



# Developing the Implicit Relational Assessment Procedure (IRAP) to Assess Evaluations of Death Within Suicidality

Thesis submitted to the Department of Psychology, Faculty of Science,  
in part fulfillment of the requirements for the degree of  
Doctor of Philosophy, Maynooth University.

Ian Hussey, B.A. (Hons.)

July 2015

Research Supervisor: Professor Dermot Barnes-Holmes

Head of Department: Dr Andrew Coogan

## Table of contents

<b>Acknowledgements</b>	iii
<b>Published sections</b>	iv
<b>Abstract</b>	vi
<b>Ethics statement</b>	viii
<b>Life of Tables</b>	ix
<b>List of Figures</b>	x
<b>List of Appendices</b>	xi
<b>Chapter 1: A review of the prediction of suicidality using self-report and implicit measures</b>	1
1.1 Definitions and theories of suicide	1
1.2 The prediction of suicidal behaviours	8
1.3 A systematic review of implicit measures and suicidal behaviours	10
1.4 A functional-analytic approach to language, suicide, and implicit attitudes	37
1.5 Overview of the current research program	49
<b>Chapter 2: Comparing the IRAP and IAT as measures of death-identity and death-evaluation in a normative population</b>	52
2.1 Method	55
2.2 Results	68
2.3 Discussion	74
<b>Chapter 3: An empirical investigation of the role of the contrast category within the IRAP</b>	78
3.1 Experiment 1	79
3.2 Method	80
3.3 Results	83
3.4 Discussion	86
3.5 Experiment 2	88
3.6 Method	91
3.7 Results	94
3.8 Discussion	97
<b>Chapter 4: Exploring death evaluations on the IRAP using mortality salience inductions and contrast category manipulations</b>	102
4.1 Experiment 1	104
4.2 Method	105
4.3 Results and Discussion	110
4.4 Experiment 2	115
4.5 Method	115
4.6 Results and Discussion	117
4.7 Experiment 3	119
4.8 Method	120
4.9 Results and Discussion	121
4.10 Experiment 4	124
4.11 Method	125
4.12 Results and Discussion	127
4.13 Post Hoc Analyses and Discussion	129

<b>Chapter 5: Comparing death evaluations on the IRAP between normative and suicidal individuals</b>	135
5.1 Method	137
5.2 Results	142
5.4 Discussion	149
<b>Chapter 6: General discussion</b>	155
6.1 Summary of findings	155
6.2 Issues raised	163
6.3 Broader implications	174
6.4 Conclusions	177
<b>References</b>	179
<b>Appendices</b>	213

## **Acknowledgements**

### **Dermot**

My heartfelt thanks for affording me such academic freedom, for helping me exorcise my mentalistic assumptions, and for keeping me on track and productive when I was inclined to run amuck in the sweet shop. I am very grateful to call you my mentor. This was great fun.

### **Yvonne**

Thank you for helping me get in touch with my values. You bring fiery compassion to a sometimes-indifferent world. Please save me a glass of Kool-Aid: I'm still thirsty.

### **John & Martina**

Thank you for the genetic and developmental endowment of your combined sharp minds; the quiet encouragement to do something with my life I value and enjoy; and the loud financial support that allowed me to pursue it. Few are gifted with such freedom and privilege.

### **Aoife, Francesca, Mary, Nicole, Sean, and everyone else**

Thank you for making a one-horse town livable. Our earnest time together as students will always be special to me.

### **The undergraduate students I worked with while conducting this research**

Having people to work with made this research infinitely more interesting. Thanks in particular to Dearbhaile Ní Mhaoileoin, Emma Fawcett, June Dennehy, and Tarah Daly.

### **The service users and staff at St Patrick's Hospital**

Thank you for your patience, support, and for generously volunteering your time. Thanks in particular to Dr. Richard Booth, Rose Keane, Georgina Mullen, Aisling Ryan and all those who participated in this research.

### **The participants at Maynooth University**

Thank you for volunteering your time. You deserved more than the chocolates you got.

### **The staff at the department of psychology**

Anne, you hold the keys to the kingdom; thank you for using them to open doors for me whenever you could. Carol and Richard, thank you for you input and advice, I thoroughly enjoyed our collaborations. Derek, Victoria, and Caroline, thank you for your kind support whenever it was needed.

### **Published sections of this thesis**

Sections of this thesis are currently published or under review. The references that correspond to each chapter are listed below. Note that some chapters contain sections from multiple papers, and material from some papers is distributed across multiple chapters.

**Chapter 1:** A review of the prediction of suicidality using self-report and implicit measures<sup>1,2,4</sup>

**Chapter 2:** Comparing the IRAP and IAT as measures of death-identity and death-evaluation in a normative population<sup>5,8</sup>

**Chapter 3:** An empirical investigation of the role of the contrast category within the IRAP<sup>7</sup>

**Chapter 4:** Exploring death evaluations on the IRAP using mortality salience inductions and contrast category manipulations<sup>6</sup>

**Chapter 5:** Comparing death evaluations on the IRAP between normative and suicidal individuals<sup>3</sup>

**Chapter 6:** General discussion<sup>1,2,3,5,6,7</sup>

### References

- <sup>1</sup>Barnes-Holmes, D., Barnes-Holmes, Y., **Hussey, I.**, & Luciano, C. (2015). Relational Frame Theory: Finding its historical and intellectual roots and reflecting upon its future development. In R. D. Zettle, S. C. Hayes, D. Barnes-Holmes, & A. Biglan (Eds.), *The Wiley Handbook of Contextual Behavioral Science* (pp. 115–128). John Wiley & Sons, Ltd. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1002/9781118489857.ch8/summary>
- <sup>2</sup>**Hussey, I.** & Barnes-Holmes, D. (under review). A systematic review of implicit measures and suicidal behaviors.

- <sup>3</sup>**Hussey, I.,** Barnes-Holmes, D., & Booth, R. (2016). Individuals with current suicidal ideation demonstrate implicit “fearlessness of death.” *Journal of Behavior Therapy and Experimental Psychiatry*, 51, 1–9. <http://doi.org/10.1016/j.jbtep.2015.11.003>
- <sup>4</sup>**Hussey, I.,** Barnes-Holmes, D., & Barnes-Holmes, Y. (2015). From Relational Frame Theory to implicit attitudes and back again: clarifying the link between RFT and IRAP research. *Current Opinion in Psychology*, 2, 11–15. <http://doi.org/10.1016/j.copsyc.2014.12.009>
- <sup>5</sup>**Hussey, I.,** Daly, T., & Barnes-Holmes, D. (2015). Life is Good, But Death Ain’t Bad Either: Counter-Intuitive Implicit Biases to Death in a Normative Population. *The Psychological Record*, 65(4), 731–742. <http://doi.org/10.1007/s40732-015-0142-3>
- <sup>6</sup>**Hussey, I.,** Dennehy, J., Fawcett, J., Barnes-Holmes, D., & Murphy, C. (under review). Negative evaluations of death on the IRAP are dependent on mortality salience and the context set by the other stimulus categories.
- <sup>7</sup>**Hussey, I.,** Mhaoileoin, D. N., Barnes-Holmes, D., Ohtsuki, T., Kishita, N., Hughes, S., & Murphy, C. (2016). The IRAP Is Nonrelative but not Acontextual: Changes to the Contrast Category Influence Men’s Dehumanization of Women. *The Psychological Record*, 66(2), 291–299. <http://doi.org/10.1007/s40732-016-0171-6>
- <sup>8</sup>**Hussey, I.,** Thompson, M., McEnteggart, C., Barnes-Holmes, D., & Barnes-Holmes, Y. (2015). Interpreting and inverting with less cursing: A guide to interpreting IRAP data. *Journal of Contextual Behavioral Science*, 4(3), 157–162. <http://doi.org/10.1016/j.jcbs.2015.05.001>

## Abstract

Over the last fifty years, psychological science can be credited with persistent efforts to prevent and treat suicidal behaviours. However, the utility of such interventions is often moderated by the ability to identify individuals who are likely to engage in suicidal behaviors ahead of time. Less progress has been made on this front; the ability to accurately predict such behaviors at an individual level remains limited. Recent evidence suggests that this may be due in part to the field's reliance on self-report measures. One promising avenue that has emerged in recent years is the objective behavioural tasks referred to as "implicit measures", which have shown greater promising predictive validity relative to existing risk factors derived from self-report measures. The current body of research sought to expand on these findings and, specifically, to explore implicit evaluations of death in both normative and suicidal individuals using the Implicit Relational Assessment Procedure (IRAP).

This thesis begins with a systematic review of the literature on implicit measures and suicide. Based on the conclusions of this review, five analogue studies were conducted exploring implicit studies to death in normative student populations. Two further experiments developed a novel experimental methodology (i.e., manipulations of the IRAP's "contrast category") in order to facilitate a more detailed understanding of what drives effects on the measure. A final experiment compared implicit death-evaluations on the IRAP between psychiatric patients with current suicidal ideation and normative controls. All studies presented participants with one or more implicit measures (IRAP and IAT) and a number of self-report measures. Generically, the IRAP presents participants with four category pairings (e.g., "life-positive", "life-negative", "death-positive", and "death-negative"), and compares the relative ease with which they respond to these pairings with "true" relative to "false". The difference in mean response-latency between the two response options is referred to as an implicit bias.

Together, these studies allowed for a more detailed interrogation of how death is evaluated between individuals with and without a history of suicidal behaviours than was previously possible. Specifically, across studies, results demonstrated the IRAP's ability to isolate specific implicit biases between categories, relative to other measures. The results from the analogue studies were used to development and assess a death-evaluation IRAP that is sensitive to mortality salience. A final experiment to conclude that suicidal ideation was found to be associated with a specific rejection of the negativity (i.e., fearlessness) of death.

Results across studies indicate that differential patterns of implicit bias between normative individuals and suicidal ideators were attributable to suicidality specifically rather than the salience of mortality more generally, thereby providing a degree of construct validity for death-evaluations on the IRAP. Importantly, these effects are consistent with leading theories of suicide (e.g., Interpersonal Theory and Integrated Motivational-Volitional model), and are in line with our stated goal to attempt to ground the effects found on implicit measures more closely with existing theory. Overall, results suggest that the assessment of implicit death-evaluations on the IRAP represents a good candidate for future research on the prospective prediction of suicidal behaviours.

## **Ethics statement**

Ethical approval was granted for this research by the National University of Ireland Maynooth Social Research Ethics Subcommittee and, where appropriate, St Patrick's Mental Health Services Research Ethics Committee (Protocol 20/11). Details are provided within each experimental chapter. Letters of ethical approval can be found in the appendices. All research was conducted within the Psychological Society of Ireland's Code of Professional Ethics (2011).

## List of Tables

Table 1.1	Key psychological risk and protective factors for suicidal behaviours	7
Table 1.2	Stimulus categories employed within implicit measures and number of publications examining each	17
Table 1.3	Prospective criterion validity of implicit measures for future suicidal behaviours in emergency psychiatric admissions	17
Table 1.4	Construct validity of implicit measures for suicidal behaviours	20
Table 2.1	Stimulus sets and category labels for the death-identity IAT	58
Table 2.2	Stimulus sets for the death-identity IRAP	60
Table 2.3	Stimulus sets for the death-evaluation IRAP	61
Table 2.4	Structure of the IAT	62
Table 2.5	Correlations between the implicit and self-report measures.	73
Table 3.1	Stimulus sets for the self-Santa and self-Hitler esteem IRAPs	82
Table 3.2	Correlations between the implicit and self-report measures	86
Table 3.3	Stimulus sets for the Gender and Agency IRAPs	91
Table 3.4	Correlations between the IRAPs' trial-types and the self-report measures	97
Table 4.1	Stimuli used in Experiment 1's death-evaluation IRAP	105
Table 4.2	Stimuli used in Experiment 2's death-evaluation IRAP	115
Table 4.3	Stimuli used in Experiment 3's death-evaluation IRAP	120
Table 4.4	Stimuli used in Experiment 4's death-evaluation IRAP	125
Table 5.1	Stimuli used in the Personal and Abstract death–evaluation IRAPs.	141
Table 5.2	Demographic and self-report data for the normative and suicidal ideation groups	143
Table 5.3	The Abstract and Personal death–evaluation IRAPs as predictors of group membership (normative vs. suicidal ideation)	146
Table 5.4	Classification table for the personal death–evaluation IRAP's "My death–negative" trial-type in predicting group membership	149

## List of Figures

Figure 1.1	Three key stimulus categories within suicide	32
Figure 2.1	The IAT trial screen	63
Figure 2.2	The four IRAP trial-types	65
Figure 2.3	Performance on the death-identity IRAP and IAT	71
Figure 2.4	Performance on the death-evaluation IRAP	71
Figure 3.1	The stimulus categories employed in the self-Hitler and self-Santa IRAPs	82
Figure 3.2	Performance on the Self-Santa and Self-Hitler IRAPs	85
Figure 3.3	The stimulus categories employed in the Gender and Agency IRAPs	90
Figure 3.4	Performance on the Gender and Agency IRAPs	95
Figure 4.1	Example “Weeks to Live” sheet for a 20-year-old female	107
Figure 4.2	Performance on Experiment 1’s death-evaluation IRAP at baseline and after mortality salience induction	113
Figure 4.3	Performance on Experiment 2’s death-evaluation IRAP at baseline and after mortality salience induction	118
Figure 4.4	Performance on Experiment 3’s death-evaluation IRAP at baseline and after mortality salience induction	123
Figure 4.5	Performance on Experiment 4’s death-evaluation IRAP at baseline and after mortality salience induction	128
Figure 4.6	The stimulus sets employed in the four IRAPs used in this chapter	130
Figure 4.7	Differences on the “my death-negative” trial-type between the four death-evaluation IRAPs, at baseline and after mortality salience induction	132
Figure 5.1	Performance on the IRAPs between the normative and suicidal ideation groups	145
Figure 5.2	Receiver Operating Characteristic curve for the personal death-evaluation IRAP’s “my death-negative” trial-type as a predictor of group membership	148
Figure 6.1	Stimulus relations within the IRAP that are assessed versus those that are commonly assumed by researchers, but not assessed by the task	173

## **List of Appendices**

Appendix A	Flowchart of Included studies	213
Appendix B	Table of excluded articles	214
Appendix C	Acceptance and Action Questionnaire II	215
Appendix D	Belief in the Afterlife Scale	216
Appendix E	Death Anxiety Scale	217
Appendix F	Depression Anxiety Stress Scales	218
Appendix G	Informed Consent Form: Maynooth University Site	219
Appendix H	Consent Form & Researcher Checklist: St. Patrick's University Hospital Site	220
Appendix I	IRAP experimenters' script	222
Appendix J	Rosenberg Self-Esteem Scale	225
Appendix K	Attitudes Towards Women Scale	226
Appendix L	Likelihood to Sexually Harass Scale	228
Appendix M	Mortality Attitudes Personality Survey	236
Appendix N	Mortality Salience Induction: Example "Weeks to Live" Sheet	237
Appendix O	Mortality Salience Induction Script	238
Appendix P	Self-Injurious Thoughts and Behaviours Interview	239
Appendix Q	Letter of ethical approval: Maynooth University	240
Appendix R	Letter of ethical approval: St. Patrick's University Hospital	241

# **CHAPTER 1: A REVIEW OF THE USE OF IMPLICIT MEASURES WITHIN THE STUDY OF SUICIDAL BEHAVIOURS**

This chapter outlines the rationale for the current program of research by arguing that suicide is a leading cause of death worldwide and that there are strong economic as well as moral arguments for addressing it. It goes on to define what is meant by suicide and suicidal behaviour, before summarising contemporary psychological theories of suicide. It then provides a critique of our ability to predict suicidal behaviours that is rooted in our heavy reliance on self-reports. Next, a new and promising area of research is then outlined, which focuses on the use of computer-based reaction time tests referred to as measures of implicit attitudes. A systematic review of the literature on implicit measures within suicidal behaviours is then provided. Based on the recommendations of this review, the empirical agenda for the current thesis is then outlined. In particular, an overarching functional-analytic approach to both suicidal behaviour and the results produced by implicit attitudes is outlined.

## **1.1 Definitions and theories of suicide**

### **Quantifying the scale of the issue**

Suicide is now recognised to be a leading cause of death worldwide. According to a recent report by the World Health Organisation, at least 800,000 individuals died by suicide in 2012 (11.4 per 100,000). To put this in perspective, this constitutes the majority of all violent deaths (WHO, 2014). In the context of Ireland, specifically, an average of 500 individuals die by suicide each year (11.3 per 100,000: National Office of Suicide Prevention, 2013). This issue also bears additional hidden costs beyond the deaths of these individuals. Recent research suggested that for each individual Irish adolescent who dies by suicide, twenty attempt suicide, hundreds are admitted to hospital for self-inflicted wounds, and thousands engage in self-harm in the community (McMahon et al., 2014; WHO, 2014). Additionally, prospective studies have shown that the presence of suicidal thoughts or

feelings increases the likelihood of death from *non-suicidal* causes as well as suicide (Suominen, Isometsä, Ostamo, & Lönnqvist, 2004). Suicide has thus been referred to as an economic and healthcare “ice-berg”, representing a burden of disease far above and beyond the deaths directly associated with it (McMahon et al., 2014). Indeed, the broad economic costs associated with suicide have been estimated to have cost Ireland just under one billion euro in 2001, or roughly 1% of GDP (Kennelly, 2007). As such, in addition to the moral and health-policy arguments for furthering our ability to predict and treat suicidality, there is therefore also a strong economic argument for doing so (see also Doessel & Williams, 2010).

