The Implicit Relational Assessment Procedure (IRAP) as a Measure of Self-Forgiveness: The Impact of a Training History in Clinical Behavior Analysis

Diana Ferroni Bast[1] · Ila Marques Porto Linares[2] · Caina Gomes[2] · Roberta Kovac[2] ·
Dermot Barnes-Holmes[3]

Abstract The current study aimed to test the effect of behavior therapy training on the assessment of self-forgiveness, focusing on the feelings or outcomes that may be associated with failing and succeeding in everyday life, using 2 Implicit Relational Assessment Procedures (IRAPs) that had been developed across a series of previously published studies. Additionally, the research explored the extent to which responding on the IRAP correlated with standardized measures of psychopathology, including depression, anxiety, stress, and a scale that was based directly on the IRAP. Forty undergraduate and postgraduate students completed the study (20 individuals who were teaching on, attending, or who had attended a course in clinical behavior analysis and 20 students from different fields). The two groups (behavior therapists and nontherapists) completed the 2 IRAPs and the explicit measures. Overall, only 1 of the 2 IRAPs, the one that targeted feelings rather than outcomes, produced clear and significant differences between the behavior therapist and nontherapist groups. This result indicated that the diverging performances were specific to the stimuli that were presented in the IRAP rather than reflecting a generic between-group difference produced by the measure itself. Furthermore, both IRAPs predicted levels of self-reported psychopathology and self-compassion. A number of potential reasons why this pattern of results emerged using the 2 IRAPs and explicit measures with these 2 groups of participants are considered.

Self-forgiveness has many definitions and interpretations, but one common feature in most, if not all, definitions involves the ability to acknowledge (rather than avoid) negative feelings and possible consequences that is deemed to involve some sort of failure, and trying to repair the wrongdoing with corrective behaviors. Examples of research on self-forgiveness in the functional-contextual literature are scarce, but very recent studies (e.g., Bast & Barnes-Holmes 2015a, b; Bast, Barnes-Holmes, & Barnes-Holmes, 2015) have investigated attitudes to self-forgiveness in terms of feelings and outcomes, related to minor failures and successes. Acknowledging feelings and outcomes in relation to self-forgiveness constitutes only one possible aspect or feature of this psychological domain.[1] Nevertheless, the authors chose to focus on this area because it was considered the first step in the process of self-forgiveness. For example, if one avoids contacting the feelings produced by some negative situation caused by oneself, there would be no necessity for self-forgiveness. Furthermore, according to Hayes, Strosahl, and Wilson (2011), accepting negative feelings or outcomes associated with an experience may contribute to psychological openness, learning, and compassion toward oneself and others. In contrast, the costs and dangers of avoidance of

✉ Diana Ferroni Bast
   dinbast@gmail.com

1 Department of Psychology, Maynooth University, John Hume
   Building, North Campus, Maynooth University, Co., Kildare, Ireland

2 Paradigma Centro de Ciências e Tecnologia do Comportamento, São
   Paulo, Brazil

3 Ghent University, Ghent, Belgium

---

[1] The term *self-forgiveness* is maintained across studies simply to orient us toward a particular psychological domain, in much the same way that the term *language and cognition* serves as a general orienting device for researchers working on Relational Frame Theory (Hayes, Barnes-Holmes, & Roche, 2001, p. 45).

negative experiences have been recognized in most systems of therapy. For example, a common assumption in Acceptance and Commitment Therapy (ACT) is that when clients attempt to avoid specific psychological events, those experiences often return and may be deemed even more distressing and dominant than before (Hayes et al., 2011).

The vast majority of studies that have aimed to explore self-forgiveness, in therapeutic or other contexts, have typically employed self-report measures or scales. For example, a questionnaire might ask participants to indicate on a scale of 1 to 5 how easily they find it to forgive themselves if they failed in some way or made some sort of mistake. In adopting such a research strategy, however, it is possible that an individual's responses might be influenced by extraneous variables. For instance, individuals might respond in a manner that aims to reduce the likelihood that they will be perceived in a negative light, by indicating a relatively low level of self-forgiveness in order not to be seen as being too soft on themselves. It is also well established in the psychological literature that individuals do not necessarily have access to highly reliable sources of information about their own psychological states, and, thus, self-reports about levels of self-forgiveness might not always reflect exactly how individuals react in their day-to-day lives in this regard.

In an effort to provide a methodology that might circumvent such problems, and could thus be used to supplement self-report measures, a number of recent studies have focused on developing the Implicit Relational Assessment Procedure (IRAP) as a measure of self-forgiveness (Bast & Barnes-Holmes, 2014, 2015a, b). Unlike self-reports, the IRAP, similar to most implicit measures, requires that participants respond quickly and accurately on a task that is designed to capture the strength of responding in the domain of interest. The metric that is often derived from the measure involves calculating the difference in speed of responding with which participants complete the task. In the context of an IRAP designed to examine self-forgiveness, for example, participants may be asked to confirm that they find it easy to forgive themselves on some blocks of trials and to deny that they find it easy to do so on other blocks of trials. The relative difference in the speed with which these two blocks are completed may be used to provide a measure of implicit self-forgiveness. In effect, more rapid responding during blocks confirming rather than denying self-forgiveness may be interpreted as a bias toward implicit self-forgiveness for that participant.

In the study reported by Bast and Barnes-Holmes (2015a), participants were presented with two different IRAPs, one targeting feelings and another targeting outcomes related to "minor" failures and successes in the day-to-day lives of the participants. For some blocks of trials on the Feelings IRAP, participants were required to respond in a pattern that was consistent with the "common sense" idea that failing produces negative feelings whereas succeeding produces positive

feelings (e.g., responding "True" to statements such as "When I fail I feel useless" and "When I succeed I feel good," and responding "False" to "When I fail I feel calm" and "When I succeed I feel frustrated"). In other blocks of trials, participants were required to respond in a pattern that was inconsistent with "common sense" (e.g., responding "False" for "When I fail I feel useless" and responding "True" for "When I succeed I feel frustrated").

The Outcomes IRAP was similar to the Feelings IRAP except that participants were asked to respond to statements that related to the outcomes of failing versus succeeding. Once again, responding in a manner that was in accordance with common sense was required on some blocks of trials (e.g., responding "True" to "Failing undermines my motivation" and "Succeeding keeps me motivated" and responding "False" to "Failing has positive consequences" and "Succeeding makes me look stupid"). Responding in the opposite direction was required on other blocks of trials (e.g., responding "False" to "Failing undermines my motivation").

The results of the study indicated that both the Feelings and Outcomes IRAPs produced response biases that were both consistent and inconsistent with common sense. For example, participants did show a bias toward confirming that succeeding produced positive feelings and outcomes but also indicated that failing produced *positive* feelings and outcomes. In contrast, the explicit measures that were employed in the study produced evidence of common-sense response biases: questions concerning failing produced negative biases and questions concerning succeeding produced positive biases. The authors offered a range of possible explanations for why the two IRAPs produced some evidence of counterintuitive response biases, and in so doing a possible weakness in the study was highlighted. Specifically, the authors suggested that generic descriptors such as "failing" and "succeeding" might not have been evocative or salient enough to elicit relatively strong emotional reactions from participants as they completed the IRAPs. Thus general statements about *failures* may not have encouraged participants to think genuinely about their own previous failures, and this allowed for the emergence of counterintuitive effects on the IRAPs.

In an effort to increase the salience of failing and succeeding before exposure to the IRAPs, a similar but more recent study was conducted (Bast & Barnes-Holmes, 2015b). In this study, participants were assigned to two different groups; one of them was positively "primed" and the other one was negatively "primed" before exposure to the IRAPs. The priming involved writing down three situations that participants considered examples in their own lives of failure (if they were assigned to the negative priming condition) or success (if they were assigned to the positive priming condition). In effect, the study sought to determine if asking one group of participants to reflect upon previous failures and another group to reflect upon previous successes would have a

differential impact on two "self-forgiveness" IRAPs—one targeting feelings and one targeting outcomes.

