Maynooth University

Developing the Implicit Relational Assessment Procedure as a

Measure of Self-Forgiveness

Thesis submitted to the Department of Psychology, Faculty of Science, in

fulfilment of the requirements for the degree of Doctor of Philosophy, Maynooth

University, Maynooth.

Diana Bast B.A.

July 2015

Head of Department: Dr. Andrew Coogan

Research Supervisor: Professor Dermot Barnes-Holmes

Table of Contents

List of Figures

List of Appendices

Publications

Sections of this thesis are currently published and under submission. The chapter number, chapter name and the corresponding citations are listed below.

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

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 Behaviour. *The Psychological Record*

Bast, D., Barnes-Holmes, D. (2015). Priming Thoughts of Failing versus Succeeding and Performance on the Implicit Relational Assessment Procedure (IRAP) as a Measure of Self-Forgiveness. *The Psychological Record.*

Bast, D., Linares, IMP; Gomes, C.; Kovac, R.; Barnes-Holmes, D. (under submission). The Implicit Relational Assessment Procedure (IRAP) as a Measure of Self-Forgiveness: The Impact of a Training History in Clinical Behavior Analysis. *The Psychological Record*

Ethical Statement

All the studies reported in this thesis were approved by the Research Ethics Committee of Maynooth University and were conducted in accordance with the University's research integrity guidelines. Additionally, the studies reported in chapters 4 and 6 conformed to the Brazilian code of professional ethic in research established by the National Council of Psychology (Brazil), and ethical approval was obtained BY WHOM?????? to conduct this research in the local jurisdiction in Brazil (Nucleo Paradigma).

The author received the support of a CAPES scholarship (Coordination for the Improvement of Higher Education Personnel), awarded by the Brazilian Ministry of Education. for this doctoratal research.

Acknowledgements

I would like to dedicate this thesis and say thanks to:

- CAPES (Coordination for the Improvement of Higher Education Personnel)

- I would like to say a huge thank to my family, especially my parents Jose Carlos Bast, Ires Ferroni and my sister Elaine Bast and my brother-in-law Marcelo Salomon for all your generous support and all the incredible things that you have made possible.

- Prof. Dermot Barnes-Holmes for your endless patience, support and teaching. You completely changed my life. Thanks for supervising me

- My partner, Ian Stewart, for your love, tenderness, help and stimulating discussions. Thanks for giving me beautiful memories that I am going to cherish for a lifetime

- Luis Manuel da Silva, Niall Mcgowan, Laura Rabelo, for all the moments that we shared, advices, encouragement and stimulating discussions. The space here is too small to describe how grateful and how important you are for me. Thank you for being my friend and for your unconditional acceptance.

- Special thanks for Claire Cullen who read in details all my thesis and gave a huge support for my viva preparation and Nigel Vahey who had a huge patience in explaining statistics for me, discussing data along my PhD and being present all the process when I most needed.

- Nanni Presti, Francesco L'Orco, Analisa Oppo, Davide, Roberta Kovac, Ila Linares, Caina, Mariana Mercaldi, thanks for believing in my project and being my colleagues in this adventure. Your contributions were invaluable. What would I do without you? It was a big privilege for me having all of you on board. I am sure that it was just a beginning of an enduring partnership.

Abstract

The current thesis set out to develop an implicit relational assessment procedure as a measure of self-forgiveness across five studies. The first study targeted the development of a measure of implicit forgiveness of self related to "minor" transgressions (mistakes, flaws, shortcomings) *versus* the forgiveness of others. The results indicated that the measure of implicit forgiveness diverged from an explicit measure designed to measure the same construct. Moreover, implicitly, participants tended to be more forgiving towards themselves than towards others, whereas at the explicit level participants tended to forgive others. The second study aimed to develop the Implicit Relational Assessment Procedure (IRAP) as a measure of response biases related to emotional reactions and expected outcomes in the context of minor failings and successes in everyday life. However, in contrast to the previous study, the IRAPs were modified to investigate forgiveness of the self rather than others. Additionally, the research explored the extent to which such implicit reactions were related to standardized measures of psychopathology, including a measure Depression, Anxiety and Stress (DASS), measures of self-compassion (Self-Compassion Scale, SCS) and a scale that was based directly on the IRAP. Once again, the results showed that there was no convergence between explicit and implicit measures. The third study was similar to the second one, except that instead of using nomographic stimuli, ideographic stimuli were used with the IRAPs that reflected problem behaviours and the feelings (and anticipated outcomes) that they evoked in each one of the participants. Although specific response biases on the IRAP correlated with psychological suffering, particularly depression and stress, the results of the second and the third study were very similar in that only a few correlations were found between the explicit and implicit measures. Due to disclosure issues, the nomographic IRAPs developed in the second study were used in the next study. The fourth study investigated the impact of two priming procedures on implicit self-forgiveness. Specifically, participants had

to write down 3 examples of failures or successes depending on priming condition (negative or positive priming); the researcher did not have access to what was written. Results indicated differences in the way in which the two types of priming impacted upon the IRAP effects, and how those effects correlated with measures of self-compassion and psychopathology. Finally, the final study aimed to investigate the impact of a history of training in behaviour therapy using the self-forgiveness IRAPs. Overall, only one of the two IRAPs, the one that targeted feelings rather than outcomes, produced clear and significant differences between the group with a history of therapy training versus a control group with no such training. The thesis concludes with a summary and a detailed discussion of the findings reported in each of the chapters. Overall, the research presented in the current thesis constitutes a first step in developing measures of the verbal behaviours involved in the psychological domain of self-forgiveness.

Chapter 1

General Introduction

The Concept of Self-Forgiveness in Psychology

The benefits of forgiving a transgression or a particular person have typically been investigated in psychology by using various self-report instruments. Results from such research have tended to confirm a positive emotional experience for the forgiver, and correlations with other generally positive psychological constructs including empathy and commitment (Worthington, 1998), trust and compassion (McCullough, Worthington, & Rachal, 1997), religiousness-spirituality (McCullough, 2001), agreeableness and emotional stability (Berry, Worthington, Parrot, Connor, & Wade, 2001; Brose, Rye, & Lutz-Zois, 2005; Leach & Lark, 2004; McCullough, Rachal, Sandage, Worthington, Brown, & Hight, 1998), self-esteem (Cardi, Milich, Harris, & Kearns, 2007), and humility (Powers, Nam, Rowatt, & Hill, 2007; Tangney, Boone, Dearing 2005).

Although much has been written about forgiveness per se, self-forgiveness has been a somewhat neglected topic in psychological research. A search conducted in October 2013 using the database PsycInfo, identified 2,794 papers when "forgiveness" was entered as the search term versus only 141 papers when "self-forgiveness" was entered. Much of the research on forgiveness in general has tended to focus on the victim and his or her reaction to a specific event or relationship (e.g., murder, abuse, betrayal, etc.) in terms of forgiving the perceived transgressor.

In contrast, studies on self-forgiveness tend to focus on the extent to which one is willing and able to forgive or excuse some wrong-doing committed by the individual him or herself. According to Thompson et al. (2005), self-forgiveness entails a reframing—a new understanding of oneself and of the offense committed that helps restore a positive self-image without condoning or excusing the offense. In short, self-forgiveness is the psychological process whereby an offender acknowledges wrongdoing following a transgression he or she committed, and without condoning or excusing it, overcomes negative sentiment toward the

self and is reconciled to the self.

The topic of self-forgiveness has only recently been studied systematically in psychology, with limited empirical research on the tendency to self-forgive (Wohl, Pychyl, & Bennet, 2010). And even the results thus far are mixed. On the one hand, some studies suggest that self-forgiveness may be deemed negative in some respects. For example, it may be related to narcissism and self-excusing (Sirois, 2004; Tangney, Boone & Dearing, 2005; Strelan, 2007). Furthermore, Vitz and Meade (2011) suggested that self-forgiveness might also be involved in the psychological problems that comprise "splitting the self", creating various problems such as a conflict of interest between the self that judges and the self that is judged. On the other hand, other studies have indicated that self-forgiveness may have positive benefits, such as facilitating self-correction when working towards a goal or value (e.g., Wohl, Pychyl, & Bennett, 2010) and may even benefit psychotherapy (e.g., Fergusson, Horwood, & Ridder, 2006). The fact that self-forgiveness is itself ill-defined may, in part, account for the conflicting evidence (Vitz & Meade, 2011).

An interesting distinction has been drawn between self-forgiveness and self-excusing. For example, Fisher and Exline (2006) argued that when measures do not consider acceptance of responsibility as a prerequisite, self-forgiveness may be considered as closer to self-excusing than to what theorists would call genuine self-forgiveness. It has further been argued that distinguishing between self-forgiveness and self-excusing may help to explain the contradictory findings related to the apparent benefits of forgiving. On the one hand, self-forgiveness has been found to be positively related to antisocial qualities (Tangney, Boone, and Dearing 2005) and to the tendency to be more blaming toward the victim (Strelan 2007; Zechmeister and Romero 2002). On the other hand, self-forgiveness also has been associated with positive outcomes. For example, Mauger et al. (1992) and Maltby et al. (2001) reported that self-forgiveness is related to mental health, in the sense that people who forgive

themselves are less depressed, introverted, anxious, and distrusted, and Hodgson and Wertheim (2007) showed that self-forgiveness is related to both mental flexibility and emotional stability and self-compassion (Neff and Pomier 2013). Perhaps the abovementioned sets of findings reflect pseudo and true forgiveness, respectively, with the former characterised by excusing one's wrongdoing and the latter by accepting and (genuinely) forgiving it. On balance, perhaps both positive and negative outcomes may be associated with genuine self-forgiveness, dependent on as yet unidentified moderating variables. Addressing this possibility, however, will require further research that would take us outside the remit of the research programme presented within the current thesis.

Self-forgiveness has many common aspects with self-compassion as defined by Neff (2003). For example, self-compassion is defined in terms of kindeness, common humanity and mindfulness as in opposition of self-judgment, isolation and over identification. In other to forgive yourself, you have to be mindful or be present to acknowledge the feelings and outcomes of a wrong-doing committed by one's self, recognise that making mistakes is not something done by a specific person but is part of humanity, not confuse the wrong-doing as part as of your own personality, show kindness as opposed to self-judgment to address the situation, and see things in a different perspective.

Although self-forgiveness has many definitions and interpretations, one common feature in most, if not all definitions, involves the ability to accept or embrace (rather than avoid) negative feelings and possible consequences that may come from an act or behaviour that is deemed to involve some sort of failure and trying to repair the wrong-doing with corrective behaviours. In highlighting this common aspect or skill involved in self-forgiveness, we are not suggesting that this is self-forgiveness *per se*, but maybe one essential aspect of self-forgiveness. In other words, it ie being argued that without acknowledging negative feelings and outcomes, there would be no necessity for forgiveness. For this reason,

feelings and outcomes that may result from failures will be explored in the studies presented in the thesis. According to Hayes, Strosahl and Wilson (2011), for example, 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 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).

How the Concept of Self-Forgiveness is used in the Current Research Programme

Before continuing it seems important to clarify our use of the term "self-forgiveness". There are no previous studies on self-forgiveness in the behaviour-analytic literature, either in general terms or in terms of implict testing. Providing a functional definition is not the aim of the current research and such an effort may be premature. Indeed, following the research reported in this thesis we may be somewhat closer to being able to provide such a definition. In the meantimne, this term is used simply to orient us towards 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). In other words, our research is not designed to provide a technical definition or psychometrically well-developed treatment of self-forgiveness as a psychological construct. Indeed, it may well be that the type of research we are pursuing will in due course render the concept of self-forgiveness largely redundant (see Hayes, Barnes-Holmes, & Wilson, 2012, for an extended discussion of the general approach to psychological science we are adopting here). For the time being, however, we will continue to use the term "self-forgiveness" because its common sense meaning seems closely related

to the psychological domain we are attempting to study. This issue will be revisited towards the end of the current chapter.

Measuring Self-Forgiveness in Psychology: Self-Report Instruments

Until relatively recently, research in social, clinical and other areas of psychology that aimed to collect data reflective of participants' attitudes relied heavily on questionnaires, interviews, and the like (e.g. focus groups). Almost all attempts to measure self-forgiveness thus far have relied exclusively on self-report scales (see Bast & Barnes-Holmes, 2014, 2015). That is, participants are typically asked to complete questionnaires and rate their agreement with items such as "I hold grudges against myself for negative things I've done" or "It is really hard for me to accept myself once I've messed up" (Yamhure, Thompson, Snyder, & Hoffman, 2005). These measures are typically completed in the absence of time-pressure, and thus participants are free to reflect at length on how to respond to each item.

Much of the research in this area has relied upon the use of self-report scales, which ask participants to indicate, for instance, their agreement or disagreement with statements pertaining to self-forgiveness. For example, the State of Self-Forgiveness Scales (SSFS; Wohl, et al. 2008) include items that ask participants to respond to statements such as, "As I consider what I did that was wrong, I feel compassionate toward myself" with responses being made on a 4-point scale: 1 = *not at all*, 2 = *a little*, 3 = *mostly*, and 4 = *completely*.

The Limitations of Self-Report

Self-report measures have potential advantages and disadvantages. On a positive note, participants may need ample time to think about a complex concept such as self-forgiveness and how it applies directly to their lives. On a negative note, this lengthy response time may permit undue influence of a self-presentation bias. For example, participants may not wish to appear excessively self-forgiving so as not to be judged as being too lenient on themselves. In

any case, self-report measures also assume that respondents have insight into their lives, and this may not be the case with regard to self-forgiveness.

In any case, it has long been recognised that attitudes expressed using self-report methodologies were subject to two generic problems or weaknesses. First, the attitudes reported by participants may be influenced by a range of contextual factors that serve to undermine the extent to which the expressed attitudes are predictive of actual behaviour (e.g., responding in a way that is perceived to be socially desirable; Hughes & Barnes-Holmes, 2011). For example, participants may respond to the question "I tend to forgive myself easily for minor wrongdoings" in a manner that does not genuinely reflect their tendency to self-forgive because they do not want to give the impression that they are willing to "go too easy" on themselves. Second, even if a participant does not attempt to respond in a deliberately socially desirable manner when answering a questionnaire, he or she may be unaware of a tendency to respond in a racially-biased manner in a different context (e.g. when interviewing candidates for a job or choosing a neighbourhood in which to live).

Existing Alternatives to Self-Report

Over the past 15 years, a range of measures that are designed to circumvent some of the problems inherent in self-report measures have been developed. These measures are often referred to as *implicit* measures, and they typically require that participants respond to stimuli in a rapid and accurate manner. Unlike self-reports, implicit measures require participants to respond quickly, thus potentially eliminating unwanted sources of contextual control and targeting biases that may be difficult for respondents to articulate or which they would wish to conceal or deny (see Barnes-Holmes, Barnes-Holmes, Stewart, & Boles, 2010; Hughes, Barnes-Holmes, & Vahey, 2012). The most popular implicit measure is known as the implicit association test (IAT; Greenwald et al. 2003), and it has been used across a wide range of areas in psychological research, including forgiveness (Goldrin 2011); gender identity (e.g.,

Greenwald and Farnham 2000), ethnic identity (Devos and Benaji 2005), academic identity (Nosek et al. 2000), self-esteem (e.g., Yamaguchi et al. 2007), clinical contexts (Egloff and Schmukle 2002), and attitudes toward death (Bassett & Dabbs 2003).

The basic idea behind most of these measures is that a participant's implicit attitudes will be revealed by a tendency to respond more or less quickly in one of two conditions, with one condition requiring responses that are congruent with a positive attitude and a second condition requiring responses that are congruent with a negative attitude. In an early IAT study on racial bias, for example, participants were required, during some blocks of trials, to categorize stimuli related to White people with positive terms and stimuli related to Black people with negative terms (Dasgupta and Greenwald 2001). The study showed that White people performed better on the former blocks of trials (White with positive and Black with negative) than on the latter (White with negative and Black with positive), thus revealing a so called pro-White racial response bias, despite reporting (on self-report measures) that they believed themselves to be nonbiased in this regard. In a recent meta-analysis across 184 independent samples ($N = 14,900$), Greenwald, Poehlman, Uhlmann, and Banaji (2009) found that implicit measures were predictive of behaviour at r of .27, compared with r of .36 for parallel explicit measures, with each possessing superior predictive utility in specific construct domains. Examples of studies that have examined self-forgiveness using implicit measures, however, are scarce. Although Goldring (2011) developed a self-concept IAT that included elements which were relevant to forgiveness per se, the emphasis was very much on forgiveness of others for transgressions towards the self, rather than self-forgiveness.

The Implicit Relational Assessment Procedure (IRAP)

One measure of implicit attitudes that has been developed relatively recently and has attracted a growing evidence base is the Implicit Relational Assessment Procedure (IRAP). In a typical IRAP, participants are asked to respond quickly and accurately to stimulus relations

presented via computer. In an IRAP designed to target attitudes to body-size, for example, a picture of a slim (rather than an over-weight) individual may be presented at the top of the computer screen with the target word "Attractive" presented immediately below and the two response options "True" and "False" presented at the bottom of the screen. During some blocks of trials participants are asked to respond as if slim individuals are attractive (in this case pressing a button that selects the "True" response option) and on other blocks of trials participants are asked to respond as if slim individuals are *not* attractive (in this case pressing a button that selects the "False" response option). A growing number of studies indicate that the relative ease with which participants can complete these two types of blocks of trials may reflect their so called implicit attitudes (e.g.,Vahey, Boles, & Barnes-Holmes, 2010; Power, Barnes-Holmes, Barnes-Holmes, & Stewart, I., 2009; Barnes-Holmes, Murtagh, Barnes-Holmes, & Stewart, 2010). For example, Roddy, Stewart and Barnes-Holmes (2010) found that responding on an IRAP that targeted implicit attitudes towards overweight individuals were significant predictors of behavioural intentions towards the overweight. In another study, Roddy, et al (2012) replicated the basic effect and showed that the IRAP effect correlated with subtle facial reactions (another measure of implicit attitudes) that indicated positive attitudes towards the pictures of slim individuals.

Unlike other measures of implicit attitudes, the IRAP emerged from a behaviour-analytic research tradition, or more specifically from Relational Frame Theory (RFT). As such, RFT adopts an exclusively functional epistemology in which scientific analysis is focused on the functional relations between the environment and behaviour that unfold across both time and context (Pepper, 1942, Hayes, 1993). This approach differs fundamentally from that typically found in psychological science, from which most if not all other measures of implicit attitudes have emerged (see Hughes & Barnes-Holmes, 2011, for a recent treatment). The traditional approach often involves working from the assumption that

measures of implicit attitudes tap into underlying mental associations stored in memory. In contrast, the functional approach sees the IRAP as targeting (relational) response probabilities that are determined by historical and current contextual variables. The functional RFT view of implicit attitudes has been formalized recently in terms of the Relational Elaboration and Coherence (REC) model (see Barnes-Holmes, Barnes-Holmes, Stewart, & Boles, 2010, Hughes, Barnes-Holmes, & Vahey, 2012).

The REC model argues that when individuals are required to respond quickly and accurately to sets of stimulus relations, the probability of the initial response on each trial of the IRAP will often be determined by the verbal and nonverbal history of the participant and current contextual variables. These initial responses have been referred to as brief and immediate relational responses (or BIRRs) and are contrasted with extended and elaborated relational responses (or EERRs). In other words, the IRAP is seen as targeting BIRRs, rather than EERRs, and the size of an IRAP effect is taken as a measure of the relative probability of the BIRR being targeted by a particular IRAP. Imagine, for example, that a strong positive IRAP effect is obtained for a trial type that requires a participant to respond "True" rather than "False" to statements such as "My failures are understandable" or "My shortcomings are acceptable". The strong IRAP effect would be taken to indicate that brief and immediate verbal responses concerning self-forgiveness are more probable than ones which deny self-forgiveness. Imagine, however, if a self-report measure was used that aimed to target verbal responses to similar types of "self-forgiveness" statements. In this case, an individual typically has more time to consider each statement and to reflect on how he or she might wish to present themselves to the researcher and such responses may be considered to be EERRs and may diverge in some cases from the BIRRs observed on the IRAP (see Barnes-Holmes et al. 2010a, for an extended discussion).

Before presenting a brief overview of the research presented in the current thesis it seems important to reflect upon the concept of self-forgiveness itself. The purpose of the current programme of research is not to provide a precise definition or psychometric measure of self-forgiveness, but to offer some starting points for beginninng that initiative. The term "self-forgiveness" is used as a verbal stimulus that serves to orient us as researchers towards a particular psychological domain. As such, "self-forgiveness" is functioning as a type of middle-level term (see Hayes, et al., 2012), which does not and almost certainly never will provide the level of precision afforded by well-defined functional-analytic abstractive principles (see Barnes-Holmes, Barnes-Holmes, & Hussey, in press; Barnes-Holmes, Hussey, McEnteggart, Barnes-Holmes, & Foody, in press). In this context, it may be tempting to argue that the IRAPs developed in the current research programme did not target self-forgiveness *per se,* but other constructs such as "self-awareness" or "self-perception", or "self-knowledge" and so on. It is difficult to disagree directly with such arguments, but it seems important to point out that such a strategy simply replaces one ill-defined middle-level term with three others. The general strategy adopted here, therefore, is to attempt to develop a measure, in this case IRAPs, that may be seen as tapping into brief and immediate relational responses. As research in this area unfolds over time the middle-level concept of self-forgiveness may become less dominant, particularly in the context of basic research studies, as more precise descriptions and explanations couched in functional-analytic abstractive principles emerge. At present, however, that remains an aspirational goal rather than an objective that any one study may achieve, and thus it seems important simply note this goal here to help bring clarity to the use of the term self-forgiveness.

Given the preliminary and exploratory nature of these studies, it was generally not possible to make any specific predictions about trial-type effects or specific correlations

between the IRAP scroes and explict measure scores. However, as the research progresses such predictions become more feasible and this will be become apparent in Experiment 5.

A Brief Overview of the Research Presented in the Current Thesis

Given the paucity of research on self-forgiveness using implicit measures, the current research programme was largely exploratory and should be seen as a first step towards a more complete and systematic analysis. It was on this basis that the current thesis sought to develop the IRAP as a measure of relational responding in the general domain of what may be described as self-forgiveness.

Chapter 2 presents the first empirical study, which was designed to measure implicit attitudes related to forgiveness of self and others. For the forgiveness IRAP, participants were presented with stimuli that referred to the mistakes, flaws and failings of self and others and were asked to respond to these as either acceptable and forgivable or as unacceptable and unforgivable. For example, for some blocks of trials participants were required to answer "True" when asked are "My mistakes forgivable" and to respond "False" to this question in other blocks; responding more quickly during the former relative to the latter blocks was taken to indicate an implicit response bias towards self-forgiveness. At the time of writing, no published study had attempted to use an implicit measure to assess the forgiveness of the minor faults, flaws and failings of self versus others, and thus no specific predictions were made. The results indicated that the measure of implicit forgiveness diverged from an explicit measure designed to measure the same construct. The key finding was that participants tended to be more forgiving towards themselves than towards others at an implicit level, but this was not the case at the explicit level; on an explicit measure participants rated their own failures as less acceptable than the failures of others. One limitation of the study reported by Bast and Barnes-Holmes (2014) concerned the extent to which responding in accordance with

self-forgiveness was potentially influenced by the contrast category of forgiveness-of-others. As a result, the authors suggested that the IRAP may not have targeted self-forgiveness *per se*

Chapter 3 presents a second study that aimed to develop the IRAP as a measure of response biases related to emotional reactions and expected outcomes in the context of minor failings and successes in everyday life. The study involved the use of two separate IRAPs, one targeting feelings and one targeting the outcomes associated with failing and succeeding. In effect, the Feelings-IRAP juxtaposed positive and negative feelings regarding failures and successes, whereas the Outcomes-IRAP juxtaposed positive and negative outcomes arising from failures and successes. Additionally, the research explored the extent to which such implicit reactions were related to standardized measures of psychopathology, including depression, anxiety, stress, and a scale that was based directly on the IRAP. The results of the two IRAPs showed biases towards confirming that success produces positive feelings and outcomes, and denying that success produces negative feelings and outcomes. The data from the self-report measures were generally consistent with common-sense in that they showed that failing produced negative biases and questions concerning success produced positive biases. Correlational analyses yielded limited evidence that the implicit and explicit measures overlapped.

Chapter 4 presents a third study that aimed to test the IRAP as a measure of self-forgiveness response biases related to emotional reactions and expected outcomes in everyday life. Unlike previous IRAPs, the stimuli were individualized in that they were based on 'problematic' and 'non-problematic' behaviours (e.g., procrastination versus keeping deadlines) that each participant reported at the beginning of the study. Specifically, participants completed two IRAPs. One (the Feelings IRAP) targeted negative and positive feelings experienced while engaging in problematic versus non-problematic behaviour. The other (the Outcomes IRAP) targeted positive and negative outcomes believed to result from

this behaviour. Participants also completed standardized measures of psychological suffering and self-compassion, as well as a questionnaire that targeted the behaviour and reactions presented in the IRAPs. While both IRAPs produced response biases that indicated that positive feelings and outcomes were related to non-problematic behaviour, neither produced clear evidence that negative feelings or outcomes were related to problematic behaviour. Furthermore, specific response biases on the IRAP (i.e., a tendency to confirm that negative actions lead to negative outcomes) correlated with psychological suffering, particularly depression and stress. The findings suggest that individualized IRAPs, even those that target minor problematic behaviour, may be predictive of psychological suffering.

Chapter 5 presents a fourth study that aimed to test the effect of positive and negative priming on the assessment of self-forgiveness using the IRAP. Additionally, the research explored the extent to which such implicit reactions were related to standardized measures of psychopathology, including depression, anxiety, stress, and a scale that was based directly on the IRAP. Participants were assigned in two groups, where they were presented with different conditions, positive and negative priming, in which they had to recall in writing three experiences of failing or succeeding; participants then completed two IRAPs, one targeting feelings and the other targeting outcomes as related to failing and succeeding behaviours. In addition, participants were asked to complete two explicit measures that were derived from the two IRAPs and another two explicit measures that targeted self-compassion and stress, anxiety and depression. The findings showed that the priming conditions affected the two IRAPs differentially. Furthermore, the IRAP that targeted feelings predicted level of self-reported psychopathology but only for participants in the positive priming condition.