### **Defining suicide**

The definition of suicide has seen much debate over the last three decades. This can be attributed to the fact that there is a wide range of stakeholders involved in the prediction, quantification, prevention, and treatment of suicide, spanning the fields of medicine, psychology, law, economics and public policy. It is therefore useful to define these terms as they apply to the current research program. A review by Silverman (2006) found that at least 15 distinct definitions of suicide are frequently cited in the literature. Common to all these definitions, however, is the idea that such behaviour must be both self-inflicted (to distinguish it from murder) and intentional (to distinguish it from accidental or natural causes). For the purpose of the current body of research, we adopt the definitions provided in a recent review by O’Connor and Nock (2014), which will be block quoted here for the purposes of precision:

Suicide is the act of an individual intentionally ending their own life. We use the general term suicidal behaviour to refer to thoughts and behaviours related to an individual intentionally taking their own life. These thoughts include the more specific outcomes of suicide ideation, which refers to an individual having thoughts about intentionally taking their own life; suicide plan, which refers to the formulation

of a specific plot by an individual to end their own life; and suicide attempt, which refers to engagement in a potentially self-injurious behaviour in which there is at least some intention of dying as a result of the behaviour. We also refer to self-harm, defined by the National Institute for Health and Care Excellence as intentional self-poisoning or self-injury, irrespective of motive. (p. 1)

It should be noted that more detailed taxonomies of suicidal behaviours do exist (e.g., O'Carroll et al., 1996), but such distinctions do not provide additional utility in the current context. Indeed, the authors of such taxonomies note that such taxonomies have not seen widespread use in academic or clinical work (Silverman, Berman, Sanddal, O'Carroll, & Joiner, 2007a).

### **Theories of suicide**

For several decades, research on suicidal behaviours operated within the descriptive framework of Beck's triad of suicidal behaviour (i.e., ideation, planning, and attempts) by attempting to find variables that are predictive of, or which differentiate between, these three categories. That is, in stark contrast with the majority of psychological research in other disciplines, research in this domain was, historically, predominantly a-theoretical in nature. From the 1990s, however, commentators began to argue that theory was needed in order to integrate the results of existing research and organise and guide future investigation (see Barzilay & Apter, 2014; Silverman, 2006). A number of theoretical accounts of suicide have since been proposed over the last two decades, including Escape from Self Theory (Baumeister, 1990), the Cry of Pain Model (J. Mark G. Williams, 1997), Psychache Theory (Shneidman, 1993), and emotional dysregulation model (Linehan, 1993; Crowell, Beauchaine, & Linehan, 2009; see also Barzilay & Apter, 2014 for a review of these theories). Critically, however, contemporary theorising has united these varied theories under the umbrella of two, arguably dominant theories: the Interpersonal Theory of suicide (IPT: Joiner, 2005; Van

Orden et al., 2010) and the Integrated Motivational-Volitional model of suicidal behaviour (IMV: R. C. O'Connor, 2011). These theories have been described as being “diathesis-stress in origin and cognitive in focus” and, in contrast to earlier theories, they attempt to account for why only a small fraction of individuals who experience suicidal ideation go on to attempt suicide (R. C. O'Connor & Nock, 2014). This section will outline the core theses of these two influential theories and highlight some of the key differences between them.

At this point it is important to note that while the current research program has been influenced *by* IPT and IMV, it is not conducted *within* either of them. Specifically, the current research program is conducted within a functional-analytic rather than social-cognitive paradigm (see (D. Barnes-Holmes & Hussey, 2015) for a paper-length treatment on the essential fracture between social-cognitive and functional-analytic theorising). As such, these descriptions of IPT and IMV will be deliberately brief, without reference to the historical foundations of, or evidence base for each.

**Interpersonal Theory of Suicide (IPT).** This theory (Joiner, 2005; Van Orden et al., 2010) posits that acute or chronic stressors in an individual’s life may produce high levels of “perceived burdensomeness” (i.e., that one is a burden on others) and “thwarted belongingness” (i.e., feeling that one is alienated from others). The co-occurrence of both perceived burdensomeness and thwarted belongingness, in addition to high levels of hopelessness regarding the potential for change of both these variables, results in suicidal ideation. The transition from suicidal ideation to suicide attempt is itself moderated by the “acquired capability for suicide”, which is subdivided into 1) a reduced fear of death and 2) an increased tolerance for physical pain. This capability is acquired through “habituation processes”, primarily through contact with physical-painful acts of self-harm.

**Integrated Motivational Volitional Model of Suicidal Behaviour (IMV).** This theory (R. C. O'Connor, 2011) argues that suicide is a behavioural repertoire that develops

across distinct “motivational” and “volitional” phases. These two phases describe the factors that influence the development of suicidal ideation and attempts, respectively. The motivational phase argues that an interplay between a biological diathesis and chronic or acute stressful circumstances causes some individuals to experience feelings of “defeat and humiliation” (i.e., perceptions of low social rank). For some individuals, these feelings progress to a sense of “entrapment” (i.e., feeling unable to escape said defeat or humiliation). This progression is moderated by the threat these circumstances pose to the sense of self (which is assessed using variables such as social problem solving and coping skills). For some individuals, this sense of entrapment then develops into suicidal ideation, where this progression is moderated by motivational factors such as thwarted belongingness and perceived burdensomeness, as well as others such as social support and personal goals for the future. In the volitional phase, the transition from suicidal ideation to suicide attempt is itself moderated by an acquired capability for suicide, but also other factors such as exposure to and imitation of the suicidal behaviour of others, having access to the means of suicide, sufficient planning, and/or impulsivity.

**Comparing and evaluating the IMV and IPT.** It is useful to consider the similarities and differences between IPT and IMV. First, whereas the IPT argues that the development of suicidal ideation is due to hopelessness about both belongingness and burdensomeness, the IMV argues suicidal ideation is instead due to the combination of both defeat and entrapment. Second, whereas the IPT argues that the movement from ideation to attempts is moderated by the acquired capability for suicide, the IMV lists the acquired capability as just one of several moderators (R. C. O’Connor & Nock, 2014). As such, the IPT can be characterised as an account of the key psychological variables that stem from interpersonal factors specifically, whereas the IMV can be characterised as a more holistic attempt to describe the psychological pathways to suicide. What is common to these theories

(and indeed most others: see Barzilay & Apter, 2014 for review) is the notion that suicide is a learned behaviour whose purpose is to escape or avoid intolerable psychological pain, in the belief that no solutions exist other than the cessation of self (i.e., the verbal rule that “the only way to stop my pain is to kill myself”).

Their differences aside, both theories have been praised as attempts to provide an integrated descriptive framework that can accommodate much of the empirical data to date. However, both have also been argued to fall short of a key goal: distinguishing between individuals who will continue to think about suicide and those who will go on to make an attempt (Klonsky & May, 2014). It is therefore useful to re-consider the aims of such theorising based on critiques of clinical psychology more generally.

Generally speaking, the purpose of theorising is to organise previous and future research to meet analytic goals (e.g., the prediction, prevention and treatment of suicide). As with any area of scientific research, a bi-directional relationship exists between theory and empirical evidence; theory is expanded or changed to fit data and new research is frequently steered by theoretical predictions. In the current context, the support for, and foundations of, IPT and IMV come from research on a very large number of psychological variables that have been implicated in suicidal behaviour. For example, 32 variables were included in a recent authoritative review (R. C. O’Connor & Nock, 2014; see Table 1.1). However, it is critical to note that a general criticism of theorising in clinical psychology is the narrow focus on construct validity over predictive utility and behaviour change, despite the fact that it is the latter two that are most relevant to clinical practice (Vervliet & Raes, 2013; see also Baker, McFall, & Shoham, 2008). It is therefore useful to consider which variables have been shown to be prospectively predictive of actual suicidal attempts or death by suicide (as opposed to those which are supported via indirect evidence). In short, we argue that the validity criterion for theories of suicide must be grounded in the ability to predict and

influence behaviour. While this approach to psychological science is best typified by the functional-analytic approach (e.g., Chiesa, 1994; Hayes & Brownstein, 1986), it has general and growing support within the study of self-harmful behaviours (Bentley, Nock, & Barlow, 2014; Nock, 2009b) and clinical psychology more generally (De Houwer, Barnes-Holmes, & Barnes-Holmes, 2015; Kazdin, 2008). The following section therefore provides a review of the evidence for the prediction of suicide.

Table 1.1. *Key psychological risk and protective factors for suicidal behaviours (reproduced from O'Connor & Nock, 2014, p. 5)*

---

<b>Personality and individual differences</b>
Hopelessness
Impulsivity
Perfectionism
Neuroticism and extroversion
Optimism
Resilience
<b>Cognitive factors</b>
Cognitive rigidity
Rumination
Thought suppression
Autobiographical memory biases
Belongingness and burdensomeness
Fearlessness about injury and death
Pain insensitivity
Problem solving and coping
Agitation
Implicit associations
Attentional biases
Future thinking
Goal adjustment
Reasons for living
Defeat and entrapment
<b>Social factors</b>
Social transmission
Modelling
Contagion
Assortative homophily (i.e., clustering of suicides among peers)
Exposure to deaths by suicide of others
Social isolation
<b>Negative life events</b>
Childhood adversities
Traumatic life events during adulthood
Physical illness
Other interpersonal stressors
Psychophysiological stress response

---

## 1.2 The prediction of suicidal behaviours

Over the last fifty years, psychological science can be credited with persistent efforts to predict and prevent suicide. For example, suicidal behaviours have been shown to be treatable via high intensity one-to-one cognitive behavioural interventions (see Tarrier, Taylor, & Gooding, 2008 for meta analysis) and low intensity population level interventions (e.g., in schools: Wasserman et al., 2015). However, the utility of treatment is often moderated by the ability to identify individuals who are likely to engage in suicidal behaviours ahead of time. Less progress has been made on this front; the ability to accurately predict such behaviours at an individual level remains limited, regardless of whether psychometric questionnaires, static risk factors (e.g., previous self-harm), or risk assessment scales are employed (De Leo, 2002; Klonsky & May, 2014; see Batterham et al., 2014; Larkin, Di Blasi, & Arensman, 2014; E. O'Connor, Gaynes, Burda, Williams, & Whitlock, 2013; Randall, Colman, & Rowe, 2011, respectively, for systematic reviews of population level research, in hospital settings, primary care, and emergency rooms). Furthermore, evidence for the prediction of suicidal behaviours within a clinically-meaningful timescale (hours, days, or weeks) is almost absent (Glenn & Nock, 2014b; Klonsky & May, 2014; Rudd et al., 2006; Silverman & Berman, 2013).

Commentators have argued that the limited utility of such measures is symptomatic of a larger issue: the almost exclusive reliance on self-report measures in clinical psychology (MacLeod, 1993; Nock et al., 2010). That is, despite the formal differences between psychometric questionnaires, structured interviews, and even risk assessment checklists, all rely on an individual's introspective access to subjective, private or internal behaviours. While the limitations of introspection are a well-trodden path within psychology (Nisbett & Wilson, 1977; Wilson, 2009), it appears that such limitations are particularly problematic within suicidal behaviours. For example, individuals who commit suicide commonly deny

experiencing suicidal ideation immediately prior to the attempt (Busch, Fawcett, & Jacobs, 2003); self-forecasts of future suicidal behaviours have very poor predictive validity (Janis & Nock, 2008), and such forecasts have low reliability across time (Eikelenboom, Smit, Beekman, Kerkhof, & Penninx, 2014). Above and beyond the typical limits of introspection seen elsewhere in psychology, individuals therefore appear to be unable, or possibly unwilling, to report their suicidal intentions. Of course, such limitations are not exclusive to the suicidal individual: clinical judgment too has been shown to have low reliability and predictive validity, with interpretations of what constitutes “high risk” differing widely between clinicians (see Berman & Silverman, 2014).

Importantly, self-report measures (e.g., of entrapment, hopelessness, suicidal ideation, and intent) do not see widespread use in clinical practice. Arguably, this speaks, in part, to the perceived utility of these measures outside of the highly rarefied context of academic research. For example, an observational study on the use of suicide risk assessment measures in English hospitals found “indiscriminate use of [unverified] risk scales in clinical services” (Quinlivan et al., 2014, p.1). Of the 32 emergency departments and mental health services surveyed, none employed any of the above measures. Only a minority of sites (31%) employed any psychometrically-tested measure (other than the above), however even when a site did employ a psychometrically-tested measure, it was most frequently a commonly-used checklist of risk factors and warning signs known as the SAD PERSONS mnemonic index (i.e., a check for the risk factors of “Male sex, Age (<19 or >45 years), Depression, Previous attempt, Excess alcohol or substance use, Rational thinking loss, Social supports lacking, Organized plan, No spouse, and [physical] Sickness”: Patterson, Dohn, Bird, & Patterson, 1983). However, despite its wide use, a recent systematic review found that the SAD PERSONS was not predictive of future suicidal behaviour (Warden, Spiwak, Sareen, & Bolton, 2014). Furthermore, the authors note that the scale was “unlikely to be a useful focus

for further study in clinical settings”, and that “ongoing clinical use of the scale could contribute to inappropriate patient management and undesired patient outcomes” (p.323). Results such as this highlight a problematic disjoint between the research activity around these measures and their perceived utility within the therapeutic community.

In response to this, some commentators have called for greater use of “objective” measures, such as the use of behavioural tasks or biological markers (Glenn & Nock, 2014a, 2014b; R. C. O’Connor & Nock, 2014; Randall et al., 2011), and the integration of the results of such measures into algorithmic risk assessment decision making processes (Claassen, Harvilchuck-Laurenson, & Fawcett, 2014; Nock, 2012). The current review therefore focuses on behavioural measures within the study of suicidal behaviours; specifically, the use of a class of procedures frequently referred to as measures of implicit associations (De Houwer, Teige-Mocigemba, Spruyt, & Moors, 2009; Nosek, Hawkins, & Frazier, 2011). Whereas self-report measures explicitly ask individuals to report their attitudes or beliefs, implicit measures typically infer their results from reaction time biases. Implicit measures circumvent the need for an individual to have introspective access to attitudes, beliefs, and future intentions, and instead rely on objective measurements of “automatic” behaviours, which are emitted outside of awareness or intentionality, under low volitional control and/or with high cognitive efficiency (De Houwer et al., 2009). At least one measure has been shown to be prospectively predictive of suicide attempts (Nock et al., 2010), as will be elaborated below. Indeed, this particular study provided the original inspiration for the current body of research. At this point, it is therefore useful to conduct a systematic review of the use of implicit measures within the study of suicidal behaviours.

### **1.3 A systematic review of implicit measures and suicidal behaviours**

Previous reviews of the use of implicit measures within clinically relevant research have been limited to those measures that specifically targeted DSM categories (e.g., Phillips,

Hine, & Thorsteinsson, 2010; Roefs et al., 2011). As such, these reviews have specifically excluded research on clinically relevant behaviours that are not specific to a diagnostic category. The degree to which implicit measures are predictive of suicidal and self-harmful behaviours is therefore unclear. Consequently, the current review sought to assess (a) the prospective criterion validity and (b) the construct validity of implicit measures within the study of suicidal behaviours. A secondary goal was to examine the validity of specific stimulus categories within the implicit measures, in order to better organize future research.

## **Methods**

**Terminology.** At least 15 distinct definitions of suicide are frequently cited in the literature (Silverman, 2006). Common to all these definitions, however, is the concept of self-inflicted, intentional death (O’Carroll et al., 1996; Silverman et al., 2007a); see also Hayes, 1992; Hjelmeland & Knizek, 1999). Studies were therefore excluded if they exclusively examined behaviours that were absent of any intent to die (i.e., “non-suicidal self-injury”). However, studies that remained agnostic to the level of intent were included (i.e., those employing the term “self-harm”: see Platt et al., 1992). The primary outcomes of interest were therefore death by suicide, suicide attempts, intent, planning, communications, gestures, ideation, and self-harm (see O’Carroll et al., 1996).

As the boundary conditions for what qualifies as an implicit measure are not universally agreed upon (De Houwer et al., 2009; Nosek et al., 2011), we limited the list of eligible procedures to those featured in a recent review (Gawronski & De Houwer, 2011). Specifically, we included: the Affective Misattribution Procedure (Payne, Cheng, Govorun, & Stewart, 2005), approach-avoidance tasks (Chen & Bargh, 1999), the Extrinsic Affective Simon Task (De Houwer, 2003), evaluative priming tasks (Fazio, Sanbonmatsu, Powell, & Kardes, 1986), the Go/No-go Association Task (GNAT: Nosek & Banaji, 2001), the Implicit Association Test (IAT: Greenwald, McGhee, & Schwartz, 1998) and its variants (e.g., Brief

IAT: Sriram & Greenwald, 2009; Single Category IAT: Karpinski & Steinman, 2006), the Implicit Relational Assessment Procedure (D. Barnes-Holmes, Barnes-Holmes, Stewart, & Boles, 2010), semantic priming tasks (Wittenbrink, Judd, & Park, 1997), and the Sorting Paired Features Task (Bar-Anan, Nosek, & Vianello, 2009). Specific procedural details will be provided only for measures that were found employed in articles that met eligibility criteria.

**Eligibility criteria.** We followed PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines (Moher, Liberati, Tetzlaff, & Altman, 2009). Articles were included if they employed at least one implicit measure from the above list, and either (a) assessed at least one form of suicide related thoughts or behaviours, and/or (b) employed an implicit measure whose validity within suicidal behaviours had been established elsewhere. This latter criterion was employed so as to allow for the inclusion of analogue studies. No exclusions were made based on the population studied (e.g., children, adolescents, adults, or older adults; clinical vs. non-clinical populations).

**Search strategy.** The PsycINFO (1967–present) and Scopus (1995–present) databases were searched using the terms ("suicid\*" OR "self-harm" OR "self-injury" OR "self-mutilation") AND ("implicit"). The semantic priming task (Fazio et al., 1986) represents the earliest development of an implicit measure among the list of those included in the current review. This was therefore deemed to represent a reasonable start date for our search. We searched for published and in-press peer-reviewed articles reporting empirical research from January 1<sup>st</sup> 1986 to June 1<sup>st</sup> 2015. Database searches were supplemented with reference lists in included articles and citations of included articles. Titles and abstracts were screened to exclude articles that did not meet eligibility criteria. Full texts of the remaining articles were then screened for eligibility. A protocol for the current review was not pre-registered.