The findings showed that the priming conditions affected the two IRAPs differentially; specifically, the priming condition appeared to affect performance on the Outcomes IRAP but not performance on the Feelings IRAP. The results also indicated that performance on the Feelings IRAP predicted the level of self-reported psychopathology (and perhaps even self-compassion), but only for participants in the positive priming condition. Specifically, a lower level of self-compassion predicted a bias toward confirming that failing produces positive feelings, but only after completing a positive priming exercise. Although the effect was marginal and was obtained for only one correlation out of eight for the Feelings IRAP, it was noteworthy because it seems somewhat counterintuitive. That is, one might expect that lower levels of self-compassion would predict that failing should produce a bias toward denying not affirming positive feelings. However, perhaps individuals low in self-compassion tended to confirm positive feelings in the context of failure because they are more avoidant of negative feelings. Of course, this was a highly speculative post hoc explanation, based on only one marginally significant correlation, but it could be seen as consistent with other findings obtained for the positive priming group, which showed a correlation between the Depression Anxiety Stress Scales (DASS) scores and the Failure-Negative trial type of the Feelings IRAP. In the latter case, higher levels of self-reported psychopathology appeared to predict reduced levels of negative bias. In other words, when participants with higher levels of stress, anxiety, and overall psychopathology had just been asked to think about previous successes in their lives, they appeared less willing to confirm that failing leads to negative feelings. Once again, this could be seen as evidence for a type of experiential avoidance. In other words, when participants are primed to embrace positive feelings (i.e., thinking about success), those who are higher in psychopathology may be more inclined to deny that failing leads to negative feelings because they tend to be more avoidant of such feelings (Costa & Pinto-Gouveia, 2013; Hayes, Luoma, Bond, Masuda, & Lillis, 2006; Kashdan, Barrios, Forsyth, & Steger 2006; Marx & Sloan, 2005; Hayes, Wilson, Gifford, Follete, & Strosahl, 1996). In summary, these findings provided potentially important new insights into the behaviors associated with self-forgiveness. Nevertheless, the findings remain preliminary, the explanations given are somewhat speculative, and, thus, continued research in this domain is required.

At this point, it appears that the self-forgiveness IRAPs that have been developed across a series of studies (Bast & Barnes-Holmes, 2014, 2015a, b; Bast, Barnes-Holmes, & Barnes-Holmes, 2015) may be sensitive to a specific behavioral history that was provided within the experimental context (i.e., positive vs. negative priming). Although

demonstrating such an effect is important in terms of establishing the validity of the IRAP as a task that is sensitive to the verbal relations associated with self-forgiveness, it is also the case that experimental priming procedures may be seen as relatively artificial or contrived. Consequently, it seemed important at this point in the research program to determine if one or both of the self-forgiveness IRAPs would prove sensitive to a potentially important feature of the participants' preexperimental history. We opted to pursue this line of inquiry by employing two groups: one who had undergone training in behavior therapy and one who had not. We reasoned that training in most forms of psychological therapy may increase levels of compassion and forgiveness of self and others given that therapy is very much focused on understanding and treating human suffering. Indeed, most forms of training in psychological therapy involve some element of increasing the therapist's ability to understand and reflect upon the perspective of other human beings, particularly clients, during the process of therapy itself. Within this process, the training may require, either implicitly or explicitly, for the therapist to reflect upon their own strengths and weaknesses, and past successes and failures, in order to empathize and better understand a client's perspective on the problems they are presenting in therapy (e.g., Tsai et al., 2009; Kohlenberg & Tsai, 1991). In this context, one might predict that individuals who have completed or are currently completing training in a form of psychological therapy, relative to non-therapists, would respond differently on IRAPs that target self-forgiveness.

In summary, given the fact that the IRAP used in this context was similar to that employed by Bast and Barnes-Holmes (2015a), it might be predicted that somewhat similar patterns would be found in the case of the non–behavior-therapy group as were found for the sample used in that study; for example, we might expect a positive bias for the success trial type and the counterintuitive bias for Failure-Positive trial type. On the other hand, it might be predicted that the behavior-therapy group would show a different pattern based on their training; for example, we might expect to find a neutral bias in comparison to the other group for the Failure-Negative, Positive-Feelings, and Positive-Outcomes trial types (e.g., not denying that failure produces negative feelings and outcomes, but instead considering failures as a learning opportunity that could facilitate approaching valued goals).

Method

Participants

Fifty-six students and teaching staff were recruited via class and department announcements from Nucleo Paradigma de Sao Paulo and through snowball sampling. Out of 56

individuals, 16 were eliminated; eight were excluded because they did not achieve the IRAP performance criteria detailed in the procedure section (four of them in achieving the criteria for the Feelings IRAP and four for the Outcomes IRAP), and another eight were eliminated due to a procedural error. In total, 40 participants remained. Of these, four individuals were currently pursuing a course in clinical behavior analysis, 12 had completed the course within the previous 2 years, and four were lecturers on the course. The remaining 20 individuals were students from different fields (e.g., law, engineering) and they functioned as a control group. Hereafter, the first group will be referred to as the behavior therapist (BT) group and the second as the non–behavior-therapist (NBT) group. Participants were between 18 and 32 years old ($M = 25$); 29 were women and 11were men, and they all completed the current study on a voluntary basis. No payment or course credits were exchanged for participation, but volunteers were offered a chocolate bar before leaving the laboratory.

The course was designed to provide postgraduate training and education in behavior analysis in all its aspects: philosophy, theory, and, specifically, the techniques employed in clinical practice. The course consists of 13 modules that cover philosophical knowledge, conceptual issues, and the methodological and technological features of behavior analysis as applied in clinical contexts. The course aims to develop the necessary skills for consistent and competent clinical practice (350 hours) and thus involves supervised clinical work (180 hours). The general strategy of the supervised practice draws on a broad functional-analytic approach, which focuses on the therapeutic setting, the analysis of verbal behavior, the therapist–client relationship, and the analysis of private events, without, however, losing an emphasis on external or environmental variables as causes of psychological events. In general, the therapist is trained to identify (a) contextual variables that create the aversive conditions associated with the clients' complaints; (b) the widespread (generalized) behavioral patterns associated with these aversive conditions/complaints; (c) the historical contexts that may have served to establish or facilitate the development of these patterns; (d) the possible effects of the client's behaviors in terms of maintaining the "problem" being reported; and (e) potential motivational variables for change.

Setting, Apparatus, and Materials

Participants completed the study in a quiet room, free from distraction. The implicit measure was presented to each participant on a standard personal computer using the IRAP-2010 program, written by the final author (an updated version of the program is available for download from www.IRAPresearch.org). Explicit measures were provided in hard-copy format.

Implicit Measure The IRAP allows for the on-screen presentation of standardized instructions, which participants can read in their own time while pressing the space bar to move between screens. The IRAP program also presents stimuli, feedback, and records and calculates measures of response accuracy and latency. The IRAP requires participants to respond quickly and accurately in ways that are deemed consistent or inconsistent with their preexperimentally established verbal relations. The basic hypothesis is that average response latencies should be shorter across blocks of trials that require responses consistent with such verbal relations than across blocks of trials that require responses that are deemed inconsistent with those relations.