Chapter 6 presents a study that aimed to test the effect of behaviour 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 the two IRAPs that had been

developed across the studies reported in the previous chapters. 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 post graduate students completed the study (20 individuals who were teaching on, attending or who had attended a course in clinical behaviour analysis and 20 students from different fields). The two groups (Behaviour Therapists and Non-Therapists) completed the two IRAPs, and the explicit measures. Overall, only one of the two IRAPs, the one that tagetted feelings rather than outcomes, produced clear and significant differences between the Behaviour Therapist and Non-Therapist 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 two IRAPs and explicit measures with these two groups of participants are considered.

Chapter 7, the final chapter of the thesis considers the implications of the findings from the previous five empirical chapters for RFT and the REC model and outlines potential challenges and opportunities that lie ahead for researchers interested in the study of self-forgiveness.

Chapter 2

A First Test of the Implicit Relational Assessment Procedure as a

Measure of Forgiveness of Self and Others

Introduction

In the present study, an IRAP was designed to measure implicit attitudes related to forgiveness of self and others. For the forgiveness IRAP, participants were presented with stimuli that referred to the mistakes, flaws and failings of self and others and were asked to respond to these as either acceptable and forgivable or as unacceptable and unforgivable. For example, for some blocks of trials participants were required to answer "True" when asked are "My mistakes forgivable" and to respond "False" to this question in other blocks; responding more quickly during the former relative to the latter blocks was taken to indicate an implicit response bias towards self-forgiveness. At the time of writing, no published study had attempted to use an implicit measure to assess the forgiveness of the minor faults, flaws and failings of self versus others, and thus no specific predictions were made. As such the study was entirely exploratory and constituting only a first step in the current doctoral research programme.

Method

Participants

Forty-seven undergraduates and post-graduates students, 21 male and 26 female, between 18-32 years old, were recruited from various departments at the National University of Ireland Maynooth and completed the current study on a voluntary basis. Participants were divided initially into three categories: a Humanities group comprised of students from the Schools/Departments of Modern Languages, Celtic Studies and Music; a Psychology group comprised of students from the Department of Psychology; an Engineering/Computer Science group comprised of students from the Departments of Electronic Engineering and Computer Science. Preliminary data analyses indicated that no significant differences emerged among the three groups and thus all analyses reported subsequently were conducted without regard to student type.

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, an up-dated version of the program is available for download from www.IRAPresearch.org. Explicit measures (IRAP Explicit Measure for Forgiveness, and other scales) 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 pre-experimentally 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.

The label stimuli consisted of 12 short statements referring to individual shortcomings or failures. Six of these statements were self-referential (i.e., "My Shortcomings are", "My failures are", "My weaknesses are," My faults are", "My flaws are," "My mistakes are") and six of the statements referred to other people's shortcomings or failures (i.e.. "Other people's shortcoming are"; "Other people's failures are"; "Other people's weaknesses are"; "Other people's faults are";"Other people's flaws are"; "Other people's mistakes are"). The target stimuli were 12 single words, six of which indicated non-acceptance (i.e., unacceptable, unforgivable, awful, terrible, embarrassing, intolerable) and a further six which indicated acceptance (i.e., okay, forgivable, normal, fine, acceptable, and tolerable). Thus each trial of the IRAP presented a label and a target stimulus that indicated one of four possible label-

target combinations, which may be described as; (i) Self-failures-unacceptable; (ii) Self-failures-acceptable; (iii) Others-failures-unacceptable; (iv) Others-failures-acceptable. Participants responded to these label-target combinations by choosing one of two response options, "True" and "False", which appeared in the bottom right- 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 stimulus, target stimulus, and both response options appeared on the screen simultaneously at the onset of each trial. The label and the target stimulus varied randomly with each trial, as did the left- and right- positions of the response options. An example of the layout of an IRAP trial is provided in Figure 1. Participants also completed another IRAP unrelated to forgiveness but the details of this IRAP are not reported here.

Explicit measure. The explicit measure consisted of a self-report scale that was derived directly from the stimuli employed with the IRAP (participants were asked to complete other scales related to a secondary research question, but these are not reported here. The scale presented 24 items that were based directly on the trials that were presented in the IRAP (see appendix 1). The scale presented the following instruction at the top of a single page:

> "*When something does not go as planned or something goes wrong in our*
>
> *lives, we often engage in some sort of evaluation of the situation and the*
>
> *people involved, including ourselves. However, the way in which we*
>
> *evaluate ourselves and others can be different. Please read the following*
>
> *sentences carefully and circle the number that best describes how much*
>
> *each statement is true for you. My shortcomings, failures, weaknesses,*
>
> *faults, flaws, mistakes are:*"

Participants were asked to give a score from 1 (not at all true) to 7 (very true) for each of the 12 target words presented in the IRAP (i.e., unacceptable, unforgivable, okay, forgivable, etc.). Immediately below these words the following instruction appeared:

"*Other people's shortcomings, failures, weaknesses, faults, flaws, mistakes are*:" And participants were again invited to score each of the words using the same seven point scale. In effect, participants were asked to indicate how acceptable or unacceptable they deemed their own failures and those of others to be using the same target words as were presented in the IRAP. For the purposes of data analysis, the scores for the words indicating "acceptability" were reversed so that higher scores indicated unacceptability of the failures associated with self or others and lower scores indicated acceptability of such failures.

Procedure

After completing consent forms, participants were asked to complete the IRAP and explicit measures (the sequence in which the implicit and explicit measures were completed was counterbalanced across participants). Depending on availability, participants were permitted to complete the implicit and explicit measures on separate days. Preliminary analyses (not reported) indicated that there were no significant effects arising from this procedural variable.

Explicit measure. For the explicit measure, the scales were simply presented to each participant 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. The researcher sought to clarify what was required of the participants if they requested any further information or help, but never stated or indicated that the researcher expected speed of responding to vary across the blocks of the IRAP.

Each trial of the IRAP presented a label stimulus, a target stimulus, and two response options (described previously). In a typical IRAP, 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' mid-screen 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.

Each block on the IRAP presented 24 trials. The trials were presented in a quasi random 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 conceptualised as involving four different trial-types (see Figures 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 IRAP, participants were required to respond in a pattern that may be described as involving a Self-Unacceptability bias (i.e. responding "True" to "Self-Failures" as "Unacceptable" and to "Others-failures" as "Acceptable" and responding "False" to "Self-Failures" as "Acceptable" and to "Others-Failures" as "Unacceptable"). In Block 2, and all subsequent even numbered blocks of the IRAP, participants were required to respond in a pattern that may be described as involving a

Self-Acceptability bias (i.e. responding "False" to "Self-Failures" as "Unacceptable" and to "Others-Failures" as "Unacceptable" and responding "True" to "Self-Failures" as "Acceptable" and to "Others-Failures" as "Unacceptable").

My shortcomings are Unacceptable		Other people's shortcomings are Acceptable	
Self-Acceptable	Self-Unacceptable	Others-Acceptable	Others-Unacceptable
Select'd' for	Select 'k' for	Select'd' for	Select 'k' for
False	True	True	False
My shortcomings are Acceptable		Other people's shortcomings are Unacceptable	
Self-Acceptable	Self-Unacceptable	Others-Acceptable	Others-Unacceptable
Select'd' for	Select 'k' for	Select'd' for	Select 'k' for
True	False	False	True

Figure 1. A schematic representation of the four trial-types from the IRAP

Each IRAP commenced with a minimum of two practice blocks. Participants were required to achieve ≥ 80% correct and a medium response latency ≤ 2000ms 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., 8 blocks in total). 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 few errors is okay" was presented before the beginning of

each block. The IRAP programme 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 and Conclusion

The IRAP

Data preparation. For the purposes of statistical analysis participants were required to

maintain an accuracy level $\geq 70\%$ correct and a median latency ≤ 3000ms on two of the three

successive pairs of the six test blocks This was different to the usual procedure, because in

analysing the IRAP data, many participants failed to reach the practice fluency criteria. It

may be that the sentence stimuli were too long to allow more rapid responding, or it may be

related to the experimenters naivite in using the IRAP procedure. According to Nicholson,

Doyles, Barnes-Holmes, & Roche (2014) the failure to reach test criteria is largely due to

instructional issues and tehse issues have yet to be fully resolved in the IRAP literature. In

any case, the data for four participants were excluded because they failed to meet the

specified 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 (this practice was adopted for two of the participants).

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., 2003).

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 from the dataset were eliminated; (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 self-unacceptability bias block from the corresponding mean latency of the self-acceptability bias 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 trial-type *D*-IRAP scores, or IRAP effects, were calculated by averaging the scores for each trial-type across the three pairs of test blocks; (ix) an overall D-IRAP socre for each IRAP was calculated by obtaining the avarage of the four D-IRAP scores; and (x) the two *D*-IRAP scores for the trial-types that involved responding to "Others' Failures" were inverted (plus scores became minus scores and minus scores became plus scores).

The same ten 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 an unacceptability bias whereas negative *D*-IRAP scores indicated an acceptability bias.

Mean scores analyses The four overall mean *D*-IRAP scores calculated across participants are presented in Figure 2. Relatively clear unacceptability biases were revealed for the two trial-types that asked participants to respond to the failures of both self and others as unacceptable. However, when participants were asked to respond to the failures of self and others as acceptable a strong bias score emerged only for the "self" trial-type and this revealed an acceptability bias (the score for the "Others" trial-type showed a marginal unacceptability bias).

The *D*-IRAP scores for the four trial-types were entered into a one-way repeated measures analysis of variance (ANOVA), and this yielded a significant effect, $F(3, 40) =$ 9.78, $p < .0001$, $\eta_p^2 = .2$. Fisher's PLSD post-hoc tests indicated that the *D*-IRAP scores for the *My-Failures-Unacceptable* trial-type differed significantly from the *My-Failures-Acceptable* trial-type (p < .0001). In addition, the *D*-IRAP effect for the *My-Failures-Acceptable* trial-type differed significantly from effects recorded for the *Others-Failures-Unacceptable* and *Others-Failures-Acceptable* trial-types (*ps* < .003); the comparison between the two latter trial-types approached significance ($p = .08$). When each of the trial-type scores was subjected to one-sample *t*-tests, three of them proved to be significant (*ps* < .03); the *Others-Failures-Acceptable* effect was non-significant. The statistical analyses thus indicated that the participants showed implicit biases towards confirming that their own failures were both unacceptable and acceptable. When asked to respond to the failures of others, participants showed a similar bias towards confirming that such failures were

unacceptable but did not show a bias towards acceptability. More informally, it appears that the participants tend to condemn the failures of self and others as unacceptable with relatively equal strength, but tend to be more "forgiving" about our own failures relative to others when responding to questions concerning acceptability.

Forgiveness IRAP
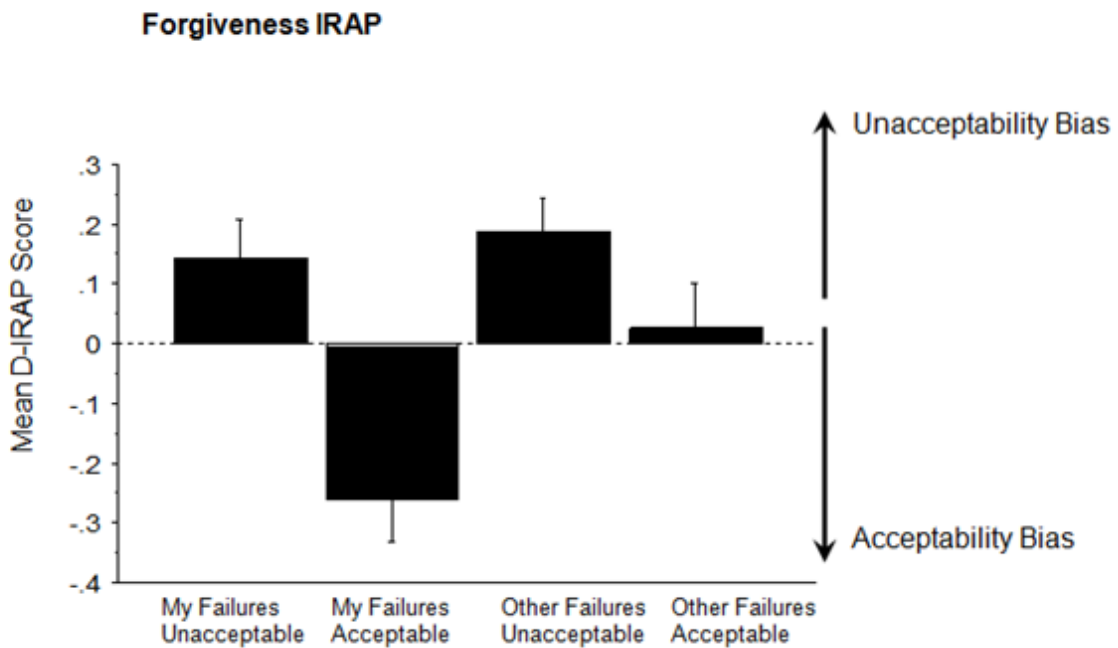
Figure 2. Mean D-IRAP- Trial-Type scores obtained for the IRAP

Explicit Measure

The ratings obtained from the explicit measure were used to calculate four separate scores, with each score mapping onto the equivalent of one of the trial-types from the IRAP. For example, the six ratings obtained for questions pertaining to the unacceptability of "My shortcomings, failures" etc were used to calculate a mean score that provided the explicit counterpart to the *My-Failures-Unacceptable* trial-type from the IRAP. As noted previously, the ratings for the words indicating "acceptability" were reversed so that higher scores indicated unacceptability of the failures associated with self and others and lower scores indicated acceptability of such failures. The overall mean scores and standard errors for each rating were as follows: *My-Failures-Unacceptable, M = 2.54, SE = .22; My-Failures-*

Acceptable, M = 2.93, *SE* = .20; *Others-Failures-Unacceptable, M* = 1.97, *SE* = .17; *Others-*

Failures-Acceptable, M = 2.56, *SE* = .20. The mean rating scores for each participant were

then entered into a one-way repeated measures ANOVA, and this yielded a significant effect,

F (3, 40) = 9.75, p < .0001, η_p^2 =0.19. Six Fisher's PLSD post-hoc tests indicated that all of

the ratings differed significantly from each other (*ps* < 0.5), except for the comparison

between questions pertaining to *My-Failures-Unacceptable* versus *Others-Failures-*

Acceptable. The results from this explicit measure thus indicated that participants tended to

rate their own failures as more unacceptable or less acceptable than the failures of others. In

contrast to the bias scores from the IRAP, therefore, there was no indication that participants

responded to the acceptability of their own failures in a more forgiving manner than they did

towards those of others. Indeed, the opposite appeared to be case – participants rated the

failures of others as more acceptable than their own failures.

  Implicit-Explicit Correlations The four *D*-IRAP scores were entered into a correlation

matrix with the explicit measures. Of the 16 correlations only one proved to be significant,

the *Others-Failures-Acceptable D*-IRAP scores with the *Self-Failures-Unacceptable* scores

from the explicit measure (*r* = -.32, *p* = .04). The negative correlation indicates that the less

implicitly accepting participants were concerning the failures of others the more explicitly

accepting they were concerning their own failures. However, this result needs to be

interpreted with caution because it constitutes only one significant correlation out of a

possible 16.

Conclusion

  The current study presented participants with an IRAP and an explicit measure that

were designed to assess forgiveness (towards self and others). The main findings of the

research indicated that participants were more "forgiving" with regard to their own failures, relative to the failures of others, at an implicit level (i.e. on the IRAP). On an explicit measure directly derived from the IRAP, however, this bias towards greater forgiveness of one's own failures was not evident. There was limited evidence for correlations between the implicit and explicit measures.

One possible criticism of the current study is that the IRAP required that participants respond in a manner that showed either Self-Failures-Acceptability/Others-Failures-Unacceptability biases in some blocks of trials or the opposite pattern in other blocks of trials (i.e., Self-Failures-Unacceptability/Others-Failures-Acceptability biases). Perhaps pitting one pattern of biases against the other in the IRAP helped to produce the differences between the implicit and explicit measures? In the study reported in the next chapter two IRAPs were employed but each targeted forgiveness of the self rather than self versus others.

Chapter 3

Developing the IRAP as a Measure of Self-Forgiveness related to

Failing and Succeeding Behaviours

Introduction

In the study presented in the previous chapter participants tended to respond more quickly when they had to confirm forgiveness of self but not others than when they had to confirm forgiveness of others but not the self. In broad terms, therefore, the IRAP data suggested that participants were more "forgiving" with regard to their own failures, relative to the failures of others, at an implicit level. Interestingly, however, the data gathered using an explicit (self-report) measure, which was directly derived from the IRAP, indicated that the same participants were more "forgiving of others' failures than they were of their own. Perhaps self-forgiveness, similar to racial bias, might be considered a socially sensitive issue because the implicit measure (in this case the IRAP) yielded a pattern of results that seemed to contradict the results obtained from an explicit (self-report) measure. To put it more bluntly, in general, we tend to be more forgiving of our own failures, shortcomings and misdemeanours than we are of those committed by others, but we are unable or unwilling to report this rather "unpleasant" feature of our characters. In drawing this conclusion, it is important to note that the study focused on the forgiveness of mistakes, flaws and failures with respect to the self and others rather than forgiveness of (serious) criminal acts or behaviours that had inflicted harm or suffering on other individuals. The study reported in the current chapter also focuses on (minor) failures rather than criminal activity. The research continued to focus on forgiveness of minor failures because such research will be directly relevant to the general population, rather than only to those who have been perpetrators or victims of serious criminal acts.

As noted previously, one limitation of the previous study identified, however, was that the IRAP (and the explicit measure) involved "pitting" forgiveness of self against the forgiveness of others. Thus, the self-forgiveness effects obtained in the study may be specific to situations in which one is asked to respond to self-forgiveness in a context in which one is

also asked about forgiveness of others. In order to assess self-forgiveness per se, it seems important to conduct a study in which only self-forgiveness is targeted (rather than forgiveness of the self versus others). This was one of the key aims of the study reported in the current chapter.

Another feature of the current study that differs from the previous study is the use of two IRAPs rather than just one. Specifically, one IRAP targets positive versus negative *feelings* that participants experience when they fail versus succeed, whereas the second IRAP targets the positive versus negative *outcomes* participants expect when they fail versus succeed. The use of these two different IRAPs was largely exploratory but was based on the commonsense assumption that sometimes a minor failure might produce negative feelings, but rationally an individual may also recognize that sometimes failures may produce positive outcomes (e.g., as in learning from one's mistakes).

Method

Participants

Sixty-four undergraduate and postgraduate students between 18 and 32 years old were recruited, via class announcements, from various departments at Maynooth University, and completed the current study on a voluntary basis. Four students were excluded because they did not achieve the IRAP performance criteria detailed in the procedure session. No payment or course credits were exchanged for participation, but volunteers were offered a chocolate bar before leaving the laboratory.

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 second author; an updated version of the program is available

for download from www.IRAPresearch.org. Explicit measures were provided in hard copy format.

Explicit measures. There were four separate explicit measures. Two measures were derived from the stimuli used with the IRAPs (see appendix 2), and the two other measures were standardized psychometric instruments targeting self-compassion (Self-Compassion Scale, SCS; Neff, 2003) and depression, anxiety and stress levels (Depression Anxiety and Stress Scale, DASS; Lovibond & Lovibond 1995). The SCS was used because it was felt that there may be some overlap between self-forgiveness and self-compassion. The DASS was used because it had been adopted successfully in previous studies that used 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 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 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", etc.). 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 as was used for the 24 "feeling" items.

Implicit Measures. Each participant was required to complete two IRAPs, one designed to target feelings and one that targeted 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, 6 of which indicated negative feelings (i.e., ","I feel Bad", "I feel Guilty", "I feel Stupid", "I feel Useless", "I feel Frustrated", "I feel Angry") and a further 6 that indicated positive feelings (i.e., "I feel Good", "I feel Strong", "I feel Energetic", "I feel Positive", "I feel Calm", "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) Failures-negative feelings, (ii) Failures-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" and "False", which appeared in the bottom right- 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 stimulus, target stimulus, and both response options appeared on the screen simultaneously at the onset of each trial. The label and the target stimulus 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 be the primary datum employed for analysis. A schematic representation of the Feelings-IRAP is presented in Fig. 3.

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 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) Failures-negative outcomes, (ii) Failures-positive outcomes, (iii) Success-negative outcomes, (iv) Success-positive outcomes. A schematic representation of the Outcomes-IRAP is presented in Fig. 4.

When I fail	When I succeed
I feel guilty	I feel good
Positive Feelings Bias Negative Feelings Bias	Positive Feelings Bias Negative Feelings Bias
Select 'd' for Select 'k' for	Select 'd' for Select 'k' for
False True	True False
When I fail	**When I succeed**
I feel good	I feel guilty
Positive Feelings Bias Negative Feelings Bias	Negative Feelings Bias Positive Feelings Bias
Select 'd' for Select 'k' for	Select 'd' for Select 'k' for
True False	True False

Figure 3. A schematic representation of the four trial-types from the Feelings- IRAP

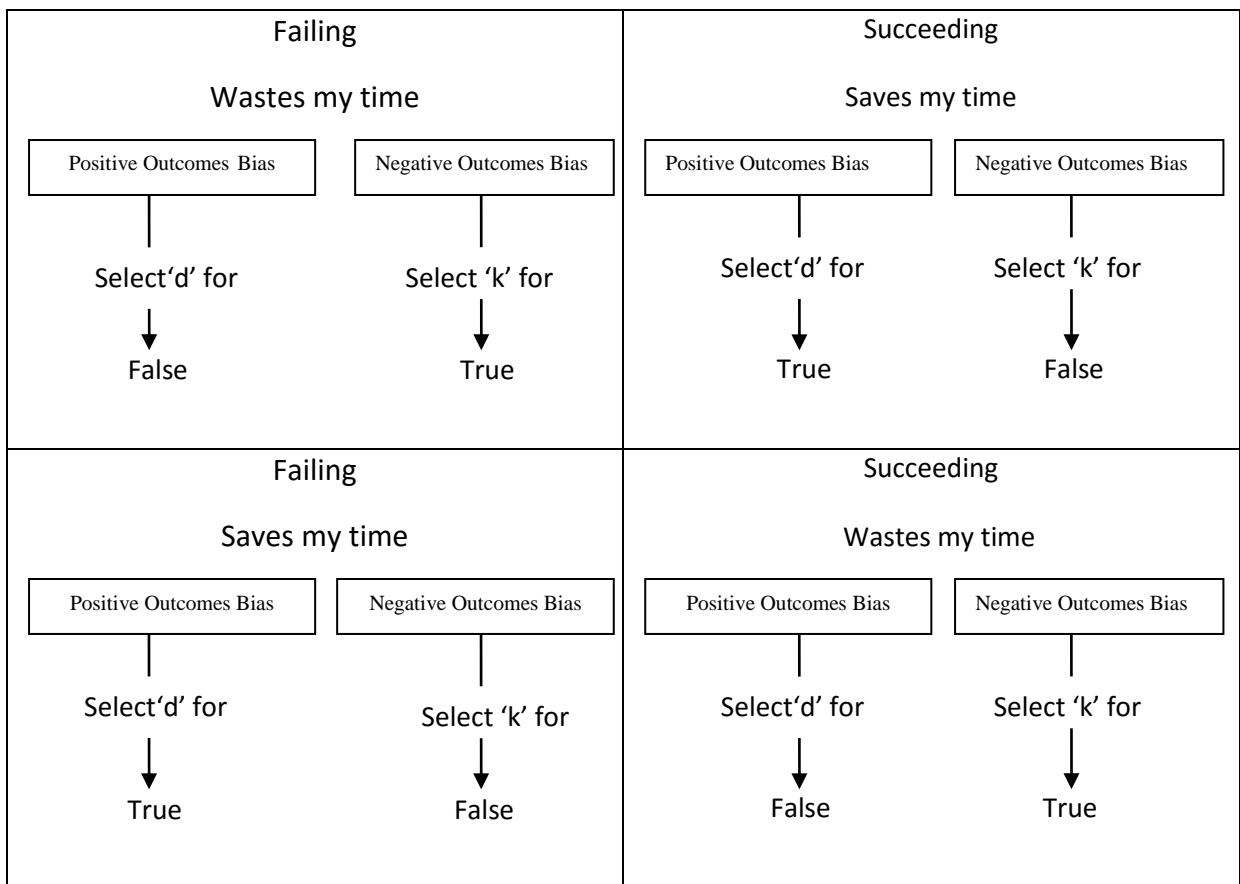| Failing	Succeeding
Wastes my time	Saves my time
Positive Outcomes Bias Negative Outcomes Bias	Positive Outcomes Bias Negative Outcomes Bias
Select 'd' for Select 'k' for	Select 'd' for Select 'k' for
False True	True False
Failing	**Succeeding**
Saves my time	Wastes my time
Positive Outcomes Bias Negative Outcomes Bias	Positive Outcomes Bias Negative Outcomes Bias
Select 'd' for Select 'k' for	Select 'd' for Select 'k' for
True False	False True

Figure 4. A schematic representation of the four trial-types from the Outcomes- IRAP

*Self Compassion Scale (*SCS; Neff, 2003) The 26-item Self-Compassion Scale (SCS) includes the 5 item Self-Kindness subscale (e.g., "I am tolerant of my own flaws and inadequacies"), the 5-item Self-Judgment subscale (e.g., "When times are really difficult, I tend to be tough on myself"), the 4-item Common Humanity subscale (e.g., "I try to see my failings as part of the human condition"), the 4-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 4-item Mindfulness subscale (e.g., "When something painful happens I try to take a balanced view of the situation"), and the 4-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 intercorrelations 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 et al. 2007).

Depression Anxiety and Stress Scale (*DASS;* Lovibond & Lovibond 1995) 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 (Antony et al. 1998; Brown et al. 1997; Clara et al. 2001; Crawford & Henry 2003). Each of the three DASS scales contains 14 items, divided into subscales of 2 to 5 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 non-specific arousal. It assesses difficulty relaxing, nervous arousal, and being easily upset/agitated, irritable/over-reactive, 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.

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 1 hour: 45 minutes to complete both IRAPs and 15 minutes to complete the explicit measures.

Explicit measures 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 measures 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–that is, 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.

Each block on the IRAP presented 24 trials. The trials were presented in a quasi random 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 conceptualised as involving four different trial-types (see Figs. 3 and 4). 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 commonsense 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 to "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 ≤ 2000ms 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 in total). 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 programme 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 and Conclusion

Preliminary data analyses indicated no significant effects arising from the procedural variable of counterbalancing the order in which the two IRAPs were presented and thus this variable was removed from all analyses.