**Synthesis of included studies.** The purpose of the current review is to assess the validity of implicit measures in predicting and understanding suicidal behaviours. The results of the included studies were organized into two categories of evidence: prospective criterion validity and construct validity. Prospective criterion validity refers to the measure's ability to predict behaviours of interest ahead of time (e.g., suicide attempts). This was quantified from contingency tables using the standard classification statistics of sensitivity, specificity, positive likelihood ratio (+LR), negative likelihood ratio (-LR), and odds ratio. Should one or more of these statistics not have been reported in the original article, they were calculated here for the purposes of comparison across studies.

In contrast, construct validity refers to the degree to which a given implicit measure assesses what it purports to, and therefore includes a measure's ability to speak to the past and present behaviour between groups or within a group across time points. This included evidence from cohort comparison studies that provide concurrent criterion validity (e.g., differences between ideators and controls) or divergent validity (e.g., differences between attempters and socially anxious individuals), intervention studies that demonstrated sensitivity to therapeutic change (e.g., pre-post intervention), and evidence from analogue studies (e.g., pre-post a-theoretically relevant manipulation). As such, although many of the included articles employed multivariate models in order to assess the relative contributions of the implicit versus self-report measures, only within and between groups comparisons were considered here. In order to make comparisons across studies, group comparisons were quantified via measures of effect size of differences between means (i.e., Cohen's *d*). In all cases, these effect sizes were calculated from the mean, standard deviations, and sample sizes reported in the original articles. While the demonstration of construct validity does not speak to a measure's ability to predict suicidal behaviour ahead of time, it can be useful in

highlighting “good candidate” measures and variables for use within future prospective research (see Nock & Banaji, 2007a; Vervliet & Raes, 2013).

The purpose of the current review is not to directly contrast the evidence for implicit measures against that of traditional self-report measures of suicidality. Indeed, the studies included in the current review typically do not make direct implicit-explicit comparisons, as might be common in other areas of implicit measures research. Indeed, there is good reason for this divergence: in many cases, particularly within construct validity studies, traditional self-report measures are used as criterion measures. As such, no direct comparisons will be made within the narrative synthesis section of this review. However, commonalities and differences are discussed in the conclusions.

## **Results**

Two hundred and fourteen records were found. Duplicates were removed and the titles and abstracts of the remaining 158 articles were compared against the eligibility criteria. The full text of the resulting 24 articles was obtained and were compared against the eligibility criteria. Thirteen articles met all criteria and were included in the final synthesis (see Appendices A and B for flowchart and excluded studies table). Because the study designs, implicit measures and criterion variables varied markedly between studies, we focused on a qualitative synthesis of their results and implications rather than conducting a meta-analysis.

Of the thirteen included articles, one employed the Go/No-go Association Task (GNAT) and the remaining twelve employed the Implicit Association Test (IAT). One of these employed both the IAT and Single Category IAT (i.e., Randall, Rowe, Dong, Nock, & Colman, 2013), and one used both the IAT and the Brief IAT (i.e., Creemers, Scholte, Engels, Prinstein, & Wiers, 2013). However, in both of these cases the authors elected to report only the results of the IAT(s), on the basis that the results of the IAT variants (i.e., Single Category

IAT and Brief IAT) were comparable to the standard IAT itself. As such, the current review will refer only to the results of the IATs included in these studies. These stimulus sets employed within the implicit measures contained in these thirteen articles were grouped into five distinct categories: “self–evaluation”, “self–escape”, “self–death”, “self–cutting”, and “self-harm–evaluation” (see Table 1.2).

It is worth noting that, whereas in many domains of research the utility of implicit measures is directly contrasted with self-report measures (see Roefs et al., 2011, for review), relatively few direct comparisons were made in the studies included in the current review. As such, this review focuses on the validity of such implicit measures rather than predominantly on the direct comparison of the relative utility of implicit versus self-report measures.

**Description of tasks and scoring procedures.** The IAT is a computer-based reaction times test that has been used in several hundred studies within a wide range of clinically relevant domains (see Roefs et al., 2011). The task requires participants to sort stimulus exemplars from four superordinate categories (e.g., death, life, self, and others: Nock et al., 2010) using two response keys, to which two of the four categories are each assigned. This pairing is then alternated across blocks. For example, one block maps both “self” and “death” to one response key and “others” and “life” to the other key, whereas the opposing block would pair “self” and “life” to one response key and “others” and “death” to the other key. In all the included articles, the difference in reaction times between the blocks on the IAT were quantified using the *D* score (Greenwald, Nosek, & Banaji, 2003), which is a variant of Cohen’s *d*.

The GNAT has several similarities to the IAT, and both are based on similar theoretical principles and design features (Nosek & Banaji, 2001). Whereas the IAT requires individuals to classify stimuli into category pairs that alternate across blocks (e.g., self–death/others–life vs. self–life/others–death), the GNAT, in contrast, asks participants to

respond or not based on whether the presented stimulus belongs to the categories or not (e.g., respond when self or death is presented vs. respond when self or life is presented: see Nosek & Banaji, 2001). Effects on the GNAT can be quantified in a number of ways (see Nosek & Banaji, 2001). In the case of the included study (Knowles & Townsend, 2012), the authors calculated difference scores between mean latencies across the two blocks (e.g., self-harm–positive vs. self-harm–negative).

Table 1.2. Stimulus categories employed within implicit measures and number of publications examining each

	Self-harm-evaluation <sup>1</sup>		Self-cutting <sup>2</sup>		Self-death <sup>3</sup>		Self-escape <sup>4</sup>		Valence <sup>5</sup>		Mood (a) <sup>6</sup>		Mood (b) <sup>6</sup>		Mood (c) <sup>7</sup>		
	Valence	Arousal	self-harm	self	others	self	others	self	others	self	others	self	others	self	others	self	others
Category 1	self-harm	self-harm	-	self	others	self	others	self	others	self	others	self	others	self	others	self	others
Category 2	-	-	-	self	others	self	others	self	others	self	others	self	others	self	others	self	others
Attribute 1	positive	arousing	arousing	images of cut skin	images of cut skin	death	death	escape	negative	depressed	depressed	anxious	anxious	shame	shame	shame	shame
Attribute 2	negative	calming	calming	images of uncut skin	images of uncut skin	life	life	stay	positive	elated	elated	calm	calm	anxiety	anxiety	anxiety	anxiety
Number of published articles	1	1	1	3*	3	8	8	3	1	1	1	1	1	1	1	1	1

Notes: Several articles employed more than one study and/or implicit measure. Four variants of the self-death stimulus set were also employed by Randall et al. (2013), however for the sake of parsimony they are not reported here. Implicit measures employed by: <sup>1</sup>Knowles & Townsend (2012); <sup>2</sup>Dickstein et al. (2015), Nock & Banaji (2007b), Randall et al. (2013); <sup>3</sup>Dickstein et al. (2015), Ellis et al. (2015), Harrison et al. (2014), Nock et al. (2010), Price et al. (2009), Price et al. (2014), Randall et al. (2013), Tang et al. (2013); <sup>4</sup>Dickstein et al. (2015); Price et al. (2009), Price et al. (2014); <sup>5</sup>Franck et al. (2007); <sup>6</sup>Glashouwer et al. (2010); <sup>7</sup>Rüsch et al. (2007). \*It is worth noting that the self-cutting IAT has also been used in four additional articles on non-suicidal self-injury that were excluded from the current review (see Appendix B).

Table 1.3. Prospective criterion validity of implicit measures for future suicidal behaviours in emergency psychiatric admissions

Study	Criterion variable	Follow-up period	n	Implicit measure	IAT D score	Group		Sensitivity	Specificity	+LR	-LR	OR [95% CI]	Evidence
						Negative event	No negative event						
Nock & Banaji (2007b)	Ideation	6 months	73	Self-cutting IAT	-	-	-	-	-	-	-	7.5 [2.1, 27.0]	✓
Nock et al. (2010)	Attempts	6 months	91	Self-death IAT	>0	7	15	.50	.81	2.56	0.62	4.1 [1.3, 13.6]	✓
Randall et al. (2013)	Self-harm	3 months	107	Self-cutting IAT	≤0	7	62	-	-	-	-	Non-significant <sup>1</sup>	✗
				Self-death IAT	>0	12	17	.42	.79	2.04	.72	3.0 [1.2, 7.5]	✓

Notes: ✓ = supportive evidence; ✗ = contradictory evidence. “-” = not reported; +LR = positive likelihood ratio; -LR = negative likelihood ratio; OR = diagnostic odds ratio; AUC = area under the curve. D scores produced by the IAT were dichotomized to indicate either overall self-death/others-life effects (D score > 0) or self-life/others-death effects (D score ≤ 0). Nock et al. (2010) recruited emergency psychiatric hospital admissions; Randall et al. (2013) recruited emergency psychiatric hospital admissions presenting with ideation or self-harm. Randall et al. (2013) reported that the self-cutting IAT was not significantly predictive of future self-harm, but did not report specific classification statistics.

**Prospective criterion validity.** Three studies were relevant to prospective criterion validity and reported the results of the implicit measure(s) in sufficient detail to allow for the calculation of classification statistics (see Table 1.3)<sup>1</sup>. All three employed the self–death IAT and/or the self–cutting IAT. First, Nock and Banaji (2007b) compared performance on the self–cutting IAT between adolescents with recent (i.e., past year) suicide attempts, suicidal ideation, and controls. Participants were contacted after a 6-month follow-up period and reassessed for suicidal ideation and self-harm within that period. Results demonstrated that the self–cutting IAT was predictive of future suicidal ideation (OR = 7.5, 95% CI [2.1, 27.0]). Suicide attempts were also assessed at follow-up but these data were not reported due to the low number of incidences in the sample.

Nock and colleagues (Nock et al., 2010) demonstrated that effects on the self–death IAT were predictive of future attempts within a 6-month follow-up period. A cut-off value of *D* score > 0 was selected on the basis that it represents a procedurally-meaningful value (i.e., the point at which participants are biased towards categorizing self with death faster than self with life, or vice versa), at which point the IAT demonstrated fair sensitivity and good specificity (see Table 1.3). Specifically, participants who demonstrated self–death effects were roughly four times more likely to make an attempt within the follow-up period (OR = 4.1, 95% CI [1.3, 13.6]). Furthermore, the IAT was found to be more predictive of future attempts than several known risk factors, including meeting the criteria for a depressive disorder, a history of multiple previous attempts, scores on the Beck Scale for Suicidal Ideation, their attending physician’s clinical judgment, or the individual’s own forecast of their likelihood to make a future attempt.

---

<sup>1</sup> A fourth study by Ellis and colleagues (2015) also employed a prospective design, however the authors did not

Finally, a study by Randall and colleagues (2013) attempted to replicate and extend the results generated by Nock et al. (2010). Four changes were made to the design: time to follow-up was shortened to 3 months, and the negative event criterion was widened from suicide attempt to self-harm. Third, a larger number of known risk factors were assessed using self-report measures; specifically level of education, comorbid depressive or psychotic symptoms, nature of self-harm (i.e., overdose vs. non-overdose), hopelessness, suicidal ideation, SAD PERSONS risk assessment (Patterson et al., 1983; Warden et al., 2014), impulsivity, a global symptom severity index, and screenings for both alcohol and drug abuse. Fourth and finally, in addition to the self–death IAT employed by Nock et al. (2010), this study also employed the self–cutting IAT, as well as four novel variants of these two stimulus sets. The similarity of these variants was such that they will not be discussed here in detail (see Randall et al., 2013 for details). An identical cut-off score was used to Nock et al. (2010). Results showed that the self–death IAT was found to be predictive of future self-harm (OR = 3.0, 95% CI [1.2, 7.5]) and the measure showed fair sensitivity and good specificity (see Table 1.3). In contrast, the self–cutting IAT and the novel variants were found not to be predictive of self-harm.

Results across the studies therefore indicate that the self–cutting IAT may possess predictive validity for suicidal ideation, but not for suicide attempts or self-harm. In contrast, the self–death IAT was shown to be predictive of future ideation upon discharge from inpatient treatment, as well as both suicide attempts and self-harm in the three-to-six months after an emergency psychiatric admission. None of the included articles assessed the prospective criterion validity of other measures of implicit associations (e.g., using the self–evaluation, self–escape or self-harm–evaluations stimulus sets, or using tasks other than the IAT).

Table 1.4. Construct validity of implicit measures for suicidal behaviours

Implicit measure	Study	Design	Group comparisons	<i>n</i>	Effect size [95% CI]	Form of validity assessed	Support evidence	
Self-harm–evaluation GNATs <i>Valence Arousal</i> Self-cutting IAT	Knowles & Townsend (2012, study 1)	Cohort comparison Cohort comparison	Recent self-harm <sup>2</sup> vs. controls	72	-0.35 [-0.84, 0.15]	Construct	×	
			Recent self-harm <sup>2</sup> vs. controls	72	0.29 [-0.2, 0.78]	Construct	×	
	Dickstein et al. (2015)	Cohort comparison	Recent attempts <sup>1</sup> vs. controls	90	0.17 [-0.25, 0.58]	Construct	×	
			Recent attempts <sup>1</sup> vs. recent non-suicidal self-injury	93	-1.06 [-1.48, -0.61]	Discriminant	× <sup>3</sup>	
			Recent ideation <sup>2</sup> vs. controls	75	0.77 [0.3, 1.24]	Construct	✓	
			Recent attempts <sup>2</sup> vs. controls	52	-1.42 [-2.07, -0.73]	Construct	✓	
	Self-death IAT	Dickstein et al. (2015)	Cohort comparison	Recent attempts <sup>1</sup> vs. controls	90	0.2 [-0.22, 0.61]	Construct	×
				Recent attempts <sup>1</sup> vs. recent non-suicidal self-injury	93	-1.82 [-2.29, -1.33]	Discriminant	× <sup>3</sup>
		Harrison et al. (2014)	Cohort comparison	Lifetime attempts vs. controls	408	0.31 [0.02, 0.61]	Construct	✓
				Current suicidal ideation (50% also with lifetime attempts): admission vs. discharge from inpatient program (mean = 6 weeks)	118	0.28 [0.02, 0.54]	Construct	✓
Price et al. (2009)		Pre-post intervention	Current ideation: baseline vs. 24 hours post pharmacotherapy (ketamine)	10	-0.37 [-1.16, 0.45]	Construct	×	
			Current ideation: baseline vs. 24 hours post pharmacotherapy (ketamine)	36	-0.06 [-0.53, 0.4]	Construct	×	
	Current ideation: baseline vs. 24 hours post active placebo		21	-0.38 [-0.98, 0.24]	-	-		
	Failure induction vs. controls		92	0.52 [0.1, 0.93]	Construct	✓		
Tang et al. (2013)	Experimental analogue	Success induction vs. controls	92	-0.57 [-0.98, -0.15]	Construct	✓		
		Recent attempts <sup>1</sup> vs. controls	90	-0.2 [-0.62, 0.21]	Construct	×		
Self-escape IAT	Dickstein et al. (2015)	Cohort comparison	Recent attempts <sup>1</sup> vs. recent non-suicidal self-injury	93	0.34 [-0.07, 0.75]	Discriminant	×	
			Current ideation: baseline vs. 24 hours post pharmacotherapy (ketamine)	10	1.37 [0.35, 2.28]	Construct	✓	
	Price et al. (2009)	Pre-post intervention	Current ideation: baseline vs. 24 hours post pharmacotherapy (ketamine)	36	0.58 [0.1, 1.05]	Construct	✓	
			Current ideation: baseline vs. 24 hours post active placebo	21	0.09 [-0.52, 0.69]	-	-	
	Franck et al. (2007)	Cohort comparison	Current ideation vs. depression	31	0.95 [0.18, 1.66]	Discriminant	✓	
			Current ideation vs. controls	31	-0.27 [-0.97, 0.44]	Construct	×	
	Glashouwer et al.	Cohort comparison	Current ideation vs. mixed clinical and nonclinical controls	2221	0.71 [0.58, 0.83]	Construct	✓	

Mood (b)	(2010)	Cohort comparison	Current ideation vs. mixed clinical and nonclinical controls	2221	0.54 [0.41, 0.67]	Construct	✓
Mood (c)	Rüsch et al. (2007)	Cohort comparison	Lifetime attempts + borderline personality disorder vs. social phobia	90	0.61 [0.16, 1.05]	Discriminant	✓
			Lifetime attempts + borderline personality disorder vs. controls	120	0.53 [0.17, 0.89]	Construct	✓

Notes: ✓ = supportive evidence; X = contradictory evidence; “-” = no evidence to date; Effect size = Cohen’s *d*. <sup>1</sup>Recent = past month; <sup>2</sup>Recent = past year; <sup>3</sup>Although significant group differences were found, the direction of effect was opposite to that predicted. NB. Nook et al. (2010) and Randall et al. (2013) did not report the results of the implicit measures in sufficient detail to allow for the calculation of between groups effect sizes, and are therefore not reported here (see Table 1.3 for predictive validity statistics).

**Construct validity.** Eleven articles that were relevant to construct validity reported the results of the implicit measure(s) in sufficient detail to allow for the calculation of effect sizes (see Table 1.4). We have chosen to group these results based on the implicit associations targeted by the implicit measure (e.g., self–death, self–cutting). Summaries of each study can be found in Table 1.4, including details of the groups comparison that were made, the sample size, and whether the results found are supportive or contradictory of the validity of the implicit measure in question.