Each participant was required to complete two IRAPs, one designed to target feelings and one designed to target expected outcomes arising from failing and succeeding. The stimuli inserted into the Feelings IRAP consisted of combinations of statements pertaining to feelings arising from failing versus succeeding. The two label stimuli consisted of the statements "When I fail" and "When I succeed." The target stimuli were 12 short statements, six of which indicated negative feelings (i.e., "I feel Bad," "I feel Guilty," "I feel Stupid," "I feel Useless," "I feel Frustrated," and "I feel Angry"), and a further six that indicated positive feelings (i.e., "I feel Good," "I feel Strong," "I feel Energetic," "I feel Positive," "I feel Calm," and "I feel Peaceful"). Thus, each trial of the IRAP presented a label and a target stimulus that indicated one of four possible label-target combinations or trial types, which may be described as (i) Failure-Negative feelings; (ii) Failure-Positive feelings; (iii) Success-Negative feelings; (iv) Success-Positive feelings. Participants responded to these label-target combinations by choosing one of two response options, True or False, which appeared in the bottom right-hand and left-hand corners of the computer screen. The two response options appeared under the prompts "select 'd' for" and "select 'k' for." The label and target stimuli and both response options appeared on the screen simultaneously at the onset of each trial. The label and target stimuli varied quasirandomly with each trial, as did the left and right positions of the response options. Participants were required to respond "True" to specific trial types on some blocks of trials and to respond "False" on other blocks of trials, and consistent with previous studies using the IRAP, the difference in average response latencies between True versus False responses was the primary datum employed for analysis. A schematic representation of the Feelings IRAP is presented in Fig. 1.

The Outcomes IRAP was similar to the Feelings IRAP except the label stimuli consisted of the single words *Failing* and *Succeeding,* and the target stimuli focused on outcomes arising from failing and succeeding. The six negative targets were "Wastes my time," "Undermines my motivation," "Has negative consequences," "Makes me look bad," "Makes me less productive," and "Makes me look stupid." The six
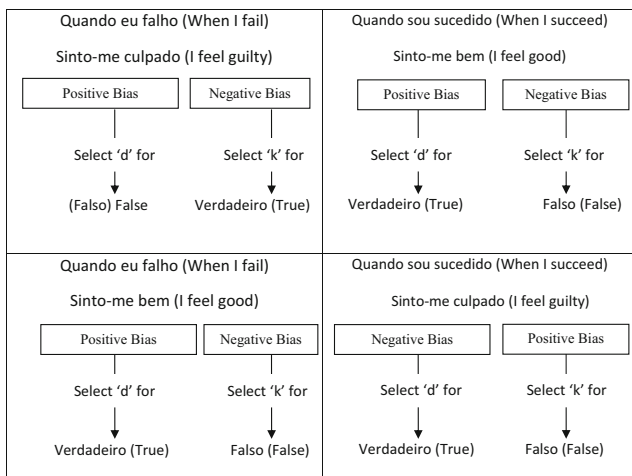
Fig. 1 A schematic representation of the four trial types from the Feelings IRAP

positive targets were "Saves me time," "Keeps me motivated," "Has positive consequences," "Makes me look good," "Makes me more productive," and "Makes me look intelligent." The four trial types for the Outcomes IRAP may thus be described as (i) Failure-Negative outcomes; (ii) Failure-Positive outcomes; (iii) Success-Negative outcomes; (iv) Success-Positive outcomes. A schematic representation of the Outcomes IRAP is presented in Figure 2.

Explicit Measures There were four separate explicit measures. Two measures were derived from the stimuli employed with the IRAPs and the two other measures were standardized psychometric instruments targeting self-compassion (the Self-Compassion Scale [SCS]; Neff, 2003) and depression, anxiety, and stress levels (the Depression Anxiety Stress Scales [DASS]; Lovibond & Lovibond, 1995).

The Self-Compassion Scale (SCS) The SCS (Neff, 2003) has been translated into Portuguese by Pinto-Gouveia and Castilho (2006). The 26-item questionnaire includes the
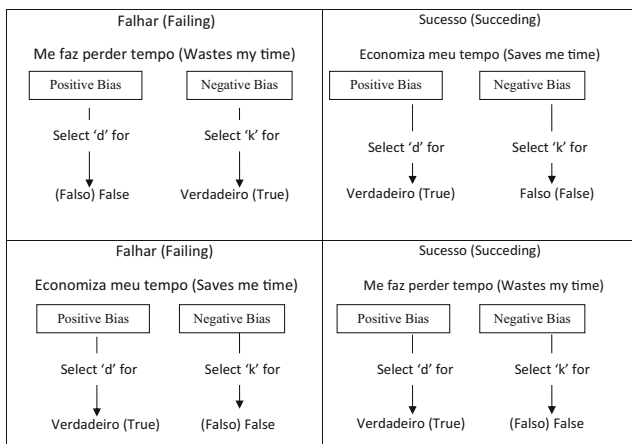

Fig. 2 A schematic representation of the four trial types from the Outcomes IRAP

five-item Self-Kindness subscale (e.g., "I am tolerant of my own flaws and inadequacies"), the five-item Self-Judgment subscale (e.g., "When times are really difficult, I tend to be tough on myself"), the four-item Common Humanity subscale (e.g., "I try to see my failings as part of the human condition"), the four-item Isolation subscale (e.g., "When I think about my inadequacies, it tends to make me feel more separate and cut off from the rest of the world"), the four-item Mindfulness subscale (e.g., "When something painful happens, I try to take a balanced view of the situation"), and the four-item Over-Identification subscale (e.g., "When I'm feeling down, I tend to obsess and fixate on everything that's wrong"). Responses are given on a 5-point scale, from *Almost Never* to *Almost Always*. Mean scores on the six subscales are then averaged (after reverse-coding negative items) to create an overall self-compassion score. 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 facets (Neff, 2003). The scale also demonstrates 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, 2003; Neff, Kirkpatrick, & Rude, 2007).

The SCS was employed because it was felt that there may be some overlap between self-forgiveness and self-compassion. The Portuguese version of the SCS has internal consistency, temporal stability, and convergent and divergent validity. The internal consistency reliability obtained was $\alpha = .89$ for the total score.

The Depression Anxiety Stress Scales (DASS) The DASS (Lovibond & Lovibond, 1995) was translated into Portuguese by Pais-Ribeiro, Honrado, and Leal (2004). The DASS was employed because it had been used successfully in previous studies that employed the IRAP as a measure of psychological constructs relevant to human mental health and well-being, such as obsessive-compulsive tendencies (e.g., Nicholson & Barnes-Holmes, 2012), depression (Hussey & Barnes-Holmes, 2012), and professional burnout (Kelly & Barnes-Holmes, 2013).

The DASS is a set of three self-report scales designed to measure the negative emotional states of depression, anxiety, and stress. The DASS was constructed not merely as another set of scales to measure conventionally defined emotional states but to further the process of defining, understanding, and measuring the ubiquitous and clinically significant emotional states usually described as depression, anxiety, and stress. Each of the three DASS scales contains 14 items, divided into subscales of two to five items with similar content. The Depression scale assesses dysphoria, hopelessness, devaluation of life, self-deprecation, lack of interest/involvement, anhedonia, and inertia. The Anxiety scale assesses autonomic

arousal, skeletal muscle effects, situational anxiety, and subjective experience of anxious affect. The Stress scale is sensitive to levels of chronic nonspecific arousal. It assesses difficulty relaxing, nervous arousal, and being easily upset/agitated, irritable/overreactive and impatient. Respondents are asked to use 4-point severity/frequency scales to rate the extent to which they have experienced each state *over the past week*. Scores for depression, anxiety and stress are calculated by summing the scores for the relevant items.

Factor-analytic studies with nonclinical (Lovibond & Lovibond, 1995) and clinical samples (Antony, Beiling, Cox, Enns, & Swinson, 1998; Brown, Chorpita, Korotitsch, & Barlow, 1997; Clara, Cox, & Enns, 2001; Crawford & Henry, 2003) have confirmed that the DASS items can be reliably grouped into three scales: (a) Depression (DASS-D), (b) Anxiety (DASS-A), and (c) Stress (DASS-S). Validity of the Portuguese adaptation presents good internal consistency, similar to the original version, as well as good convergent and discriminant validity. Cronbach's alpha ranges from .83 to .93 (Depression = .93 ; Anxiety = .83; Stress = .88; Pais-Ribeiro, Honrado, & Leal, 2004).