The Feelings-IRAP

Data preparation. For the purposes of statistical analysis, participants were required to maintain an accuracy level $\geq 75\%$ correct and a median latency $\leq 2,000$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 (this practice was adopted for 18 IRAPs related to feelings).

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. 2003). Calculating D-IRAP scores for each participant who met the criteria for all six test blocks involved the same ten steps described in the previous study and for participants who met the criteria for two of the three pairs of test blocks except the algorithm was adjusted accordingly (i.e., 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/outcomes bias whereas negative *D*-IRAP scores indicated a negative

feelings/outcomes bias.

Mean scores analyses. The four overall mean *D*-IRAP scores calculated across

participants are presented in Fig. 5 (upper left panel). A relatively strong positive bias was

revealed for the *Success-Positive feelings* trial type, but this was not the case for the

remaining trial types. However, the *Success-negative feelings* trial type also produced a

positive bias, whereas the biases for the two *Failure* trial types went in opposite directions

(Figs.3 and 4 indicate which responses on the IRAP were deemed to show positive versus

negative bias).

The *D*-IRAP scores for the four trial types were entered into a one-way repeated

measures analysis of variance (ANOVA), and this yielded a significant effect, $F(3, 59) =$

18.973, p .<.0001, η_p^2 = .0.243. Fisher's PLSD post-hoc tests yielded five significant

differences (*ps* < .004) among the four trial types, with only the *Failure-Positive* versus

Success-Negative trial type comparison producing a nonsignificant effect (p > .6). When each

of the four trial type scores was subjected to one-sample *t* tests, only the *Failure-Positive*

feelings trial type did not differ significantly from zero (*p* = .1; remaining *ps* < .05).In

general, therefore, the statistical analyses supported the descriptive statistics presented in Fig.

5.

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 (for 13 of 60 IRAPs

related to outcomes, the data were calculated from two, rather than three, pairs of test blocks).

Mean scores analyses. The four overall mean *D*-IRAP scores calculated across

participants are presented in Fig. 5 (upper right panel). The general pattern of biases observed

for the Feelings-IRAP was also observed for the Outcomes-IRAP. When the *D*-IRAP scores

were entered into a one-way repeated measures ANOVA, it proved to be significant, F (3, 59) = 20.340, $p <.0001$, $\eta_p^2 = .279$. Fisher's PLSD post-hoc tests produced a similar pattern to that recorded for the Feelings-IRAP: five significant differences ($ps < .002$) among the four trial-types, with only the *Failure-Positive* versus *Success-Negative* trial type comparison producing a nonsignificant effect ($p > .6$). When each of the four trial type scores was subjected to one-sample t tests all four proved to be significantly different from zero ($ps <$ .05). Once again, therefore, the statistical analyses supported the descriptive statistics presented in Fig. 5.

Explicit Measures

The scale based on the Feelings-IRAP. The ratings obtained from the explicit measure that was derived from the Feeling-IRAP (hereafter referred to as the Explicit-Feelings scale) were used to calculate four separate scores, with each score mapping onto the equivalent of one of the trial types 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 7-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. 5 (lower left panel). The two subscales that mapped onto the two Success trial types from

the IRAP both produced responses that indicated a positive bias, but the levels of bias were reversed relative to the IRAP scores (i.e., stronger for *Success-Negative feelings* than for *Success-Positive feelings*). The two subscales that mapped onto the two Failure trial types from the IRAP both yielded negative bias, which contrasts with the pattern observed for the IRAP, which produced a negative bias for the *Failure-Negative feelings* trial type but a positive bias for the *Failure-Positive feelings* trial-type. The mean rating scores for each participant from the Explicit-Feelings scale were entered into a one-way repeated measures ANOVA, and this yielded a significant effect, $F(3, 59) = 254.161$, $p < .0001$, $\eta_p^2 = 0.811$. Six Fisher's PLSD post hoc tests indicated that all of the ratings differed significantly from each other ($ps < .0001$); four one-sample t tests indicated that the ratings for each subscale differed significantly from zero ($ps < .0001$). The statistical analyses thus supported the descriptive statistics presented in Fig. 5.

The scale based on the Outcomes-IRAP. The data from the explicit measure that was derived from the Outcomes-IRAP (hereafter referred to as the Explicit-Outcomes scale) were used to calculate four separate scores, with each score mapping onto the equivalent of one of the trial types from the IRAP. The data were transformed in the same way as for the Explicit-Feelings scale, and are presented in Fig. 5 (lower right panel). The pattern of results is broadly similar to those observed for the *Explicit-Feelings scale,* and they contrast with those obtained from the Outcomes-IRAP in much the same way as the results contrast with each other across the two Feelings measures. When the mean rating scores for each participant were entered into a one-way repeated measures ANOVA, this yielded a significant effect, $F(3, 59) = 137.56$, $p < .0001$, $\eta_p^2 = 0.70$. Six Fisher's PLSD post-hoc tests indicated that all of the ratings differed significantly from each other ($ps < .03$), and four one-sample *t* tests indicated that the ratings for each subscale differed significantly from zero ($ps < .0001$). Once again, the statistical analyses thus supported the descriptive statistics presented in Fig. 5.

Self compassion scale. According to Neff (2003), average overall self-compassion scores tend to be around 3.0 on the 1-5 scale, so the overall score can be interpreted accordingly: An overall score of 1-2.5 indicates low self-compassion; 2.5-3.5 indicates moderate self-compassion; and 3.5-5.0 indicates high self-compassion. In the present study, it was found similar results to the previous study conducted in the previous chapter, in which the mean overall score was 2.81 (*SD* = .36); an average score indicating moderate self-compassion for the current sample.

Depression, Anxiety and Stress Scale. The interpretation of the DASS is based primarily on the use of cut-off scores. Lovibond & Lovibond (1995) presented severity ratings from *normal* to *extremely severe* on the basis of percentile scores, with 0–78 classified as *normal*, 78–87 as *mild*, 87–95 as *moderate*, 95–98 as *severe*, and 98–100 as *extremely severe*. In the current study, the mean overall score for the DASS was 27.033 (*SD* = 17.84), indicating that the sample fell well below the cut-off between *normal* and *mild* severity.

Implicit-Explicit Correlations

The Feelings-IRAP and Explicit Measures. The four *D*-IRAP scores were entered into a correlation matrix with the total and subscale scores from each of the three explicit measures (the scores obtained from SCS, DASS, and the scale based on the Feelings-IRAP). Of the 60 correlations, just 1 proved to be significant; increased positive bias on the *Fail-positive feeling* IRAP trial type predicted reduced positive bias ratings on the *Success-negative feeling* subscale from the Explicit-Feelings scale (*r* = -.29, p = .03). In other words, confirming that failing produces positive feelings at an implicit level related to denying that success produces negative feelings at an explicit level. However, given the large number of correlations involved this one significant effect should be viewed with caution.

The Outcomes-IRAP and Explicit Measures The four *D*-IRAP scores were entered into a correlation matrix with the total and subscale scores of the three explicit measures (the

scores obtained from SCS, DASS and the Explicit Outcomes-IRAP), 4 of the 60 correlations

proved to be significant. Three of these involved relationships between the IRAP and the

Explicit-Outcomes scale. The first correlation indicated that increased negative bias on the

Fail-negative outcomes IRAP trial type predicted increased positive bias ratings on the

Success-positive outcomes subscale from the Explicit-Outcomes measure ($r = -.35$, $p = .005$).

In other words, an increase in confirming that failing produces negative outcomes at an

implicit level related to an increase in confirming that success produces positive outcomes at

an explicit level. The second correlation indicated that increased positive bias on the *Fail-*

positive outcomes IRAP trial type predicted decreased negative bias ratings on the *Fail-*

positive outcomes subscale ($r = .25$, $p = .05$). That is, an increase in confirming that failing

produces positive outcomes at an implicit level related to a reduction in denying that failing

produces negative outcomes at an explicit level. The third correlation indicated that increased

positive bias on the *Success-positive outcomes* IRAP trial type predicted increased positive

bias ratings on the *Success-Positive Outcomes* subscale ($r = .35$, $p = .006$). In effect, an

increase in confirming that success produces positive outcomes at an implicit level related to

an increase in confirming that success produces positive outcomes at an explicit level. The

fourth and final significant correlation was obtained between the *Fail-negative outcomes*

IRAP trial type and the *Mindfulness* subscale of the SCS ($r = .31$, $p = .02$), indicating that

denying that failure produces negative outcomes predicts higher levels of self-reported

mindfulness. The number of correlations between the implicit and explicit measures thus

remained relatively low (only 4 out of 60), but perhaps warrant more attention than the single

correlation that was obtained with the Feelings-IRAP.
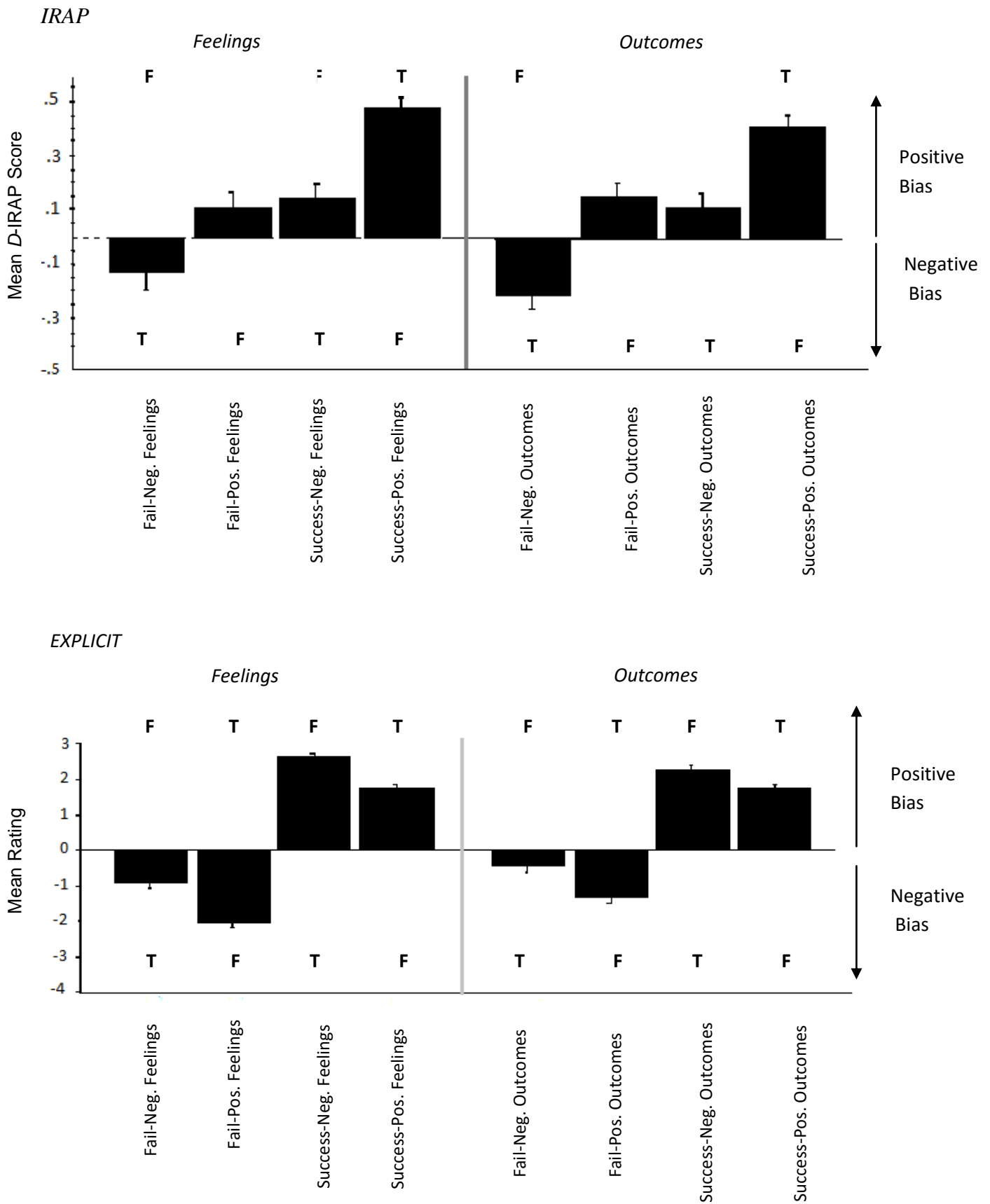
Figure 5. Mean D $_{\text{IRAP-p-Trial-Type Scores}}$ obtained on the IRAP and the mean rating obtained on the explicit measures for the feelings and outcomes related to failing and succeeding. The letters "T" and "F" indicate the direction of the response biases ("True" and "False", respectively) that were recorded by the measures.

Conclusion

The current study presented participants with two separate IRAPs—one targeting feelings and the other targeting outcomes in relation to failing and succeeding. In addition, participants were asked to complete two explicit measures that were derived from the two IRAPs and another two explicit measures that targeted compassion and psychopathology. The results arising from the two trial types that focused on "success" for both IRAPs were broadly consistent with "common-sense" conclusions in that all of the IRAP effects yielded positive bias effects. The IRAP effects for the two trial types that focused on "failure," however, were not so straightforward. Although the trial types that targeted failure and negative feelings, or failure and negative outcomes, produced negative biases, the trial types that targeted failure and positive feelings/outcomes both produced positive biases. Interestingly, the explicit measures that were designed to map onto the trial types from the two IRAPs produced biases that were all consistent with "common-sense" conclusions–questions concerning failing produced negative biases and questions concerning success produced positive biases.

The pattern of biases observed between the IRAPs and the explicit measures derived from the IRAP trial-types differed in another way. Specifically, although the IRAP effects for the two "Success" trial types were both positive (for both IRAPs), the effects for the *Success-positive feelings* and *Success-positive outcomes* trial types were considerably stronger than the effects for the two respective *Success-negative* trial types; this pattern was the opposite of that observed for the two explicit measures. In the latter case, the *Success-Negative* subscales produced stronger positive bias ratings than the *Success-Positive* subscales.

At the current time, it remains unclear why the correlations were so weak and/or few in number. It could be argued that the IRAP and the scales of SCS and DASS are part of different constructs. However, this cannto help explain the lack of correlation between the IRAPs and the scales based on the IRAP itself. Perhaps the implicit and explicit measures

simply targeted relational responses that were under different forms of contextual control and thus they failed to correlate. More specifically, the REC model assumes that the IRAP effect, when produced under appropriate time pressure, is driven largely by immediate and relatively brief relational responses, whereas explicit measures reflect extended and coherent relational networks. The core of the REC model explanation for the impact of increased time pressure on the divergence between implicit and explicit measures is that it is assumed that participants usually "reject" their immediate and brief relational responses if they do not cohere with their more extended relational responding. In addition, the REC model predicts that the divergence between implicit and explicit "socially sensitive"attitudes should increase with greater time pressure on the IRAP, because participants have less time to engage in elaborated relational responding (Barnes-Holmes et al (2011). Thus, it may be that different relational repertoires were tapped into in the different testing contexts acorss the implict and explicit test.

Another, or perhaps additional, reason for the lack of correlation between the implicit and explicit measures is that the statements pertaining to failure (versus successes) were simply not evocative or salient enough to elicit relatively strong emotional reactions in many, if not most, of the participants. In other words, general statements about failing versus succeeding did not encourage participants to recall or genuinely think about their own previous failures and successes, and thus the absence of any consistent relationships with levels of psychopathology or self-compassion would be expected. In any case, to address this later possibility, one strategy might be to ask participants to provide examples of failures or shortcomings that were specific to them and then insert these into 'individualized' IRAP. This strategy was adopted in the study reported in the next chapter.

Chapter 4

Developing an Individualized Implicit Relational Assessment

Procedure (IRAP) as a Potential Measure of Self-Forgiveness

related to Negative and Positive Behaviour

Introduction

The study reported in the previous chapter raised a number of issues concerning the extent to which the IRAPs they employed could be seen as providing measures of implicit self-forgiveness. As noted above, the trial-types that targeted failure and positive feelings/outcomes both produced positive biases -- intuitively, one would expect participants to deny (more quickly than confirm) that failures are positive. Furthermore, the study reported a lack of correlations between the implicit and explicit measures. One probable reason for such findings might be that the statements of failure and success employed in the IRAP were simply not evocative or salient enough to elicit relatively strong emotional reactions from participants. In other words, general statements about failing versus succeeding did not encourage participants to genuinely think about their own previous failures and successes. The current study attempted to address this issue by employing a similar method and procedure to that of the previous study, except that instead of using generic words referring to 'failures' and 'successes', ideographic IRAPs were created based on the answers given by participants on the questionnaire that assessed behaviour problems (see Appendix 3).

In the research reported here participants were first asked to provide examples of failures in their own lives, and these were subsequently inserted into individualized IRAPs that again targeted feelings and outcomes. The use of individualized IRAPs constituted a specific attempt to enhance the salience and meaning of the failures targeted by the IRAPs, and provides the first study in which non-generic, or ideographic IRAPs were employed (but see Vahey, Barnes-Holmes, Barnes-Holmes & Stewart, 2009, for an IRAP study in which participants' own names were used as response options). In one sense, therefore, the current study could be seen as a partial replication of the previous study but adopting ideographic IRAPs so that the failure-related stimuli presented in the implicit measure might evoke

stronger emotional responses. Would adopting this strategy produce intuitively predictable IRAP effects on the *Failure-Positive* trial-types (negative rather than positive biases)? Furthermore, would employing ideographic IRAPs produce any correlational evidence that the implicit measures were predicting criterion variables that may be related to the psychological domain of self-forgiveness?

Method

Participants

Participants were recruited from two different cities and countries. Initially, we were interested in exploring potential differences between two different cultures (Brazil and Ireland), but preliminary data analyses indicated that no significant group differences emerged, and thus all analyses reported subsequently were conducted without regard to site. The sample included 21 undergraduate and postgraduate students from Maynooth University; aged between 18-35 years, 9 males and 12 females, and 23 postgraduate students from Nucleo Paradigma Analise do Comportamento, Sao Paulo, Brazil; aged between 18-35 years, 10 males and 13 females. Participants from Ireland were recruited via class announcements, from various departments at Maynooth University, and participants from Brazil were recruited via class announcements at Nucleo Paradigma and e-mails. Of the 44 volunteers, only 26 participants completed the study, 13 from each site, with the remaining participants failing to meet the performance criteria on one or both IRAPs (see *Procedure* section). The attrition rate was approximately 40% (±20% for each sample). Although relatively high, some IRAP studies have recorded rates of up to 50% rates of attrition (see Nicholson, Doyle, Barnes-Holmes, & Roche, 2014). As an aside, since the current work was conducted, a meta-analysis has indicated that a sample of 29 participants or more is required for first-order correlations to achieve a statistical power of .80 when testing the criterion validity of clinically-focused IRAP effects (Vahey, Nicholson, & Barnes-Holmes, 2015). Given that

achieving power at .80 is an ideal rather than a strict requirement, the current data set (with 26 participants) would not be considered particularly underpowered.

Participants completed the current study on a voluntary basis and no payment or course credits were exchanged for participation, but volunteers were offered a chocolate bar before leaving the laboratory or experimental room.

Setting, Apparatus, and Materials

Participants completed the study in either an experimental cubicle or an office. The implicit measure was presented to each participant on a laptop using the IRAP 2009 program.

Explicit Measures. All the questionnaires were presented in hard copy format. These included the Problem Behaviours Questionnaire (designed specifically for the present study) that asked participants to identify currently problematic behaviours; these then formed the basis for the stimuli presented in each individualized IRAP. Participants also completed two scales used in the previous study, the Self-Compassion Scale (SCS; Neff, 2003; translation and adaptation by Castilho & Pinto-Gouveia, 2006) and the Depression Anxiety and Stress Scale (DASS; Lovibond & Lovibond, 1995; translation and adaptation by Pais-Ribeiro, Honrado, & Leal, 2004). The scales were presented in English for all participants in Ireland and in Portuguese for all participants in Brazil.

The Problem Behaviours Questionnaire (PBQ). This questionnaire was used to identify behaviours that participants felt were problematic and would like to reduce, avoid or change (see Appendix 3). The questionnaire consisted of 12 items, preceded by the following instruction:

> "*The questions below are brief and are designed to gauge your experiences of behaving in ways that you don't want, didn't plan to, or don't like. For example, you might find that you can't resist sweet things when you're on a diet. You promise yourself that you won't have it*

before, but when the opportunity presents itself, you just eat it anyway.

Then, maybe afterwards, you are filled with guilt and so you make the

same promise for the next time and hope that on that occasion you might

be more successful."

The first item asked participants to identify a 'problem' behaviour they did not like or had promised they would try to reduce or avoid. If participants indicated the absence of any problem behaviours at present, they were thanked and debriefed, and had no further involvement in the study.

The second item sought one or two examples of the problem behaviour (subsequent questions pertained to this or related behaviours, but did not ask participants to specify exactly which example of the behaviour they were referring to). The third question assessed the frequency of such behaviours as "Daily", "Weekly", "Monthly", or "Rarely", while the fourth ascertained whether these behaviours were repetitive. The fifth question asked whether participants believed other people engage in broadly similar problem behaviours as "Never", "Hardly Ever", "Sometimes", "Frequently", or "Always". The sixth item assessed how participants felt during such behaviours (i.e., Hopeless, Angry, Stupid, Helpless, Regretful, Frustrated, Out of control, Weird, Others), while the next two questions determined what participants believed other people feel when engaging in problematic behaviours. The ninth item explored potential reasons why the behaviour continues, while Question 10 asked participants to indicate how hard they had tried to change from 0% (Not tried very hard) to 100% (Tried very hard). Question 11 asked participants if they thought they would be in the same behavioural trap forever, while the last question asked if participants had any idea of how their situation might be changed.

A single score for the PBQ was calculated by assigning points to the majority of the items from the questionnaire (see Appendix 3). The total scores could range from 3 to 17

points, with higher scores indicating that participants self-rated the problem behaviour as more severe. Severity was defined loosely in terms of frequency of occurrence, efforts made to change the behaviour, and how problematic participants reported the behaviour to be, relative to the problem behaviour of others. That is, the more frequent the behaviour, the greater the effort to change it, and the more problematic the behaviour was deemed relative to others -- the more severely "the problem" was categorized. The Portuguese version was translated from English to Portuguese by the author of this thesis and reviewed by another Brazilian researcher specialized in behaviour analysis, who was proficient in English and Portuguese (Dr. Regina Wielenska).

Implicit Measures. Implicit attitudes concerning feelings and outcomes related to each participant's problem behaviour were measured using two IRAPs, one designed to assess feelings and another designed to assess outcomes. As noted previously, the stimuli that were inserted into the IRAP were based, in part, on a participant's responses on the PBQ. Because the current research was focused on self-forgiveness, the questionnaire targeted so called problem behaviours that participants wished to stop or reduce in some way, which thus required some level of self-forgiveness when they failed to do so. In effect, the stimuli that were ideographically selected for the IRAP for each participant consisted of descriptors of the problematic or negative behaviours that they reported in the questionnaire. The implicit measures also required that non-problematic or positive descriptors of behaviours be inserted into the IRAPs, but the questionnaire did not explicitly ask participants to provide examples of these. Therefore, the researchers often simply inserted descriptors of behaviours that would be deemed the opposite of the problem behaviours; these were typically highly intuitive or obvious. For example, the opposite of eating unhealthy food would be eating healthy food; the opposite of drinking too much alcohol would be reducing alcohol. Some participants did in fact identify a "positive" behaviour in the questionnaire, which contrasted with their

problem behaviours, and when this occurred these were used to create the appropriate stimuli for the IRAP. For example, if a participant wrote for question 2 "I spend too much time on the internet, when I should focus on my essays", the contrast category for the IRAPs referred to studying as an example of a positive or non-problematic behaviour.

In creating stimuli that were deemed the opposite of the problem behaviours, the characteristics of the sample were also considered. For example, many of the participants were students and thus if "Facebook" was identified as a negative behaviour, "Studying" was used as a positive behaviour. Although no formal checks were used to determine if participants agreed that that the positive behaviours were indeed perceived to be opposite to the negative behaviours, it is important to note that completing an IRAP successfully requires that participants can discriminate relatively easily between the two categories of stimuli that are employed.

The complete list of stimuli employed for each participant in the current study is available in the Appendix 4.

The Feelings-IRAP. As noted above, the two label stimuli that were inserted into the Feelings-IRAP were based on each participant's responses to the PBQ. For example, for one participant the IRAP juxtaposed spending too much time on Facebook rather than studying. In this case, the label stimuli were "Facebook makes me feel" and "Studying makes me feel". The target stimuli presented in all of the Feelings-IRAPs consisted of six negative emotions (i.e., bad/mal; guilty/culpado; stupid/estúpido; useless/inútil; frustrated/frustrado; and angry/nervoso) and six positive emotions (i.e., good/bem; strong/forte; wise/ sábio; in control/no controle; calm/calmo; and peaceful/ em paz). Each Feelings-IRAP comprised four possible label-target combinations: Negative Action-Negative Feeling; Negative Action-Positive Feeling; Positive Action-Negative Feeling; and Positive Action-Positive Feeling.

Participants responded to these with one of two response options "True" or "False", operated through the "d" and "k" keys.

Responses on each IRAP trial were defined as either consistent or inconsistent with the participants' responses to the PBQ. Consider again the participant who identified spending too much time on Facebook as problematic. When the IRAP trial presented the label "Facebook makes me feel" with the target "Guilty", choosing the response option "True" would be defined as a questionnaire-consistent response, but choosing "False" would be defined as a questionnaire-inconsistent response (see Figure 6 for a schematic representation of an example of the Feelings IRAP).