**Self-harm–evaluation.** This refers to measures that employed the categories “self-harm” (e.g., cutting, hanging, overdose, self-harm, suicide) and a pair of evaluative categories (e.g., positive vs. negative). One article investigated the construct validity of implicit self-harm–evaluations by comparing individuals with and without a recent history (i.e., past year) of self-harm (Knowles & Townsend, 2012). While this article contained two studies, the results of Study 2 were not reported in sufficient detail to allow for the calculation of between-groups effect sizes here (i.e., no means or standard deviations were reported). As such, only the results of Study 1 are considered here. Two GNATs targeted the valence (positive vs. negative) and arousal (e.g., arousing vs. calming) of self-harm. Individuals with and without a recent history of self-harm did not produce different results on either the self-harm–evaluation GNAT ( $d = -0.35$ , 95% CI [-0.84, 0.15]) or the self-harm–arousal GNAT ( $d = 0.29$ , 95% CI [-0.20, 0.78]). Results therefore do not support the concurrent validity of self-harm–evaluations on the GNAT. However, due to the lack of additional studies using either this measure or stimulus set, it is not possible to determine whether it is the measure and/or the implicit associations targeted that lack validity here.

A key limitation of this study was the way in which effects on the GNAT were quantified. Previous research has often employed the sample  $D$  algorithm developed for the IAT and used within all the other studies in this review. This scoring technique minimizes the

impact of extraneous variables such as differences in average responding speed (see Greenwald et al., 2003). The choice to rely on difference scores alone may have obscured otherwise important effects on the measure (see Whelan, 2008).

**Self-cutting.** This refers to measures that employed the categories “self”, “other”, pictures of cut skin and pictures of uncut skin, originally employed by Nock & Banaji (2007a). Two articles investigated the construct validity of implicit self-cutting associations, both using the IAT. Nock and Banaji (2007b) compared adolescents with a recent history (i.e., past year) of attempts, those with ideation but not attempts, and typically developing controls. The ideation group produced stronger self-cutting effects than controls ( $d = 0.77$ , 95% CI [0.30, 1.24]), and the attempts group produced strong self-cutting effects than either ideators ( $d = 0.82$ , 95% CI [0.18, 1.45]) or controls ( $d = 1.42$ , 95% CI [-2.07, -0.73]).

Dickstein and colleagues (2015) performed a similar study using slightly more stringent inclusion criteria. Individuals were recruited to three mutually exclusive groups: adolescents with a recent attempt (i.e., past month) but no lifetime history of non-suicidal self-injury, recent non-suicidal self-injury (i.e., past month) but no lifetime attempts, and typically developing controls. The attempts group did not differ from the control group ( $d = 0.17$ , 95% CI [-0.25, 0.58]), in contrast to Nock and Banaji (2007a). The attempts group did produce lower self-cutting effects than the non-suicidal self-injury group ( $d = -1.06$ , 95% CI [-1.48, -0.61]). This might suggest that self-cutting effects are more strongly related to non-suicidal self-injury than suicidal behaviours. Evidence for the validity of implicit measures of self-cutting associations across these two articles is therefore mixed. This may be related to whether overlapping or mutually exclusive groups are recruited, or to the recency of the self-injurious behaviours prior to assessment.

**Self-death.** This refers to IATs that employ the categories “self”, “other”, “death” and “life”, as originally employed by (Nock et al., 2010). Six articles investigated the construct

validity of implicit measures of self-cutting associations using the IAT. In the same study as discussed above, Dickstein and colleagues (2015) also compared differences on a self-death IAT. Counter-intuitively, the suicide attempts group's effects were comparable to the normative group ( $d = 0.20$ , 95% CI [-0.22, 0.61]), and the non-suicidal self-injury group produced weaker self-life effects than the attempters ( $d = -1.82$ , 95% CI [-2.29, -1.33]). These results therefore run contrary to what was predicted. In contrast, a larger study on university students did find the expected group differences, however (Harrison, Stritzke, Fay, Ellison, & Hudaib, 2014). Individuals with a lifetime history of attempts were found to produce significantly weaker self-life effects than those with no history of attempts ( $d = 0.31$ , 95% CI [0.02, 0.61]).

Three studies compared performance before and after a therapeutic intervention. In two studies, one of them a double-blinded RCT, Price and colleagues (2014; 2009) assessed psychiatric inpatients with current suicidal ideation 2 hours before and 24 hours after a ketamine pharmacotherapy. Self-death effects on the IAT did not change in either case ( $d = -0.37$ , 95% CI [-1.16, 0.45];  $d = -0.06$ , 95% CI [-0.53, 0.40], respectively). In contrast, changes have been found across multi-week inpatient psychiatric interventions. Ellis and colleagues (2015) recruited psychiatric patients with current suicidal ideation, 50% of whom also had lifetime attempts. Participants were assessed at admission and again within a week of discharge (average stay = 6 weeks). Self-life effects increased between the two time points ( $d = 0.28$ , 95% CI [0.02, 0.54]). However, it should be noted that the authors did not provide details of the specifics of the interventions received by participants, and their heterogeneity across participants. It is therefore unclear whether such changes are due to the greater length of the intervention, or due to the addition of a psychotherapeutic component to the intervention.

Finally, one experimental analogue study demonstrated that self–death effects are sensitive to participants’ immediate psychological context (Tang, Wu, & Miao, 2013). Participants who completed a task designed to induce rumination about personal failure immediately before the IAT produced stronger self–death effects than controls ( $d = 0.52$ , 95% CI [0.10, 0.93]), whereas participants who completed a comparable task intended to induce rumination about personal success produced the opposite pattern ( $d = 0.57$ , 95% CI [0.15, 0.98]). These results therefore speak to both the sensitivity of such effects to the individual’s immediate emotional context, and their malleability via verbal intervention (this point will be expanded upon in a following section). Together, results from these six studies provide generally supportive evidence for the construct validity of self–death associations within suicidal behaviours (with the notable exception of Dickstein et al., 2015).

***Self–escape.*** This refers to measures that employed the categories “self”, “other”, “escape”, and “stay”. In the same two studies discussed above, Price and colleagues (2014, 2009) assessed individuals with current suicidal ideation 2 hours before and 24 hours after a ketamine psychopharmacological intervention. In contrast to the pattern of effects found on the self–death IATs, self–escape effects decreased across treatment in both studies ( $d = 1.37$ , 95% CI [0.35, 2.28];  $d = 0.58$ , 95% CI [0.10, 1.05], respectively). Dickstein and colleagues (2015) also included a self–escape IAT. However, results demonstrated that individuals with recent attempts (and no non-suicidal self-injury) did not differ from controls ( $d = -0.20$ , 95% CI [-0.62, 0.21]) or from those with recent non-suicidal self-injury (and no attempts;  $d = 0.34$ , 95% CI [-0.07, 0.75]). Results therefore provide mixed evidence for the construct validity of self–escape effects.

***Self–evaluation.*** This refers to measures that employed the categories “self”, “other”, and a pair of evaluative responses (e.g., positive vs. negative: see Table 1.2). Four forms of evaluation have been employed across three articles, including valence and three different

mood contrasts (i.e., depressed/elated, anxious/calm, and shame/anxiety). One study compared psychiatric patients with current ideation and depression, depression and no ideation, and healthy controls (Franck, De Raedt, Dereu, & Van den Abbeele, 2007). Unexpectedly, the ideation and depression groups did not differ on the IAT ( $d = 0.95$ , 95% CI [0.18, 1.66]), whereas the depression without ideation group produced stronger self-positive effects ( $d = -0.27$ , 95% CI [-0.97, 0.44]). Such results are not supportive of either the construct or discriminant validity of implicit self-evaluations on the IAT.

More positive evidence has been found elsewhere, using self-mood evaluations. Glashouwer and colleagues (2010) recruited a large and heterogeneous sample of individuals who were and were not seeking help for a mental health complaint ( $n = 2221$ ). Participants with current ideation demonstrated stronger self-depressed ( $d = 0.71$ , 95% CI [0.58, 0.83]) and self-anxious effects ( $d = 0.54$ , 95% CI [0.41, 0.67]). Comparably, Rüsç et al. (2007) demonstrated that women who met the criteria for borderline personality disorder who had a lifetime history of suicide attempts demonstrated stronger self-shame (relative to self-anxious) effects than both women with social anxiety and no attempts ( $d = 0.53$ , [0.17, 0.89]) and healthy controls ( $d = 0.53$ , 95% CI [0.17, 0.89]). These results therefore provide support for both the construct and discriminant validity of self-mood associations, but perhaps not self-valence associations. However, no research to date has examined the sensitivity of such effects to change through therapeutic or analogue interventions.

## **Discussion**

Two studies have demonstrated that self-death associations on the IAT are prospectively predictive of suicidal ideation, suicide attempts, and self-harm, within follow-up periods of three to six months (Nock et al., 2010; Randall et al., 2013). In both cases, the authors noted that the self-death IAT was more predictive than more traditional self-report measures, including suicidal ideation, hopelessness, clinical judgment, the individual's

forecast of their own behaviours, impulsivity, and the SADS risk rating scale. Furthermore, the self–death IAT demonstrated fair sensitivity (.42 to .50) and good specificity (.79 to .81) in both studies. Given that a key limitation of previous research using self-report measures is the low specificity of their predictions (e.g., Harriss & Hawton, 2005), these results are therefore encouraging. Additionally, these results are notable by their ability to predict future self-harmful behaviours within clinically meaningful time scales (i.e., months rather than years).

Evidence for the self–cutting IAT is more mixed however, with one study finding that it is predictive of suicidal behaviours at six months (Nock & Banaji, 2007b), and another finding that it is not predictive of self-harm at three months (Randall et al., 2013). It should be noted that a number of articles that are outside the remit of the current review have demonstrated the predictive and construct validity of implicit self–cutting associations within non-suicidal self-injurious behaviours more specifically (see Appendix B). As such, measures of self–cutting associations may be more relevant to behaviours without suicidal intent. No research to date has examined the prospective predictive validity of implicit associations between self and evaluations, self and escape, or evaluations of self-harm. Thus, the prospective criterion validity of such measures is unclear at this time.

While the use of such approaches therefore requires much further attention, results from the self–death IAT in particular are consistent with the argument for moving away from unreliable clinical judgments of suicide risk and towards the use of objective measures (Glenn & Nock, 2014b). Specifically, a algorithmic combination of self-report measures and the results of implicit measures may provide a more accurate metric of future suicide risk within a clinically meaningful timeline than can (a) the results of any individual measure alone, or (b) clinical judgement within this specific domain. Furthermore, Randall and colleagues' (2013) demonstration of the additional predictive validity of combining the

results of implicit measures with those from self-reported risk factors represent the potential of more algorithmic decision-making approaches to assessing risk (Claassen et al., 2014; Nock, 2012; Prinstein, 2008). While this requires replication using an *a priori* model, implicit measures appear to represent a promising avenue of research for the prospective prediction of suicidal and self-harmful behaviours, and therefore warrant further empirical study and clinical attention.

How such future research might be organized is relatively less clear, however, given the mixed evidence for these various “good candidates” (e.g., of self-evaluation and self-escape associations). The following sections therefore highlight several points that emerged from the qualitative synthesis that may help to guide future research.

**Psychological context at time of assessment.** As noted above, the results generated by Nock and colleagues (2010) and Randall and colleagues (2013) suggest that implicit self-death associations (rather than self-life associations) on the IAT are predictive of future self-harmful behaviour. In these studies, participants were assessed immediately following an emergency psychiatric admission, in many cases after surviving a suicide attempt. Given that previous research has demonstrated that reactions to survival are predictive of future attempts (Bhaskaran et al., 2014), it is important to note that this research cannot speak to whether these performances on the IAT are a function of the immediate consequences of an emergency psychiatric admission, or whether such results also generalize to other assessment contexts. Indeed, the results provided by Tang and colleagues (2013) demonstrate that an individual’s immediate level of distress influences the results on such measures. This would be consistent with evidence for the differential activation theory of hopelessness and suicidality (see J. M. G. Williams, van der Does, Barnhofer, Crane, & Segal, 2008), which suggests that suicidality is better characterized by a specific pattern of response to distress. Future research should therefore investigate whether the validity of such implicit associations

is dependent on individuals' current suicidal ideation or intent, or perhaps psychological distress more generally. This could be done using more frequent longitudinal assessment (e.g., momentary assessment methods: Nock, Prinstein, & Sterba, 2009; Shiffman, Stone, & Hufford, 2008) or within experimental analogues that manipulate, for example, participant's acute mood state (e.g., Gemar, Segal, Sagrati, & Kennedy, 2001; Hussey & Barnes-Holmes, 2012) or mortality salience (e.g., Bassett & Dabbs, 2003).

**Linking implicit measures to theories of suicide.** Despite the encouraging results being generated using implicit measures and their frequent citation in the wider field as a promising avenue of research for the field of suicidology (see Randall et al., 2011; Wenzel & Beck, 2008), they have not yet received significant theoretical attention. Specifically, the predictive validity of the effects found on implicit measures stands apart from most other psychological variables because such effects are not directly derived from or accommodated by any leading theories of suicidality (e.g., Interpersonal Theory: Joiner, 2005; Integrated Motivational Volitional model: O'Connor, 2011). Of course, it is important to recognize that there exists a wealth of theory around measures of implicit associations more generally (De Houwer & Moors, 2007; Gawronski & Bodenhausen, 2006, 2011). However, to date, no theoretical work has attempted to directly connect the results of these measures with models of suicidal behaviour.

The deleterious impact of this "gap" between theory and empirical research can be observed within the existing research. At present, it is difficult to account for why a given IAT stimulus set was found to have predictive validity but others were not. For example, Randall and colleagues (2013) compared the predictive validity of the self-death IAT (Nock et al., 2010) against five new variants, which were created on the basis of face validity (i.e., suicide-method identity, self-harm identity, and mixed distress-death-suicide identity IATs). However, when the results demonstrated that only the original self-death IAT was predictive

of future self-harm, no theoretical explanation was offered for why the five other IAT variants were not, despite the fact that all six possessed good face validity. Without the ability to connect the results of these measures to theory, it is difficult to organize existing and future research. Furthermore, failure to do so arguably risks returning to the largely a-theoretical approach to the study of suicide that existed before the 1990s (Barzilay & Apter, 2014; Nock, 2009a).

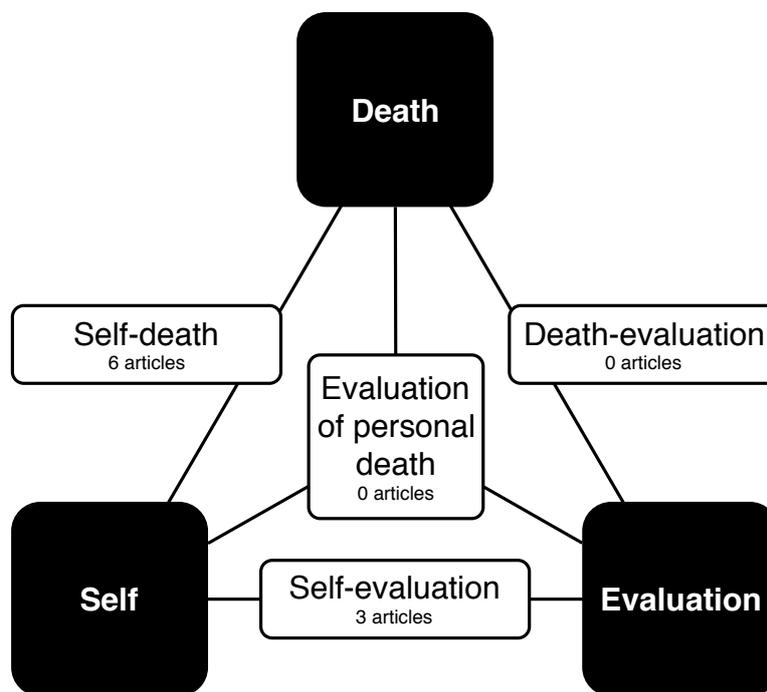
Closer alignment between existing theory and assessment using implicit measures would arguably be facilitated by the ability to ascribe effects on such measures to specific implicit associations. That is, due to the “relative” nature of the IAT’s score, it is not possible to determine whether effects on the so-called “self–death” IAT were actually driven by “self–death”, “self–life”, “others–death”, and/or “others–life” associations (De Houwer, 2002; Nosek, Greenwald, & Banaji, 2005). However, the question of whether a given individual is implicitly motivated to move “away from life” or “towards death” would seem to have important theoretical and possibly therapeutic implications for what established and maintains suicidal behaviour. Future research might therefore employ other “non-relative” measures of implicit attitudes, such as the Go/No-go Association Task (Nosek & Banaji, 2001), Implicit Relational Assessment Procedure (D. Barnes-Holmes, Barnes-Holmes, et al., 2010), or Sorting Paired Features Task (Bar-Anan et al., 2009).

**Which implicit attitudes to target.** As discussed previously, despite both possessing comparable face validity, self–death associations were shown to be predictive of future self-harmful behaviour, whereas self–cutting associations were not (Randall et al., 2013). This speaks to a generic issue within the use of implicit measures; that is, how researchers should select stimulus categories in a theory-driven way (e.g., Nosek et al., 2005; Robinson, Meier, Zetocha, & McCaul, 2005). Specifically, self-harmful behaviours such as cutting are a defining feature of suicide attempts, but they are not specific to them: self cutting behaviours

can be absent of suicidal intent and engaged in for the purposes of emotional regulation (Bentley et al., 2014; Klonsky, 2009). In contrast, death is the ultimate intended consequence of suicidal behaviours, by definition, and IATs targeting self–death associations have shown good predictive validity for future suicidal behaviours. As such, the most fruitful avenues for future research may lie in the targeting of the intended consequence of suicidal behaviour (e.g., death), which is specific to suicidality, rather than the non-specific forms these behaviours take (e.g., cutting, hanging, etc.). While the preceding conceptualization is somewhat speculative, it may provide organizing power that has been somewhat absent to date. Furthermore, it generates specific predictions: for example, the current review demonstrated that implicit self–evaluations differ between individuals with and without a history of suicidal behaviour. However, given that implicit self–evaluations play a role in many diagnostic categories of mental health complaints (see Roefs et al., 2011 for review), implicit self–evaluations should be less predictive of future suicidal behaviour than self–death associations. Future research could therefore examine the relative predictive validity of such measures through the lens of the specificities of these stimulus categories to suicidality (e.g., intended outcomes vs. the form that behaviours may take). By this logic, the category “death” may be a more fruitful target for research than “self-injury” given that the former is specific to suicide (i.e., is its intended consequence) whereas the latter is not specific to suicidal behaviours (e.g., non-suicidal self-harm). This issue of theory-driven stimulus category selection is expanded upon further below.