The Scale Based on the Feelings and Outcomes IRAPs The remaining two explicit measures that were derived directly from the stimuli employed with the IRAPs were designed specifically for the current study, and they were used to record the feelings and outcomes that participants expected when they experienced either success or failure in their lives. The first 12 items asked participants to indicate how they felt when they failed in some way, with the first six items targeting negative feelings (e.g., "When I fail in some way I feel bad") and the next six targeting positive feelings (e.g., "When I fail in some way I feel good"). The next 12 items asked participants to indicate how they felt when they succeeded in some way, with the first six items again targeting negative feelings (e.g., "When I succeed in some way I feel bad") and the next six targeting positive feelings (e.g., "When I succeed in some way I feel good"). Participants were asked to give a score from 1, which was marked as *Completely False,* to 7, which was marked as *Completely True.* The number 4 was marked as *Neither True nor False.* The numbers 2, 3, 5, and 6 thus gave participants the opportunity to indicate that the relevant statement was somewhat false or true along a graded continuum. In effect, participants were asked to indicate to what extent they experienced positive or negative feelings following failures and successes using the same target words as were presented in the Feelings IRAP.

The next 24 items were similar to the previous 24, except they focused on the *outcomes* of failing and succeeding, using the target stimuli employed with the Outcomes IRAP (e.g., "Failing wastes my time," "Succeeding makes me more productive"). Thus, the first 12 items targeted the outcomes of failing, with the first six items focusing on negative outcomes and the next six focusing on positive outcomes. The remaining 12 items focused on the outcomes related to succeeding, with six items each focusing on negative and positive outcomes, respectively. Participants were again invited to score each of the 24 items using the same 7-point scale that was used for the 24 "feelings" items.

Procedure

After completing consent forms, participants were asked to complete the IRAPs followed by the explicit measures. The order in which the two IRAPs were presented was counterbalanced across participants. Each session took approximately one hour: 45 minutes to complete both IRAPs and 15 minutes to complete the explicit measures.

Explicit Measure For the explicit measures, the scales were simply presented to participants and they were asked to complete them in their own time. Participants were instructed to read each item carefully and to ask for clarification from the researcher if anything seemed unclear.

Implicit Measure Participants were guided to a small room equipped with a computer. The room was free from excessive noise and other distractions (e.g., participants were asked to switch off their mobile phones while they completed the IRAP). Instructions were first given to participants by the researcher, who provided a description of the trials, while demonstrating how to respond on the computer keyboard to the stimuli appearing on screen. Participants were asked to respond quickly and accurately to all tasks irrespective of whether or not they considered their responses to be consistent or inconsistent with their established beliefs about failing and succeeding. After the instructions, the researcher was available to answer any remaining questions from participants. Nevertheless, at no point did the researcher indicate that differential response accuracies or latencies were expected across different blocks of trials of the IRAP—participants were simply asked to respond as quickly and accurately as possible throughout the task.

Each trial of the IRAP presented a label stimulus, a target stimulus, and two response options (described previously). Choosing the response option deemed to be correct for that particular block of trials removed all stimuli from the screen for a 400 ms interval before the next trial was presented. Choosing the response option deemed incorrect for that particular block of trials produced a red X midscreen directly below the target stimulus. The IRAP program only proceeded to the next trial when the correct response option for that particular block of trials was selected. Note, the left-right locations of the two response options varied randomly across trials.

Each block on the IRAP presented 24 trials. The trials were presented in a quasirandom order with the following constraints: each of the 12 target stimuli appeared twice, once with each of the two types of label stimuli. The IRAP trials are typically conceptualized as involving four different trial types (see Figs. 1 and 2). The randomization algorithm ensured that within each block of 24 trials, the four IRAP trial types were each presented six times.

In Block 1, and all subsequent odd numbered blocks of the Feelings IRAP, participants were required to respond in a pattern that was consistent with the "common-sense" idea that failing produces negative feelings whereas succeeding produces positive feelings (e.g., responding "True" to "When I fail I feel bad" and "When I succeed I feel good"). In Block 2, and all subsequent even numbered blocks of the Feelings IRAP, participants were required to respond in a pattern that was inconsistent with the common-sense position (e.g., responding "False" to "When I fail I feel bad" and "When I succeed I feel good"). Similar patterns were required for the Outcomes IRAP (e.g., responding "True" to "Failing undermines my motivation" and "Succeeding keeps me motivated" across odd numbered blocks but responding "False" to these questions across all even numbered blocks).

Each IRAP commenced with a minimum of two practice blocks. Participants were required to achieve ≥80 % correct and a median response latency ≤2,000 ms for each of the two practice blocks. If participants failed to achieve these performance criteria, a message appeared on screen informing them that the criteria had not been met and they were invited to complete the two practice blocks again. Participants were permitted four exposures to the pairs of practice blocks (i.e., eight blocks). If the criteria were not met after the fourth exposure, participants were invited to return later that day or on a subsequent day to try it again (no participant failed to achieve the practice criteria on the second attempt). When participants met the criteria on a pair of practice blocks, they continued immediately to a fixed set of six test blocks; these were similar to the practice blocks except that no performance criteria were applied in order to proceed across successive pairs of blocks. However, accuracy and average latency were presented at the end of each block in order to encourage participants to maintain relatively accurate and rapid responding. In addition, the instruction, "This is a test—go fast. Making a few errors is okay," was presented before the beginning of each block. The IRAP program automatically recorded response accuracy (e.g., based on the first response emitted on each trial) and response latency (e.g., the time [in milliseconds] between the onset of the trial and the emission of a correct response) for each participant on every trial. Upon completion of all practice and test blocks, the following message appeared on screen: "Thank you. That is the end of this part of the experiment. Please report to the experimenter."

Results

The Feelings IRAP

Data Preparation For the purposes of statistical analysis, participants were required to maintain an accuracy level ≥75 % correct and a median latency ≤2,000 ms on two of the three successive pairs of the six test blocks. The data for four participants were excluded because they failed to meet these criteria. If a participant maintained the criteria across all six blocks, all of the data were used to calculate the D-IRAP scores (described subsequently). If a participant failed to maintain the criteria on one successive pair of the test blocks, the data for those blocks were discarded and the D-IRAP scores were calculated from the remaining two pairs of test blocks.

Consistent with the majority of previous IRAP studies, the data were transformed into D-IRAP scores. The D transformation functions 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 (Greenwald et al., 2009).

Calculating D-IRAP scores for each participant who met the criteria for all six test blocks involved the following nine steps: (i) only response-latency data from test blocks were used; (ii) latencies above 10,000 ms were eliminated from the dataset; (iii) all data for a participant were removed if he or she produced more than 10 % of test-block trials with latencies less than 300 ms; (iv) 12 standard deviations for the four trial types were computed: four for the response-latencies from Test Blocks 1 and 2, four from the latencies from Test Blocks 3 and 4, and a further four from Test Blocks 5 and 6; (v) 24 mean latencies for the four trial types in each test-block were calculated; (vi) difference scores were calculated for each of the four trial types, for each pair of test blocks, by subtracting the mean latency of the common-sense-consistent block from the corresponding mean latency of the common-sense-inconsistent block; (vii) each difference score was divided by its corresponding standard deviation from Step 4, yielding 12 D-IRAP scores, one score for each trial type for each pair of test blocks; (viii) four overall D-IRAP scores, one for each trial type, were calculated by averaging the three scores from each pair of test blocks; and (ix) the two D-IRAP scores for the trial types that involved responding to Failure-Negative Feelings and Failure-Positive Feelings were inverted (plus scores became minus scores and minus scores became plus scores).

