). Do you really know what you believe? Developing the implicit relational assessment procedure (IRAP) as a direct measure of implicit beliefs. *Irish Psychologist, 32*(7), 169–177.

Barnes-Holmes, D., Barnes-Holmes, Y., Stewart, I., & Boles, S. (2010). A sketch of the implicit relational assessment procedure (IRAP) and the relational elaboration and coherence (REC) model. *Psychological Record, 60*(3), 527–542.

Bast, D., & Barnes-Holmes, F. (2014). A first test of the implicit relational assessment procedure as a measure of forgiveness of self and others. *Psychological Record, 64*(2), 253–260.

Bast, D., Linares, F., Gomes, I., Kovac, M., & Barnes-Holmes, P. (2016). The implicit relational assessment procedure (IRAP) as a measure of self-forgiveness: The impact of a training history in clinical behavior analysis. *Psychological Record, 66*(1), https://doi.org/10.1007/s40732-016-0162-7 177-19.

Brown, K. W., & Ryan, R. M. (2003). The benefits of being present: Mindfulness and its role in psychological well-being. *Journal of Personality and Social Psychology, 84*(4), 822–848. https://doi.org/10.1037/0022-3514.84.4.822.

Finn, M., Barnes-Holmes, D., & McEnteggart, C. (2018). Exploring the single-trial-type-dominance-effect in the IRAP: Developing a differential arbitrarily applicable relational responding effects (DAARRE) model. *Psychological Record, 68*(1), 11–25. https://doi.org/10.1007/s40732-017-0262-z.

Forsyth, D. R. (2008). Self-serving bias. *International encyclopedia of the social sciences: Vol. 7*. Detroit, USA: Macmillan Reference.

Gilbert, P. (2010). *The compassionate mind: A new approach to life's challenges*. Oakland, CA: New Harbinger Publications.

Gilbert, P., Mcewan, K., Matos, M., & Rivis, A. (2011). Fears of compassion: Development of three self-report measures. *Psychology and Psychotherapy: Theory, Research and Practice, 84*(3), 239–255. https://doi.org/10.1348/147608310X526511.

Greenwald, A., & Banaji, M. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotype. *Psychological Review, 1*, 4–27.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology, 74*, 1464–1480.

Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the implicit association test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology, 97*(1), 17–41.

Hughes, S., Barnes-Holmes, D., & Vahey, N. (2012). Holding on to our functional roots when exploring new intellectual islands: A voyage through implicit cognition research. *Journal of Contextual Behavioral Science, 1*(1–2), 17–38. https://doi.org/10.1016/j.jcbs.2012.09.003.

Hussey, I., Thompson, M., Mcenteggart, C., Barnes-Holmes, D., & Barnes-Holmes, Y. (2015). Interpreting and inverting with less cursing: A guide to interpreting IRAP data. *Journal of Contextual Behavioral Science, 4*(3), 157–162.

Karpinski, A. (2004). Measuring self-esteem using the implicit association test: The role of the other. *Personality and Social Psychology Bulletin, 30*(1), 22–34. https://doi.org/10.1177/0146167203258835.

Kavanagh, D., Roelandt, A., Van Raemdonck, L., Barnes-Holmes, Y., Barnes-Holmes, D., & McEnteggart, C. (2019). *The on-going search for perspective-taking IRAPs: Exploring the potential of the natural language-IRAP*. The Psychological Record1–24.

Leary, R. L. (2004). *The curse of the self: Self-awareness, egotism, and the quality of human life*. Oxford ; New York: Oxford University Press.

Lee, R., Robbins, S., & Hill, C. (1995a). Measuring belongingness: The social connectedness and the social assurance scales. *Journal of Counseling Psychology, 42*(2), 232–241.

Lee, R., Robbins, S., & Hill, C. (1995b). The relationship between social connectedness and anxiety, self-esteem, and social identity. *Journal of Counseling Psychology, 45*(3), 338–345.

Muris, P., & Petrocchi, N. (2017). Protection or vulnerability? A meta-analysis of the relations between the positive and negative components of self-compassion and psychopathology. *Clinical Psychology & Psychotherapy, 24*(2), 373–383. https://doi.org/10.1002/cpp.2005.

Neff, K. a (2003a). Self-compassion: An alternative conceptualization of a healthy attitude toward oneself. *Self and Identity, 2*(2), 85.

Neff, K.b. (2003b). The development and validation of a scale to measure self-compassion. *Self and Identity, 2*(3), 223.

Neff, K. D. (2008). Self-compassion: Moving beyond the pitfalls of a separate self-concept. In J. Bauer, & H. A. Wayment (Eds.). *Transcending self-interest: Psychological*

Explorations of the quiet ego (pp. 95–105). Washington DC: APA Books.

Neff, K. D. (2009). The role of self-compassion in development: A healthier way to relate to oneself. *Human Development, 52*, 211–214.

Neff, K. D. (2016). Does self-compassion entail reduced self-judgment, isolation, and over-identification? A response to muris, otgaar, and Petrocchi. *Mindfulness, 7*(3), https://doi.org/10.1007/s12671-016- 0531-y 791-79.

Neff, K. D., Pisitsungkagarn, K., & Hsieh, Y. (2008). Self-compassion and self-construal in the United States, Thailand, and Taiwan. *Journal of Cross-Cultural Psychology, 39*(3), 267–285. https://doi.org/10.1177/0022022108314544.

Neff, K. D., Rude, S. S., & Kirkpatrick, K. L. (2007). An examination of self- compassion in relation to positive psychological functioning and personality traits. *Journal of Research in Personality, 41*(4), 908–916. https://doi.org/10.1016/j.jrp.2006.08.002.

Neff, K. D., Tóth-Király, I., Yarnell, L. M., Arimitsu, K., Castilho, P., Ghorbani, N., & Mantzios, M. (2018). Examining the factor structure of the self-compassion scale in 20 diverse samples: Support for use of a total score and six subscale scores. *Psychological Assessment, 31*(1), 1–19. https://doi.org/10.1037/pas0000629.

Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.). *Measures of personality and social psychological attitudes* (pp. 17–59). San Diego, CA: Academic Press.

Roddy, S., Stewart, I., & Barnes-Holmes, D. (2009). Anti-fat, pro-slim, or both? Using two reaction time based measures to assess implicit attitudes to the slim and overweight. *Journal of Health Psychology, 15*, 416–425.

Schmukle, S. C., & Egloff, B. (2005). A latent state-trait analysis of implicit and explicit personality measures. *European Journal of Psychological Assessment, 21*(2), 100–107.

Schwarz, N., & Clore, G. L. (1983). Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. *Journal of Personality and Social Psychology, 45*, 513–523.

Tirch, D. D. (2010). Mindfulness as a context for the cultivation of compassion. *International Journal of Cognitive Therapy, 3*(2), 113–123.

Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin, 133*(5), 859–883. https://doi.org/10.1037/0033-2909.133.5.859.

Weber, S. J., & Cook, T. D. (1972). Subject effects in laboratory research: An examination of subject roles, demand characteristics, and valid inference. *Psychological Bulletin, 77*(4), 273–295. https://doi.org/10.1037/h0032351.