On a related point, without a strong theoretical rationale for the selection of specific stimulus categories (e.g., death, self, evaluation), it is difficult to know whether research using these categories is sufficient, or whether future research should target other as yet unspecified categories. For example, implicit evaluations of the future, given the established link between hopelessness and suicidality (see Meites, Deveney, Steele, Holmes, & Pizzagalli,

2008), implicit attitudes towards death as an escape, or implicit evaluations of death (see Bassett & Dabbs, 2003). Upon reflection, the lack of research on the death-evaluations is particularly surprising, given that evaluations of death play a key role in much of the existing theoretical and empirical work on suicide. For example, both leading contemporary theories of suicide posit that an individuals' movement from ideation to attempts involves a change in how they evaluate death, such as a loss of fear of death (see Joiner, 2005; O'Connor, 2011). This is supported by a wealth of empirical research using self-report measures. For example, Muehlenkamp and Gutierrez (2004) found that individuals with a history of self-harm reported significantly more repulsion to life and attraction to death than those with no such history (see also Brown, Steer, Henriques, & Beck, 2005; Ferrara, Terrinoni, & Williams, 2012; Kovacs & Beck, 1977; Orbach et al., 1991; Osman et al., 2000).



*Figure 1.1.* Three key stimulus categories within suicide (self, death, and evaluation), and the links that can be drawn between them to be assessed within a given implicit measure (self-death, death-evaluation, self-evaluation, and evaluation of personal death).

In order to illustrate this gap in the existing literature, Figure 1.1 (above) illustrates the possible links that can be drawn between the stimulus categories “death”, “self”, and “evaluation”. The concepts that these categories represent all have important roles within theories of suicide, and previous research has targeted different combinations of them (e.g., self-death and self-evaluation) within implicit measures. As demonstrated by this review, “self-death” IATs have been employed within six published articles on suicidal behaviours, and “self-evaluation” IATs were employed within three articles. Despite the clear theoretical and empirical role of evaluations of death within suicidality, the current review did not find any articles that targeted evaluations of death using an implicit measure. As such, closer scrutiny of implicit evaluations of death (e.g., loss of fear of death in suicidal individuals) within suicidality might represent one potential way in which the results generated by implicit measures could be more closely linked with existing theories of suicidal behaviour. Future research, including that within the current body of research, should therefore investigate implicit death-evaluations.

An important extension of the above point is that no research has explored more complex combinations of these stimulus categories. For example, theories of suicide (such as the IPT and IMV) refer to the interaction between all three concepts: i.e., (1) evaluations of (2) a person’s own (3) death. No research included within the current review has examined more combinations of more than two such categories. This is likely for methodological rather than conceptual reasons: it is important to note that, until relatively recently, it was widely accepted that implicit measures were necessarily associative in nature, and could therefore only assess the link between two categories (e.g., “self-death”). However, more recently, implicit measures have been developed that can assess semantically complex propositions such as “my death will be painful” (e.g., the Implicit Relational Assessment Procedure: D.

Barnes-Holmes, Barnes-Holmes, et al., 2010; Relational Responding Task: De Houwer, Heider, Spruyt, Roets, & Hughes, 2015).

The development of such measures would appear to open up a large number of new possibilities for future research using implicit measures within suicidality, including within the current body of work. First, more than two suicide-relevant concepts might be included in a given implicit measure (e.g., evaluation of personal death). Second, rather than merely assuming the presence of associations between categories, the nature of the link between concepts are linked can be examined. For example, one might expect important differences between measures that include the propositions “*I am* dead” (i.e., actual death) versus “*I want to be* dead” (i.e., desired death), even though both notionally meet the description of “self-death”-related measures.

**Theoretically-driven stimulus category selection.** The above section argues that future research should target death-evaluations. Any attempt to do this will require the creation and selection of novel stimulus sets for use within the implicit measure(s). This highlights a broader point regarding how stimuli categories are generally to be selected: a practically infinite number of possible options exist, but it is not possible to exhaustively assess and compare their utility. To use an analogy, the field of psychometrics has well-elaborated (though not universal) guidelines for the creation of new measures, such as how items should be generated (e.g., derived from qualitative work and/or by expert panels), refined (e.g., readability ratings) and reduced (e.g., factor analysis, internal consistency; see DeVellis, 2012). However, the literature on the construction and refinement of the stimulus sets employed within implicit measures is comparatively sparse, being generally limited to outcome assessments of a measure’s internal consistency, test-retest reliability, and predictive utility (e.g., Greenwald, Poehlman, Uhlmann, & Banaji, 2009; Nosek et al., 2005).

Researchers are provided with relatively little guidance on generic and *a priori* strategies that are useful in creating a “good candidate” stimulus set.

To take a concrete example, implicit measures (e.g., the IAT and IRAP) require the use of pairs of categories. For example, although a researcher may be interested in implicit death-identity (e.g., using the categories “death” and “self”), the death-identity IAT requires the inclusion of two further categories (e.g., “life” and “others”: Nock et al., 2010). These pairs of stimulus categories are commonly selected on the basis of being orthogonal or opposing (e.g., “death” vs. “life”, “self vs. “others”). Robinson, Meier, Zetocha and McCaul (2005) referred to this stimulus category selection strategy as one of “obvious opposites” (p.208), and argued that it is tacitly employed within most studies. Indeed, all of the studies included in the above review that employed such a measure (i.e., the IAT) appear to have employed this strategy (i.e., death vs. life, self vs. other, positive vs. negative, happy vs. depressed, escape vs. stay, self-harm vs. no self-harm: see Table 1.2). More generally, the majority of clinically relevant research using such implicit measures can also be said to employ this approach (see Roefs et al., 2011; Vahey, Nicholson, & Barnes-Holmes, 2015 for systematic reviews).

Despite the prevalence of this strategy, researchers rarely explicitly state this stimulus category selection strategy of choosing “obvious opposite” categories. Of course, this may be because this strategy is universal enough to not require explication. However, while we agree that it is a useful starting point, we would argue that it fails to specify with sufficient precision (a) which “obvious opposite” category should be selected given that there are often many options, or (b) what constitutes an obvious opposite and how this could be empirically assessed. This is most easily illustrated using an example. Imagine that the categories that a researcher is most interested in are “self” and “suicide”, *à la* Nock and colleagues (2010). Certainly, one category that is “obviously opposite” to “suicide” would be “life”, insofar as

they are opposite along the dimension of an organism's *animus*. However, it is also worth considering that the category "natural death" (e.g., cancer, old age, car crash, heart attack) would also be opposite to "suicide", albeit along the dimension of whether that death was intentional and self-inflicted. Simply specifying that category pairs should be "obvious opposites" does not distinguish between these two possibilities, nor does it make any *a priori* suggestion as to whether one would be superior to the other.

The lack of solid theorising around how stimulus sets tend to be selected (in the descriptive sense), and indeed how they should be to optimise their utility (in the proscriptive sense), may contribute the field's reliance on relatively a-theoretical, brute empiricist research using implicit measures, as discussed above. We therefore suggest that an explication and systematic analysis of these stimulus selection strategies may help improve the utility of implicit measures within suicide, and indeed more generally.

**Limitations.** The main limitation of this systematic review, as with many similar overviews, is that outcome definitions are not the same across studies. In particular, the use of non-standardized definitions and subtypes suicidal and self-harmful behaviours presents conceptual difficulties for the integration of results. This point has been made repeatedly elsewhere (Silverman, 2006). Future research should be clear about what definition is being adopted (e.g., O'Carroll et al., 1996; Platt et al., 1992; Silverman et al., 2007a), and how this affects the design, interpretations of results and/or implications of a given study. A second limitation of the current review was that although language of publication did not form an exclusion criterion, it might be the case that that articles published in languages other than English were not detected by the systematic search.

With regard to the limitations of individual articles included in the review, it must be noted that the majority of studies assessing construct validity relied on self-report measures to assess the presence or absence of suicidal behaviours (e.g., rather than physical records of

hospital admissions). Given that this review is predicated in part on avoiding the unreliability of such self-reports, this must be acknowledged as a potential systematic limitation to such research. Nevertheless, despite their limitations, self-reports of suicidal behaviours represent the current gold standard within suicide risk assessment.

**Conclusion.** This review located 13 studies on the use of implicit measures within suicidal behaviours. Self–death associations on the IAT appear to be prospectively predictive of suicidal and self-harmful behaviour at 3-to-6 months. While self-cutting associations may be predictive of ideation, they were found not to be predictive of actual self-harm. Implicit measures therefore appear to represent a promising avenue of research for the field, given the particular limitations of self-report measures within this domain. However, evidence for the construct validity of implicit self-cutting, self-death, self–escape, self–evaluative, and self-harm–evaluative associations is generally mixed. As such, it is difficult to determine which represent “good candidates” for future prospective research and which do not based on construct validity alone. Building stronger links between the design and results of implicit measures and existing theories of suicidal behaviour may help organize future research, which has been largely a-theoretical to date. In particular, future research (including the current body of research) should consider (a) the influence of immediate psychological context on performances on these measures, (b) which implicit attitudes are targeted, using what measure, and on what theoretical basis, and (c) how the stimulus categories, individually and in combination, are selected to target the desired implicit attitudes.

#### **1.4 A functional-analytic approach to language, suicide, and implicit attitudes**

The previous section concluded by arguing that there is a key disconnection between the promising data being generated by implicit measures and existing theories of suicide. The current section represents an attempt to bridge this gap through an overarching functional-analytic framework that is capable of accounting for complex human behaviour, including

both suicidal behaviour and the effects seen on implicit measures. Specifically, this section discusses how Relational Frame Theory (RFT: Hayes, Barnes-Holmes, & Roche, 2001) can account for the fact that humans can learn from events they have not directly experienced, which has implications for our ability to account for suicide. Subsequently, commentary is provided on the theoretical bridge between RFT and the effects produced on implicit measures, through which the current research program was conceptualised and interpreted. In doing so, one implicit measure that was designed specifically for use within RFT research, the Implicit Relational Assessment Procedure (IRAP: D. Barnes-Holmes et al., 2006; D. Barnes-Holmes, Barnes-Holmes, et al., 2010), is described and its merits argued for.

### **Links between suicide and the ability “to language”**

The concept of death plays a prominent role in human society, both in terms of valued ends (e.g., religion and the search for meaning in life: Kastenbaum, 2000; Neimeyer, Wittkowski, & Moser, 2004) and suffering (e.g., suicide as an escape from intolerable psychological pain: Baumeister, 1990). From a functional-analytic perspective, our ability to conceptualize death and engage in suicidal behaviour is inherently interesting because death cannot be directly experienced or consequated. Instead, death is verbally constructed, for example, through the metaphor of sleep. Furthermore, suicidal behaviour is defined in part by an *intent* to die, and is therefore guided by a conceptualized future (Hjelmeland & Knizek, 1999). As such, it is therefore somewhat difficult to account for the development of suicidal behaviour in terms of a direct learning history (i.e., direct contingencies). This is not merely a pedantic or definitional point: the ability to verbally construct the future through language therefore appears to play a key role within suicidal behaviours. Indeed, suicide appears to be unique to humans among all known life (Hayes, 1992; Preti, 2011). Furthermore, suicide appears to be absent in non-verbal humans (such as in cases of severe mental retardation) and is severely diminished in cases of severe cognitive impairment (e.g., dementia, acquired brain

injury, etc.; Harris & Barraclough, 1997). Loosely put, we therefore broadly agree with Heidegger when he argued that only humans die, whereas plants and animals merely perish (Heidegger, 1962, p. 291).

Importantly, the centrality of the ability “to language” has arguably been overlooked by most theories of suicide (e.g., the IMV and IPT). Specifically, such theories commonly assume, rather than account for, the presence of key behavioural repertoires such as language, reasoning and planning, and therefore the ability conceptualise future, self, and death (see also Fiedler, 2014 who has made this point regarding psychological theory more generally). If theories of suicide are to be linked across domains (e.g., to other clinically-relevant theories) and levels of analysis (e.g., to basic psychological processes), an overarching theoretical account is therefore required.

### **Relational Frame Theory**

RFT represents a functional-analytic account of language and higher cognition more generally in terms of “emergent” or “derived” responding that occurs in the absence of a direct history of reinforcement. As a species, humans are unique in their capacity to learn to spontaneously derive novel relations between objects and events that have not been directly trained, and which are unrelated to the formal properties of those stimuli. RFT defines this ability as arbitrarily applicable relational responding. A large body of work supports the notion that this type of behaviour represents an overarching operant response class; that is, a learned behaviour that we acquire early on in our development through our interaction with the verbal community (see Hughes & Barnes-Holmes, 2016 for a recent overview).

The simplest unit of analysis posited by RFT is the relational frame, which is defined by three properties: mutual entailment, combinatorial entailment, and the transformation of stimulus function (Hayes et al., 2001). For the purposes of illustration, imagine that an individual is trained, either through contacting contingencies or via instruction, that one

stimulus (A) is the *same* as a second stimulus (B), and B is the *same* as a third stimulus (C). Here, mutual entailment refers to the relation that spontaneously emerges between A and B in the absence of explicit training. Specifically, when trained that A is the *same* as B, humans will also derive that B is the *same* as A without any additional training. Combinatorial entailment refers to the relations that emerge between multiple mutually entailed stimuli. Thus, if A is same as B and B is same as C, then humans will spontaneously derive that A is same as C, and C is same as A. Furthermore, the psychological functions of such mutually or combinatorially related stimuli are transformed in accordance with the stimulus relation (i.e., transformation of function). For example, if an aversive function is established for the A (e.g., an electric shock), both B and C will other stimuli will also acquire the aversive functions of A, despite the fact that they were never directly coordinated with a shock (Dougher, Augustson, Markham, Greenway, & Wulfert, 1994). Importantly, relations other than “sameness” (i.e., coordination) can be established between stimuli, such as “different to” (distinction), “opposite to” (opposition), “more than” (comparison), “comes after” (temporal), “contains” (hierarchy), etc. This represents a key difference between RFT and previous functional-analytic accounts of symbolic meaning (e.g., Stimulus Equivalence: Sidman, 1994). As such, psychological functions are *transformed through* such relations and not merely *transferred between* them. For example, if an individual is trained that A is *less than* B, and C is *more than* B, and B is coordinated with an electric shock, then that individual will be *less* avoidant of A than B, and *more* avoidant of C than B (Dougher, Hamilton, Fink, & Harrington, 2007). Hughes summarised this succinctly: “in short, Relational Frame Theory posits a rather simple notion – that complex human behaviour reflects the learned and contextually controlled ability to arbitrarily relate one stimulus to another” (Hughes, 2012, p. 22)<sup>2</sup>.

---

<sup>2</sup> It is important to note a distinction between relational responding with regard to the formal properties of a

Importantly, this unit of analysis (i.e., relational framing) has shown utility in constructing and deconstructing many of the behavioural phenomena that characterise human higher cognition, such as instructional control and rule following (O’Hora & Barnes-Holmes, 2004), perspective taking (McHugh, Barnes-Holmes, & Barnes-Holmes, 2004), analogy and metaphor (D. Barnes-Holmes et al., 2005; Stewart, Barnes-Holmes, & Roche, 2004), the emergence of self and awareness (Dymond & Barnes, 1997; Foody, Barnes-Holmes, & Barnes-Holmes, 2012). It suffices to say that RFT appears to be at a stage of development where it allows for the functional-analytic conceptualization and modelling of many of complex human behaviours (see Hayes et al., 2001; Hughes & Barnes-Holmes, 2016).

### **RFT’s ability to account for suicidal behaviours**

While the current body of research is not directly concerned with developing and assessing an RFT model of suicidal behaviours, it nonetheless seems important to consider, if only briefly, how the framework can also account for the fact that suicide appears to be unique to verbally-able humans. Existing theories of suicide have derived suicidal behaviour as functioning as an escape from intolerable psychological pain (e.g., Baumeister, 1990; Joiner, 2005; R. C. O’Connor, 2011; Shneidman, 1993). However, in doing so, such theories fail to specify how such escape responses (i.e., “away” from life and “towards” death) are established and strengthened, given that death itself cannot be contacted or consequated. To illustrate this point, consider the following: one could argue that individuals can learn about the consequences of others’ suicides, for example through the personal experience of losing a loved one or by more indirect means such as reading about suicide in literature or the media. The point that we wish to make here is that this while such discussions might point to

---

stimulus (e.g., whether one coin is “larger” than another) and “arbitrarily applicable” relational responding with regard to a stimulus’s abstract properties (e.g., whether one coin is “worth more” than another). The former can reliably be observed in both humans and nonhuman animals, whereas the former cannot be convincingly demonstrated in nonhumans. Arbitrarily applicable relational responding therefore represents a key disruption of the “continuity assumption” regarding the cognitive ability across humans and non-humans (see Hughes & Barnes-Holmes, 2014 for an article-length discussion).

potential environmental sources for such learning, they do not provide a specific account of the generic underlying learning process or mechanism for such things as “learning via the experience of others”. We therefore suggest that RFT provides an account of this “missing link”, for example through derived escape functions. For example, in emphasizing the fact that suicidal behaviour cannot be established or strengthened by direct contingencies, and within a larger discussion on the nature of what constitutes “intentional” behaviour from a functional-analytic perspective, Hayes (1992) suggests that suicidal behaviours participate in verbal rules, such as “when I die, my suffering will stop” (p. 116). This is entirely consistent with the descriptions of suicide as constituting an intentional escape response (Baumeister, 1990; Hjelmeland & Knizek, 1999). Thus, our purpose here is therefore not to suggest that the content of existing theories of suicide should be modified, but only to note that there is potential to integrate the predictions of these theories within a broader, functional-analytic framework. Additionally, such a framework may also encompass the effects generated within implicit measures, which will now be discussed.