The same nine steps were followed for participants who met the criteria for two of the three pairs of test blocks, except the algorithm was adjusted accordingly (e.g., 8, rather than 12, standard deviations for the four trial types were computed in

Step iv). Once the foregoing data transformation was complete, positive *D*-IRAP scores indicated a positive feelings bias whereas negative *D*-IRAP scores indicated a negative feelings bias.
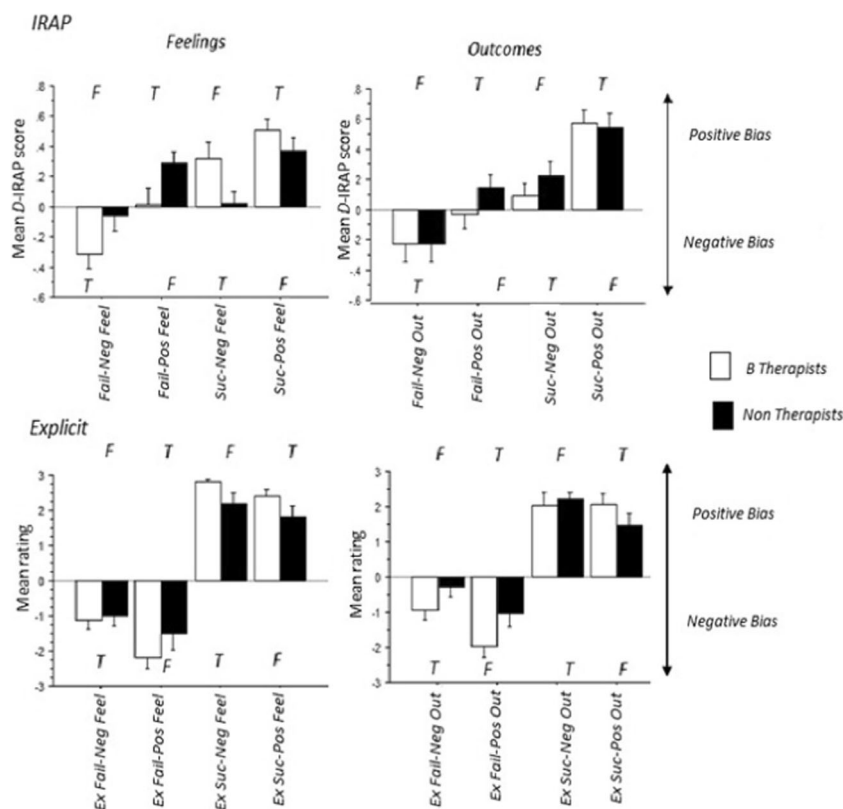
Mean Scores Analyses The four overall mean *D*-IRAP scores calculated across participants, for the BT and NBT groups, are presented in the upper left panel of Fig. 3. The relative size of the *D*-IRAP scores differed markedly across the two groups for three of the trial types. Specifically, negative, neutral and positive biases were recorded for the therapists across the Failure-Negative, Failure-Positive, and Success-Negative trial types, whereas neutral biases were recorded for non–behavior therapists across Failure- and Success-Negative Feelings and a positive bias was recorded for Failure-Positive Feelings. Both groups produced positive biases on the remaining Success-Positive trial types.

The *D*-IRAP scores for the four trial types were entered into a two-way mixed repeated measures analysis of variance (ANOVA), and this yielded a nonsignificant main effect for group ($p = .69$). The main effect for trial type was significant, $F(3, 114) = 14.625$, $p < .0001$, $\eta_p^2 = .28$, as was the interaction between trial type and group $F(3, 114) = 3.313$, $p = .02$, $\eta_p^2 = .08$. Given the significant interaction, a series of follow-up tests were conducted. Four between-group one-way ANOVAs were nonsignificant,

although three of them approached significance, Failure-Negative, $F(1, 38) = 2.820$, $p = .1$, $\eta^2 = .07$; Failure-Positive, $F(1, 38) = 2.808$, $p = .1$, $\eta^2 = .07$; and Success-Negative, $F(1, 38) = 3.590$, $p = .07$, $\eta^2 = .09$ (all remaining $ps > .27$). Two within-group ANOVAs both yielded significant effects: BT group, $F(3, 57) = 11.394$, $p < .0001$, $\eta_p^2 = .37$; NBT group, $F(3, 57) = 5.524$, $p < .002$, $\eta_p^2 = .22$. Fisher's PLSD post hoc tests for the BT group yielded significant or marginally significant effects for five of the comparisons among the four trial types ($ps < .08$), and one nonsignificant effect for the comparison between the Success-Positive and Success-Negative trial types ($p > .14$). For the NBT group, three of the post hoc tests were significant: Failure-Negative versus Failure-Positive ($p = .01$), Failure-Negative versus Success-Positive ($p < .001$), and Success-Negative versus Success-Positive ($p = .006$). The comparison between the Failure-Positive and Success-Negative trial types approached significance ($p = .09$), with the remaining two tests yielding nonsignificant results ($ps > .26$).

When each of the eight trial-type scores for the two groups were subjected to one-sample *t* tests, three of the tests yielded significance ($ps < .01$) for the BT group (the Failure-Positive Feelings trial type was nonsignificant, $p = .75$). For the NBT group, two of the tests, Failure-Positive and Success-Positive, were significant ($ps < .004$; remaining $ps > .5$).

Fig. 3 Mean *D* IRAP-p-Trial-Type Scores obtained on the IRAP and the mean rating obtained on the explicit measures for feelings and outcomes related to failing and succeeding, according to the groups. The letters T and F indicate the direction of the response biases (True and False, respectively) that were recorded by the measures

The Outcomes IRAP

Data Preparation The same general procedures for data preparation that were applied to the Feelings IRAP were applied to the data from the Outcomes IRAP. The data for four participants were excluded because they failed to meet these criteria.

Mean Scores Analyses The four overall mean D-IRAP scores calculated across participants are presented in Fig. 3 (upper right panel). The general pattern of biases did not differ substantively between the groups. Note, however, that the Failure-Positive trial type produced a very weak negative bias for the BT group but a positive, if relatively, modest bias for the NBT group. For the remaining three trial types, both groups produced negative biases for the Failure trial types and positive biases for two Success trial types.

When the D-IRAP scores were entered into a two-way mixed repeated measures ANOVA, it yielded a nonsignificant main effect for group ($p = .41$) and a nonsignificant interaction effect ($p = .42$). The main effect for trial type was significant, $F(3, 114) = 32.86$, $p < .0001$, $\eta_p^2 = .46$. Given the absence of any significant main or interaction effect for group, the data were collapsed across groups and post hoc comparisons of the four trial types yielded five significant effects (all ps < .01); only the Failure-Positive versus Success-Negative comparison was nonsignificant ($p > .11$). When each of the eight trial-type scores for the two groups were subjected to one-sample t tests, two of the four tests yielded significance (ps < .03) for the BT group on the Failure-Negative and Success-Positive trial types (all remaining ps > .16), and three of the tests were significant (ps < .04) for the NBT group; the exception was the Failure-Positive trial type ($p = .10$). Overall, therefore, the inferential statistics supported the conclusions arising from the descriptive analyses of the data provided in Fig. 3.