Facebook makes me feel		Studying makes me feel	
Guilty		Wise	
Questionnaire-Consistent	Questionnaire-Inconsistent	Questionnaire-Consistent	Questionnaire-Inconsistent
Select 'd' for	Select 'k' for	Select 'd' for	Select 'k' for
↓	↓	↓	↓
True	False	True	False
Facebook makes me feel		Studying makes me feel	
Wise		Guilty	
Questionnaire-Consistent	Questionnaire-Inconsistent	Questionnaire-Consistent	Questionnaire-Inonsistent
Select 'd' for	Select 'k' for	Select 'd' for	Select 'k' for
↓	↓	↓	↓
False	True	False	True

Figure 6. A schematic representation of an example of the four trial-types from The Feelings IRAP

The Outcomes-IRAP. The two label stimuli that were inserted into the Outcomes IRAP were also based on each participant's responses to the PBQ. Using the previous example, the two label stimuli for the participant were "Facebook" and "Studying". The target stimuli in the Outcomes-IRAP were 12 outcomes arising from these two activities. Six referred to negative outcomes (i.e., wastes my time/perco tempo; undermines my confidence/fico menos confidante; undermines my success/sou mal sucedido; makes me less focused/fico menos focado; makes me lazy/torno-me preguiçoso; reduces my concentration/reduzo minha concentração). The remaining six stimuli referred to positive outcomes (i.e., a good use of my time/aproveito meu tempo; increases my confidence/aumento minha confiança; increases my success/aumento meu sucesso; helps me focus/fico focado; makes me productive/torno-me produtivo; helps my concentration/melhoro minha concentração).

The Outcomes-IRAP comprised of four possible label-target combinations: Negative Action-Negative Outcome; Negative Action-Positive Outcome; Positive Action-Negative Outcome; and Positive Action-Positive Outcome. Once again, responses on each IRAP trial were defined as either consistent or inconsistent with each participant's responses to the PBQ. Thus, given the current Facebook example, if the IRAP presented "Facebook" with "wastes my time", choosing "True" was defined as a questionnaire-consistent response, but choosing "False" was defined as questionnaire-inconsistent.

Procedure

The procedure was very similar to the previous experiment, except that participants were given all the questionnaires first and as soon as these were completed (in approximately 10-15 mins), a second session was scheduled for at least two days later thereby providing sufficient time for the researcher to read the questionnaires and prepare the ideographic IRAPs. During the next session, participants completed the two IRAPs, taking approximately

45-60 minutes. The order of the presentation of the Feelings- and Outcomes-IRAPs was counterbalanced between participants. All participants commenced both IRAPs with a questionnaire-consistent block of trials because previous studies have typically found little evidence for block-order effects (see Barnes-Holmes, Barnes-Holmes, Stewart, & Boles, 2010). The criteria for the IRAP was the same as for the previous experiment (e.g., 80% and 2000 ms).

In Block 1, and all subsequent *odd* numbered blocks of the Feelings-IRAP, participants were required to respond in a pattern that was questionnaire-consistent (e.g., responding "True" to "Facebook makes me feel guilty"). In Block 2, and all subsequent *even* numbered blocks of the Feelings IRAP, participants were required to respond in a questionnaire-inconsistent pattern (i.e., responding "False" to "Facebook makes me feel guilty"). Similar patterns were required for the Outcomes IRAP (e.g., responding "True" to "Facebook wastes my time" across odd numbered blocks, but responding "False" to this question across all even numbered blocks).

Results and Conclusion

Explicit Measures

The Problem Behaviours Questionnaire. This questionnaire was used primarily to identify problem behaviours that could then be used to create IRAP stimuli specific to the self-identified problem behaviours of individual participants. As noted in the Method section, a total score was derived from the questionnaire as a metric of the severity of the behaviours thus identified. The mean overall score for the sample was 9.93 ($SD = 2.413$).

Self Compassion Scale. According to Neff (2003), average overall self-compassion scores tend to be around 3.0 on the 1-5 scale, so the overall score can be interpreted accordingly: An overall score of 1-2.5 indicates low self-compassion; 2.5-3.5 indicates

moderate self-compassion; and 3.5-5.0 indicates high self-compassion. In the present study, we found similar results to the previous research described in the chapter 3 in which the mean overall score was 3.05 (minimum = 1.87 and maximum = 4; SD = .76); an average score indicating moderate self-compassion for the current sample.

Depression Anxiety Stress Scale. The interpretation of the DASS is based primarily on the use of cut-off scores. Lovibond & Lovibond (1995) presented severity ratings from 'normal' to 'extremely severe' on the basis of percentile scores, with 0–78 classified as 'normal', 78–87 as 'mild', 87–95 as 'moderate', 95–98 as 'severe', and 98–100 as 'extremely severe'. In the current study, the mean overall score for the DASS was 25.74 (minimum = 5 and maximum = 72, SD = 19.11), indicating that the sample fell well below the cut-off between 'normal' and 'mild' severity.

Implicit Measures

Data Preparation. The data preparation was done following the same criteria from the previous study reported in Chapter 3. Participants were required to maintain an accuracy level ≥ 75% correct and a median latency ≤ 2000 ms on two of the three successive pairs of the six test blocks. If a participant maintained the criteria across all six blocks, all of the data were used to calculate the *D*-IRAP scores. 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. This practice was adopted for 6 participants with the Feelings-IRAP and 7 participants with the Outcomes-IRAP (for three of the seven participants this practice was applied to both IRAPs).

Preliminary analyses of variance indicated no significant effects arising from the procedural variable of counterbalancing the order in which the two IRAPs were presented, and thus this variable was removed from all subsequent analyses (*ps* > .21 for the Feelings-IRAP and *ps* > .60 for the Outcomes-IRAP).

The Feelings-IRAP

Mean Scores Analyses. The four overall mean *D*-IRAP scores calculated across participants are presented in the Figure 7 (upper panel). The letters "T" and "F" in the Figure indicate the direction of the IRAP effect in terms of responding "True" or "False", respectively. Thus, for example, the IRAP effect for the *Negative-Action/Negative-Feelings* trial-type indicates that participants tended to respond *True* more quickly than *False,* whereas the opposite was the case for the *Positive-Action/Negative-Feelings* trial-type (i.e., responding *False* more quickly than *True*).

Figure 7 for the Feelings-IRAP shows that relatively clear positive biases were revealed for the two *Positive-Action* trial-types, but this was not the case for the two *Negative-Action* trial-types, one of which showed a negative feelings bias and the other a near zero effect. The *D*-IRAP scores for the four trial-types were entered into a one-way repeated measures analysis of variance (ANOVA), and this yielded a significant effect, $F(3, 26) = 6.00, p = .001, \eta_p^2 = .187$. Fisher's PLSD post-hoc tests yielded three significant differences ($ps < .009$); two between the *Negative-Action/Negative-Feeling* trial-type and the two *Positive-Action* trial-types, and a third between the *Negative-Action/Positive-Feeling* and the *Positive-Action/Positive-Feeling* trial-types. When each of the four trial-type scores was subjected to one-sample *t*-tests, only the two *Positive-Action* trial-type scores proved to be significantly different from zero ($ps = .03$). The statistical analyses thus indicated that the participants showed implicit biases towards confirming that their positive actions tended to elicit positive feelings; the effects for the two *Negative-Action* trial-types were less clear cut, although the *Negative-Action/Negative-Feeling* trial-type differed significantly from the two *Positive-Action* trial-types.

FEELINGS IRAP
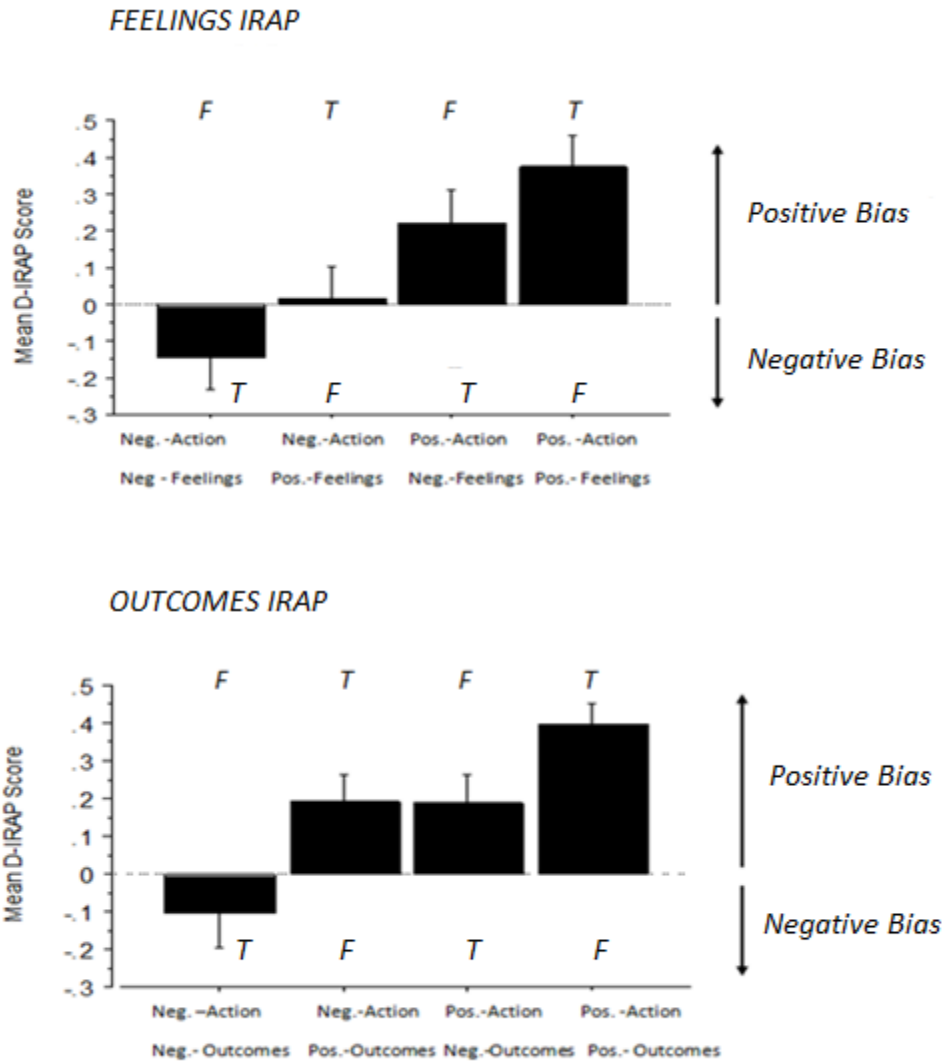
OUTCOMES IRAP

Figure 7. Mean *D* IRAP-Trial-Type Scores obtained for the Feelings and Outcomes
IRAPs. The letters "T" and "F" indicate the direction of the IRAP effect in
terms of responding "True" or "False", respectively.

The Outcomes-IRAP

    Mean Scores Analyses. The four overall mean *D*-IRAP scores calculated across

participants are presented in the Figure 7 (lower panel). Positive biases were revealed for the

two *Positive-Action* trial-types and for the *Negative-Action/Positive-Outcome* trial-type; a

relatively small negative bias was revealed for the *Negative-Action/Negative-Outcome trial-*

type. The *D*-IRAP scores for the four trial-types were entered into a one-way repeated

measures ANOVA, which proved to be significant, $F(3, 26) = 6.50$, $p = .0005$, $\eta_p^2 = .218$.

Fisher's PLSD post-hoc tests indicated that the *D*-IRAP scores for the *Negative-Action/Negative-Outcome* trial-type differed significantly from the other three trial-types (*ps* <.02). The two comparisons between the *Positive-Action/Positive-Outcome* trial-type and the *Negative-Action/Positive-Outcome* and *Positive-Action/Negative-Outcome* trial-types approached significance (*ps* < .08). When each of the trial-type scores was subjected to one-sample *t*-tests, only the effect for the *Negative-Action/Negative-Outcome* trial-type failed to reach significance (remaining three *ps* < .02). The statistical analyses thus indicated that the participants showed implicit biases towards confirming that their positive actions tended to produce positive outcomes; interestingly the effect for the *Negative-Action/Positive-Outcome* trial-type yielded a similar result (i.e., a positive bias). Similar to the Feelings IRAP, however, the *Negative-Action/Negative-Outcome* trial-type yielded a negative but relatively weak effect.

Implicit-Explicit Correlations

   The Feelings-IRAP and Explicit Measures. The four *D*-IRAP scores were entered into a correlation matrix with each of the three explicit measures. The total and subscale scores for the SCS and DASS were entered into the matrix. Of the 48 correlations (12 correlations for each trial-type of the IRAP) none proved to be significant.

   The Outcomes IRAP and Explicit Measures. When the four trial-type scores for the Outcomes IRAP were entered into a correlation matrix three of the 48 correlations proved to be significant. Specifically, the greater the negative outcomes bias on the *Negative-Action/Negative-Outcome* trial-type, the higher the Depression (*r* = -.40, *p* = .03), Stress (*r* = -.65, *p* = .0001); and total DASS scores (*r* = -.526, *p* = .004). In effect, the stronger participants responded to negative actions as producing negative outcomes, the higher the level of self-reported depression, stress, and general psychological suffering.

Conclusion

In terms of the direction and strength of the IRAP effects obtained in the current study, similar to the previous experiment, negative biases for the *Failure-Positive* trial-types were found. In this sense, employing ideographic stimuli in the IRAPs did not change the counter-intuitive result reported previously. Possibly, the counter-intuitive nature of the IRAP data and the lack of correlations between the Problem Behaviour Questionnaire and the IRAPs might be related to the generally trivial or moderate nature of the problem behaviour reported on the questionnaire (e.g., watching television, breaking a diet, and so on). In other words, perhaps many if not most participants, who were selected from a normative sample of the population, simply reported relatively minor "problem" behaviour in the context of the current study, rather than disclosing something more serious (if indeed there was something more serious to disclose). Thus, when participants completed either the Feelings or Outcomes IRAPs these failed to evoke relatively negative biases because the specified problem behaviour was simply too minor or trivial. In an effort to circumvent this potential problem a type of "priming" procedure was employed in the next study in which participants were asked to write down three situations related to failure or success depending on the group conditions that they were assigned before doing the same IRAPs and the other explicit measures described in Chapter 3. However, in order to assure that participants would feel comfortable in reporting their various personal circumstances, the experimenter and participants agreed before the experiment that the paper in which they would write down their personal scenarios would be shredded in the experimental setting, this assuring that no one could know what had been written down.

On balance, it is worth noting that significant correlations were found between performance on the Outcomes IRAP and psychological suffering as measured by the DASS. Although the number of correlations was small relative to the size of the matrix, they were

specific to one trial-type; that is, the more strongly participants responded to negative actions as producing negative outcomes, the higher the level of self-reported depression, stress, and general suffering. Of course, this result could still be due to error variance because so many correlations were conducted. But it does suggest, if only tentatively, that the Outcomes IRAP was tapping into something important, even if many of the participants were reporting relatively trivial problem behaviours in the initial questionnaire. It is also worth emphasizing that the correlations between the DASS and the Outcomes IRAP were obtained using ideographic stimuli, which suggests that although different stimulus sets were used across participants the IRAPs were tapping into a broadly similar response class.

In continuing with the current programme of research there appear to be many lines of potential inquiry. A strategy adopted in the next study reported in the Chapter 5 was to ask participants to think about some examples of failures or successes in their lives before completing the IRAPs to determine if this impacts on their performances and correlations with explicit measures. For example, would implicit self-forgiveness increase or decrease if participants had just spent a few minutes beforehand contemplating some examples of failures in their own lives?

Chapter 5

Priming Thoughts of Failing versus Succeeding and Performance on the Implicit Relational Assessment Procedure (IRAP) as a Measure of Self-Forgiveness

Introduction

In the previous study reported in Chapter 4, one of the possible reasons for the counter-intuitive nature of the IRAP data and the lack of correlations between the Problem Behaviour Questionnaire and the IRAPs was that it could be related to the generally trivial or moderate nature of the problem behaviour reported on the questionnaire (e.g., watching television, breaking a diet, and so on). In other words, perhaps many if not most participants, who were selected from a normative sample of the population, simply reported relatively minor "problem" behaviour in the context of the current study, rather than disclosing something more serious (if indeed there was something more serious to disclose). Thus, when participants completed either the Feelings or Outcomes IRAPs these failed to evoke relatively negative biases because the specified problem behaviour was simply too minor or trivial.

One way in which to encourage participants to think about previous failures (or successes) would be to present them with a type of priming task in which they are required to spend some time thinking about relevant examples from their own lives. The current study adopted this approach; specifically, participants were "primed" before exposure to the IRAPs to reflect upon either previous failures or successes in their personal lives. To avoid possible problems concerning disclosure, although participants were required to write down examples of failures or successes, this material was destroyed before they left the experiment (i.e., no one but the participant knew what he or she had written). In effect, the current 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.

Given the assumption that historical and current contextual variables have been shown to impact upon IRAP performances (e.g., Cullen, Barnes-Holmes, Barnes-Holmes, & Stewart, 2009), it was predicted that priming participants to contact their personal

assessments of failures or successes would influence the IRAP performances in some way. Given the exploratory nature of the current research, however, explicit predictions were not made.

Method

Participants

One hundred undergraduate and post-graduate students, aged between 18- 45 years (M=23), 44 female and 37 male, were recruited via class announcements, from various departments at the National University of Ireland Maynooth (NUIM), and completed the current study on a voluntary basis. Participants were divided randomly in two groups (50 in each group) with one group being designated a "Positive Priming" group and the other a "Negative Priming" group. Nineteen of the 100 participants were excluded because they did not achieve the IRAP performance criteria (see *Procedure* section). Twelve of the excluded participants were from the Positive Priming group and seven were from the Negative Priming group. No payment or course credits were exchanged for participation, but volunteers were offered a chocolate bar before leaving the laboratory.

Setting, Apparatus, and Materials

The setting, apparatus and materials were replicated from the study described in the Chapter 3.

Implicit measures. Each participant was required to complete the same two IRAPs used in the previous study reported in Chapter 3, one designed to target feelings and one that targeted expected outcomes arising from failing and succeeding. The stimuli inserted into the Feelings and Outcomes-IRAP consisted of combinations of statements pertaining to feelings arising from failing *versus* succeeding (see Chapter 3).

Explicit measures. The explicit measures used here were the same used in the study reported in the Chapter 3, two measures were derived from the stimuli used with the IRAPs

and the two other measures were standardized psychometric instruments targeting self-compassion (Self-Compassion Scale, SCS; Neff, 2003) and depression, anxiety and stress levels (Depression Anxiety and Stress Scale, DASS; Lovibond & Lovibond, 1995).

Procedure

The procedure was almost a replica from the study reported in Chapter 3, except for the priming procedure. After completing consent forms, participants received a 5-minute task related to positive or negative priming, and were then asked to complete the same IRAPs, followed by the same explicit measures employed in Chapter 3. The order in which the two IRAPs were presented was counterbalanced across participants. Each session took approximately one hour: 45 minutes to complete the priming procedure and both IRAPs and 15 minutes to complete the explicit measures.

Priming. Participants in the positive priming group were given a pen and paper and were asked to write down in as much detail as they could remember three situations from their past in which they had been successful or had achieved something of which they were particularly proud. Each participant was allowed approximately 5 minutes to complete this writing task. Participants assigned to the Negative Priming group were given a similar task but were asked to use three situations from their past in which they had failed or not achieved something that was important to them. Participants were told that they should feel free to write anything that came to mind and were reassured that the experimenter would not read what they wrote and would shred the paper in front of them as soon as they finished writing. It was assumed this would encourage participants to reflect more openly on their successes or failures because they would not be required to disclose the details to the experimenter or anyone else. To facilitate a feeling of "privacy" or non-disclosure all participants completed the priming task individually in an experimental cubicle. For ethical reasons, the positive priming task was given to participants in the Negative Priming group at the end of the

experiment, that is, when participants had completed the two IRAPs and all of the questionnaires.

Results and Conclusion

Preliminary data analyses indicated no significant effects arising from the procedural variable of counterbalancing the order in which the two IRAPs were presented and, thus, this variable was removed from all analyses. The data preparation for Feelings and Outcomes-IRAP was identical of the study reported in the Chapter 3. If a participant maintained the criteria across all six blocks, all of the data were used to calculate the D-IRAP scores. 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. This practice was adopted for 10 Feeling-IRAPs related to the Positive Priming group and 4 Feeling-IRAPs related to the Negative Priming Group and 8 Outcome-IRAPs from Positive Priming group and for 6 Outcome-IRAPs from the Negative Priming group.

Feelings-IRAP

Mean scores analyses. The four overall mean D-IRAP scores calculated across participants, for the positive and negative priming groups, are presented in the upper left panel of Figure 8. The direction and relative size of the D-IRAP scores did not differ markedly across the two priming conditions. In both cases, a relatively strong positive bias was revealed for the *Success-Positive feelings* trial-type, but this was not the case for the remaining trials-types -- for *Success-Negative* and *Failure-Positive* the effects indicated weaker positive bias, and for *Failure-Negative* the effects were again relatively weak but in a negative direction.

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 non-significant main

effect for priming (p =.35). The main effect for trial-type was significant, $F(3, 79) = 21.881$, p .<.0001, $\eta_p^2 = .665$, but the interaction with the priming variable was not (p = .47). When Fisher's PLSD post-hoc tests were applied to the differences among the trial-types (collapsed across the two priming conditions) they yielded five significant differences ($ps < .002$) among the four trial types, with only the *Failure-Positive* versus *Success-Negative* trial-type comparison producing a non-significant effect ($p = .34$). When each of the eight trial-type scores for the positive and negative priming groups was subjected to one-sample t-tests, three of the tests yielded significance ($ps < .02$) for the Negative Priming group (the *Failure-Positive feelings* trial-type was marginally significant, $p = .07$). For the Positive Priming group, only one trial type, *Success-Positive feelings,* differed significantly from zero (p <.0001).

Outcomes-IRAP

Mean scores analyses. The four overall mean D-IRAP scores calculated across participants are presented in Figure 8 (upper right panel). The general pattern of biases observed for the Outcomes-IRAP differed somewhat from the patterns observed for the Feelings-IRAP. Specifically, the D-IRAP effects for the *Failure-Negative* and *Success-Negative* trial-types did not differ greatly between the two priming conditions (similar to the Feelings-IRAP), but the effects for the *Failure-Positive* and *Success-Positive* trial-types were markedly different; for the former trial-type positive priming produced a far stronger D-IRAP score than negative priming but the reverse was true for the latter trial-type.

When the D-IRAP scores were entered into a two-way mixed repeated measures ANOVA, it yielded a non-significant main effect for priming ($p = .55$), but a significant effect for trial type, $F(3, 237) = 28.64$, p <.0001, $\eta_p^2 = .27$, and critically a significant interaction between priming and trial-type, $F(3, 237) = 3.99$, $p = .008$, $\eta_p^2 = .05$. Given the

significant interaction, four follow-up between-groups ANOVAs were used to determine if any of the trial-types yielded significant differences between the two priming conditions. The *D*-IRAP effects did not differ significantly for the *Failure-Negative* (p = .39) and *Success-Negative* (*p* = .94) trial-types, but significant differences were obtained for the *Failure-Positive, F*(1, 79) = 6.89, *p* = .01, η_p^2 = .08, and *Success-Positive* trial-types, *F*(1, 79) = 5.67, *p* = .02, η_p^2 = .07. When each of the eight trial-type scores for the two priming groups was subjected to one-sample *t*-tests, only the *Failure-Negative* trial-type failed to reach significance for the positive priming condition (ps ≤ .02 for the remaining three trial types). For the negative priming group, only the *Failure-Positive* trial-type was non-significant (*ps* ≤ .02 for the remaining three trial types). In summary, the priming variable appeared to impact upon performances on the Outcomes-IRAP, in a manner not observed for the Feelings-IRAP. Specifically, negative priming appeared to weaken positive bias on the *Failure-Positive* trial-type relative to the *Success-Positive* trial-type, but the bias scores for the positive priming condition were quite similar across the two trial-types.

Explicit Measures

   The scale based on the Feelings-IRAP. The overall mean ratings obtained from the Explicit-Feelings scale are presented in Figure 8 (lower left panel). In general, similar to the IRAP, the two subscales that mapped onto the two 'Success' trial types produced positive bias with the negative priming condition producing stronger effects than the positive priming condition. Unlike the IRAP, however, the effects were slightly stronger for the *Success-Negative* than for the *Success-Positive* subscale. The two subscales that mapped onto the two 'Failure' trial types yielded negative bias for both priming conditions for the *Failure-Positive* relation; for the *Failure-Negative* relation positive priming also produced a negative bias, whereas negative priming produced a positive but weak effect.

The mean rating scores for each participant from the Explicit-Feelings scale were entered into a two-way mixed repeated measures analysis of variance (ANOVA), and this yielded a significant effect for priming, F (1, 79) = 12.87, p = .0006, η_p^2 = 0.14, and a significant effect for trial-type, F (3, 237) = 137.4, p < .0001, η_p^2 = 0.63, but the interaction was non-significant (p > 1.4). Four follow-up between-groups ANOVAs indicated that each of the bias scores for the negative priming condition produced more positive (or less negative) bias than the positive priming condition for three of the scales (ps < .03), with the effect for the *Success-Positive* subscale approaching significance (p = .08). Six Fisher's PLSD post-hoc tests comparing the four sub-scales with each other (data collapsed across priming conditions) indicated that all of the ratings differed significantly from each other (ps ≤ <.02). When each of the eight subscale scores for the positive and negative priming groups was subjected to one-sample t-tests, the ratings for each subscale differed significantly from zero (ps < .0001) for the positive priming group, but for the negative priming group only the *Success-Negative and Success-Positive* subscales were significant (p < .0001; two remaining ps ≤ .51).

The scale based on the Outcomes-IRAP. The data from the explicit measure that was derived from the Outcomes-IRAP were used to calculate eight separate scores, with each score mapping onto the equivalent of one of the trial types from the IRAP. The data were transformed in the same way as for the Explicit-Feelings scale, and are presented in Figure 8 (lower right panel).

Similar to the Outcomes-IRAP, the two subscales that mapped onto the two 'Success' trial types produced positive bias, with the negative priming condition producing stronger effects than the positive priming condition. Unlike the IRAP, however, the difference between the two sub-scales was not particularly large. The two subscales that mapped onto the two 'Failure' trial types yielded weak positive bias for the negative priming condition.

For the positive priming condition the effects were in opposite directions; positive for the *Failure-Negative* subscale and negative for the *Failure-Positive* subscale.