### **Bridging the gap between RFT and the effects produced on implicit measures**

As noted above, RFT has allowed for the conceptualization of complex behavioural phenomena, including those on which modern theories of suicide are predicated (e.g., the ability to language and reason; the construction of self, death, and evaluation: see Foody et al., 2012; Hayes, 1992; Hughes, 2012, respectively). However, our ability to capture this relational responding “in-flight”, as it is emitted by an individual, was until recently quite limited. Existing paradigms within behavioural psychology, such as matching-to-sample (MTS), provided a binary or dichotomous outcome. For example, an individual in a given study would either demonstrate the derived acquisition of fear of spiders (following training and testing with MTS) or she would not (Smyth, Barnes-Holmes, & Forsyth, 2006). This inevitably invited a binary or dichotomous way of thinking about relational framing itself.

Rarely did we ask about the relative strength, probability, or persistence of relational responding.

As discussed above, the standard derived relational responding RFT study involved training, for example, A to B and B to C, before testing for A-C relations, and then concluding that the frame of interest (e.g., coordination or distinction) was present in that individual's behavioural repertoire (e.g., Dymond & Barnes, 1995). Instead, the research methods invited a binary or dichotomous way of thinking about relational frames: they were either demonstrated or they were not. This was entirely consistent with the definition of a relational frame as something that emerges in the absence of a history of direct reinforcement. In the natural environment, however, novel or emergent relational responses rarely occur in the absence of reinforcement, either by a listener in the verbal community or by achieving internal verbal coherence (e.g., when reasoning or problem solving). The definition of a relational frame as unreinforced, emergent and completely derived behaviour – and the fact that relational framing in the natural environment is almost certainly reinforced at high rates – creates both a conceptual tension and a methodological conundrum. At some point, it therefore seems important to draw a line under the need to engage in nothing but demonstration work (which requires that framing is exhibited in the absence of reinforcement) and to accept that relational framing does provide a reasonably adequate functional analysis of human language and cognition. Doing so, however, raises the methodological question of how to capture relational framing “in flight”, as it actually occurs in the natural environment, rather than simply demonstrate relational framing in the laboratory.

Imagine, for example, that one wishes to pose questions about the relative strengths of specific verbal relations rather than their mere presence or absence. For example, between the words “dog” and “canine”. One could conduct a study in which verbally sophisticated adults

were tested using a match-to-sample (MTS) task to see if they would match these words. However, if participants were simply asked to match “dog” and “canine” over and over again they would likely continue to do so for as long as they were asked to do so and had some reason to continue to participate in the study. Technically, the problem we face when attempting to study the strength of verbal relations is that they have such long histories and they participate in very extensive and coherent verbal networks. Indeed, although extinction curves are traditionally the metric by which strength of responding was measured with non-verbal organisms (Catania, 1998), the extinction of specific verbal relations in the absence of experimentally programmed reinforcement is difficult to obtain. Indeed, this perseveration in the absence of obvious sources of reinforcement is, of course, one of the defining features of verbal behaviour (Hayes, Brownstein, Haas, & Greenway, 1986; Hayes, Brownstein, Zettle, Rosenfarb, & Korn, 1986).

Around the time that the seminal RFT volume was being written (Hayes et al., 2001), cognitive and social psychology developed and became interested in a class of methodologies referred to in Section 1.2 as measures of implicit attitudes. The IAT, in particular, provided the inspiration for how to advance this debate. Specifically, the IAT contributed the simple strategy of requiring participants to emit a specific pattern of relational responding in one context, and to emit an opposing pattern in another context, and to obtain some measure of the relative fluency between the two patterns. Rather than different combinations of category pairings, as in the IAT, a measure of the relative strength of relational responding might therefore compare responding to a single category pair along two orthogonal relations. For example, to say in one context that “death” and “self” are “similar”, and in another context to say that they are “different”. If participants show greater behavioural fluency in responding to “death” and “self” as “similar” rather than “different”, it would seem reasonable to conclude that the former relational response is, *ipso facto*, at a greater relative strength than the latter. It

is this simple conceptual idea of obtaining a fluency differential via forced responding between opposing relational response classes that lies at the heart of the Implicit Relational Assessment Procedure (IRAP: D. Barnes-Holmes et al., 2006; D. Barnes-Holmes, Barnes-Holmes, et al., 2010).

Specifically, the IRAP involves presenting pairs of stimuli to participants on a computer screen. Participants respond to blocks of these stimulus pairings, and are required to respond as accurately and quickly as possible according to following one of two responding rules across blocks (e.g., “respond as if life is positive and death is negative” vs. “respond as if life is negative and death is positive”). In short, the IRAP compares the relative ease (i.e., speed) with which participants respond according to one rule relative to the other under accuracy and latency pressure.

From a functional-analytic perspective, the most probable response class is that which is emitted faster on average, by definition (Catania, 1973), and is therefore deemed to be consistent with the individual’s learning history. For example, if an individual is faster to respond to “death” and “self” with “similar” relative to “different”, this is interpreted as indicating that she has a deeper history of coordinating self with death. It is therefore worth reiterating that the IRAP represents the application of traditional operant principles to the assessment of response strength. This extension of RFT to include the effects generated on implicit measures, and furthermore, to delineate between the results generated on implicit measures versus self-report measures, is referred to as the Relational Elaboration and Coherence model (REC model: D. Barnes-Holmes, Barnes-Holmes, et al., 2010; Hughes, Barnes-Holmes, & Vahey, 2012).

Briefly, the REC model asserts that measures such as the IRAP capture behaviours that reflect relatively *brief and immediate relational responding* (BIRR), whereas the lack of time pressure involved in completing a self-report measure allows for more *extended and*

*elaborated relational responding* (EERR) that coheres with other relational responses in the individual's behavioural repertoire<sup>3</sup>. For example, hypothetically speaking, a suicidal individual might produce BIRRs within an IRAP that reflect evaluations of life as negative and death as positive; given a learning history of life being coordinated with extensive psychological suffering, and their verbal construction of death as an escape from this suffering. However, when asked to self-report their attitudes to life and death (e.g., on the Suicide Cognitions Scale: Ellis & Rufino, 2015) their responses are likely to represent relatively more EERR-like behaviours, which may or may not coordinate with their BIRRs. For example, on the one hand, the individual might self-report that they evaluate life negatively (e.g., that they feel their life has been a failure). In this case their BIRRs and EERRs coordinate, as their BIRRs cohere with their broader history. In contrast however, when asked via a self-report they may *deny* that they evaluate death positively (e.g., as a source of relief). This might be due to a history of social punishment associated with communicating suicidal behaviours, or an interpersonal context that does not facilitate such disclosures (see Pompili, Girardi, Ruberto, Kotzalidis, & Tatarelli, 2005 for a review of the evidence on emergency room staff's negative evaluations of individuals who self-harm). In this case, the individual's BIRRs and EERRs diverge, due to the incoherence between the BIRRs and their broader repertoire.

From this perspective, implicit measures may represent a way to circumvent the fact that suicidal individuals often appear to be either unable or unwilling to self-report their suicidal history and future intentions, due to the fact that the more BIRR-like behaviour observed within implicit measures is less contaminated by EERR-like behaviours such as self-presentation biases than responses on self-reports. However, in making this point, two

---

<sup>3</sup> It should be noted that BIRRs and EERRs are descriptive terms. Recent theorizing elsewhere has postulated several properties of arbitrarily applicable relational responding that may be useful in analyzing these behaviours (e.g., levels of derivation, complexity and coherence) and how these interact in a multidimensional framework (see D. Barnes-Holmes, Y. Barnes-Holmes & Hussey, in press; Hughes et al., 2012).

key issues should be noted. First, from the perspective of RFT and the REC model, the BIRRs tapped by implicit measures do not reflect “what people really think”: neither BIRRs nor EERRs reflect more “genuine” performances because from a functional-analytic perspective neither modality measures a proxy for a mental state or event, such as associations in memory. That is, there is no Cartesian “attitude” or “belief” to be measured. Rather, both BIRRs and EERRs simply reflect behaviour-in-context under different sources of contextual control. Second, RFT makes no *a priori* assumption that BIRRs are a superior predictor of behaviour in some other context than EERRs, given that the functional-analytic approach does not see such private behaviours as causal agents. Rather, the utility of assessing “behaviour–behaviour” relations (Hayes & Brownstein, 1986), such as that between performance on an IRAP and the probability of future suicide attempts, simply represents the examination of probabilistic functional relations. In summary, the RFT account of implicit attitudes explicitly employs the terms “implicit” and “attitude” as heuristic for behaviour that is produced within tasks that arrange the environment in such a way as to tap relatively brief and immediate relational responses and limit the influence of more extended and elaborated relational responses. This is done with no *a priori* assumption that the effects shown on such tasks are inherently predictive of other (usually more EERR-like) behaviours of interest, only that they provide sufficiently good contextual control over responding within them that such behaviour-behaviour relations may then be examined.

The current body of research employs an RFT approach to implicit attitudes throughout. For example, we refer to the strength of relational responding (in the functional analytic sense) rather than associative strengths (in the cognitive sense). While an in-depth discussion of the similarities and differences between functional-analytic (i.e., relational) and cognitive-mechanistic (i.e., associative) accounts of implicit attitudes is beyond the scope of

the current research, it should be noted that his topic has received extensive treatment elsewhere (see Hughes, Barnes-Holmes, & De Houwer, 2011; Hughes et al., 2012).

Although the predictive utility of measures of BIRR-like behaviour is not assumed *a priori*, the IRAP has nonetheless performed well in this regard. A recent meta-analysis (Vahey et al., 2015) of 15 studies examining the IRAP's ability to predict clinically-relevant criterion effects showed a moderate effect size ( $\bar{r} = .45$  [95% CI .23 to .67]), across domains such as depression (Hussey & Barnes-Holmes, 2012; Remue, De Houwer, Barnes-Holmes, Vanderhasselt, & De Raedt, 2013), anxiety (Kishita, Muto, Ohtsuki, & Barnes-Holmes, 2014), OCD (Nicholson & Barnes-Holmes, 2012), eating disorders (Parling, Cernvall, Stewart, Barnes-Holmes, & Ghaderi, 2012), substance abuse (Parling et al., 2012), and paraphilia (Dawson, Barnes-Holmes, Gresswell, Hart, & Gore, 2009).

Finally, for the purposes of comparison, it is useful to briefly consider a key difference between the IAT and IRAP in order to appreciate why the current body of work has elected to employ the IRAP over a more commonly used implicit measure. A stated goal of the current body of work is to more closely align the use of implicit measures with existing theories of suicidality. In the service of this, we considered that the interpretation of effects on a given implicit measure must serve to aid this alignment. In this regard, the IAT does not provide adequate specificity due to the relativity of its effects. For example, the IAT effect can be interpreted as the strength of one pattern of categorisation (e.g., death goes with positive and life goes with negative) relative to another (e.g., death goes with negative and life goes with positive). In contrast, trial-type effects on the IRAP can be interpreted as the strength of asserting a specific proposition (e.g., "death is negative") relative to rejecting that proposition. This ability to tie effects on the IRAP back to specific proposition rather than overall patterns of response bias affords a greater degree of conceptual clarity over the

interpretation of its effects, and therefore a greater ability to connect with and test theories of suicidality.

### **1.5 Overview of the current research program**

The above systematic review concluded with several recommendations for future work, which served to guide the current research. First, in order to attempt to link performance on such measures more closely with existing theories of suicide, the current research was conceptualized from the perspective of RFT, given that this can potentially account for both suicidal behaviours and the effects found on implicit measures in functional-analytic terms. Second, in line with our above review's conclusion that future research should more carefully consider which measure is employed, and on what theoretical basis, we elected to employ the IRAP as it meets three key criteria: (a) it is widely recognized as an implicit measure (e.g., Gawronski & De Houwer, 2011; Nosek et al., 2011); (b) unlike most other implicit measures, its construction and the interpretation of its results emerged from RFT (D. Barnes-Holmes, Barnes-Holmes, et al., 2010); and finally (c) it produces four separate scores that refer to the strength of four separate classes of BIRRs (i.e., is a "non-relative" measure of implicit attitudes). Third, in order to attempt to link implicit measures to theories of suicide through a focus on which implicit attitudes are targeted, the current research focuses on evaluations of death. It should be noted that this is therefore the first empirical research to examine implicit evaluations of death between individuals with and without a history of suicidal behaviours. Fourth, in line with the need for more careful consideration of the influence of immediate psychological context while completing implicit measures, we have conducted a number of analogue studies in which the salience of mortality was manipulated across time points. Fourth, and finally, the current research developed and employed a novel IRAP experimental methodology (i.e., manipulations of the IRAP's "contrast category") in line with the need to better understand how stimulus categories,

individually and in combination, produce effects on the task and thus influence our theoretical conclusions.

The primary aim of the current program of research was to develop an IRAP that could assess differences in implicit evaluations of death between individuals based on their history of suicidal behaviours. To this end, variations of a death-evaluation IRAP were employed across six empirical studies, employing both experimental analogue and cohort comparison designs. Two further experiments attempted to provide a “proof of concept” of the contrast category manipulation method.

Chapter 2 presents a first study on the use of a death-evaluation IRAP within a normative population, alongside the established death-identity IAT and a death-identity IRAP. The utility of the IRAP’s four separate bias scores (in contrast to an overall relative bias score, as in the IAT) is therefore explored. This study also includes a number of self-report measures in order to assess whether effects on the implicit measures are related to self-reported attitudes to death (e.g., death anxiety, belief in the afterlife) or proxies of non-normative attitudes to death (e.g., suicidality) such as depression and hopelessness. The somewhat unexpected results (i.e., the presence of death-positive effects and absence of death-negative effects) of this study highlighted two potential issues that served to guide later experiments: (a) although the trial-types are procedurally separate, the degree to which behaviour on one trial-type is influenced by the contents of the others is not currently known, and (b) the salience of death may be low in a population of young, healthy students.

The two experiments contained in Chapter 3 represent attempts to provide “proof of concept” that the contents of one stimulus category in the IRAP (e.g., “life”) influences behaviour on other trial-types (e.g., “death-negative”). In doing so, this chapter attempts to develop such “contrast category manipulations” as a generic experimental methodology to aid the refinement of IRAP stimulus sets in order to meet analytic goals.

Having demonstrated the viability of such contrast category manipulations in the previous chapter, Chapter 4 then applies the contrast category manipulation method to the interrogation of implicit death-evaluations in normative participants across four separate experiments. Additionally, in order to explore the sensitivity of four variations of death-evaluation IRAPs to the salience of mortality, each experiment assessed participants before and after a standard mortality salience induction. A series of post hoc analyses were used to examine the interaction between the inductions and the contrast category manipulations, in order to select, by systematic comparison, the stimulus set that showed most utility in assessing implicit evaluations of death.

Chapter 5 contains a final experiment that directly compares implicit evaluations of death between normative and suicidal individuals. Additionally, in order to better connect the design and results of such measures with existing theory, the presence of a reference to self is manipulated across two IRAPs (i.e., an abstract death IRAP vs. a personal death IRAP). Finally, Chapter 6 provides a summary of the results and includes a discussion of a range of empirical and conceptual issues.

## **CHAPTER 2: COMPARING THE IRAP AND IAT AS MEASURES OF DEATH-IDENTITY AND DEATH-EVALUATION IN A NORMATIVE POPULATION**

Abstract: The current chapter represents an exploratory study that used the IAT and IRAP to examine implicit attitudes to death in a normative student population. An IAT and an IRAP targeted implicit death-identity, and a second IRAP targeted implicit death-evaluations. Results suggest that the self-life/others-death effects found on the IAT in previous research may indeed be related specifically to self-life biases on the IRAP. Additionally, death-positive effects were found on the death-evaluation IRAP. These effects were unexpected given the normative nature of the sample, and were not correlated with depression, hopelessness, death anxiety or belief in the afterlife. Effects on the death-evaluation IRAP therefore appear to require further scrutiny.

As noted in Chapter 1, previous research has shown that IATs that target implicit attitudes to self and death (i.e., using the categories self, others, death, & life) are prospectively predictive of self-harm and suicide attempts over and above established risk factors. These include clinical judgment, self-report measures of impulsivity and hopelessness, a variety of routine risk assessment tools (e.g., the previously discussed SAD PERSONS assessment, Patterson et al., 1983), and the individual's own self forecast (see Nock et al., 2010; Randall et al., 2013). However, it is important to note that the IAT does not permit the identification of individual, independent response biases. Indeed, the creators of the IAT have been assiduous in specifying that it is a measure of the relative (rather than absolute) strength of associations between categories. This has been argued for both conceptually (e.g., Greenwald et al., 1998; (Nosek, Greenwald, & Banaji, 2007) and empirically (e.g., Nosek et al., 2005; Pinter & Greenwald, 2005) on numerous occasions. For example, these authors point out that an IAT that includes the stimuli "self", "others", "life",

and “death” (e.g., Nock et al., 2010) must be interpreted as a measure of the relative strength of associations between all four categories. For example, while somewhat verbose, these biases should be interpreted as either towards “self and life relative to others and death” or “self and death relative to others and life”.