Explicit Measures

The Scale Based on the Feelings IRAP The ratings obtained from the explicit measure that was derived from the Feelings IRAP (hereafter referred to as the Explicit-Feelings scale) were used to calculate four separate scores, with each score mapping onto the equivalent trial type from the IRAP. For example, the six ratings obtained for questions pertaining to the subscale targeting "When I fail in some way, this produces negative feelings" were used to calculate a mean score that provided the explicit counterpart to the Failure-Negative Feelings trial type from the IRAP. For the purposes of data analysis, 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 indicated a positive bias and all negative scores indicated a negative bias. As noted previously, participants responded on a 7-point scale for each item, from 1

indicating *Completely False* to 7 indicating *Completely True*, with 4 indicating *Neither False nor True*. For the purposes of graphical representation, responses on this seven-point scale were recoded from -3 (instead of 1) to +3 (instead of 7); a score of 4 was recoded as 0.

The overall mean ratings obtained from the Explicit-Feelings scale are presented in Fig. 3 (lower left panel). The two subscales that mapped onto the two Success trial types produced positive biases for both groups and, unlike the IRAP, the effects were relatively similar across the two groups. The two subscales that mapped onto the Failure trial types yielded negative biases, The mean rating scores for each participant from the Explicit-Feelings scale were entered into a two-way mixed repeated-measures ANOVA, and it yielded a nonsignificant effect for group ($p > .5$) and interaction ($p > .17$). The main effect for trial type was significant, $F(3, 114) = 101.73$, $p < .0001$, $\eta_p^2 = .12$. When Fisher's PLSD post hoc tests were applied to the differences among the trial types (collapsed across the groups), they yielded five significant differences (ps < .001) among the four trial types, with only the Success-Negative versus Success-Positive trial-type comparison producing a nonsignificant effect ($p = .17$). When each of the eight trial-type scores for the BT and NBT groups were subjected to one-sample t tests, all eight were significant (ps < .003).

The Scale Based on the Outcomes-IRAP The data from the explicit measure that were derived from the Outcomes IRAP (hereafter referred to as the Explicit-Outcomes scale) were used to calculate eight separate scores, with each score mapping onto the equivalent trial type from the IRAP. The data were transformed in the same way as for the Explicit-Feelings scale and are presented in Fig. 3 (lower right panel).

Similar to the Outcomes IRAP, the two subscales that mapped onto the two Success trial types produced positive biases for both groups. The two subscales that mapped onto the two Failure trial types yielded negative biases for both groups. When the ratings were entered into a two-way mixed repeated-measures ANOVA, the main effect of group and interaction were both nonsignificant (ps > .16), but the main effect of trial type was significant, $F(3, 114) = 57.487$, $p < .0001$, $\eta_p^2 = .60$. When Fisher's PLSD post hoc tests were applied to the differences among the trial types (collapsed across the two groups), they yielded five significant differences (ps < .01) among the four trial types, with only the Success-Positive versus Success-Negative trial-type comparison producing a nonsignificant effect ($p = .27$). When each of the eight trial-type scores for the two groups were subjected to one-sample t tests, they each yielded significance (ps < .01), except for the Failure-Negative trial type ($p = .18$) for the NBT group. Overall, therefore, the two explicit measures that were derived directly from the IRAP did not produce any

significant effects that indicated a difference between the two groups.

Self-Compassion Scale (SCS) According to Neff (2003), average overall self-compassion scores tend to be around 3.0 on the 1 to 5 scale, so the overall score can be interpreted accordingly: 1 to 2.5 indicates low self-compassion, 2.5 to 3.5 indicates moderate self-compassion, and 3.5 to 5.0 indicates high self-compassion. In the present study, both groups produced overall mean self-compassion scores in the moderate range (BT group, $M = 3.21$, $SD = .73$; NBT group, $M = 2.76$, $SD = .59$). An independent t test indicated that the difference between the two groups was significant, $t = 2.124$, $p = .04$, suggesting that the BT group possessed higher levels of self-compassion relative to the NBT group.

Depression Anxiety Stress Scales (DASS) The interpretation of the DASS is based primarily on the use of cutoff scores. Lovibond and Lovibond (1995) presented severity ratings from *normal* to *extremely severe* on the basis of percentile scores, with 0 to 78 classified as *normal,* 78 to 87 as *mild,* 87 to 95 as *moderate,* 95 to 98 as *severe,* and 98 to 100 as *extremely severe.* In the current study, the mean overall score for the DASS was similar between the two groups: 22.1 ($SD = 17.07$) for the BT group and 30.90 ($SD = 16.62$) for NBT group, indicating that both samples fell well below the cutoff for *normal.* Independent t tests yielded no difference in the overall DASS score between the groups. At a subscale level, the groups did show differences on the Anxiety and Stress subscales ($ps < .05$), but not on the subscale for Depression ($p \geq .6$).

Implicit–Explicit Correlations

The Feelings IRAP and Explicit Measures Two correlation matrices were created, one for each group. For each matrix, the four D-IRAP scores were entered with the scores from each of the three explicit measures (the scores obtained from SCS, DASS, and the scale based on the Feelings IRAP). The main correlations are shown Table 1

The Feelings IRAP and SCS Of the 56 correlations (28 for each group), three were significant (or marginally so) for the BT group: Failure-Positive Feelings with Self-Compassion Average, Common Humanity, and Self-Kindness. In other words, an increased bias toward confirming that failing produces positive feelings predicted lower levels of overall Self-Compassion, Common Humanity, and Self-Kindness.

For the NBT group, three correlations also were significant (or marginally so)*,* Success-Posit*ive* Feelings with Isolation and Self-Kindness, and Failure-Negative Feelings with Common Humanity. In other words, a bias toward confirming that success produces positive feelings predicted lower levels

of Isolation and increased levels of Self-Kindness; a bias toward disconfirming that failure leads to negative feelings predicted increased levels of Common Humanity.

The Feelings IRAP and DASS Of the 16 correlations for the BT group (the four trial types with the four DASS scores), none were significant. Of the 16 correlations obtained for the NBT group, just one was significant: Failure-Positive Feelings and Depression; that is, increased bias in confirming that failing produces positive feelings predicted higher levels of self-reported depression.

The Feelings IRAP and Explicit-Feelings Scale In correlating the IRAP scores with the explicit scales, the analyses focused on the relationship between the IRAP trial type that mapped onto the relevant subscale of the explicit measure. None of the eight correlations across the two groups were significant.

The Outcomes IRAP and explicit measures Similar to the Feelings IRAP, the four D-IRAP scores from the Outcomes IRAP were entered into two correlation matrices (one for each group) with the three explicit measures.

The Outcomes IRAP and SCS Out of 28 correlations for the BT group, just one was significant: Failure-Negative Outcomes and Self-Judgment; that is, a bias toward disconfirming that failing leads to negative outcomes predicted lower levels of self-judgment. For the NBT group, out of 28 correlations, three were significant (or marginally so); Self-Judgment with Failure-Positive Outcomes, Success-Positive Outcomes, and Success-Negative Outcomes. In other words, increased biases toward confirming that failure and success lead to positive outcomes, and disconfirming that success leads to negative outcomes, predicted lower levels of self-judgment.

The Outcomes IRAP and DASS Of the 16 correlations for the BT group (the four trial types with the four DASS scores), five correlations were significant (or marginally so): Failure-Negative Outcomes with Stress and Success-Positive Outcomes with Depression, Anxiety, Stress, and the DASS total score. The first correlation indicates that increasing bias toward disconfirming that failure produces negative outcomes predicts increased levels of self-reported stress. The remaining correlations indicate that increasing bias toward confirming that success leads to positive outcomes predicts lower levels of self-reported psychopathology generally. Of the 16 correlations obtained for the NBT group, none were significant.

The Outcomes IRAP and Explicit-Outcomes Scale None of the eight correlations across the two groups were significant.