When mean rating scores for each participant were entered into a two-way mixed repeated measures ANOVA, this yielded significant main effects for priming condition, $F(1, 79) = 6.82$, $p = .01$, $\eta_p^2 = 0.07$ and trial-types, $F(3, 237) = 424.76$, $p < .0001$, $\eta_p^2 = 0.30$ and a significant interaction, $F(3, 237) = 2..94$, $p = .03$, $\eta_p^2 = 0.03$. Given the significant interaction, four follow-up between-groups ANOVAs were used to determine if any of the trial-types yielded significant differences between the two priming conditions. The subscale ratings did not differ significantly for the *Failure-Negative* ($p = .81$) and *Success-Negative* ($p = .11$) subscales, but significant differences were obtained for the *Failure-Positive*, $F(1, 79) = 7.34$, $p = .008$, $\eta_p^2 = .085$, and *Success-Positive* subscales, $F(1, 79) = 5.37$, $p = .02$, $\eta_p^2 = .06$. When each of the eight subscales for the positive and negative priming groups was subjected to one-sample t-tests, seven of the ratings differed significantly from zero ($ps < .05$); the ratings for the *Failure-Positive* subscale in the negative priming condition was non-significant ($p > .76$).
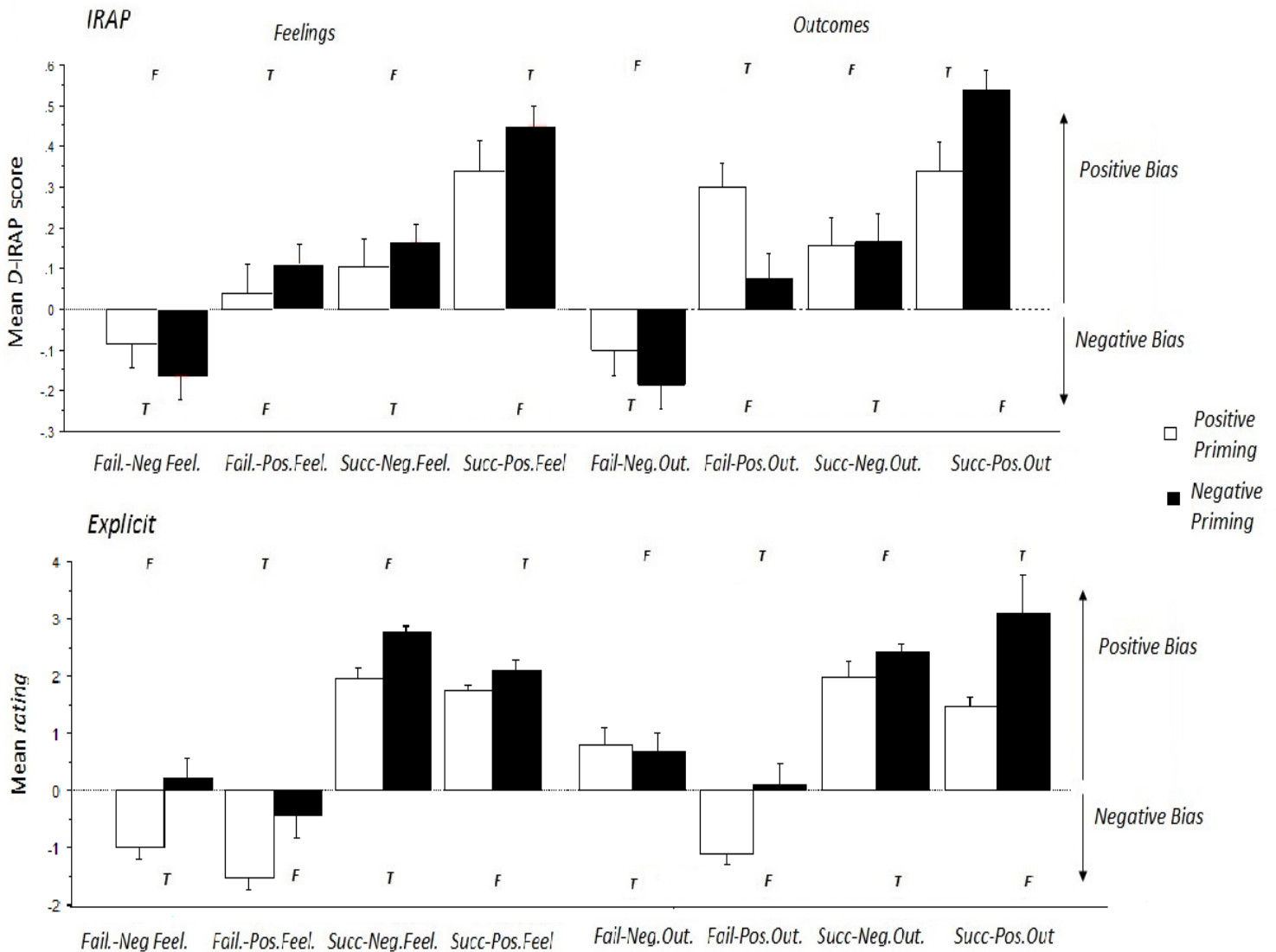
Figure 8. Mean D $_{\text{IRAP-p-Trial-Type Scores}}$ obtained on the IRAP and the mean rating obtained on the explicit measures for feelings and outcomes related to failing and succeeding, according to the priming conditions. The letters 'T' and 'F' indicate the direction of the response biases ('True' and 'False', respectively) that were recorded by the measures.

Self compassion scale (SCS). In the present study, both priming conditions produced overall mean self-compassion scores that fell in the moderate range (positive priming = 3.39, *SD* = .87; negative priming 2.80, *SD* = .79). An independent *t*-test indicated that the difference between the two priming conditions was significant, *t* = 3.08, *p* = .002, suggesting that positive relative to negative priming increased levels of self-compassion.

Depression, Anxiety and Stress Scale (DASS). In the current study, the mean overall score for the DASS between the two groups was similar, 28.497 (*SD* = 25.58) for the positive

74

priming condition and 24.84 (SD = 20.473) for negative priming, indicating that both samples fell well below the cut-off between 'normal' and 'mild' severity. An independent *t*-test indicated no significant difference between the two priming conditions for the overall DASS measure, nor for the three subcategories ($ps \geq .22$)

Implicit-Explicit Correlations

The Feelings-IRAP and explicit measures. The four *D*-IRAP scores were entered into a correlation matrix with the scores from each of the three explicit measures (the scores obtained from *SCS, DASS and* the scale based on the Feelings-IRAP) for the both positive and negative priming groups.

The Feelings IRAP and SCS. Of the 8 correlations, 4 for the positive priming and 4 for the negative priming group, just one proved to be marginally significant (for positive priming): *Fail-Positive Feelings IRAP* trial type with *Self-Compassion* ($r = -.316, p = .053$). That is, a lower level of self-compassion predicted a bias towards confirming that failing produces positive feelings (but only after completing a positive priming exercise).

The Feelings IRAP and DASS. Of the 16 correlations for the positive priming group (the four trial-types with the four DASS scores), three correlations proved to be significant (or marginally so)*; Fail-Negative Feelings* and *Anxiety* ($r = .31, p = .06$), *Fail-Negative Feelings* and *Stress* ($r = .4, p < .01$), and *Fail Negative Feelings* and overall *DASS* ($r = .32$, p = .04). In each case, therefore, higher levels of self-reported psychopathology predicted reduced levels of negative bias on the IRAP for the *Failure-Negative* trial-type. Of the 16 correlations obtained for the negative priming group none of the correlations were significant ($rs < .26, ps > .9$). The priming variable thus appeared to impact upon the relationship between the Feelings IRAP and the explicit measure of psychopathology.

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 sub-scale of the explicit measure. Of the eight correlations (four for the positive priming group and four for the negative priming group), none of them proved to be significant. The correlation for the *Fail-Negative Feeling* correlation for the negative priming group did approach significance, however ($r = 2.8$, $p = .07$).

  The Outcomes-IRAP and explicit measures. The four *D*-IRAP scores were entered into a correlation matrix with the total and subscale scores of the three explicit measures (the *SCS, DASS* and the Explicit Outcomes-Scale). None of the 48 correlations across the explicit measures proved to be significant.

Conclusion

  In terms of the mean trial-type scores recorded with each IRAP, the priming condition appeared to affect performance on the Outcomes-IRAP but not performance on the Feelings-IRAP**.** Specifically, significant differences emerged between the two priming conditions for the Failure- and Success-Positive trial-types, with a weaker bias towards confirming that failing produces positive outcomes but a stronger bias towards confirming that succeeding produces positive outcomes in the negative priming condition.

  The effects for the two (explicit measure) scales that were derived from the IRAPs yielded results that were relatively consistent across the two priming conditions. Exposure to positive relative to negative priming appeared to increase negative bias for the two *Failure* sub-scales and decrease positive bias for the two *Success* sub-scales, with the clear exception of the *Fail-Negative* scale for the Outcomes measure, which yielded a non-significant difference across the two priming conditions. In other words, it appears that positive relative to negative priming led participants to rate failing more negatively and rate succeeding less positively.

  Another possibly interesting finding that emerged in the current study was a marginally significant correlation between the bias scores obtained on the *Fail-Positive*

Feelings IRAP trial type and the explicit measure of *Self-Compassion*. Specifically, a lower level of self-compassion predicted a bias towards confirming that failing produces positive feelings (but only after completing a positive priming exercise).

A possibly related finding was the pattern of correlations that was obtained for the positive priming group between the DASS scores and the *Fail-Negative* trial-type of the Feelings-IRAP. That is, 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 (for a detailed discussion about possible reasons why those pattern of correlations emerged both for SCS and DASS after the positive priming, see General Discussion, Chapter 7).

The findings from the current study indicated that manipulating the experimental histories of the participants with a relative brief priming procedure, impacted upon the IRAP effects. However, one empirical question relates to how the implicit attitudes of a group dealing with failures maintains in the long term? In other words, in contrast to the priming condition in which participants had to think about their failures briefly before responding on the IRAP, it may be interesting to ask how an extended history of dealing with personal impact upon implict attiudes? Thus, in the next and final study of the current research programme IRAP data were collected from participants who had undergone extensive training in clinical behaviour analysis, verses a control group who had not completed such training. It was expected that, as part of their training, clinical behaviour therapists would learn to acknowledge the negative feelings that failure may cause and to put things and events in a different persepctive, including their own failures and the failures of their clients. More specifically, the study aimed to determine if such pre-experimental behavioural histories impact in predictable and reliable ways on implicit self-forgiveness biases. If such effects are

found it may even be possible to begin to measure the relative impact of such training using implicit measures, such as the IRAP.

Chapter 6

The Implicit Relational Assessment Procedure (IRAP) as a

Measure of Self-Forgiveness: The Impact of a Training History in

Clinical Behaviour Analysis

Introduction

In the previous experiments, the studies were somewhat exploratory and the findings were revlealing divergences between implicit attitudes and explicit attitudes. One remarkable characteristic result found from the first to the third study, was the lack of correlation between the IRAP scores and the explicit scales. In particular, there was no correlation between the failure-positive feelings/oucomes tryal-type and the scale based on the IRAP One of the possible explanations given for this outcome was that the generic word "failure" failed to evoke strong relational responses due to the fact that in our culture people are not encouraged to think on a regular basis about their failures in comparison to their successes. Similarly, trivial failures described in Chapter 4 may also have failed to evoke strong relational responses for similar reasons, they could have not been strong enough or hidden a more important "behaviour problem". However, in the last experiment employing priming, participants were encouraged to think about failures just before the IRAPs and scales were administered. In this case, however, the confidentiality of the participant's private experience regaridng failure was assured and so it may have been more likely that participants actually did confront significant failures during that exercise.

As an exploratory piece of research, it was difficult to make very specific predictions regarding the outcomes of the tests and correlations among the measures. As was found in the previous study, some trial types correlated (e.g. Failure-Negative Feelings and Success Positive and Negative Feelings trial types, with the exception of the Failure-Positive Outcomes for the negative priming group), but Failure-Positive Feelings/Outcomes did not. Of particular interest here, is the differences in the strength of the biases observed across the two groups. For example, the Failure-Positive Feelings trial-type was non-significant for both groups, but for the Failure-Positive Outcomes trial-type only the positive priming group showed a significant effect. In other words, after thinking about their successes, a neutral bias

was found in relation to Failure-Positive Feelings trial-type, but a positive bias for Failure-Negative Outcomes trial. In addition, further correlations were found between the IRAP and self-compassion and psychopathology across the two priming conditions. These may be suggestive of patterns of experiential avoidance, and this matter will be returned to in the General Discussion.

At this point, it appears that the self-forgiveness IRAPs that have been developed across a series of studies may be sensitive to a specific behavioural history that was provided within the experimental context (i.e., positive versus 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 programme to deterimine if one or both of the self-forgiveness IRAPs would prove sensitive to a potentially important feature of the participants' pre-experimental history.

This line of inquiry was pursued by employing two groups – one who had undergone training in behaviour therapy and a group who had not. It was reasoned that training in most forms of psychological therapy may increase levels of compassion and forgivness 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 empathise and better understand a client's perspective on the problems they are presenting in therapy (e.g., Kohlenberg & Tsai, 1991; Tsai, Callaghan, Kohlenberg, Follette, & Darrow, 2009). 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 in Bast & Barnes-Holmes (2015) study, it might be predicted that somewhat similar patterns of results would be found using a non-behaviour therapist sample. For instance, a positive correlation was found between positive feelings and outcomes and success, as well as (counter intuitively) between failure and positive feelings. On the other hand, it might be predicted that the behaviour therapist group would exhibit some differences from these previous group based on their training. For example, it might be expected that a neutral bias for the trial types Failure-Negative (Positive Feelings and Outcomes) might be found, in contrast to previous groups (i.e, not denying that failure produces negative feelings and outcomes, but instead considering it as a learning opportunity that could facilitate approaching their 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, 8 were excluded because they did not achieve the IRAP performance criteria detailed in the procedure section and another 8 were eliminated due to a procedural error. The remaining sample of 20 participants consisted of four individuals who were currently pursuing a course in clinical behaviour analysis, 12 individuals who had completed the course within the previous two years and four individuals who were lecturers on the course. The course was designed to provide postgraduate training and education in Behaviour Analysis in

all its aspects: philosophy, theory and specifically in 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 behaviour 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 behaviour, 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's complaints; (b) the widespread (generalised) behavioural 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 behaviours in terms of maintaining the "problem" being reported and; (e) potential motivational variables for change.

The remaining 20 individuals were students from different fields (e.g. law, engineering etc) and they functioned as a control group. Hereafter, the first group will be referred to as the Behaviour Therapist (BT) group and the second as the Non-Therapist (NBT) group. Participants were between 18-32 years old ($M = 25$), 29 women and 11 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.

Setting, Apparatus and Materials

Implicit measures. Each participant was required to complete two IRAPs, one designed to target feelings and a second one that targeted expected outcomes arising from failing and succeeding. The stimuli inserted into the Feelings-IRAP were similar to those used in the previous study, except that they were translated to Portuguese by the researcher who and another Brazilian researcher who is also proficient in both languages (English and Portuguese) and it consisted of combinations of statements pertaining to feelings arising from failing *versus* succeeding. 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.

Explicit measures. The same two measures were derived from the stimuli used with the IRAPs (translated to Portuguese by the same researchers who translated the IRAPs) and the two other measures in the Portuguese version used in the study reported in the Chapter 4, targeting self-compassion (Self-Compassion Scale, SCS; translated and adapted by Castilho & Pinto-Gouveia, 2006) and depression, anxiety and stress levels (Depression Anxiety and Stress Scale, DASS; translated and adapted by Pais-Ribeiro, Honrado, & Leal, 2004).

Procedure

The procedure was the similar to that employed in the study reported in the Chapter 3; that is, 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.

Results and Conclusion

The data preparation for the Feelings and Outcomes-IRAP was identical to previous studies.

Feelings-IRAP

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 Figure 9. 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 *Fail-Negative, Fail-Positive,* and *Success-Positive* trial-types, whereas neutral, positive and weak biases were recorded for the non-therapists across these trial-types. Both groups produced positive biases on the remaining *Success-Positive* trial-type.

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 non-significant 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 each proved to be non-significant, 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 one was marginally significant, *Success-Negative*, $F (1, 38) = 3.590$, $p = .07$, $\eta^2 = .09$ (remaining $p > .27$). 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 non-significant effect for the comparison between the *Success-Positive* and *Success-Negative* trial-types ($p > 1.4$). For the NBT group three of the post-hoc tests were significant; *Failure-Negative* versus *Failure-Positive* ($p = .01$), *Failure-Negative* versus *Success-Positive* ($p < .001$), *Success-Negative* versus *Success-Positive* ($p = .006$). The remaining tests yielded non-significant results ($ps > .09$).

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 non-significant, $p = .75$). For the NBT group, two of the tests, *Failure-Positive* and *Success-Positive,* were significant ($ps < .004$; remaining $ps > .5$).

Outcomes-IRAP

Mean scores analyses. The four overall mean D-IRAP scores calculated across participants are presented in Fig. 9 (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-type 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 non-significant main effect for group ($p = .41$) and a non-significant 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 *Fail-Positive* versus *Success-Negative* comparison was non-significant ($p > .11$). 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 *Fail-Negative* and *Success-Positive* trial-types (remaining $ps > .16$), and three of the tests were significant ($ps < .04$) for the NBT group; *Fail-Positive* trial-type ($p = .10$). Overall, therefore, the inferential statistics supported the conclusions arising from the descriptive analyses of the data provided in Figure 9.
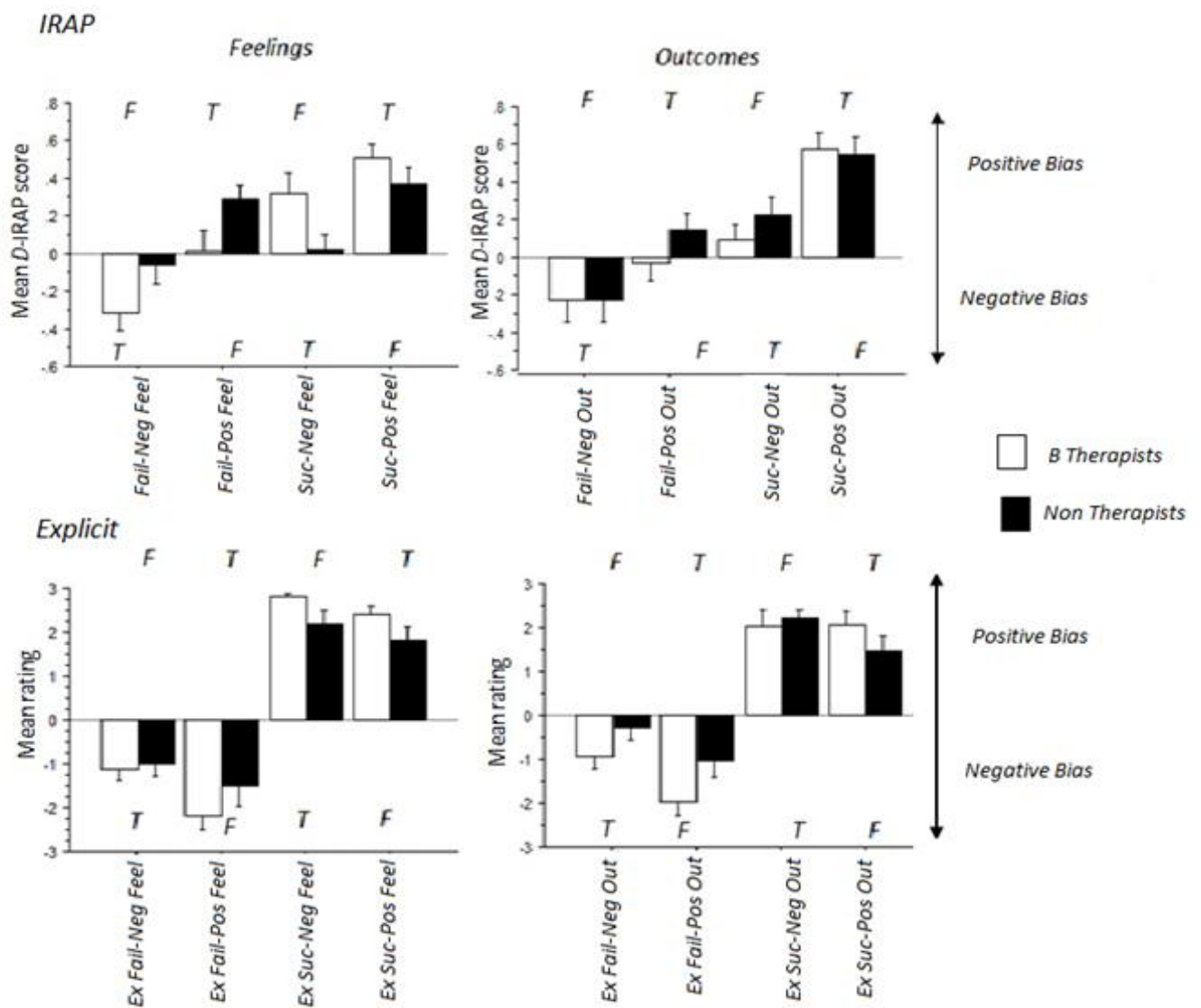
Figure 9. Mean D $_{\text{IRAP-p-Trial-Type Scores}}$ obtained on the IRAP and the mean rating obtained on the explicit measures for feelings and outcomes related to failing and succeeding, according to the groups. The letters 'T' and 'F' indicate the direction of the response biases ('True' and 'False', respectively) that were recorded by the measures.

The scale based on the Feelings-IRAP. The overall mean ratings obtained from the Explicit-Feelings scale are presented in Fig. 9 (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 and the two subscales that mapped onto the *Success* trial-types yielded positive 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 non-significant effect for group ($p > .5$) and for its interaction with trial-type ($p > .17$). 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 non-significant 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 were transformed in the same way as for the Explicit-Feelings scale, and are presented in Fig. 9 (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 effect for group and its interaction with trial-type were both non-significant ($ps > .16$); but the main effect for 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 non-significant 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). In the present study, both groups produced overall mean self-compassion scores that fell 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 relative to the NBT group possessed higher levels of self-compassion.

Depression, Anxiety and Stress Scale (DASS). In the current study, the mean overall score for the DASS between the two groups was similar, 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 cut-off between 'normal' and 'mild' severity. Independent *t*-tests yielded significant differences between the groups for the Anxiety and Stress sub-scales (*ps* < .05), a difference that approached significance on the overall DASS score (*p* = .10), but little evidence of any difference on the sub-scale for Depression (*p* ≥ .6).

Implicit-Explicit Correlations

The Feelings-IRAP and explicit measures. Two correlation matrices were created, one for each group of participants. 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 Feelings IRAP and SCS. Of the 56 correlations, 28 for the BT group and 28 for the NBT group, three proved to be significant (or marginally so) for the BTs*; Failure-Positive Feelings* with Self-Compassion Average (*r* = -.46, *p* = .037); Common Humanity (*r* = -.43, *p* = .06) and Self-Kindness (*r* = -.42, *p* = .06). In other words, an increased bias towards confirming that failing produces positive feelings predicted higher levels of Self-Compassion, Common Humanity and Self-Kindness.

For the NBT group, three correlations also proved to be significant (or marginally so)*, Success-Positive Feelings* with Isolation (*r* = -.43, *p* = .06) and Self Kindness (*r* = .39, *p* =

.08), and *Failure-Negative Feelings* with Common Humanity ($r = -.38$, $p = .09$). In other words, a bias towards confirming that success produces positive feelings predicted lower levels of Isolation and increased levels of Self-Kindness; a bias towards 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 proved to be significant. Of the 16 correlations obtained for the NBT group just one correlation was significant, *Failure-Positive Feelings* and Depression ($r = .54$, $p < .01$), 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 sub-scale of the explicit measure. None of the eight correlations across the two groups proved to be 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 of participants) with the three explicit measures.

The Outcomes IRAP and SCS. Out of 28 correlations for the BT group, just one proved to be significant; *Failure-Negative Outcomes* and Self-Judgement ($r = -.44$, $p = .04$); that is, a bias towards disconfirming that failing leads to negative outcomes predicted lower levels of self-judgement. For the NBT group, out of 28 correlations, three proved to be significant (or marginally so); Self-Judgement with *Failure-Positive Outcomes* ($r = -.38$, $p = .09$), *Success-Positive Outcomes* ($r = -.39$, $p = .08$) and *Success-Negative Outcomes* ($r = -.42$, $p = .05$). In other words, increased biases towards confirming that failure and success lead to positive outcomes, and disconfirming that success leads to negative outcomes, predicted lower levels of self-judgement.

The Outcomes IRAP and DASS. Of the 16 correlations for the BT group (the four trial-types with the four DASS scores) five correlations proved to be significant (or marginally so)*; Failure-Negative Outcomes* with *Stress* ($r = .48$, $p = .03$); *Success-Positive Outcomes* with Depression ($r = -.502$, $p = .02$), Anxiety ($r = -.445$, $p = .04$); Stress ($r = -.419$, $p = .06$); and *DASS* Total ($r = -.54$, $p = .01$). The first correlation indicates that increasing bias towards disconfirming that failure produces negative outcomes predicts increased levels of self-reported stress. The remaining correlations indicate that increasing bias towards 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 proved to be significant.

The Outcomes IRAP and Explicit Outcomes Scale. None of the eight correlations across the two groups proved to be significant.

Conclusion

The performances of the two groups differed considerably across three of the trial-types on the Feelings-IRAP, but the Outcomes-IRAP yielded little evidence of any clear between-group differences. 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 in the explicit measures 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.

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 correlational analyses between the Feelings-IRAP and the explicit measures for the BT group indicated that a tendency towards confirming that failing produces positive feelings was associated with higher levels of Self-Compassion, Common Humanity and Self-Kindness. 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 appear to make intuitive sense.

The only remaining significant correlation between the Feelings-IRAP and the explicit measures was obtained for the NBT group, who showed that confirming that failing produces positive feelings is associated with higher levels of self-reported depression. This latter finding might be seen as counter-intuitive because it indicates higher levels of depression in individuals who confirm that failing leads to positive feelings.

For the BT group, the correlational analyses between the Outcomes-IRAP and the Self-Compassion scales 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-Judgement. Interestingly, for the NBT group, three of the correlations were significant (or marginally so), with the results indicating that lower levels of self-judgement 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. 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 proved to be significant. This result is consistent with the previous studies and suggests once again that the IRAPs were tapping into responses towards succeeding and failing that are not captured readily with explicit self-report measures of the responses targeted by the IRAPs.