Whereas the IAT presents participants with all four categories on each trial (e.g., self, other, life, and death) and examines the relative ease of categorization (e.g., self-life and others-death vs. self-death and others-life), the IRAP only ever presents exemplars from one category in a pair (e.g., either self or others and life or death) on each trial, and requires participants to respond in opposing directions across blocks (e.g., “similar” vs. “different”). As such, four separate bias scores are produced, one for each “trial-type” (e.g., self-life, self-death, others-life, and others-death). This ability to separate out the strength of individual category pairings may serve to uncover subtle effects that might be obscured within the IAT’s overall bias score (D. Barnes-Holmes, Barnes-Holmes, et al., 2010). A recent meta-analysis demonstrated that the IRAP has been used to explore a variety of clinically relevant domains, including self-esteem, depression, OCD, and substance abuse (see Vahey et al., 2015 for meta analysis). However, this represents the first study to explore the utility of the IRAP in exploring attitudes to death.

This study tests the assumption made with previous research, whether explicitly or implicitly, that normative participants do indeed demonstrate specific “self-life” and/or “self-not-death” biases (e.g., Dickstein et al., 2015; Harrison et al., 2014; Nock et al., 2010; Price et al., 2014, 2009; Randall et al., 2013; Tang et al., 2013). Participants completed both a death-identity IAT and a death-identity IRAP that was created from the same stimuli. It was expected that a sample of normative participants would demonstrate overall “self-life/others-death” effects on the IAT, but that effects on the IRAP would load onto a two specific trial-types: an assertion of “self-life” and a rejection of “self-death”.

A second IRAP was also included to explore implicit evaluations of death. One of the key findings produced by the review of the literature in Chapter 1 was that, somewhat surprisingly, no research to date has examined evaluations of death using implicit measures. Specifically, while previous research on implicit attitudes to death has explored the associations between the concepts of (a) self and evaluation (e.g., “self-evaluation”: Dickstein et al., 2015; Franck et al., 2007; Glashouwer et al., 2010; Price et al., 2014, 2009), and (b) self and death (i.e., “death-identity”: (i.e., “death-identity”: Dickstein et al., 2015; Harrison et al., 2014; Nock et al., 2010; Price et al., 2014, 2009; Randall et al., 2013; Tang et al., 2013), no work has explored the association between (c) death and evaluation. This is somewhat surprising, given the centrality of evaluations within many psychological theories that focus on death. For example, Terror Management Theory (Greenberg, Pyszczynski, & Solomon, 1986) is predicated on the assumption that awareness of one’s own mortality is highly aversive, and argues that this serves as the motivator for humans’ ubiquitous need for meaning and self-esteem. Comparably, many theories of non-normative attitudes to death (e.g., suicidality) attempt to account for how life and/or self acquires aversive properties in the face of unbearable psychological suffering, often with explicit reference to an acquired loss of fear of death (e.g., the Interpersonal Theory of Suicide: Joiner, 2005; the Integrated Motivational-Volitional Model of Suicidal Behavior: (R. C. O’Connor, 2011). This idea that death is aversive or evaluated negatively within ‘normative’ individuals, and that decreased fear of death is associated with suicidality, has been supported by a relatively large number of studies using self-report questionnaires (Ribeiro et al., 2014a). In summary, given the centrality of evaluations of death to many theories, and the recent emphasis on the relative utility of implicit measures to traditional self-report methods (see Randall et al., 2011), it is therefore somewhat surprising that no research to date has examined evaluations of death

using implicit measures. Consequently, a second IRAP was included, which assessed evaluations of death (i.e., life, death, positive, & negative).

Finally, in addition to the death-identity IAT, death-identity IRAP and death-evaluation IRAP, a number of self-report measures were included in order to explore the self-report correlates of such implicit attitudes. Self-report measures of depression and hopelessness were included, due to their known association with non-normative attitudes to death (i.e., suicidality: see Brown, Beck, Steer, & Grisham, 2000). A number of additional exploratory self-report measures were included to assess beliefs in the afterlife, fear of death, and the relationship between an individual and their aversive mental content.

## 2.1 Method

### Participants

Forty-two undergraduate students (25 female, 17 male) aged between 18 and 51 years old ( $M = 27.8$ ,  $SD = 9.6$ ) were recruited from the student population at the National University of Ireland Maynooth. Inclusion criteria were self-reported fluent English, normal or corrected to normal vision, age 18-65, and full use of both hands. Participants reported having completed between 0 and 9 IRAPs previous to the current study ( $M = 2.60$ ,  $SD = 3.08$ ). It should be noted that, due to ethical approval constraints, self-reports of history of suicidal behaviour were not collected. As such, the current study was conducted within a normative sample with an uncertain history of suicidal behaviours, rather than a strictly “non-suicidal” sample. A range of self-report measures was employed to establish the normativity of the sample (see below). Handedness was not assessed in this or any of the subsequent studies reported in this body of work, on the basis that both the IRAP and IAT require participants to responding to the same stimuli pairings using an equal number of both the left and right response keys across blocks. Indeed, previous large scale studies using the IAT have demonstrated no effect for self-reported handedness (e.g., Greenwald & Nosek, 2001).

## **Ethical considerations**

Ethical approval was granted by Maynooth University's Social Research Ethics Subcommittee (see Appendix Q for confirmation letter). Participants were informed as to the nature and purpose of the study prior to participation. Furthermore, they were made aware that participation was entirely voluntary, that they could cease participation at any time without giving a reason, and that all data would be irrevocably anonymised immediately after participation. Written informed consent was obtained from all participants (see Appendix G for consent form). Upon completion of the tasks, participants were fully debriefed on the nature and purpose of the experiment and provided with an opportunity to ask questions, which were then fully answered by the researcher. During this debrief, the researcher made each participant aware of the free student counselling service available on campus, whose contact details were included at the bottom of their copy of the consent form, should they experience any form of post-experimental distress.

## **Measures**

Self-report measures included an assessment of individuals' attitudes to death and dying using the Death Anxiety Scale (Templer, 1970), and belief in what comes after death using the Belief in the Afterlife Scale (Osarchuk & Tatz, 1973). Both depressive symptoms and hopelessness were assessed, based on their known relationship with suicidal behaviours, using the Depression Anxiety Stress Scales (Lovibond & Lovibond, 1995) and Beck Hopelessness Scale (Beck, Weissman, Lester, & Trexler, 1974), respectively. Finally, we included the Acceptance and Action Questionnaire II, a measure of psychological flexibility, as an exploratory measure (Bond et al., 2011).

**Acceptance and Action Questionnaire II.** The AAQ-II (see Appendix C) is a 7-item measure of psychological flexibility (Bond et al., 2011). That is, the ability to change or persist in behaviour in the service of valued ends (e.g., "Worries get in the way of my

success”). Psychological flexibility is a key process within Acceptance and Commitment Therapy (Hayes, Strosahl, & Wilson, 1999). Participants respond on a 7-point Likert scale from 1 (never true) to 7 (always true), with possible scores ranging from 7 to 49. Internal consistency was excellent in the current sample (Cronbach’s  $\alpha = .90$ ).

**Beck Hopelessness Scale.** The BHS is a 20-item self-report measure of an individual’s hopelessness over the past week, and has been shown to be a significant predictor of death by suicide in longitudinal studies (Beck, Steer, Kovacs, & Garrison, 1985; Beck et al., 1974). Each item has a binary (True/False) response format and is scored from 0-20, where higher scores represent greater levels of hopelessness. Internal consistency was good ( $\alpha = .86$ ). Please note that this scale is proprietary and therefore not included in the Appendices.

**Belief in the Afterlife Scale.** This 7-item self-report questionnaire assesses individuals’ beliefs in the afterlife using items such as “Earthly existence is the only existence we have” (Osarchuk & Tatz, 1973: see Appendix D). Participants respond using a 1 (disagree) to 10 (agree) scale and is scored from 0 to 70, where higher scores represent greater belief in the afterlife. It should be noted that specific beliefs about the afterlife (e.g., whether it will involve punishment or reward) are not examined, only whether participants believe death to be extinction of self. Internal consistency was good in the current sample ( $\alpha = .89$ ).

**Death Anxiety Scale.** This 17-item self-report questionnaire (see Appendix E) assesses fear and anxiety around the act of dying and the finality of death using items such as “I am very much afraid to die” and “I often think about how short life really is” (Templer, 1970). It employs a true/false answer format. Internal consistency was found to be good in the current sample ( $\alpha = .70$ ).

**Depression Anxiety and Stress Scale.** This 7-item depression subscale (see Appendix F) asks participants about depressive symptoms in the past week, and uses a 0 (did not apply to me at all) to 3 (applied to me very much, or most of the time) response format. This subscale has been shown to correlate highly with other well-established depression scales such as the Beck Depression Inventory II ( $r = .74$ ; Lovibond & Lovibond, 1995). Internal consistency was excellent in the current sample ( $\alpha = .92$ ).

**Death-identity IAT.** The Implicit Association Test (Greenwald et al., 1998) is a computer-based test that assesses reaction time biases. The current study employed the death-identity stimulus set developed by Nock and colleagues (2013) which has been employed in multiple previous studies (e.g., Dickstein et al., 2015; Harrison et al., 2014; Nock et al., 2010; Price et al., 2009; Price et al., 2014; Randall et al., 2013; Tang et al., 2013). One small modification was made to the stimulus set: due to ethical constraints on asking university students about suicide, and in order to assess attitudes to death generally rather than suicide specifically, the stimulus “suicide” was substituted for the word “death”. The stimuli employed in the four categories (i.e., self, others, life, and death) are presented in Table 2.1. All procedural details of the IAT were standard (see Nosek et al., 2007 for methodological review).

Table 2.1. *Stimulus sets and category labels for the death-identity IAT*

Self	Others	Death	Life
Myself	Them	Death <sup>1</sup>	Alive
My	They	Die	Live
Mine	Theirs	Funeral	Thrive
I	Their	Lifeless	Survive
Self	Other	Deceased	Breathing

Note: <sup>1</sup>Substituted the word “Suicide” as used by Nock et al. (2010)

**IRAPs.** The structure of the IRAP employed in the current study was as described by Barnes-Holmes, Barnes-Holmes and Stewart (D. Barnes-Holmes, Barnes-Holmes, et al., 2010) see below). The task was programmed in Visual Basic 6, and the “2012” version of the program was used (D. Barnes-Holmes & Hussey, 2012). Several parameters were identical across both IRAPs, and thus will be summarized here. Participants were provided with up to 4 pairs of practice blocks. If participant met both the mastery criteria on both blocks in a pair of practice blocks (i.e., accuracy  $\geq$  80% and median latency  $\leq$  2000ms) they were immediately moved to the test blocks. If they failed to meet the mastery criteria after four practice block pairs, the task ended. In accordance with the majority of previous research, three pairs of test block pairs were completed (e.g., Hussey & Barnes-Holmes, 2012; Rönspies et al., 2015). The latency feedback message was set to “!”. The location of the response options remained static throughout the task.

***Death-identity IRAP.*** The stimuli employed within the IRAP were drawn from those used within the IAT in order to make the two measures maximally comparable. The IRAP program required an even number of stimuli per category. As such, one additional stimulus was added to each category (i.e., “me”, “others”, “dead” and “living”: see Table 2.2). The four trial-types were therefore “self-death”, “self-life”, “others-death”, and “others-life”. The response options were set to “Similar” and “Different”. Specific responding rules were presented on screen before each block. Responding rule A was “Please answer as if you associate yourself with death and others with life,” and responding rule B was “Please answer as if you associate yourself with life and others with death”. The order in which individuals were exposed to these blocks (i.e., rule A first vs. rule B first) was counterbalanced between participants in both IRAPs.

Table 2.2. *Stimulus sets for the death-identity IRAP*

Label stimuli		Target stimuli		Response options	
Myself	Them	Death <sup>1</sup>	Alive	True	False
My	They	Die	Live		
Mine	Theirs	Funeral	Thrive		
I	Their	Lifeless	Survive		
Self	Other	Deceased	Breathing		
Me <sup>2</sup>	Others <sup>2</sup>	Dead <sup>2</sup>	Living <sup>2</sup>		

*Note:* <sup>1</sup>Substituted the word “Suicide” as used by Nock et al. (2010), <sup>2</sup>Additional stimuli added to meet requirements of the IRAP.

***Death-evaluation IRAP.*** This IRAP attempted to target relational responding around the valence of life and death. Labels 1 and 2 were set to “living” and “dying”, respectively, target 1 stimuli were set to positive-valence-high-arousal words (i.e., enjoyable, exciting, lovely, great, pleasant, and satisfying), and target 2 stimuli were set to negative-valence-high-arousal words (i.e., awful, distressing, hurtful, horrible, painful: see Table 2.3). The four trial-types were therefore “life-positive”, “life-negative”, “death-positive”, and “death-negative”. The response options were set to “True” and “False”. These words were selected following consultation of the Affective Norms for English Words battery (ANEW: Bradley & Lang, 1999) and the agreement of two researchers acquainted with the research area. Responding rule A was “Please answer as if life is pleasant and death is painful” and responding rule B was “Please answer as if life is painful and death is pleasant”. Due to their greater complexity, the internal consistencies of the two IRAPs are reported in the results section.

Table 2.3. *Stimulus sets for the death-evaluation IRAP*

Label stimuli		Target stimuli		Response options	
Living	Dying	Enjoyable	Awful	True	False
		Exciting	Distressing		
		Great <sup>1</sup>	Hurtful		
		Lovely	Horrible		
		Pleasant	Painful		
		Satisfying	Upsetting		

*Note:* Label stimuli appear at the top of the screen and target stimuli in the middle of the screen. <sup>1</sup>Vernacular usage of the word “great” in Ireland is most frequently as a synonym of positively valenced words (e.g., “excellent”) rather than denoting quantity or rank (cf. “substantial” and “prominent”).

### Procedure

All experimental sessions were conducted in individual experimental cubicles. This was done in a one-to-one setting with a trained researcher. After obtaining informed consent was, the researcher performed a verbal assessment of all inclusion criteria. Participants completed the self-report measures first, followed by the death-identity IAT, death-identity IRAP, and death-evaluation IRAP. The order of the three implicit measures was fully counterbalanced between participants. No remuneration was offered.

**Death-identity IAT.** The IAT consisted of two instruction screens followed by seven blocks, each of which consisted of a number of trials (see Table 2.4 on next page). The pre-block instruction screens contained the following written instructions:

“For this portion of the study, words will appear one at a time in the middle of the screen. Classify those words into groups which will be designated with labels appearing on the top half of the screen. All words belonging to the groups on the left will be classified with the “e” key. All words belonging to the groups on the right will be classified with the “i” key. Classify the words as quickly as possible while making as few mistakes as possible. Accuracy and speed are both important. Pay close attention to the group labels, they will change from block to block. Direct any questions to the experimenter.”

“For the next portion of this study, you will be asked to classify words into the categories of DEATH and LIFE, as well as words related to ME and NOT ME. The words related to each of the categories are shown below. Remember, when the word in the centre corresponds to the category on the left, you will use the “e” key, and when the word in the centre corresponds to the category on the right, you will use the “i” key. Classify the words as quickly as possible while making as few mistakes as possible.”

Table 2.4. *Structure of the IAT (adapted from Nosek et al., 2007, p.268).*

Block	Number of trials	Categories assigned to the “e” key	Categories assigned to the “i” key
1	20	Self	Other
2	20	Death	Life
3	20	Self + death	Other + life
4	40	Self + death	Other + life
5	40	Other	Self
6	20	Other + death	Self + life
7	40	Other + death	Self + life

*Note:* The order of presentation of the category pairings in blocks 3, 4, 6 and 7 (i.e., self-death/others-life vs. self-life/others-death) was counterbalanced between participants.

Before each block, the message “Check categories - Press space bar when ready” appeared at the bottom of the screen. A trial was defined as the time in milliseconds from the onset of a stimulus to the emission of a correct response. The numbers of trials in each block and the classes of stimuli presented in each are presented in Table 2.4. The stimulus categories (self, other, death, life) remained on the top left and top right of the screen throughout each block. Self and other related words were presented in white, whereas death and life related words were presented in green. Each trial presented the to be categorized stimulus in the middle of the screen. Participants responded using the “e” and “i” keys. If an incorrect response was emitted, a red “X” was displayed below the stimulus, and a correct

response was required before proceeding to the next trial (see Figure 2.1). After each trial the stimulus in the middle of the screen was cleared for an inter-trial interval of 250ms.



*Figure 2.1.* The IAT trial screen. The categories “life” and “death” alternated positions between blocks 4 and 5, whereas the “me” and “not me” categories were static. The red “X” appeared only if an incorrect response was emitted.

**IRAPs.** Participants were verbally instructed in how to complete the IRAP in several stages using a pre-written script (see Appendix I). In contrast to much previously published research (e.g., Nicholson & Barnes-Holmes, 2012), no additional written or on-screen instructions were provided. The experimenter’s verbal instructions for the death-identity IRAP contained the following key points, which were delivered before the participant completed the first practice block. If a participant indicated a lack of clarity around any point, as the researcher worked through the script, that point was reiterated and clarified to the participant’s satisfaction. The instructions for the death-evaluation IRAP were identical other than the specific stimuli that were referred to.

1. Participants were instructed that they would be presented with pairs of words related to “self”, “others”, “death” and “life”, and would be asked to respond to those pairs as being “Similar” or “Different”.

2. They were informed that, unlike a questionnaire that asked for their subjective opinion, this behavioural task simply required that they follow a rule, and this rule would be provided on screen.
3. Next, they were instructed that the rule would swap after each block, that there were only two rules, and that they would be reminded of the rule for the following block on screen.
4. It was emphasized that they were to initially go as *slowly* as they needed to get as many trials as possible ‘right’ according to the rule, and that they would naturally become faster with practice. Furthermore, it was emphasized to each participant that they must learn how to be accurate before they could learn to go both quickly and fluently. Once they had learned to be accurate they should then naturally learn to speed up.
5. Finally, they were then informed that they would complete pairs of practice blocks until they learned to meet accuracy and speed criteria that would be presented at the end of the block. Once these were met on both blocks within a pair, they would then complete three pairs of test blocks.