Table 1 Main correlations between the IRAP trial types, Self-Compassion Scale (SCS), Depression Anxiety Stress Scales (DASS)

IRAP	Group	Trial Type	SCS	DASS	r	p
Feelings	BT	**Failure-Positive**	**SCS Average**		**-.46**	**.037**
		Failure-Positive	Common Humanity		-.43	.06
		Failure-Positive	Self Kindness		-.42	.06
	NBT	Success-Positive	Isolation		.43	.06
		Success-Positive.	Self-Kindness		.39	.08
		Failure-Negative	Common Humanity		-.38	.09
		Failure-Negative		**Depression**	**.54**	**<. 01**
Outcomes	BT	**Failure-Negative**	**Self-Judgment**		**-.44**	**.04**
		Failure-Negative		**Stress**	**.48**	**.03**
		Success-Positive		**Depression**	**-.50**	**0.2**
		Success-Positive		**Anxiety**	**-.44**	**0.4**
		Success-Positive		Stress	-.41	.06
		Success-Positive		**DASS Total**	**-.54**	**.01**
	NBT	Failure-Positive	Self-Judgment		-.38	.09
		Success-Positive	Self-Judgment		-.39	0.8
		Success-Negative	**Self-Judgment**		**-.42**	**0.5**

Note. BT = behavior therapist, NBT = non–behavior therapist. The boldface items indicate significative *p* values

Discussion

The primary aim of the current study was to determine if participants who had been exposed to a training history in behavior therapy, with a focus on clinical behavior analysis, would respond differently from a control group on IRAPs that targeted expected feelings and outcomes arising from failing and succeeding. Both groups were also required to complete two explicit measures that were derived from the two IRAPs and another two explicit measures that targeted self-compassion (the SCS) and psychopathology (the DASS).

The performances of the two groups differed considerably across three of the trial types on the Feelings IRAP, but not on the Outcomes IRAP. On the Feelings IRAP, the BT group, relative to the controls, produced response biases that indicated that failing generates more negative feelings and succeeding produces more positive feelings. The effects for the two explicit-measure scales that were derived from each of the IRAPs yielded results that were relatively consistent across the two groups, and statistical analyses failed to indicate any significant between-group differences. In general, the direction of the ratings for the two groups were very polarized in terms of failure and success; that is, both groups provided negative ratings in relation to failures and positive ratings in relation to success. Overall, therefore, only the Feelings IRAP produced indications of differences between the BT and NBT groups and trial types. The fact that only one of the two IRAPs suggested a between-group difference indicates that the diverging performances were specific to the stimuli that were presented in the IRAPs (in this case expected

feelings) rather than a generic group difference produced by the measure per se.

With respect to the two explicit measures that focused on self-compassion and psychopathology, the BT group reported significantly higher levels on both instruments relative to the control group. The reason for this difference remains unclear at the current time. However, one possible explanation might be that therapy training had encouraged participants in the BT group to observe their own feelings and physical reactions, which perhaps sensitized them to the types of concepts and terms employed in the DASS and SCS and subsequently impacted upon their responding to these scales relative to the NBT group. On balance, the overall effects for the DASS, and the depression subscale, were nonsignificant; furthermore, for each of the three subscales, the means were well below the cutoff for normal levels of anxiety, stress, and depression for both groups.

Nevertheless, it is interesting that the IRAP that targeted feelings rather than outcomes yielded significant differences between the two groups who also differed in terms of self-compassion and psychopathology. Perhaps the word *feelings* in the IRAP possessed specific psychological functions for the BT participants who reported higher levels of self-compassion and psychopathology (relative to the NBT group). It is possible, for example, that undergoing training in psychotherapy increases levels of stress, anxiety, and general levels of compassion (for both self and others) relative to training in other areas, which may heighten the salience of the word *feelings* in the IRAP. Of course, future research will need to pursue this line of inquiry, but it does indicate the potential value in employing relatively specific measures of implicit response

biases in clinically relevant research (see Vahey, Nicholson, & Barnes-Holmes, 2015).

The correlational analyses between the Feelings IRAP and the explicit measures for the BT group indicated that a tendency toward confirming that failing produces positive feelings was associated with lower levels of overall Self-Compassion, Common Humanity and Self-Kindness. These correlations replicate the findings of the previous study (Bast & Barnes-Holmes, 2015b) in which Failure-Positive Feelings was negatively correlated with overall self-compassion for the positive priming group. In that study, it was suggested that having just been encouraged to feel positively toward the self during the positive priming task, individuals low in self-compassion tended to confirm positive feelings in the context of failure, possibly because they were more avoidant of negative feelings. The same pattern is seen again in this study, most obviously in the context of the BT group.

For the NBT group, the correlational analyses indicated that confirming that success produces positive feelings was associated with lower levels of Isolation and increased levels of Self-Kindness. In addition, the analyses indicated that disconfirming that failure leads to negative feelings was associated with increased levels of Common Humanity. All of these correlational effects make intuitive sense.

The only remaining significant correlation between the Feelings IRAP and the explicit measures was obtained for the NBT group, which showed that confirming that failing produces positive feelings is associated with higher levels of self-reported depression. This latter finding might be seen as counterintuitive because it indicates higher levels of depression in individuals who confirm that failing leads to positive feelings. On balance, this result might reflect a tendency toward experiential avoidance, which has been associated with a broad range of psychopathological reactions (Hayes et al., 2011). In other words, claiming that failing makes one feel positive could reflect a type of psychological inflexibility that is designed to avoid negative feelings, which, in the long run, produces the very emotion one is seeking to control (Hayes et al., 1996).

The correlations for both groups appear to provide some support for the Feelings IRAP as a measure of the broadly defined concept of self-compassion. For example, although it may appear counterintuitive to associate failing with positive feelings (or to deny an association with negative feelings), the tendency to do so was indicative of lower levels of self-compassion, particularly for the BT group and for the positive priming group in previous research (Bast & Barnes-Homes, 2015b). To put it another way, it makes sense that self-compassion can have an important role in the acceptance of the negative feelings that failures could produce.

For the BT group, the correlational analyses between the Outcomes IRAP and the SCS yielded only one significant effect, but, again, it appeared to support the validity of the IRAP in that disconfirming that failing leads to negative outcomes was associated with lower levels of Self-Judgment. Interestingly, for the NBT group, three of the correlations were significant (or marginally so), with the results indicating that lower levels of self-judgment are associated with confirming that failure and success lead to positive outcomes and disconfirming that success leads to negative outcomes. Thus, it seems that lower levels of self-judgment may reduce the negative impact of failures and increase the positive impact of success at the implicit level.

The correlational analyses between the Outcomes IRAP and the DASS scales yielded no significant relationships for the NBT group, but five of the results were significant (or marginally so) for the BT group. Specifically, the results indicated that disconfirming that failure produces negative outcomes predicted increased stress, with the remaining correlations indicating that confirming that success leads to positive outcomes predicts lower levels of psychopathology generally. The latter correlations make intuitive sense, but the first correlation seems less obvious—why would denying that failure produces negative outcomes predict stress? Perhaps this counterintuitive result provides another example of the possible role of experiential avoidance. That is, denying that failure produces negative outcomes might reflect a tendency to avoid events or experiences that are deemed unpleasant or stressful in some way. Moreover, as the literature on experiential avoidance suggests, the very act of trying to avoid stressful situations (or control negative emotional content more generally) may serve to create the stress that one is paradoxically seeking to avoid (Hayes et al., 1996). Again, of course, this interpretation is highly speculative, but future research might pursue this line of inquiry. For example, it would be interesting to ask participants to complete self-forgiveness IRAPs before and after exposure to some form of stressor to determine its potential impact on the IRAP measures and their correlations with measures of psychopathology (e.g., see Hussey & Barnes-Holmes, 2012, for an example of this research strategy in the context of assessing dysphoria before and after a mood-induction procedure).