In general, the topography of the graphs for the implicit and explicit attitudes were very similar to those shown in Figure 5 for the NBT group, as was predicted. In contrast, however, further correlations with psychometrical measures were found. One possible explanation for this is that this study was conducted in Brazil with a different sample, whereas the study reported in Chapter 3 was conducted in Ireland. In relation to the BT group, it was expected that a neutral bias (in comparison to the NBT group) would be found for the trial types Failure-Negative and Positive (Feelings and Outcomes). This is because, this group may be expected to have a different perspective in relation to failures (e.g. acknowledging that failure produces negative feelings/outcomes, but at the same time, that one can learn from this experience). However, the neutral bias was observed only in the Feelings IRAP. Thus, it appears that the outcomes of failures for the BT group were not always positive. Interestingly, this trial-type also correlated with the explicit IRAP and it was correlated with stress measures on the DASS.

Importantly, the performance of the BT group was the only one that showed convergence in one of the IRAPs. That is there was a statisticla convergence between the Outcomes IRAP and the scale based on the IRAP. In anaysing the results, it could be argued that we should observe more convergences between implicit and explicit measures with a BT sample, because training in clinical behaviour therapy may improve "awareness". However, it may be that this is not necesarilly the case, or it could be the case that the IRAP captured only a subtle change of attitudes in the students under training. However, this is an empirical question worthy of investigation in further studyies. Specifically, it would be interestign to examine whether or not implicit attitudes towards self-forgiveness differed among fully qualifed therapists and trainees, as a function of clinical experience.

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.

Chapter 7

General Discussion

Overall Summary and Discussion of the Research Programme

The overall objective of the research presented in the current thesis was to develop a measure of the verbal behaviours that are would be considered relevant to the psychological domain of self-forgiveness, as measured using methods that may be described as direct (self-reports) and indirect (the IRAP). The thesis presents five studies, which focused primarily on refining the IRAP a measure of self-forgiveness. In the following section of this final chapter a brief summary of each of the studies will be presented. Subsequently, the findings from each of the studies will be discussed in more detail.

Summaries of the Five Studies

Study 1, Chapter 2

The first study reported in the current thesis was largely exploratory and sought to begin the development of a measure of implicit forgiveness of self related to "minor" transgressions (mistakes, flaws, shortcomings) *versus* the forgiveness of others. Forty-seven students completed a scale designed to measure forgiveness of self and others using the implicit relational assessment procedure (IRAP). The results indicated that the measure of implicit forgiveness diverged from an explicit measure designed to measure the same construct. The key finding was that participants tended to be more forgiving towards themselves than towards others at an implicit level, but this was not the case at the explicit level; on an explicit measure participants rated their own failures as less acceptable than the failures of others. Overall, the findings supported the general thesis that it may be useful to supplement explicit measures of forgiveness with implicit measures in future research.

Study 2, Chapter 3

The second study aimed to develop the IRAP as a measure of response biases related to emotional reactions and expected outcomes in the context of minor failings and successes in everyday life. Additionally, the research explored the extent to which such implicit

reactions were related to standardized measures of psychopathology, including depression, anxiety, stress, and a scale that was based directly on the IRAP. Sixty undergraduates completed two IRAPs and the explicit measures. The pattern of biases observed across the implicit and explicit measures, diverged, and the correlations between the two types of measures were either absent or relatively weak. The results suggested that implicit measures may provide an additional source of information concerning self-forgiveness beyond that provided by explicit self-report measures per se.

Study 3, Chapter 4

The third study also aimed to test the IRAP as a measure of self-forgiveness response biases related to emotional reactions and expected outcomes in everyday life. In contrast to the previous two studies, and indeed other IRAP research, the stimuli were individualized in that they were based on 'problematic' and 'non-problematic' behaviours (e.g., procrastination versus keeping deadlines) that each participant reported at the beginning of the study. Specifically, participants completed two IRAPs. One (the Feelings IRAP) targeted negative and positive feelings experienced while engaging in problematic versus non-problematic behaviour. The other (the Outcomes IRAP) targeted positive and negative outcomes believed to result from this behaviour. Participants also completed standardized measures of psychological suffering and self-compassion, as well as a questionnaire that targeted the behaviour and reactions presented in the IRAPs. While both IRAPs produced response biases that indicated that positive feelings and outcomes were related to non-problematic behaviour, neither produced clear evidence that negative feelings or outcomes were related to problematic behaviour. Furthermore, specific response biases on the IRAP (i.e., a tendency to confirm that negative actions lead to negative outcomes) correlated with psychological suffering, particularly depression and stress. The findings suggest that individualized IRAPs, even those that target minor problematic behaviour, may be predictive of psychological

suffering.

Study 4, Chapter 5

This fourth aimed to test the effect of positive and negative priming on the assessment of self-forgiveness with the IRAP as related, again, to emotional reactions and expected outcomes in everyday life. Additionally, the research explored the extent to which such implicit reactions were related to standardized measures of psychopathology, including depression, anxiety, stress, and a scale that was based directly on the IRAP. Eighty one undergraduates were assigned to two groups, where they were presented with different conditions, positive and negative priming, in which participants had to recall in writing three experiences of failing or succeeding; participants then completed two IRAPs, one targeting feelings and the other targeting outcomes as related to failing and succeeding behaviours. In addition, participants were asked to complete two explicit measures that were derived from the two IRAPs and another two explicit measures that targeted self-compassion and stress, anxiety and depression. The findings showed that the priming conditions affected the two IRAPs differentially. Furthermore, the IRAP that targeted feelings predicted level of self-reported psychopathology but only for participants in the positive priming condition. As discussed in more detail below, the findings provide tentative evidence for experiential avoidance.

Study 5, Chapter 6

The fifth and final study aimed to test the effect of behaviour 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 the two IRAPs that had been developed across the previously reported 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 post graduate students completed the study (20 individuals who were teaching on, attending or who had attended a course in clinical behaviour analysis and 20 students from different fields). The two groups (Behaviour Therapists and Non-Therapists) completed the two IRAPs, and the explicit measures. Overall, only one of the two IRAPs, the one that targeted feelings rather than outcomes, produced clear and significant differences between the Behaviour Therapist and Non-Therapist 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 two IRAPs and explicit measures with these two groups of participants are considered later in the current chapter.

Discussion of the Individual Studies

Study 1

Perhaps one of the most interesting findings of Study 1 was that participants indicated high levels of forgiveness towards self, but not others, in the context of failure at an implicit level, but at an explicit level they appeared relatively forgiving towards both self and others (i.e., on the explicit measure the mean ratings were all below 3 on a 7-point scale with higher scores indicating unacceptability). Such a finding appears to be generally consistent with previous research with implicit measures that have shown such instruments to be sensitive to socially sensitive response biases that may remain "hidden" on explicit measures (e.g. Goldring, 2011; Power et al., 2009; Cullen & Barnes-Holmes, 2008). That is, it seems likely that it would be deemed socially unacceptable to be less forgiving concerning the failures of others relative to failures attributed to oneself. If this tendency on explicit measures reflects, even in part, the impact of socially desirable responding it seems important that future work

in this area include the use of implicit measures. Study 1 constituted a first step in this direction.

A related, but admittedly tentative finding was that the less accepting participants were, implicitly, concerning the failures of others the more accepting they were, explicitly, concerning their own failures. In other words, an implicit bias towards judging others negatively appeared to predict a "softer" (explicit) attitude towards the self. While this single correlation must be interpreted with caution it does suggest that the tendency not to forgive others may be associated with a tendency to forgive the self more easily.

It is important to note that the IRAP data failed to correlate significantly with the explicit measure derived directly from it (across 15 of the 16 correlations). In other words, even when an explicit measure and the IRAP contain items that are mapped onto each other, it is possible that they are tapping into different classes of behaviour or behaviours that are under distinct sources of contextual control (see Hughes, Barnes-Holmes, & Vahey, 2012, for a detailed treatment of this issue from a relational-frame theory perspective). Once again, therefore, the current findings highlight a possibly important role that implicit measures can play in assessing forgiveness of self and others in general terms.

It is interesting that the IRAP yielded a relatively strong acceptability bias for the *My-Failures-Acceptable* trial-type, but this was not observed for the *Others-Failures-Acceptable* trial-type. One post-hoc explanation is that individuals are generally "well practiced" in "making excuses" for their own minor mistakes and flaws (e.g. arriving late for a meeting, underperforming in an exam, etc), and thus, insofar as the IRAP is sensitive to the relative strength of behaviours that occur frequently in an individual's repertoire (see Hughes et al., 2012), we might expect to see a self-forgiveness bias. In contrast, individuals are typically less well practiced at making excuses for the minor failures of others, if for no other reason than we are not usually required to do so. Thus the absence of a strong bias on the *Others-*

Failures-Acceptable trial-type of the IRAP makes sense. Perhaps a future study could test this explanation by requiring participants to engage in a task that requires them to provide many excuses or explanations for the mistakes, flaws, and errors of other individuals before completing the current IRAP – in effect, would exposure to such a task produce an acceptability bias on the implicit measure?

Study 2

Study 2 presented participants with two separate IRAPs—one targeting feelings and the other targeting outcomes in relation to failing and succeeding. In addition, participants were asked to complete two explicit measures that were derived from the two IRAPs and another two explicit measures that targeted compassion and psychopathology. The results arising from the two trial types that focused on "success" for both IRAPs were broadly consistent with "common-sense" conclusions in that all of the IRAP effects yielded positive bias effects. The IRAP effects for the two trial types that focused on "failure," however, were not so straightforward. Although the trial types that targeted failure and negative feelings, or failure and negative outcomes, produced negative biases, the trial types that targeted failure and positive feelings/outcomes both produced positive biases. Interestingly, the explicit measures that were designed to map onto the trial types from the two IRAPs produced biases that were all consistent with "common-sense" conclusions–questions concerning failing produced negative biases and questions concerning success produced positive biases.

The pattern of biases observed between the IRAPs and the explicit measures derived from the IRAP trial-types also differed in another way. Specifically, although the IRAP effects for the two "Success" trial types were both positive (for both IRAPs), the effects for the *Success-positive feelings* and *Success-positive outcomes* trial types were considerably stronger than the effects for the two respective *Success-negative* trial types; this pattern was the opposite of that observed for the two explicit measures. In the latter case, the *Success-*

Negative subscales produced stronger positive bias ratings than the *Success-Positive* subscales.

At this point, therefore, it seems clear that the implicit and explicit measures produced diverging patterns of responding. Of course, such divergence is quite common in the literature on implicit attitudes and cognition (McConell and Leibold 2001; Payne et al. 2008), and, thus the current findings are hardly unique. Nevertheless, given the exploratory nature of the current research it does seem important to consider why such divergences were observed. As noted in the introduction to the current thesis, the IRAP was designed to measure BIRRs, whereas explicit measures are seen as typically targeting EERRs. Although both types of relational responding may overlap in many contexts, there are situations in which it has been argued they may not (see Hughes et al.2012, for a detailed discussion). Lack of overlap may be seen, for example, in so-called socially sensitive domains (e.g., Barnes-Holmes et al. 2010b). Another possible reason for lack of overlap, however, might be traced to the relative probability of the relational responses targeted by the measures. In the context of the current study, it may be that the *Fail-positive* trial types for both IRAPs targeted BIRRs that were relatively improbable, in the sense that participants had rarely been asked in the past to consider if failing produced positive feelings or outcomes for them. In the simple absence of a relevant history of such BIRRing activity, perhaps participants simply showed a tendency to respond "True" rather than "False", and thus the counterintuitive positive bias emerged for these trial types.

A broadly similar explanation may be applied to the pattern of results obtained across the other three trial types from the IRAP. For example, although the effects for the *Fail-negative* trial-types both produced intuitively predictable negative biases, they were relatively weak (when compared to the effects for the *Success-positive* trial-types). It may be that the tendency to avoid discussing or even thinking about our failures, relative to our successes, in

day to day life means that BIRRs related to failures may be at relatively low strength, which is reflected in the current IRAP effects. In contrast, most of us are willing to enjoy thinking about, if not talking about, our successes, which may account for the relatively strong IRAP effects observed for the *Success-positive* trial types. The relatively weak effects observed for the *Success-negative* trial types may be explained by the fact that we are rarely asked to consider if succeeding had a negative impact on us in some way.

One objection that may be raised at this point is that the foregoing explanation for the Success-negative and Fail-positive trial types are similar, but the results are quite different. Specifically, the direction of the effect for the Success-negative trial-type, although weak, is intuitively incorrect (i.e., failure is positive). In other words, the absence or low probability of relevant BIRRs in the case of Success-negative responses may help to explain why the IRAP effects for this trial type were weak but in the intuitively correct direction, but why would weak or absent BIRRs for the Fail-positive trial type produce effects that were also weak but in the counterintuitive direction? One possibility is that in the absence of strong BIRRs relating the label and target stimuli, the stimulus which participants observed just before choosing one of the two response options came to dominant the IRAP performances. That is, when positive words were presented (during the Fail-positive trial types), a bias for responding "True" emerged, but when negative words were presented (during the Success-negative trial-types), a bias for responding "False" emerged. Note, that in offering this explanation, we are assuming that the word *True* would be categorized more readily as positive and *False* would be categorized more readily as negative (see Blanton and Jaccard, 2006, for detailed argument pertaining to the interpretation of absolute scores from implicit measures).

An alternative explanation that could accommodate these findings might be that success is very rarely negative and thus denying that success leads to negative feelings and

outcomes was observed on the IRAPs. In contrast, failure may not always be seen as negative, in the sense that one can learn something of value or importance from mistakes or errors, and, thus, the very small positive bias was observed on the IRAPs. Of course, the foregoing post-hoc explanations remain highly speculative, but they do suggest possibly interesting directions for future research that aims to better understand the role of implicit biases in the domain of self-forgiveness.

The current study failed to obtain strong evidence that the IRAPs predicted responses on the explicit measures. Only one of 60 correlations calculated for the feelings IRAP proved to be significant, with 4 of the 60 correlations reaching significance for the outcomes IRAP. Interestingly, the four correlations obtained for the latter IRAP might be considered intuitively predictable. For example, an increase in confirming that failing produces negative outcomes at an implicit level related to an increase in confirming that success produces positive outcomes at an explicit level. Furthermore, denying that failure produces negative outcomes predicted higher levels of mindfulness subscale of SCS. Nevertheless, it would be unwise to read too much into such a low number of correlations, and perhaps at this point it is safer simply to conclude that, in general, the IRAPs failed to predict the responses on the explicit measures.

At the current time, it remains unclear why the correlations were so weak and/or few in number. Perhaps the implicit and explicit measures simply targeted relational responses (i.e., BIRRs versus EERRs) that were under different forms of contextual control and thus they failed to correlate. Another, or perhaps additional, reason for the lack of correlation between the implicit and explicit measures is that the statements pertaining to failure (versus successes) were simply not evocative or salient enough to elicit relatively strong emotional reactions in many, if not most, of the participants. In other words, general statements about failing versus succeeding did not encourage participants to recall or genuinely think about

their own previous failures and successes, and thus the absence of any consistent relationships with levels of psychopathology or self-compassion would be expected. In any case, to address this later possibility, the next study reported in the thesis involved asking participants to provide examples of failures or shortcomings that were specific to them and then insert these into 'individualized' IRAP.

In conclusion, Study 2 indicated that implicit measures designed to target self-forgiveness produced patterns of response biases that diverged from those obtained using explicit measures that were specifically designed to map onto the implicit measures. In general, there was extremely limited evidence for correlations between the implicit and explicit measures, although the IRAP targeting outcomes (rather than feelings) yielded four rather just one significant effect. The current findings are broadly consistent with the results reported in the previous study, except that the current study used measures that did *not* involve 'pitting' forgiveness of self against that of others. As such, it appears that IRAPs may be used to measure BIRRs that are related to self-forgiveness *per se,* rather than forgiveness of self relative to others.

Study 3

The main findings arising from Study 3 showed that participants produced response biases, at an implicit level, that indicated positive actions generally produce positive feelings. Interestingly, however, the IRAP data did not provide strong evidence that negative actions lead to negative feelings. The same general pattern was obtained with the Outcomes IRAP; positive actions lead to positive outcomes, but negative actions do not lead to negative outcomes. In fact, for the *Negative-Action/Positive-Outcome* trial-type participants produced a response bias that was significantly different from zero in a *positive* direction.

In terms of the direction and strength of the IRAP effects obtained in the current study, similar to Study 2, we did not find negative biases for the *Failure-Positive* trial-types.

In this sense, employing ideographic stimuli in the IRAPs did not change the counter-intuitive result reported previously. On balance, perhaps the lack of an effect was due to the fact that only the label stimuli were genuinely ideographic. That is, the label stimuli inserted into the IRAPs were based on the PBQ but the target stimuli were not (i.e., the latter were generic across all participants). A future study might pursue this issue by employing label and target stimuli based on participants' self-reports. We shall return to the issue of using ideographic stimuli below.

Possibly, the counter-intuitive nature of the IRAP data and the lack of correlations between the Problem Behaviour Questionnaire and the IRAPs might be due to factors broached in Chapter 3, That is, because participants reported only moderate problem behaviour on the questionnaire (e.g., too much facebook, television, etc.), the stimuli failed to evoke a relatively negative bias in the trial-type Negative Actions-Negative Feelings/Outcomes.

Although there were many similarities in the findings across the experiments,, it is worth noting that significant correlations were found between performance on the Outcomes IRAP and psychological suffering as measured by the DASS; that is, the more strongly participants responded to negative actions as producing negative outcomes, the higher the level of self-reported depression, stress, and general suffering. This tentatively indicates that even though the participants were reporting trivial problems, the Outcomes IRAP was tapping into something important.

On the one hand, the failure to find any correlations between the current IRAPs and the Self-Compassion scale could be seen as undermining the claim that the IRAPs are relevant to the domain of self-forgiveness (insofar as self-forgiveness is seen as a critical component of self-compassion). As noted above, perhaps the relatively trivial nature of the problem behaviours that were reported by a normative sample may help to explain the lack of

correlations. Indeed, it is worth noting that in the fifth study reported in the current thesis evidence for correlations between a self-forgiveness IRAP and the self-compassion scale was found with a sample of participants who were training as behaviour therapists (Bast, Barnes-Holmes, Presti, Dell'Orco, Carnevali, Oppo, Kovac, Linares, 2014). Such a finding is consistent with the argument that the sample, rather than the instrument, may be the critical variable.

The largest trial-type effect for both IRAPs was produced by the Positive-Positive relation (the fourth bar in the two graphs), whereas the Negative-Negative relation (the first bar in the two graphs) was relatively weak (and non-significant) by comparison. At the present time, it remains unclear why these differential effects emerged. One possible explanation is that the valence of the stimuli presented for the Positive-Positive trial-types was more easily associated with the "True" response option, whereas the valence of the stimuli presented for the Negative-Negative trial-types was more easily associated with the "False" response option (assuming that "True" is more positively valenced than "False"). If this was the case, then any response bias towards "True" when confirming that negative actions produce negative feelings or outcomes may have been reduced somewhat by a competing bias to associate negatively valenced stimuli with the negatively valenced response option ("False").[1]

A closely related explanation might appeal to a general positivity bias to which the IRAP may be sensitive (see Barnes-Holmes et al 2010, p. 75-76). For example, all things being equal, in natural language interactions speakers tend to emphasize the positive over the negative, reporting for instance that a glass is half full rather than half empty (see Dodds et al,

---

[1] One argument might be that the IRAP effects were stronger for the Positive-Action trial-types because they relied to some extent on the researchers' interpretations in creating the relevant stimuli. That is, the contrast category involved inserting label stimuli that were deemed to be the opposite of the negative actions specified in the PBQ. Although this point certainly applies to the Feelings-IRAP it does not apply to the Outcomes-IRAP. In the latter case, the effects for the *Positive-Actions/Negative-Outcomes* and *Negative-Actions/Positive-Outcomes* trial-types were almost identical.

2015, for evidence that this effect is observed across numerous languages). Given that the IRAP was specifically designed to capture differential probabilities (or biases) in patterns of verbal or relational responding that are found in natural language (Barnes-Holmes, Hayden, Barnes-Holmes, & Stewart, 2008) it seems reasonable to assume that such biases may also be reflected in IRAP performances. Indeed, one would hope so if the IRAP is to be considered a measure of the response patterns found in natural language. In fact, it might even be important to capture such positivity biases if they feed into the criterion variables that one aims to predict with the IRAP (see Vahey et al. 2015). In the case of the current study, for example, it was the *Negative-Negative* trial-type of the Outcomes-IRAP that correlated with human suffering. Perhaps the interaction, or response competition, between a general positivity bias and a bias towards confirming that negative actions lead to negative outcomes were jointly responsible for the observed correlations with the DASS measure? This of course remains an empirical issue.

An alternative explanation for the fact that the largest trial-type effect was produced by the Positive-Positive relation might be that positive actions generally always produce positive feelings and outcomes, and thus the IRAP effects were relatively strong for these trial-types. In contrast, negative actions do not always produce negative feelings and outcomes, particularly in the context of the types of relatively minor "negative" actions that the current IRAPs were targeting. Indeed, many of the "problem" behaviours that the participants listed, and which were used in the IRAPs, involved activities that would also be deemed enjoyable and could in principle lead to positive outcomes. For example, spending too much time on Facebook might have been listed as a problem, but of course it was also being identified as an enjoyable way to spend one's leisure time, and could in principle lead to good outcomes, such as connecting with friends. Perhaps, therefore, any bias towards

responding negatively to these trial-types was moderated by the positive properties of the problem behaviour itself.

Additional research will be needed to determine if any or all of the foregoing explanations for the weaker Negative-Negative trial-types effects are valid. At this point, however, it is worth noting that the correlations between the explicit measures of psychological suffering and the IRAPs all occurred for the Negative-Negative trial-type of the Outcomes IRAP. It appears, therefore, that at least one of the response biases observed on the IRAPs could not be explained solely by a tendency to associate negatively valenced stimuli or a general positivity bias (because the correlations with suffering occurred for the Outcomes but not the Feelings IRAP). It is also worth noting that the effects for the Negative-Positive trial-type across the two IRAPs also differed substantively -- for the Feelings IRAP the effect was close to zero but it was significantly positive for the Outcomes IRAP. A simple explanation in terms of participants associating the valence of the trial-type stimuli with the response options, or a general positivity bias, thus seems untenable because the IRAP effects differed depending on whether they were targeting feelings or outcomes.

Indeed, the contrast between the two IRAPs for the Negative-Positive trial-type could be explained in the following way. For the Feelings IRAP the negative actions may have produced some element of response competition between "True" and "False" because the listed actions produce both positive and negative feelings. For example, eating too much chocolate produces feelings of pleasure but also guilt. In contrast, the Outcomes IRAP might have failed to produce this type of competition because the outcomes were not particularly strong or salient, given the relatively minor problems that the participants had identified. As a result, the effect for the Negative-Positive trial-type may have reflected a tendency to respond to trivial problem behaviour as not having negative outcomes, and the positive bias reflected a tendency to coordinate the positively valenced target stimuli with the positively valenced

response option (i.e., "True"). In any case, in future studies in this area it may be important to ensure that stimuli inserted into the IRAPs that refer to problem behaviours are appropriately negatively valenced.

Another interesting finding of the current study is related to the high attrition rates in comparison to previous the two previously reported studies, in which generic IRAPs were employed. It seems likely that the attrition rate could be attributed, at least in part, to the use of ideographic IRAPs. A regular practice adopted when using nomothetic IRAPs is to conduct a pilot study to test the effects of the stimuli employed in the IRAP and, if necessary change the stimuli to reduce potential attrition rates. However, in the case of ideographic IRAPs, conducting such pilot work with each individual would be extremely difficult and perhaps unwise in terms of maintaining experimental fidelity. Consequently, incurring relatively large attrition rates could be a risk in any study that attempts to use ideographic IRAPs. Indeed, this possibility should be considered carefully in any future study that attempts to employ or develop IRAPs that are even more ideographic than the current versions (i.e., ones that generate both label and target stimuli, and perhaps even response options, via ideographic means).

In continuing with the current programme of research there appear to be many lines of potential inquiry. First, perhaps future studies could generate individualized IRAPs that employed both label and target stimuli that were provided by the participants, rather than just the label stimuli, as was the case in the current study. Second, perhaps future studies could employ a type of priming task in which participants are asked to reflect on their own failures and successes without having to disclose that information to the researcher. In adopting this strategy, participants may be more likely to focus on perhaps less trivial or mundane failures (ones they prefer not to discuss or make public), and a nomothetic IRAP could then be used

to assess the impact of such a priming task. This was the strategy that was adopted in the next study reported in the current thesis.

Notwithstanding the questions that remain to be answered concerning possible explanations for the current set of results, and potential directions for future research, it is important to recognize that this was the first study in which the stimuli inserted into an IRAP were derived from personalized self-reports from individual participants. Although much work remains to be done, the ideographic IRAPs appeared to be sensitive to a common class of behaviours across individuals, and as a result the implicit-explicit correlations yielded a highly specific effect – the Negative-Negative trial-type from the Outcomes IRAP predicted higher levels of psychological suffering. This suggests that "personalizing" an IRAP with stimuli that capture even minor problems with which participants are currently struggling may make it more sensitive as a potential measure of human suffering than the self-forgiveness IRAPs that were employed in the previous two studies in which such correlations were not obtained.

Study 4

Study 4 presented two different conditions, positive and negative priming, in which participants had to recall in writing three experiences of failing or succeeding before completing two IRAPs and then the explicit measures. In terms of the mean trial-type scores recorded with each IRAP, the priming condition appeared to affect performance on the Outcomes-IRAP but not performance on the Feelings-IRAP**.** Specifically, significant differences emerged between the two priming conditions for the Failure- and Success-Positive trial-types, with a weaker bias towards confirming that failing produces positive outcomes but a stronger bias towards confirming that succeeding produces positive outcomes in the negative priming condition. The fact that the priming variable impacted on one IRAP but not the other indicates that the effects were specific to the targeted domain (i.e., outcomes

versus feelings), rather than constituting a general influence on self-forgiveness IRAP performances per se. This finding is consistent with other research that has shown that two separate IRAPs, which target different aspects of the same clinical domain, may produce different outcomes (Nicholson & Barnes-Holmes, 2012).

How might we explain the differential priming effects that we observed here for the Outcomes-IRAP? First, we should note that the effects for the positive priming condition were quite similar across the two trial-types but the difference was substantive for the negative priming condition (i.e., a relatively weak effect for the *Failure-Positive* trial-type but a relatively strong effect for the *Success-Positive* trial-type). It appears, therefore, that negative priming had a considerable impact on the response biases observed across these two trial-types, whereas positive priming did not. Although somewhat speculative, perhaps asking participants to think about previous failures in the negative priming condition may have evoked events from the past where failure did *not* lead to positive outcomes, thus weakening any bias towards responding to examples of failure that did lead to something positive (e.g., failing to get a particular job led to getting another better job in the future). In contrast, the large positive bias observed for the *Success-Positive* trial-type in the negative priming condition may have emerged because the evaluative functions of success were increased by having just recently thought about failures, which produced no positive outcomes. In other words, perhaps success is valued more highly in the context of having recently thought about failure. As an aside, it is worth noting that adherence measures were not employed in the current study to determine the extent to which participants believed that they were successful in thinking about previous successes or failures in their lives – it may be useful for future studies that employ priming tasks to include such measures.

The effects for the two (explicit measure) scales that were derived from the IRAPs yielded results that were relatively consistent across the two priming conditions. Exposure to

positive relative to negative priming appeared to increase negative bias for the two *Failure* sub-scales and decrease positive bias for the two *Success* sub-scales, with the clear exception of the *Fail-Negative* scale for the Outcomes measure, which yielded a non-significant difference across the two priming conditions. In other words, it appears that positive relative to negative priming led participants to rate failing more negatively and rate succeeding less positively. One simple but tentative explanation of these effects is that having recently thought about previous successes in one's life, failure may be perceived to be more negative, but successes less positive due to a contrast effect. Similarly, thinking about failures may cause one to respond to failure less negatively and success more positively due to the contrast effect. In any case, whatever the explanation for these effects, it is clear that the priming conditions impacted quite significantly on the explicit measures that were derived from the two IRAPs.

Another possibly interesting finding that emerged in Study 4 was a marginally significant correlation between the bias scores obtained on the *Fail-Positive Feelings IRAP* trial type and the explicit measure of *Self-Compassion*. Specifically, a lower level of self-compassion predicted a bias towards 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 is worthy of note because it seems somewhat counter-intuitive. That is, one might expect that lower levels of self-compassion would predict that failing should produce a bias towards denying not affirming positive feelings. Or more informally, if an individual is relatively low in self-compassion then surely failure would be seen in a more negative light? The fact that the counter-intuitive correlation only emerged for the positive priming condition could be important here, however. Perhaps individuals with low self-compassion have a greater tendency to "protect themselves" against negative feelings because in the absence of

compassion such feelings are more threatening to "the self" (Baumeister, Bushman, & Campbell, 2000; Raskin, Novacek, & Hogan, 1991; Watson & Hickman, 1995). If the positive priming increased the need for such protection, at least temporarily while completing the IRAP, the correlation makes some sense. In effect, having just been encouraged to feel positive towards the self during the positive priming task, 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 is a highly speculative post-hoc explanation, based on only one marginally significant correlation, but it is highlighted here because it could be seen as consistent with other aspects of the current findings (discussed in the next paragraph).

A possibly related finding was the pattern of correlations that was obtained for the positive priming group between the DASS scores and the *Fail-Negative* trial-type of the Feelings-IRAP. That is, 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 et al., 2006; Marx & Sloan, 2005; Hayes, Wilson, Strosahl, Gifford, & Follete, 1996). Again, this post-hoc explanation remains rather speculative but it is broadly consistent with the correlation discussed above.

It is also worth noting that the only suggestion of a significant correlation between the IRAPs and the scales based on the implicit measures was obtained for participants who were

exposed to the negative priming condition (for the *Fail-Negative Feelings* trial-type); none of the eight correlations even approached significance given positive priming. In effect, the implicit and explicit measures may have captured response biases that overlapped more in the context of negative than positive priming. Although tentative, this supports the argument that exposure to positive priming may have increased an implicit bias towards experiential avoidance to a greater extent than exposure to negative priming. That is, when participants were encouraged to feel more positively about themselves, they reacted more strongly against embracing negative feelings during a subsequent IRAP, particularly if they were relatively high in measures of psychopathology.

Overall, the findings from Study 4 are interesting because they indicate that performance on an IRAP that is designed to target self-forgiveness may predict self-reported levels of psychopathology (and perhaps even self-compassion). The results of Studies 1 and 2 failed to find any such predictive relationships. Critically, however, the correlations obtained in the Study 4 only emerged for a specific priming condition and with the Feelings-IRAP. Furthermore, some of the effects appear, at first blush, to be rather counter-intuitive. That is, higher levels of psychopathology predicted reduced levels of negative bias, but only in the positive-priming condition. As noted above, however, this type of result could be seen as consistent with the argument that higher levels of experiential avoidance (associated with higher levels of psychopathology) may be at play here. Although such an explanation must remain speculative until further research is conducted, the present findings appear to move us closer towards a more sophisticated understanding of self-forgiveness using implicit measures.

Study 5

The primary aim of Study 5 was to determine if participants who had been exposed to a training history in Behaviour Therapy, with a focus on clinical behaviour analysis, would

respond differently from a control group on IRAPs that were designed to target expected feelings and outcomes arising from failing and succeeding. The performances of the two groups differed considerably across three of the trial-types on the Feelings-IRAP, but the Outcomes-IRAP yielded little evidence of any clear between-group differences. 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 clear and significant differences between the BT and NBT groups. The fact that only one of the two IRAPs produced a between-group difference suggests that the diverging performances were specific to the stimuli that were presented in the IRAP (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, and perhaps sensitized them to the types of concepts and terms employed in the DASS and SCS, which then 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 non-significant;

furthermore, for each of the three subscales the means were well below the cut-off for normal levels of anxiety, stress and depression for both groups.

Nevertheless, it is interesting that the IRAP, which 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 may well serve to increase levels of stress and anxiety, and general levels of compassion (for both self and others), relative to training in other areas, and this served to 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 towards confirming that failing produces positive feelings was associated with higher levels of Self-Compassion, Common Humanity and Self-Kindness. 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 appear to make intuitive sense.

The only remaining significant correlation between the Feelings-IRAP and the explicit measures was obtained for the NBT group, who showed that confirming that failing produces positive feelings is associated with higher levels of self-reported depression. This latter finding might be seen as counter-intuitive 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 towards experiential avoidance, which has been associated with a broad range of psychopathological reactions (Hayes et al, 2011). In other words, claiming that failing makes you 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, Wilson, Gifford, Follette, & Strosahl,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 counter-intuitive to associate failing with positive feelings (or to deny an association with negative feelings), the tendency to do so was indicative of increased levels of self-compassion, particularly for the therapist group. Or to put it another way, it makes sense that increased levels of self-compassion would reduce the negative impact of failures on how we feel when they occur during our day to day lives (Tirch, Schoendorf, Silberstein, 2014).

For the BT group, the correlational analyses between the Outcomes-IRAP and the Self-Compassion scales 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-Judgement. Interestingly, for the NBT group, three of the correlations were significant (or marginally so), with the results indicating that lower levels of self-judgement 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 counter-intuitive 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. And 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 remains 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 proved to be significant. This result is consistent with the previously reported studies in the current thesis and suggests once again that the IRAPs were tapping into responses towards 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, clear differences emerged between the two groups in their performances on 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 participants 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 less likely to be seen across the two groups. Of course, this post-hoc explanation must remain highly speculative at the current time, but 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, 2011). Given the current findings future studies might attempt to target specific verbal histories using relevant IRAPs. For example, would 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 CBT? And would the strength of the IRAP effects correlate with potentially important variables, such as stress and professional burnout (see Kelly & Barnes-Holmes, 2014, for a relevant example in the context of teachers working children with learning disabilities).

Conclusion

The research programme presented in the current thesis aimed to develop the IRAP as a measure of the verbal or relational responding that may be considered directly relevant to the psychological domain of self-forgiveness.

It was not the aim of this thesis to provide a functional definition on self-forgiveness. Rather, one of the aims of this thesis was to measure an aspect of self-forgiveness in relation to failures, acknowledging the negative feelings and outcomes that it might cause, and correlating this with self-forgiveness and psychopathology.. Thus. for the most part the studies presented in the thesis were largely exploratory, and so it was premature to make any kind of prediction in relation to the possible relationships that may be found between the IRAP scores and explicit measures. Many of the correlations found might be said to correspond with common-sense, For example, it was often observed that according to test outcomes, failures produce negative feelings and success produce positive feelings. However, there were also some counter-intuive findings (e.g., participants confirmed implicitly that failure produces positive feelings). In general, it seems that when participants were exposed to the negative priming, the bias in the Failure-Positive Feelings trial-type was weaker in comparison to the other group, and in the BT group the bias was weaker then the bias produced in the negative priming group. It might be expected that the bias of the BT group in the Failure- Positive Outcomes would also be neutral, but the sample consisted largely of trainees and most likely their failures did not produce positive outcomes. In contrast, an experienced therapist may have less negative outcomes from their professioanl activity, but this is a empirical question.

Based on the studies reported in the thesis, outcome predictions might be tentatively made in future studies. For instance, the general pattern of the bias has already now been replicated with participants from Ireland and Brazil (see Fig 5 and Fig 9). Except for one

counter intuitive finding (from the Failure-Positive Feelings trial-type), the bias of the correlations between the IRAP and the explicit IRAP questionnairre corresponded topographically (although only a few significant correlations were found in the various studies). One possible explanation for this lack of correlations may be the way in which the stimuli in the questionaries were arragend (e.g., in clusters, first positive feelings, later negative feelings). Even though the IRAP was capable of arranging stimuli in random clusters, it may be worth investigating the use of questionnaire with different arrangements of questions. It is also worth noting that the patterns of bias started to change when participants were exposed briefly to the details of their own failures in Experiment 4, as well as for a behaviour therapist group, that had presumably been trained to deal with failures on a more regular bases in comparison to the normative population.

Hypothethically, the perfect "self-forgiveness" of failures pattern observable on the IRAP outcomes, would be a balance (neutral bias) on all trial-types. That is, this would indicate that for that individual failure is neither positive nor negative and success is neither positive nor negative. Failure is not taken too personally and proivides leanring opportunities, and success is neither taken too personally and is seen as potentially dangerous if identified with too strongly (i.e., the flip-side of taking failure too personally is taking success too personally).

On balance, the five studies reported herein yielded evidence that the domain of self-forgiveness may indeed be "captured" to some extent by IRAPs that targeted the feelings and outcomes arising from minor failings in everyday life. Interestingly, there was repeated evidence of divergence between the IRAPs and the explicit self-report measures that were derived from the implicit measures. However, there was also evidence that the response biases produced on some of the IRAPs correlated with specific measures of self-compassion and human suffering more generally. Furthermore, the nature of some of these correlations

suggested, if only tentatively, that experiential avoidance may be implicated in so called implicit self-forgiveness. Clearly, the research reported in the current thesis constitutes only the first step in attempting to better understand the psychological processes involved in the domain of implicit self-forgiveness, but the current research does provide a solid foundation upon which to develop subsequent experimental and conceptual analyses in this regard.

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.

Hughes, S., Barnes-Holmes, D., & De Houwer, J. (2011). The dominance of associative theorizing in implicit attitude research: Propositional and behavioral alternatives. *The Psychological Record*, *61*(3), 6.

Barnes-Holmes, D., Murtagh , L., Barnes-Holmes, Y., & Stewart, I. (2010). Using the Implicit Association Test and the Implicit Relational Assessment Procedure to measure attitudes toward meat and vegetables in vegetarians and meat-eaters. *The Psychological Record, 60,* 287-306.

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

Barnes-Holmes, D., Murphy, A., Barnes-Holmes, Y., & Stewart, I. (2011). The Implicit Relational Assessment Procedure: Exploring the impact of private versus public contexts and the response latency criterion on pro-white and anti-black stereotyping among white Irish individuals. *The Psychological Record*,*60*(1), 4.

Barnes-Holmes, D., Hayden, E., Barnes-Holmes, Y., & Stewart, I. (2008). The Implicit Relational Assessment Procedure (IRAP) as a response-time and event-related-potentials methodology for testing natural verbal relations: A preliminary study. *The Psychological Record*, *58*, 497–515.

Barnes-Holmes, Barnes-Holmes, Y., & Hussey, I. (in press). Relational frame theory: Finding
    it historical roots and reflecting upon its future development. In R. Zettle, S. C. Hayes,
    D. Barnes-Holmes, & A. Biglan (Eds), *The Handbook of Contextual Behavioral
    Science*. Hoboken, NJ: Wiley-Blackwell.

Barnes-Holmes, Y., Hussey, I., McEnteggart, C., Barnes-Holmes, D., & Foody, M. (in press).
    Scientific ambition: The relationship between relational frame theory and middle-level
    terms in acceptance and commitment therapy. In R. Zettle, S. C. Hayes, D. Barnes-
    Holmes, & A. Biglan (Eds), *The Handbook of Contextual Behavioral Science*.
    Hoboken, NJ: Wiley-Blackwell.

Bassett, J. F., & Dabbs, Jr, J. M. (2003). Evaluating explicit and implicit death attitudes in
    funeral and university students. *Mortality*, *8*(4), 352-371.

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.

Baumeister, R. F., Bushman, B. J., & Campbell, W. K. (2000). Self-Esteem, Narcissism, and
    Aggression Does Violence Result From Low Self-Esteem or From Threatened
    Egotism? *Current directions in psychological science*, *9*(1), 26-29.doi: 10.1111/1467-
    8721.00053

Berry, J.W., Worthington Jr, E.L., Parrot III, L., O'Connor, L.E., & Wade, N.G. (2001).
    Dispositional forgiveness: Development and construct validity of the Transgression
    Narrative Test of Forgiveness (TNTF). Personality and Social Psychology Bulletin; 27,
    1277-1290. doi: 10.1177/0146167201227100004

Blanton, H., & Jaccard, J. (2006). Arbitrary metrics in psychology. *American
    Psychologist*, *61*(1), 27.

Brown, T. A., Chorpita, B. F., Korotitsch, W., & Barlow, D. H. (1997). Psychometric

   properties of the Depression Anxiety Stress Scales (DASS) in clinical samples.

   Behaviour Research and Therapy, *35*, 79-89.

Brose, L.A., Rye, M.S., & Lutz-Zois, C. (2005) Forgiveness and personality traits.

   Personality and Individual Differences 39, 35-46. Doi: 10.1016/j.paid.2004.11.001

Buechsel, R (2010). D*evelopment of an implicit measure of dispositional forgiveness.*

   (Doctoral dissertation). Retrieved from Pro Quest Dissertations and Thesis. (UMI No

   3418958)

Cardi, M., Milich, R., Harris, M.J., & Kearns, E. (2007). Self-esteem moderates the response

   to forgiveness instructions among women with a history of victimization. Journal of

   Research in Personality, 41, 804-819.

Castilho, P., & Gouveia, J. P. (2011). Auto-Compaixão: Estudo da validação da versão

   portuguesa da Escala da Auto-Compaixão e da sua relação com as experiências

   adversas na infância, a comparação social e a psicopatologia.*Psychologica*, (54), p-203.

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 Psychopathol

   Beheavior, *23*, 61-67.

Costa, J., & Pinto-Gouveia, J. (2013). Experiential avoidance and self-compassion in chronic

   pain. *Journal of Applied Social Psychology*, *43*(8), 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.

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of

psychopathology. *Journal of Consulting Psychology, 24*(4) 349-354. doi:

10.1037/h0047358.

Cullen, C., & Barnes-Holmes, D. (2008). Implicit pride and prejudice: A heterosexual

phenomenon? In T. G. Morrison & M. A. Morrison (Eds.), *Modern Prejudice* (pp. 195-

223). New York: Nova Science

Cullen, C., Barnes-Holmes, D., Barnes-Holmes, Y., & Stewart, I. (2009). The Implicit

Relational Assessment Procedure (IRAP) and the malleability of ageist attitudes. *The

Psychological Record*, *59*(4), 591-620.

Dasgupta, N., & Greenwald, A. G. (2001). On the malleability of automatic attitudes:

combating automatic prejudice with images of admired and disliked individuals.

Journal of personality and social psychology, *81*(5), 800.

Devos, T., & Banaji, M. R. (2005). American= white? *Journal of personality and social

psychology*, *88*(3), 447.

De Figueiredo, J. M. (1980). Some methodological remarks on transcultural interviewing on

psychopathology. Journal of Social Psychiatric, 26 (4), 280-308

Dodds, P. S., Clark, E. M., Desu, S., Frank, M. R., Reagan, A. J., Williams, J. R., Mitchell,

Lewis, Harris, K.D., Kloumann,I.M., Bagrow, J.P., Megerdoomian, K., McMahon,

M.T., Tivnan, B.F., & Danforth, C. M. (2014). Human language reveals a universal

positivity bias. *arXiv preprint arXiv:1406.3855*.

Egloff, B., & Schmukle, S. C. (2002). Predictive validity of an Implicit Association Test for

assessing anxiety. *Journal of personality and social psychology*, *83*(6), 1441.

Exline, J. J., & Baumeister, R. F. (2000). Expressing forgiveness and repentance.

Forgiveness: Theory, Research, and Practice, 133-155.

Fisher, M.L., & Exline, J.J. (2006). Self-forgiveness versus condoning: The importance of accepting responsibility. *Self and Identity, 5,* 127-146.

Fergusson, D. M., Horwood,J. L., & Ridder, E. M. (2006). Abortion in young women and subsequent mental health. *Journal of Child Psychology and Psychiatry*, *47*(1), 16-24.

Gangdev, P. (2009). Forgiveness: A note for psychiatrists. *Indian Journal of Psychiatry*, *51*(2), 153. Retrieved from EBSCOhost. (Acession No 38697124)

Gawronski, B., & De Houwer, J. (in press). Implicit measures in social and personality psychology. In H. T. Reis, & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology (2nd edition)*. New York: Cambridge University Press

Gilbert, P. (2009). *The compassionate mind*. Constable & Robinson.

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

Goldring, J (2011). *The development of an implicit association test for measuring forgiveness.* (Doctoral dissertation). Retrieved from Adelaide Research & Scholarship. (On line accession hhtp://hdl.handle.net/2440/67104)

Golijani-Moghaddam, N., Hart, A., & Dawson, D. L. (2013). The implicit relational assessment procedure: emerging reliability and validity data. *Journal of Contextual Behavioral Science*, *2*(3), 105-119.

Greenwald, A. G., & Farnham, S. D. (2000). Using the Implicit Association Test to measure self-esteem and self-concept. *Journal of Personality and Social Psychology, 79,* 1022-1038.

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and Using the

Implicit Association Test: I. An Improved Scoring Algorithm. *Journal of Personality*

and Social Psychology, *85*, 197-216.

Greenwald, A. G. (2003). On doing two things at once: III. Confirmation of perfect

timesharing when simultaneous tasks are ideomotor compatible. *Journal of*

Experimental Psychology: Human Perception and Performance, *29*(5), 859.

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.

Hayes, SC (1993). Analitic goals and the varieties of scientific contextualism. In: S.C. Hayes

L.J. Hayes, H.W. Reese, & T.R. Sarbin (Eds.), *Varieties of scientific contextualism*

(pp./11-27). Reno, NV: Context Press

Hayes, S.C., Barnes-Holmes, D. & Roche, B. (Eds.). (2001). *Relational Frame Theory: A*

Post-Skinnerian account of human language and cognition. New York: Plenum Press.

ISBN 0-306-46600-7

Hayes, S. C., Barnes-Holmes, D., & Wilson, K. G. (2012). Contextual behavioral science:

Creating a science more adequate to the challenge of the human condition. *Journal of*

Contextual Behavioral Science, *1*(1), 1-16.

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. doi: 10.1016/j.brat.2005.06.006

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. Retrieved from EBSCOhost. (Acession No 1996-07086-005)

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

Hodgson, L. K., & Wertheim, E. H. (2007). Does good emotion management aid forgiving? Multiple dimensions of empathy, emotion management and forgiveness of self and others. *Journal of Social and Personal Relationships, 24*, 931-949

Hoyt, W. T., Fincham, F. D., McCullough, M. E., Maio, G., & Davila, J. (2005). Responses to interpersonal transgressions in families: Forgivingness, forgivability, and relationship-specific effects. *Journal of Personality and Social Psychology*, *89*, 375-394

Hoyt, W.T., & McCullough, M.E. (2005). Issues in the multimodal measrument of forgiveness. In E.L. Worthington Jr (Ed.), *Handbook of forgiveness* (pp.109-124). New

Hughes, S., Barnes-Holmes, D., & De Houwer, J. (2011). The dominance of associative theorising in implicit attitude research: Propositional and behavioral alternatives. *The Psychological Record, 61*, 465–498

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,* 17-38.

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*(4), 573-582.

Jones, E E., Sigall, H. (1971). The bogus pipeline: A new paradigm for measuring affect and attitude. *Psychological Bulletin, 76.* 349-364

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*(9), 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*(1), 17-28.

Kohlenberg, R. J., & Tsai, M. (1991). *Functional analytic psychotherapy* (pp. 169-188).

Springer US.

Leach, M. M., & Lark, R. (2004). Does spirituality add to person. ality in the study of trait

forgiveness? Personality and Individual Differences, 37, 147-156. doi:

10.1016/j.paid.2003.08.007

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*(3), 335-343.

Maltby, J., Macaskill, A. and Day, L (2001). Failure to forgive self and others: a replication

and extension of the relationship between forgiveness, personality, social desirability

and general health. *Personality and Individual Differences*. *181*, 1-6.

Mauger,P.A, Perry.J.E., Freeman, T.,Grove, D.C., McBride, A.G., Mckinney,K. (1992). The

measurement of forgiveness. Preliminary research. *Journal of Psychology and

Christianity, 11*, 170-1

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.doi: 10.1016/j.brat.2004.04.004

McCullough, M. E. (2001). Forgiveness: Who does it and how do they do it? *Current Directions in Psychological Science, 10*, 194-197

McCullough, M. E., Worthington, E. L., Jr., & Rachal, K. C. (1997). Interpersonal forgiving in close relationships. *Journal of Personality and Social Psychology, 73*, 321-336.

McCullough, M.E., Rachal, K.C., Sandage, S.J., Worthington, E.L., Jr., Brown, S.W., & Hight, T.L. (1998). Interpersonal forgiving in close relationships: II. Theoretical elaboration and measurement. *Journal of Personality and Social Psychology, 75,* 1586-1603

Neff, K. D. (2003). The development and validation of a scale to measure self-compassion. *Self and identity*, *2*(3), 223-250.

Neff, K. D., Pommier, E. (2013). The relationship between self-compassion and other-focused concern among college undergraduates, community adults, and practicing meditators. *Self and Identity, 12(2),* 160-176.

Neff, K. D., Kirkpatrick, K. L., & Rude, S. S. (2007). Self-compassion and adaptive psychological functioning. *Journal of research in personality*, *41*(1), 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*(3), 922-930.

Nicholson, E., Hopkins-Doyle, A., Barnes-Holmes, D., & Roche, R. A. (2014). Psychopathology, Anxiety or Attentional Control: Determining the variables which predict IRAP performance. *The Psychological Record*, *64*(2), 179-188.

Nieuwenhuijsen, K., De Boer, A. G. E., Verbeek, J. H. A. M., Blank, R. W. B., & Dijik, F. J.
H. (2003).The Depression Anxiety Stress Scales (DASS): Detecting anxiety disorder
and depression in employee absent from work because of mental health problems.
Occupational Environmental Medicine, *60*, 177-182.

Nosek, B.A.; Banaji, M.R., Greenwald, A.G.(2002) Math = male, me = female, therefore
math ≠me *Journal of Personality and Social Psychology, 83* (1), pp. 44–59

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.
Psychologica, 36, 235-246.

Payne, B. K., Burkley, M. A., & Stokes, M. B. (2008). Why do implicit and explicit attitude
tests diverge? The role of structural fit. *Journal of personality and social
psychology*, *94*(1), 16.

Pepper, S. C. (1942). *World hypotheses: A study in evidence*. Berkeley, CA: University of
California Press.

Pinto-Gouveia, J., & Castilho, P. (2006). Escala de Auto-Compaixão. *Tradução e adaptação
para a população portuguesa (non-published manuscript provided by the authors)*.

Power, P. M., Barnes-Holmes, D., Barnes-Holmes, Y., & Stewart, I. (2009). The Implicit
Relational Assessment Procedure (IRAP) as a measure of implicit relative preferences:
A first study. *The Psychological Record, 59,* 621-640.

Powers, C., Nam, R., Rowatt, W. C., Hill, P. (2007). Associations between humility, spiritual transcendence, and forgiveness. In R. Piedmont (Ed.), Research in the Social Scientific Study of Religion (pp. 74-94). Herndon, VA: Brill Academic Publishers

Raskin, R., Novacek, J., & Hogan, R. (1991). Narcissistic self-esteem management. *Journal of Personality and Social Psychology*, *60*(6), 911.doi: 10.1037/0022-3514.60.6.911

Roddy, S., Stewart, I., & Barnes-Holmes, D. (2010). 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.

Roddy, S., Stewart, I., & Barnes-Holmes, D. (2012). Facial reactions reveal that slim is good but fat is not bad: Implicit and explicit measures of body-size bias. *European Journal of Social Psychology.*41(*6*), 688-694

Sirois, F.M. (2004). Procrastination and counterfactual thinking: Avoiding what might have been. *British Journal of Social Psychology, 43*, 269-286.

Strelan, P. (2007). The prosocial, adaptive qualities of just world beliefs: Implications for the relationship between justice and forgiveness. *Personality and Individual Differences*, *43*(4), 881-890.

Tangney, J.P., Boone, A.L., & Dearing, R. (2005). Forgiving the self: conceptual issues and empirical findings. In E.L. Worthington Jr (Ed), *Handbook of forgiveness* (pp.143-158). New York: Brunner-Routledge.

Tangney, J.,Boone, A.L.,& Dearing, R.L. (2005). *Shame and guilt*. New York: Guilford Press.

Thompson, L.Y., Snyder, C.R., Hoffman,L., Michael,S.T, Rasmussen, H.N., Billings, L.S. (2005). Dispositional forgiveness of self, others, and situations. *Journal of Personality*, *73*, 313-359.

Tirch, D., Schoendorff, B., & Silberstein, L. R. (2014). *The ACT practitioner's guide to the science of compassion: Tools for fostering psychological flexibility*. New Harbinger Publications

Tsai, M., Callaghan, G. M., Kohlenberg, R. J., Follette, W. C., & Darrow, S. M. (2009). Supervision and therapist self-development. In *A Guide to Functional Analytic Psychotherapy* (pp. 1-32). Springer US.