With respect to the correlational analyses for both the Feelings and Outcomes IRAPs and the explicit rating scales that were derived from them, none of the correlations were significant. This result is consistent with our previous research (Bast & Barnes-Holmes, 2014, 2015a) and suggests once again that the IRAPs were tapping into responses toward succeeding and failing that are not captured readily with explicit self-report measures of the responses targeted by the IRAPs. On balance, the fact that the current research yielded correlations with established psychometric instruments for self-compassion and psychopathology does indicate that the IRAPs may be capturing potentially important response biases.

As noted above, differences emerged between the two groups and trial types in the Feelings but not the Outcomes IRAP (no clear between-group differences emerged on the IRAP-derived explicit measures). At the present time, it remains unclear why the Feelings IRAP appeared to separate the groups, whereas the other measures did not. On balance, it might be expected that an educational and professional history involving therapeutic theory and practice may increase the salience or importance of human feelings, relative to a history of education/training in other areas (e.g., law and engineering). Thus, the repeated appearance of the word *feelings* in an IRAP may well have served to evoke relatively strong or specific psychological functions for the BT group that were not evoked for the NBT controls. With respect to the Outcomes IRAP, however, the word *feelings* does not appear on any trial, and thus the difference in the educational histories of the two groups would be far less important and differences across the two groups less likely to be seen. While this post hoc explanation is highly speculative, it is consistent with the general notion that verbal histories are important in determining performance on the IRAP and other implicit measures (Barnes-Holmes, Barnes-Holmes, Stewart, & Boles, 2010). Given the current findings, future studies might attempt to target specific verbal histories using relevant IRAPs. For example, it would be interesting to investigate if the IRAPs designed to assess verbal relations concerning the concepts of "acceptance" versus "control" of feelings and emotions yield different results with individuals trained in different types of therapy, such as ACT versus traditional cognitive behavior therapy (CBT). Another empirical question that could be addressed in future studies is if the strength of the IRAP effects correlate with potentially important variables, such as stress and professional burnout (see Kelly & Barnes-Holmes, 2013, for a relevant example in the context of teachers working with children with learning disabilities).

References

Antony, M. M., Bieling, P. J., Cox, B. J., Enns, M. W., & Swinson, R. P. (1998). Psychometric properties of the 42-item and 21-item versions of Depression Anxiety Stress Scales in clinical groups and a community sample. *Psychological Assessment, 10*, 176–181.

Bast, D., & Barnes-Holmes, D. (2014). A first test of the Implicit Relational Assessment Procedure as a measure of forgiveness of self and others. *The Psychological Record, 64*(2), 253–260.

Bast, D., & Barnes-Holmes, D. (2015a). Developing the Implicit Relational Assessment Procedure (IRAP) as a measure of self-forgiveness related to failing and succeeding behaviors. *The Psychological Record, 65*(1), 189–201.

Bast, D., & Barnes-Holmes, D. (2015b). Priming thoughts of failing versus succeeding and performance on the Implicit Relational Assessment Procedure (IRAP) as a measure of self-forgiveness. *The Psychological Record, 65*(4), 667–678.

Bast, D., Barnes-Holmes, Y., & Barnes-Holmes, D. (2015). Developing an individualized Implicit Relational Assessment Procedure (IRAP) as a measure of self-forgiveness related to negative and positive behavior. *The Psychological Record, 65*(4), 717–730.

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. *The Psychological Record, 60*, 527–542.

Brown, T. A., Chorpita, B. F., Korotitsch, W., & Barlow, D. H. (1997). Psychometric properties of the Depression Anxiety Stress Scales (DASS) in clinical samples. *Behavior Research and Therapy, 35*, 79–89.

Clara, I. P., Cox, B. J., & Enns, M. W. (2001). Confirmatory factor analysis of the Depression Anxiety Stress Scales in depressed and anxious patients. *Journal of Psychopathology and Behavioral Assessment, 23*, 61–67.

Costa, J., & Pinto-Gouveia, J. (2013). Experiential avoidance and self-compassion in chronic pain. *Journal of Applied Social Psychology, 43*, 1578–1591.

Crawford, J. R., & Henry, J. D. (2003). The Depression Anxiety Stress Scales (DASS): Normative data and latent structure in a large non-clinical sample. *British Journal of Clinical Psychology, 42*, 111–131.

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*, 17.

Hayes, S. C., Barnes-Holmes, D., & Roche, B. (2001). *Relational frame theory: A post-Skinnerian account of human language and cognition*. New York: Plenum Press.

Hayes, S. C., Luoma, J. B., Bond, F. W., Masuda, A., & Lillis, J. (2006). Acceptance and commitment therapy: Model, processes and outcomes. *Behaviour Research and Therapy, 44*(1), 1–25.

Hayes, S. C., Strosahl, K. D., & Wilson, K. G. (2011). *Acceptance and commitment therapy: The process and practice of mindful change*. New York: Guilford Press.

Hayes, S. C., Wilson, K. G., Gifford, E. V., Follette, V. M., & Strosahl, K. (1996). Experiential avoidance and behavioral disorders: A functional dimensional approach to diagnosis and treatment. *Journal of Consulting and Clinical Psychology, 64*(6), 1152–1168.

Hussey, I., & Barnes-Holmes, D. (2012). The Implicit Relational Assessment Procedure as a measure of implicit depression and the role of psychological flexibility. *Cognitive and Behavioral Practice, 19*, 573–582.

Kashdan, T. B., Barrios, V., Forsyth, J. P., & Steger, M. F. (2006). Experiential avoidance as a generalized psychological vulnerability: Comparisons with coping and emotion regulation strategies. *Behaviour Research and Therapy, 44*, 1301–1320.

Kelly, A., & Barnes-Holmes, D. (2013). Implicit attitudes towards children with autism versus normally developing children as predictors of professional burnout and psychopathology. *Research in Developmental Disabilities, 34*, 17–28.

Kohlenberg, R. J., & Tsai, M. (1991). *Functional analytic psychotherapy.* New York: Springer Verlag US.

Lovibond, P. F., & Lovibond, S. H. (1995). The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behaviour Research and Therapy, 33*, 335–343.

Marx, B. P., & Sloan, D. M. (2005). Peritraumatic dissociation and experiential avoidance as predictors of posttraumatic stress symptomatology. *Behaviour Research and Therapy, 43*(5), 569–583.

Neff, K. D. (2003). The development and validation of a scale to measure self-compassion. *Self and Identity, 2*, 223–250.

Neff, K. D., Kirkpatrick, K. L., & Rude, S. S. (2007). Self-compassion and adaptive psychological functioning. *Journal of Research in Personality, 41*, 139–154.

Nicholson, E., & Barnes-Holmes, D. (2012). Developing an implicit measure of disgust propensity and disgust sensitivity: Examining the role of implicit disgust propensity and sensitivity in obsessive-compulsive tendencies. *Journal of Behavior Therapy and Experimental Psychiatry, 43*, 922–930.

Pais-Ribeiro, J., Honrado, A., & Leal, I. (2004). Contribuição para o estudo da adaptação portuguesa das escalas de ansiedade depressão stress de Lovibond e Lovibond (Portuguese adaptation of Anxiety Depression Stress Scale of Lovibond and Lovibond). *Psychologica, 36*, 235–246.

Pinto-Gouveia, J., & Castilho, P. (2006). *Escala de Auto-Compaixão. Tradução e adaptação para a população portuguesa (Self-Compassion Scale. Translation and adaptation for Portuguese population).* Unpublished manuscript, provided by the authors.

Tsai, M., Kohlenberg, R. J., Kanter, J. W., Kohlenberg, B., Follette, W. C., & Callaghan, G. M. (2009). *A guide to functional analytic psychotherapy.* New York: Springer.

Vahey, N. A., Nicholson, E., & Barnes-Holmes, D. (2015). A meta-analysis of criterion effects for the Implicit Relational Assessment Procedure (IRAP) in the clinical domain. *Journal of Behavior Therapy and Experimental Psychiatry, 48*, 59–65.