Vahey, N., Boles, S., & Barnes-Holmes, D. (2010). Measuring adolescents' smoking-related social identity preferences with the Implicit Relational Assessment Procedure (IRAP) for the first time: A starting point that explains later IRAP evolutions. *International Journal of Psychology and Psychological Therapy, 10*, 453-474.

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.

Vahey, N. A., Barnes-Holmes, D., & Barnes-Holmes, Y. Stewart, I. (2009). A first test of the Implicit Relational Assessment Procedure (IRAP) as a measure of self-esteem: Irish prisoner groups and university students. *The Psychological Record*, *59*, 371-388.

Vitz, P. C., & Meade, J. M. (2011). Self-forgiveness in psychology and psychotherapy: A critique. *Journal of Religion and Health*, *50*(2), 248-263.

Watson, P. J., & Hickman, S. E. (1995). Narcissism, self-esteem, and parental nurturance. *Journal of Psychology*, *129*, 61-74.doi: 10.1080/00223980.1995.9914948

Wilson, T. D., & Dunn, E. (2004). Self-knowledge: Its limits, value, and potential for improvement. *Annual Review of Psychology, 55,* 493-518.

Wohl, M. J. A., DeShea, L., & Wahkinney, R. L. (2008). Looking within: The state forgiveness scale.*Canadian Journal of Behavioural Science, 40,* 1-10. doi:10.1037/0008-400x.40.1.1.1.

Wohl, M. J., Pychyl, T. A., & Bennett, S. H. (2010). I forgive myself, now I can study: How self-forgiveness for procrastinating can reduce future procrastination. *Personality and Individual Differences*, *48*(7), 803-808.

Worthington Jr, E.L. (1998). The pyramid model of forgiveness: Some interdisciplinary speculations about unforgiveness and the promotion of forgiveness. In E.L. Worthington Jr (Ed), *Dimensions of Forgiveness* (107-138). Radnor, Pennsylvania: Templeton Foundation Press.

Yamaguchi, S., Greenwald, A. G., Banaji, M. R., Murakami, F., Chen, D., Shiomura, K., & Krendl, A. (2007). Apparent universality of positive implicit self-esteem. *Psychological Science*, *18*(6), 498-500.

Yamhure, Thompson, L., Snyder, C. R., & Hoffman, L. (2005). Heartland Forgiveness Scale.

Zechmeister JS, Romero C. (2002) Victim and offender accounts of interpersonal conflict: Autobiographical narratives of forgiveness and unforgiveness. *Journal of Personality and Social Psychology, 82*, 675–686

Appendix 1: *Scale based on the Forgiveness of Self and Others IRAP*

When something does not go as planned or something goes wrong in our lives, we often engage in some sort of evaluation of the situation and the people involved, including ourselves. However, the way in which we evaluate ourselves and others can be quite different. Please read the following sentences carefully and circle the number that best describes how much each statement is true for you.

Not at all like me	A little bit like me	Moderately like me	Quite a bit like me	Extremely like me
0	1	2	3	4

My Shortcomings are:

1	**Unacceptable**		0	1	2	3	4
2	**Okay**		0	1	2	3	4
3	**Forgivable**		0	1	2	3	4

4	Normal	0	1	2	3	4
5	Awful	0	1	2	3	4
6	Terrible	0	1	2	3	4
7	Fine	0	1	2	3	4
8	Embarrassing	0	1	2	3	4
9	Intolerable	0	1	2	3	4

My Failures are:

1	Unacceptable	0	1	2	3	4
2	Okay	0	1	2	3	4
3	Forgivable	0	1	2	3	4
4	Normal	0	1	2	3	4
5	Awful	0	1	2	3	4
6	Terrible	0	1	2	3	4
7	Fine	0	1	2	3	4
8	Embarrassing	0	1	2	3	4
9	Intolerable	0	1	2	3	4

My Weaknesses are:

1	Unacceptable	0	1	2	3	4
2	Okay	0	1	2	3	4
3	Forgivable	0	1	2	3	4
4	Normal	0	1	2	3	4
5	Awful	0	1	2	3	4
6	Terrible	0	1	2	3	4

7	Fine	0	1	2	3	4
8	Embarrassing	0	1	2	3	4
9	Intolerable	0	1	2	3	4

My Faults are:

1	unacceptable	0	1	2	3	4
2	Okay	0	1	2	3	4
3	Forgivable	0	1	2	3	4
4	Normal	0	1	2	3	4
5	Awful	0	1	2	3	4
6	Terrible	0	1	2	3	4
7	Fine	0	1	2	3	4
8	Embarrassing	0	1	2	3	4
9	Intolerable	0	1	2	3	4

My Flaws are:

1	unacceptable	0	1	2	3	4
2	Okay	0	1	2	3	4
3	Forgivable	0	1	2	3	4
4	Normal	0	1	2	3	4
5	Awful	0	1	2	3	4
6	Terrible	0	1	2	3	4
7	Fine	0	1	2	3	4
8	Embarrassing	0	1	2	3	4

9	Intolerable	0 1 2 3 4

My Mistakes are:

1	Unacceptable	0 1 2 3 4
2	Okay	0 1 2 3 4
3	Forgivable	0 1 2 3 4
4	Normal	0 1 2 3 4
5	Awful	0 1 2 3 4
6	Terrible	0 1 2 3 4
7	Fine	0 1 2 3 4
8	Embarrassing	0 1 2 3 4
9	Intolerable	0 1 2 3 4

Other People's Shortcomings are:

1	unacceptable	0 1 2 3 4
2	Okay	0 1 2 3 4
3	Forgivable	0 1 2 3 4
4	Normal	0 1 2 3 4
5	Awful	0 1 2 3 4
6	Terrible	0 1 2 3 4
7	Fine	0 1 2 3 4
8	Embarrassing	0 1 2 3 4
9	Intolerable	0 1 2 3 4

Other People's Failures are:

1	unacceptable	0	1	2	3	4
2	Okay	0	1	2	3	4
3	Forgivable	0	1	2	3	4
4	Normal	0	1	2	3	4
5	Awful	0	1	2	3	4
6	Terrible	0	1	2	3	4
7	Fine	0	1	2	3	4
8	Embarrassing	0	1	2	3	4
9	Intolerable	0	1	2	3	4

Other People's Weaknesses are:

1	unacceptable	0	1	2	3	4
2	Okay	0	1	2	3	4
3	Forgivable	0	1	2	3	4
4	Normal	0	1	2	3	4
5	Awful	0	1	2	3	4
6	Terrible	0	1	2	3	4
7	Fine	0	1	2	3	4
8	Embarrassing	0	1	2	3	4
9	Intolerable	0	1	2	3	4

Other People's Faults are:

1	unacceptable	0	1	2	3	4
2	Okay	0	1	2	3	4

3	Forgivable		0	1	2	3	4
4	Normal		0	1	2	3	4
5	Awful		0	1	2	3	4
6	Terrible		0	1	2	3	4
7	Fine		0	1	2	3	4
8	Embarrassing		0	1	2	3	4
9	Intolerable		0	1	2	3	4

Other People's Flaws are:

1	Unacceptable		0	1	2	3	4
2	Okay		0	1	2	3	4
3	Forgivable		0	1	2	3	4
4	Normal		0	1	2	3	4
5	Awful		0	1	2	3	4
6	Terrible		0	1	2	3	4
7	Fine		0	1	2	3	4
8	Embarrassing		0	1	2	3	4
9	Intolerable		0	1	2	3	4

Other People's Mistakes are:

1	Unacceptable		0	1	2	3	4
2	Okay		0	1	2	3	4
3	Forgivable		0	1	2	3	4
4	Normal		0	1	2	3	4

5	Awful	0 1 2 3 4
6	Terrible	0 1 2 3 4
7	Fine	0 1 2 3 4
8	Embarrassing	0 1 2 3 4
9	Intolerable	0 1 2 3 4

Appendix 2:*The scale based on the Feelings and Outcomes-IRAP*

Generic Feeling IRAP Scale

The questions below are brief and are designed to gauge your experience of failing or succeeding in some way.

For each of the following statements, please indicate how true it is for you when you have to face an experience of failure or success, marking an X, using the following scale

When I fail in some way, I feel:

1	2	3	4	5	6	7
Completely False			Neither false nor true			Completely true

	1	2	3	4	5	6	7
Bad							
Guilty							

Stupid							
Useless							
Frustrated							
Angry							
Good							
Strong							
Energetic							
In control							
Calm							
Peaceful							

When I succeed in some way, I feel:

1	2	3	4	5	6	7
Completely False			**Neither false nor true**			**Completely true**

	1	2	3	4	5	6	7
Bad							
Guilty							
Stupid							
Useless							
Frustrated							
Angry							

Good						
Strong						
Energetic						
In control						
Calm						
Peaceful						

Generic Outcomes IRAP Scale

The questions below are brief and are designed to gauge what you think the consequences or outcomes of failing or succeeding in some way might be for you.

In my opinion, failing:

1	2	3	4	5	6	7
Completely False			**Neither false nor true**			**Completely true**

	1	2	3	4	5	6	7
Wastes my time							

Undermines my motivation							
Has negative consequences							
Makes me look bad							
Makes me less productive							
Makes me look stupid							
Saves my time							
Keeps me motivated							
Has positive consequences							
Makes me look good							
Makes me more productive							
Makes me look intelligent							

In my opinion, succeeding:

1	2	3	4	5	6	7
Completely False			**Neither false nor true**			**Completely true**

	1	2	3	4	5	6	7
Wastes my time							
Undermines my motivation							
Has negative consequences							
Makes me look bad							

Makes me less productive							
Makes me look stupid							
Saves my time							

Appendix :

Has positive consequences							
Makes me look good							
Makes me more productive							
Makes me look intelligent							

Appendix 3: *The Problem Behaviours Questionnaire and scoring*

The Problem Behaviours Questionnaire

The twelve questions below are brief and are designed to gauge your experiences of behaving in ways that you don't want, didn't plan to, or don't like. For example, you might find that you can't resist sweet things when you're on a diet. You promise yourself that you won't have it before, but when the opportunity presents itself, you just do it anyway. Then, maybe afterwards, you are filled with guilt and so you make the same promise for the next time and hope that on that occasion you might be more successful.

1. Do you ever do things that you don't like or had promised yourself that you wouldn't do? *Please circle one response*

Yes   (1 point)   No (0 points)

2. Please could you give one or two examples of the types of things you do?

3. As an estimate, how often do you think that you engage in this or any other type of behaviour that you have tried not to do? *Please circle one response*

Daily (4 points)   *Weekly* (3 points)   *Monthly* (2 points)   *Rarely* (1 point)

4. If you do things you didn't intend to or don't like, do you find that these are always the same things over and over? *Please circle one response.*

Yes (1 point)   No (0 points)

5. Do you think that **other people** do things that they try not to? *Please circle one response*

Yes (0 points)   No (1 point)

6. As an estimate, how often do you think **other people** engage in this or any other type of behaviour that they have tried not to do? *Please circle one response*

Daily (1 point)   *Weekly* (2 points)   *Monthly* (3 points)   *Rarely* (4 points)

7. Below is a list of feelings that might show up for you after you have done something you hoped not to do. *Please tick ALL* that apply to you and feel free to add any others that are not listed here

Guilty   **Hopeless**   **Helpless**   **Others:** _____

Angry   **Regretful**   **Frustrated**

Stupid   **Out of control**   **Weird**

8. Do you think that other people feel the same as you when they do unwanted things, or are you more sensitive or more self-critical? *Please circle one response*

a) Others probably feel the same (0 points)

b)   I am probably more sensitive/more self-critical (1 point)

9. When you think about the reasons why you might continue to do these things, what do you come up with? Please try and summarise below what you have concluded about your own actions.

10. How hard would you say that you have tried to change this type of behaviour, even if it doesn't appear to have worked? *Please place an X at one point along the line*

_____

0%   **25%**   **50%**   **75%**   **100%**
Not tried very hard   Have tried somewhat   Tried very hard
(0 point)   (1 point)   (2 points)   (3 points)   (4 points)

11. Do you think you will be caught in the same sort of loop forever? *Please circle one response*.

Yes (1 point)   No (0 point)

12. Have you any thoughts on how your situation might be changed? *Please summarise these thoughts or solutions here.*

Appendix 4: *Table with target and sample stimuli of all participants*

	Feeling IRAP		Outcomes IRAP	
Sample	*Facebook makes me feel*	*Studying makes me feel*	*Facebook*	*Studying*
Target	Bad	Good	Wastes my time;	A good use of my time
	Guilty	Strong	Undermines my confidence;	Increases my confidence
	Stupid	Wise	Undermines my	
	Useless	In control		Increases my

	Frustrated Angry	Calm Peaceful	success; Makes me less focused; Makes me lazy; Reduces my concentration	success Helps me focus Makes me productive Helps my concentration
Sample	*Junk food makes me feel* *Over-sleeping makes me feel*	*Good food makes me feel* *Rising early makes me feel*	*Junk food* *Rising early*	*Good food* *Rising early*
Target	Bad Guilty Stupid Useless Frustrated Angry	Positive Strong Energetic In control Calm Peaceful	Has negative consequences Undermines motivation Undermines confidence Makes me tired Makes me unhealthy Makes me lazy	Has positive consequences Increases motivation Increases confidence Gives me energy Makes me healthy Keeps me active
Sample	*Bad food makes me feel*	*Good food makes me feel*	*Bad Food*	*Good Food*
Target	Bad Guilty Stupid Useless Frustrated Angry	Good Strong Energetic In Angry control Calm Peaceful	Increases my weight, Makes me unhealthy Undermines my confidence Makes me tired Makes me lazy Makes me unattractive	Controls my weight Increases my confidence Makes me healthy Gives me energy Makes me attractive Helps me

				concentrate
Sample	*Over-sleeping makes me feel*	*Rising early makes me feel*	*Over-sleeping*	*Rising early*
Target	Bad Guilty Stupid Useless Frustrated Angry	Positive Strong Energetic In control Calm Peaceful	Has negative consequences Undermines motivation Undermines confidence Makes me tired Makes me unhealthy Makes me lazy	Has positive consequences Increases motivation Increases confidence Gives me energy Makes me healthy Keeps me active
Sample	*Sweets Soft drinks*	*Fruit Healthy drinks*	*Sweets Soft drinks*	*Fruits Healthy drinks*
Target	Bad Guilty Stupid Hopeless Frustrated Angry	Good Strong Energetic Positive Happy Peaceful	Increase my weight Make me unhealthy Undermine my confidence Make me tired Make me lazy Makes me unattractive	Control my weight Increase my confidence Make me healthy Give me energy Make me attractive Help concentration

Sample	*Fatty food makes me feel*	*Healthy food makes me feel*	*Fatty Food*	*Healthy Food*
Target	Bad

Guilty
Stupid

Useless

Frustrated

Angry | Good

Strong

Energetic

In control

Calm

Peaceful | Increases my weight

Makes me unhealthy

Undermines my confidence

Makes me tired

Makes me lazy

Makes me unattractive | Controls my weight

Increases my confidence

Makes me healthy

Gives me energy

Makes me attractive

Helps me concentrate |
| **Sample** | *Too much smoking* | *Controlled smoking* | *Too much smoking* | *Controlled smoking* |
| **Target** | I feel Bad

I feel Guilty
I feel Stupid

I feel Useless

I feel Frustrated

I feel Angry | I feel Good

I feel Strong

I feel Energetic

I feel Positive

I feel Calm

I feel Peaceful | Wastes money

Makes me unhealthy

Undermines my confidence

Makes me tired

Makes me lazy

Makes me unattractive | Saves money

Increases my confidence

Makes me healthy

Gives me energy

Makes me attractive

Maintains my motivation |
| **Sample** | *Over drinking makes me feel* | *Controlled drinking makes me feel* | *Over drinking* | *Controlled drinking* |

Target	Bad	Good	Increases my weight	Maintains my weight
	Guilty	Strong	Makes me unhealthy	Increases my confidence
	Stupid	Energetic	Undermines my confidence	Makes me healthy
	Useless	Positive	Makes me tired	Gives me energy
	Frustrated	Calm	Makes me lazy	Makes me attractive
	Angry	Peaceful	Makes me unattractive	Helps me concentrate
Sample	*Working too much*	*Giving myself time off*	*Working too much*	*Giving myself time off*
Target	Annoyed	Satisfied	Has negative consequences	Has positive consequences
	Vulnerable	Positive	Traps me	Free me
	Stressed	Healthy	Wastes my time	A good use of my time
	Tired	Energetic	Undermines my motivation	Maintains my motivation
	Sad	Attractive	Makes me unfulfilled	Makes me fulfilled
	Unattractive	Motivation	Reduces self confidence	Builds self-confidence
Sample	*Eating badly*	*Eating normally*	*Eating badly*	*Eating normally*
	Too much alcohol	*Controlled drinking*	*Too much alcohol*	*Controlled drinking*

Target	Unbalanced	Balanced	Increases my weight	Maintains my weight
	Guilty	Strong	Makes me unhealthy	Increases my confidence
	Stupid	Energetic	Undermines my confidence	Makes me healthy
	Useless	Positive	Makes me tired	Gives me energy
	Frustrated	Calm	Makes me lazy	Makes me attractive
	Angry	Peaceful	Makes me unattractive	Helps me concentrate
Sample	*Unhealthy food*	*Healthy food*	*Unhealthy food*	*Healthy food*
	Drinking too much	*Controlled drinking*	*Drinking too much*	*Controlled drinking*
Target	Bad	Good	Increases my weight	Maintains my weight
	Guilty	Strong	Makes me unhealthy	Increases my confidence
	Stupid	Energetic	Undermines my confidence	Makes me healthy
	Useless	Positive	Makes me tired	Gives me energy
	Frustrated	Calm	Makes me lazy	Makes me attractive
	Angry	Peaceful	Makes me unattractive	Helps me concentrate
Sample	*Too much smoking*	*Controlled smoke*	*Too much smoking*	*Controlled smoke*
Target	I feel Bad	I feel Positive	Undermines my confidence	Saves money
	I feel Stupid	I feel Peaceful	Makes me unhealthy	Gives me energy
	I feel Useless	I feel Calm	Wastes money	Increases my confidence
	I feel Frustrated	I feel Energetic	Makes me	Maintains my

	I feel Angry I feel Guilty	I feel Strong I feel Good	unattractive Makes me tired Makes me lazy	motivation Makes me attractive Makes me healthy
Sample	*Spending too much makes me feel*	*Spending wisely makes me feel*	*Spending wisely*	*Too much spending*
Target	Guilty Angry Bad Useless Stupid Frustrated	Peaceful Calm Strong In control Energetic Good	Frees me Increases confidence Strengthens relationships Avoids problems Gives me choices Makes me balanced	Makes me foolish Traps me Undermines confidence Creates difficulties Spoils relationships Creates difficulties
Sample	*Deixando de exercitar-me sinto; Perder a cabeça me faz sentir*	*Fazendo exercício me sinto; Controlando me sinto*	*Deixando de exercitar-me Perdendo a cabeça*	*Fazendo exercício Controlando-me*

Target	Mal,Culpado,Estupido, Remorso,Frustrado, Raiva	Bem**,** Forte**,** Animado, Segura,Calmo, Energico	Crio dificuldades Acaba com minha confiança Me aprisiono Comprometo a saúde Desperdiço meu tempo Fico vulnerável	Evito problemas Aumento minha confiança Me liberto Me dá alternativas Aproveito meu tempo Fico equilibrado
Sample	*Evitando dirigir me sinto*	*Dirigindo me sinto*	*Evitando dirigir*	*Dirigindo*
Target	Raivoso,Fora do controle, Frustrado, Culpado, Mal, Com remorso	Bem,Calmo,Fort e, Livre,Satisfeito, No controle	Dinimuo minha confianca, Pareco ruim, Diminuo meu sucesso, Desperdico tempo, Fico preocupado, Perco oportunidades	Fico resolvido, Fortaleco minhas relacoes, Uso melhor o tempo, Me liberto, Aumento meu sucesso, Aumento minha confianca,
Sample	*Faltar na academia me faz sentir*	*Ir para academia me faz sentir*	*Faltar na academia*	*Fazer exercício*

Target	Mal, Culpado, Estupido, Remorso, Frustrado Raiva	Bem, Forte, Animado, No controle, Calmo, Energico,	Cria dificuldades Acaba com minha confiança Me aprisiona Comprometo a saúde Desperdiça meu tempo Gasto dinheiro à toa	Evita problemas Aumenta minha confiança Me liberta Me dá alternativas Aproveito meu tempo Fico equilibrado
Sample	*Reclamar da pessoa me faz sentir; Comer poucos vegetais me faz sentir*	*Ser assertivo me faz sentir; Comer mais vegetais me faz*	*Reclamar da pessoa; Comer poucos vegetais*	*Ser assertivo; Aceitar a pessoa*
Target	Mal, Remorso, Estupido, Chateado, Frustrado, Aborrecido	Bem, Forte, Resolvido, No controle, Calmo, Tranquilo	Têm consequencias negativas Me aprisiona As coisas ficam iguais Fico pouco atraente Não me satisfaz Nao e saudavel	Têm consequencias positivas Me liberta Atinjo minha meta Fico mais atraente Me satisfaz Melhoro qualidade de vida
Sample	*Procrastinando eu me sinto* *Burlando a dieta eu me sinto*	*Focando eu me sinto* *Seguindo a dieta eu me sinto*	*Burlar a dieta* *Procrastinar*	*Seguir a dieta* *Focar na tarefa*

Target				
Target	Mal Culpada Decepcionada Chateada Frustrada Estúpida	Bem Forte Energetica No controle Calma Tranquila	Não é saudável Diminui minha confiança Dificulta meu sucesso Pareço ruim Aumento minha preocupação Fico presa	Atinjo minha meta Aumento minha confiança Aumento meu successo Pareço determinada Torno-me produtiva Me liberta
Sample	*Não ser assertivo me faz sentir*	*Ser assertivo me faz sentir*	*Não ser assertivo*	*Ser assertivo*
Target	Mal Culpado Estupido Chateado Frustrado Raivoso	Bem Forte Resolvido No controle Calmo Tranquilo	Têm consequencias negativas Me aprisiona As coisas ficam iguais Diminui minha motivação Não me satisfaz Diminui minha auto-confiança	Têm consequencias positivas Me liberta Problemas são resolvidos Mantém minha motivação Me satisfaz Cria auto-confiança
Sample	*Deixando de exercitar-me sinto;* *Perder a cabeça me faz sentir*	*Fazendo exercício me sinto;* *Controlando me sinto*	*Deixando de exercitar-me;* *Perdendo a cabeça*	*Fazendo exercício;* *Controlando-me*

| **Target** | Mal, Culpado, Estupido, Remorso, Frustrado, Raiva, | Bem, Forte, Animado, Segura, Calmo, Energico, | Crio dificuldades

Acaba com minha confiança

Me aprisiono

Comprometo a saúde

Desperdiço meu tempo

Fico vulnerável | Evito problemas

Aumento minha confiança

Me liberto

Me dá alternativas

Aproveito meu tempo

Fico equilibrado |
|---|---|---|---|---|
| **Sample** | *Burlando a dieta me sinto* | *Mantendo a dieta me sinto* | *Burlar dieta* | *Manter dieta* |
| **Target** | Mal

Culpado

Estúpido

Fora do controle

Frustrado

Raivoso | Bem

Forte

Positivo

No controle

Calmo

Disposto | Aumenta meu peso

Não é saudável

Diminui a auto-confiança

Me torna cansado

Me torna preguiçoso

Me torna menos atraente | Atinjo a meta

Aumenta a auto-confiança

Me faz saudável

Me dá energia

Me torna atraente

Ajuda na concetração |
| **Sample** | *Cobrando eu me sinto* | *Relevando eu me sinto* | *Cobrando* | *Relevando* |

Target			
Target	Mal Fora do controle Triste Desesperançosa Frustratada Chata	Bem Em paz Disposta Legal Calma Resolvida	Crio dificuldades Diminuo minha confiança Dificulta a mudança Pareço ruim Aumento minha preocupação Desgasto a relação	Evito problemas Aumento minha confiança Aumento meu sucesso Pareço ponderada Fico sossegada Mantenho a paz
Sample	*Não ser assertiva me faz sentir*	*Sendo assertiva eu me sinto*	*Não ser assertiva*	*Sendo assertiva*
Target	Arrependida Culpada Estupida Chateada Frustrada Raivosa	Bem Forte Resolvida No controle Calma Tranquila	Têm consequencias negativas Me aprisiona As coisas ficam iguais Diminui minha motivação Não me satisfaz Diminui minha auto-confiança	Têm consequencias positivas Me liberta Problemas são resolvidos Mantém minha motivação Me satisfaz Cria auto-confiança
Sample	*Comendo muito chocolate me sinto*	*Diminuindo chocolate me sinto*	*Muito chocolate*	*Diminuindo chocolate*

Target	Mal	Bem	Aumenta meu peso	Mantenho o peso
	Culpado	Forte	Não é saudável	Aumenta a auto-confiança
	Estúpido	Positivo	Diminui a auto-confiança	Me faz saudável
	Fora do controle	No controle	Me torna dependente	Equilibro a dieta
	Frustrado	Calmo	Desequilibro a dieta	Me torna atraente
	Raivoso	Disposto	Fico pouco atraente	Fico equilibrado
Sample	*Sair com a pessoa me faz sentir*	*Ser assertiva me faz sentir*	*Sair com a pessoa*	*Ser assertiva*
Target	Mal	Bem	Me aprisiona	Problemas são resolvidos
	Culpada	Forte	Não me satisfaz	Têm consequencias positivas
	Estupida	Resolvida	Diminui minha auto-confiança	Cria auto-confiança
	Chateada	No controle	Têm consequencias negativas	Me liberta
	Frustrada	Calma	As coisas ficam iguais	Mantém minha motivação
	Aborrecida	Tranquila	Diminui minha motivação	Me satisfaz
Sample	*Procrastinando eu me sinto*	*Focando eu me sinto*	*Procrastinar*	*Focar na tarefa*

Target	Frustrado	Calmo	Desperdiça meu tempo	Administro melhor o tempo
	Aborrecido	No controle	Diminui minha confiança	Aumenta minha confiança
	Desesperançoso	Energetico	Dificulta meu sucesso	Aumenta meu successo
	Com remorso	Bem	Me faz parecer ruim	Me faz parecer determinado
	Mal	Forte	Aumenta minha preocupação	Torno-me produtivo
	Culpado	Tranquilo	Reduz minha concentração	Ajuda minha concentração