The IRAP task consisted of up to three pairs of practice blocks and exactly three pairs of test blocks. Each block included 1) a pre-block rule screen, 2) 24 trials, and 3) a post-block feedback screen. The pre-block rule screen contained the responding rule for the forthcoming block (rule A or rule B, see above), as well as the instructions “Try to get as many as possible ‘right’ according to the rule. If you go over time on any trial “!” will appear. If you get one wrong an “X” will appear – press the correct response to continue.” Each trial presented participants with one label stimulus at the top of the screen, one target stimulus in the middle of the screen, and the response options on the bottom left and bottom right of the screen. The correct response option differed depending on the trial-type and alternated between the blocks,

in line with the rules presented before that block (see Figure 2.2). Participants responded using the “d” and “k” keys. If an incorrect response was emitted, a red “X” was displayed below the stimulus, and a correct response was required before proceeding to the next trial. After each trial the stimulus in the middle of the screen was cleared for an inter-trial interval of 400 ms. The post-block feedback screen displayed both the participant’s percentage accuracy and median latency performance on the previous block and the mastery criteria (i.e., accuracy  $\geq 80\%$  and median latency  $\leq 2000$  ms.).

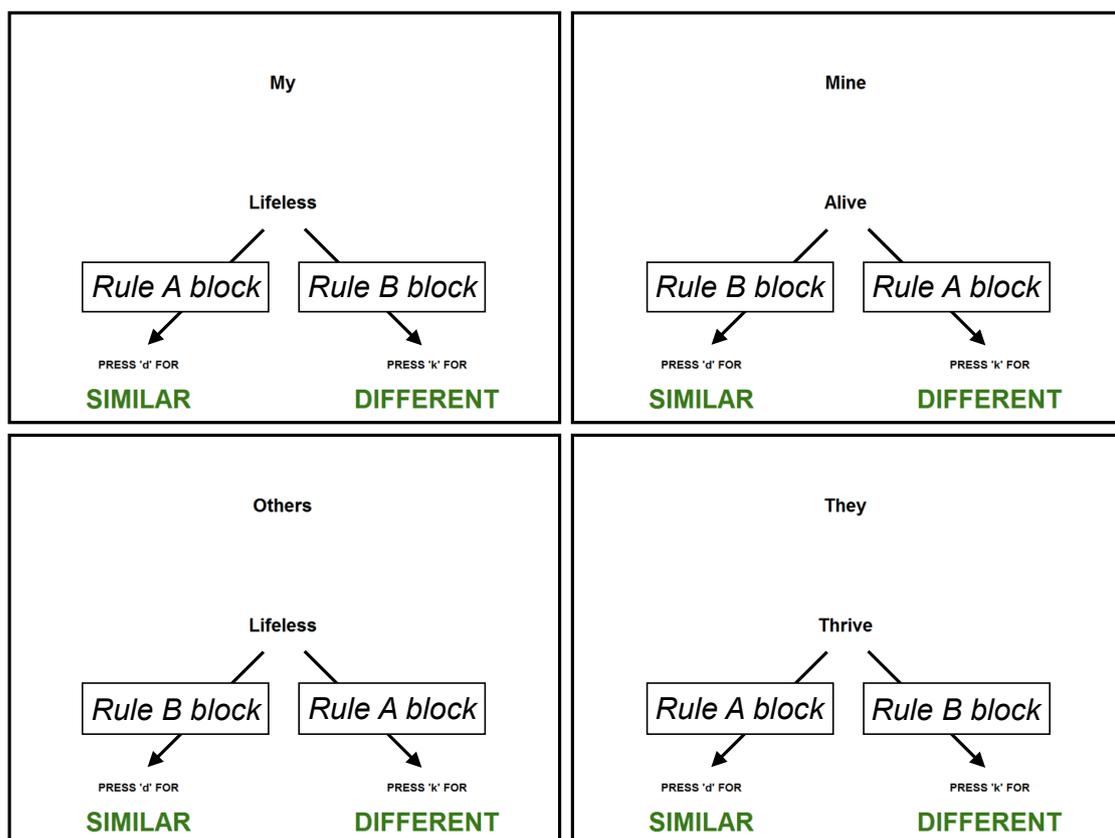


Figure 2.2. The four IRAP trial-types. Superimposed boxes and arrows indicate the correct and required responses on each of the two blocks (rule A vs. rule B).

**Data processing.** The primary datum produced by both the IAT and IRAP is reaction time in millisecond from the onset of a trial to the first correct response. Effects on both the IAT and IRAP are defined as the latency difference between the two blocks in a block pair

(e.g., rule A block vs. rule B block). The IAT provides one overall bias score (e.g., “life-self/death-others”), whereas the IRAP provides four, one for each of the trial-types (e.g., “life-self”, “life-others”, “death-self”, and “death-others”). The effects on both measures were quantified using the *D* score, which has been found to limit the impact of extraneous variables such as responding speed and age (Greenwald et al., 2003). Briefly, *D* scores were calculated as follows. First, latencies above 10,000 ms. were removed. Then, for each pair of blocks, *D* is equal to the difference between mean block A reaction times and mean block B reaction times, divided by the standard deviation of all the reaction times included in both rule A and rule B blocks. In the current body of work, reaction times from rule B blocks were always subtracted from rule A blocks. There are some differences in the calculation of *D* between the IAT and IRAP, which will now be discussed.

***IAT data processing.*** *D* scores were calculated separately for the practice block pair (blocks 3 & 4) and test block pair (blocks 6 & 7) before being averaged to create a final *D* score. Previous research has demonstrated that excluding participants’ IAT data based on accuracy or latency (e.g., accuracy  $\geq$  80%, median latency  $\leq$  2000ms) criteria does not significantly increase the reliability or validity of the measure (Nosek et al., 2007). As such, no such data exclusion criteria were applied here. Positive *D* scores represented quicker responding on rule A blocks (i.e., where self was coordinated with death and others was coordinated with life) relative to rule B blocks (i.e., where self was coordinated with life and others was coordinated with death). This can be interpreted as a “self-death/others-life” effect. In contrast, negative *D* scores represented the opposite pattern of responding, where responding on rule B blocks was quicker than on rule A blocks. This can be interpreted as a “self-life/others-death” effect.

***IRAP data processing.*** As with the majority of previous research using the IRAP, one *D* score was calculated for each of the IRAP’s four individual trial-types (D. Barnes-Holmes,

Barnes-Holmes, et al., 2010). Practice block data were not included in the analysis.

Accuracies on the IRAP are typically lower than the IAT. As such, in order to ensure that IRAP effects were derived from performances that involved the targeted patterns of stimulus control, *D* scores were excluded if a participant failed to maintain the mastery criteria in the test blocks. Although a number of specific exclusion strategies could in principle be applied, the current study adopted the method that is most commonly employed by previous research (e.g., Nicholson and Barnes-Holmes, 2012), which equally balances the two goals of (a) removing unwanted performances, and (b) minimizing attrition. Following these authors' approach, *D* scores from IRAP test blocks that failed to meet criteria were excluded from the analysis in the following manner:

1. If accuracy on one or both test blocks within a pair was < 78% and/or median latency was > 2000ms, then the four *D* scores from that test block pair were excluded.
2. If only one of a participant's three test block pairs were excluded in this manner, the final *D* scores were calculated by averaging the *D* scores across the two remaining test block pairs. *D* scores for three participants were calculated on this basis.
3. If more than one of a participant's three test block pairs were excluded in this manner, all of the data from that particular IRAP was excluded from the analysis. IRAP data for three participants were removed on this basis.<sup>4</sup>

It should also be noted that, for consistency, these same criteria were adopted for all studies within the current body of research. Following convention, *D* scores for two of the

---

<sup>4</sup> It should be noted that this exclusion strategy is applied at the task level rather than the participant level in this and all subsequent experiments. That is, a participant may have a single IRAP excluded from the analysis but the other included. As such, the above list of exclusions represents the total number of IRAP performances that were excluded and not the number of excluded participants. For this reason, the relationship between the numbers of recruited participants, the exclusions performed, and the degrees of freedom in the analyses do not have a 1:1 correspondence.

four trial-types were inverted. In this case, trial-types 3 and 4 on both the death-identity and death-evaluation IRAPs were inverted so as to create a common axis within each IRAP. Specifically, for the death-identity IRAP, the others-life and others-death trial-types were multiplied by -1. Positive *D* scores on this IRAP therefore represented “death” or “not-life” effects, whereas negative *D* scores represented “life” or “not-death” effects, depending on whether the stimuli within that trial-type referred to life or death, respectively. For example, if a positive *D* score was found on trial-type 1 this would be interpreted as a “self-death” effect (i.e., participants responded to self and death with “similar” more rapidly than with “different”); in contrast, if a negative *D* score was found it would be interpreted as a “self-not-death” effect (i.e., participants responded to self and death with “different” more rapidly than with “similar”). Axes were therefore arranged in a similar manner to the IAT in order to allow for comparison. That is, on both measures, positive scores represented pro-death effects and negative scores represented pro-life effects.

A comparable set of inversions was applied to the death-evaluation IRAP: *D* scores for the “death-positive” and “death-negative” trial-types were multiplied by -1. This IRAP employed “True” and “False” as response options and, as such, positive *D* scores represented “positive” or “not-negative” effects, whereas negative *D* scores represented “negative” or “not-positive” effects (see Hussey, Thompson, McEnteggart, Barnes-Holmes, & Barnes-Holmes, 2015 for an in depth-discussion of the interpretation of IRAP trial-type effects).

## 2.2 Results

### Self-report measures

The sample reported normative levels of depression ( $M = 4.6, SD = 4.8$ ), hopelessness ( $M = 4.7, SD = 4.0$ ), psychological flexibility ( $M = 19.9, SD = 8.4$ ), and death anxiety ( $M = 8.1, SD = 2.0$ ). The sample was roughly equally divided by their belief in the afterlife ( $N = 12$  low belief,  $N = 12$  ambivalent belief,  $N = 20$  high belief;  $M = 35.5, SD = 17.3$ ). Results from

the self-report measures therefore indicated that the sample could be considered to represent normative levels of psychopathology and attitudes to death.

### **Death-identity IAT**

After latencies above 10,000 ms were excluded, mean response time on the IAT was found to be 1132 ms ( $SD = 915$ ). The sample produced an overall “self-life/others-death” effect on the IAT, as hypothesized ( $M = -0.34$ ,  $SD = 0.43$ ; see Figure 2.3). That is, participants were faster to pair self with life and others with death than self with death and others with life. A one sample  $t$ -test showed that this effect was significantly different from zero,  $t(41) = -5.2$ ,  $p < 0.001$ . Participants were therefore faster to categorize stimuli when self was coordinated with life and others was coordinated with death relative to when self was coordinated with death and others was coordinated with life. However, as discussed previously, the nature of the IAT precludes the ability to determine which specific relational responses drove this effect (i.e., self-life, self-death, others-life, others-death). IAT  $D$  scores were not correlated with age, depression, hopelessness, psychological flexibility, belief in the afterlife, or death anxiety (all  $ps > .05$ , see Table 2.5).

When IAT  $D$  scores were binarized using zero as a cut-off score, as employed by previous research (Nock et al., 2010; Randall et al., 2013), 32 participants were shown to have overall “self-life/others-death” effects (i.e.,  $D$  scores  $\leq 0$ ) and 7 participants were shown to have overall “self-death/others-life” effects (i.e.,  $D$  scores  $> 0$ ). The majority of the sample (82%) therefore demonstrated effects on the IAT that were consistent with what previous research has classified as low risk of attempted suicide or self-harm, as might be expected within a normative sample. Regardless of scores on the IAT or any of the self-report measures, all participants were fully debriefed and made aware of the availability of the free student counselling service and provided with details via the consent form. Ethical approval

precluded the discussion of suicidality; therefore debriefings were not tailored to the individual.

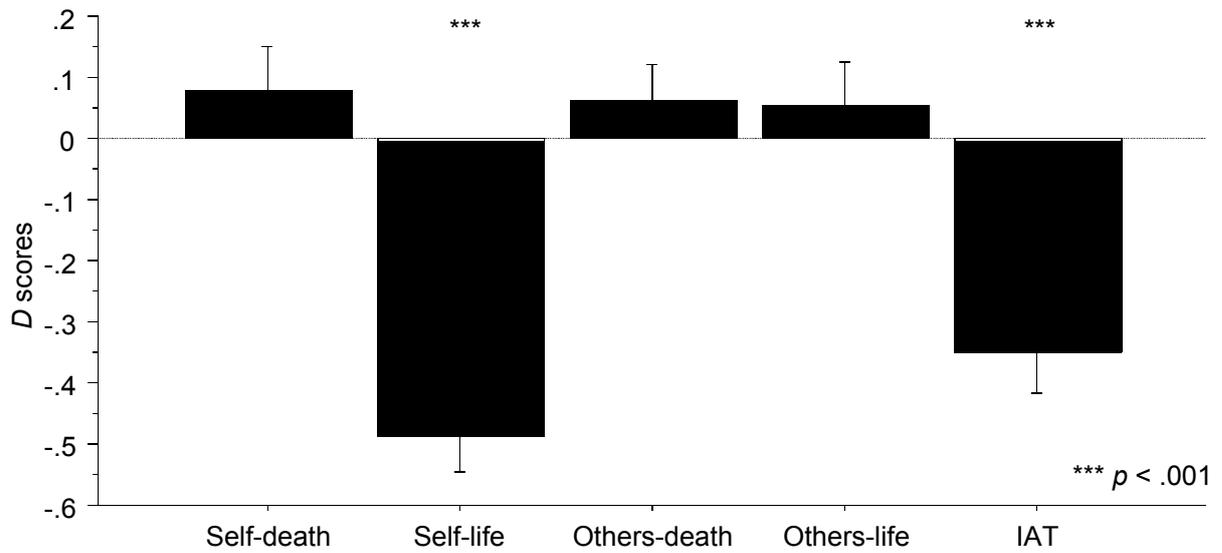


Figure 2.3. Performance on the death-identity IRAP and IAT. Positive scores represent “death” biases, whereas negative scores represent “life” biases. Bars 1-4 represent mean *D* scores on the IRAP trial-types. Bar 5 represents mean IAT *D* scores. Error bars represent standard errors.

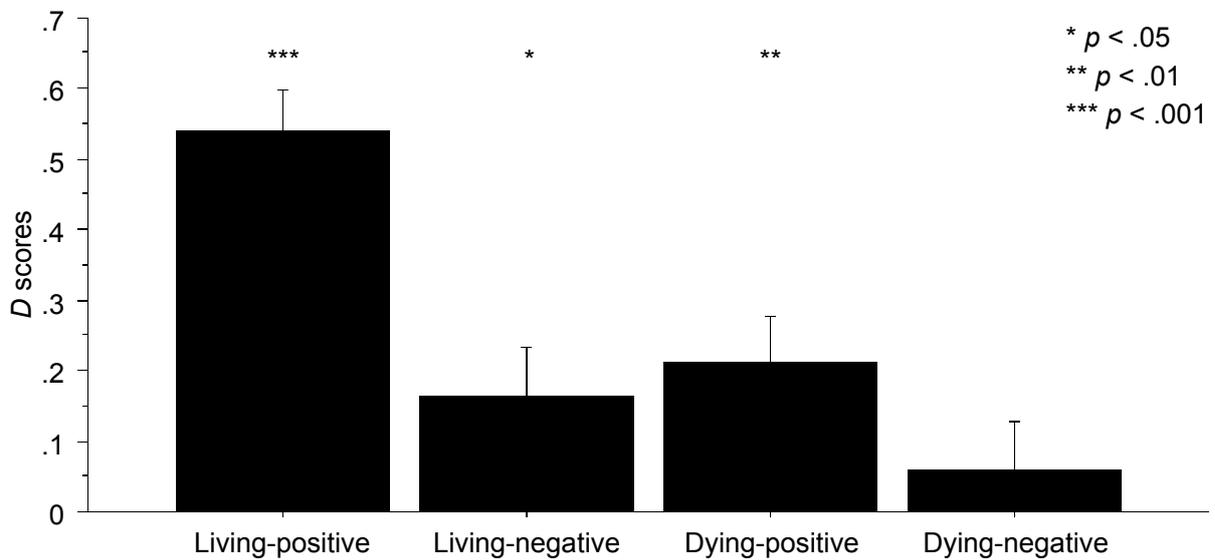


Figure 2.4. Performance on the death-evaluation IRAP. Positive scores represent “positive” or “not-negative” evaluative biases, whereas negative scores represent “negative” or “not-positive” evaluative biases. Bars represent mean *D* scores on each IRAP trial-type. Error bars represent standard errors.

## Death-identity IRAP

After latencies of 10,000 ms were removed, mean response time on the death-evaluation IRAP was found to be 1285 ms (SD = 650). Mean scores for the four death-identity IRAP trial-types and the death-identity IAT are presented in Figure 2.3, which demonstrates participants' strong "self-life" biases (i.e., participants were faster to respond to "self" and "life" with "similar" relative to "different":  $M = -.49$ ,  $SD = .37$ ). No strong effects were found on the other three trial-types (self-death:  $M = .08$ ,  $SD = .44$ ; others-death:  $M = .06$ ,  $SD = .37$ ; others life:  $M = .05$ ,  $SD = .44$ ). It should be noted that the IAT is included in this figure only for the purposes of visual comparison, and was not included in the following ANOVA. A within-subjects ANOVA confirmed significant differences between the IRAP trial-types  $F(3, 40) = 15.9$ ,  $p < .0001$ ,  $\eta^2 = .30$ . Follow-up Bonferroni-corrected one sample  $t$ -tests demonstrated a significant "self-life" effect ( $p < 0.0001$ ), but no other effects (all  $ps > .27$ ). This result therefore provides evidence for the assumptions made with previous research, whether explicitly or implicitly (e.g., Nock et al., 2010; Dickstein et al., 2015) that normative participants do indeed demonstrate "self-life" effects *specifically*, as opposed to "self-not-death" biases.

IRAP  $D$  scores were not correlated with age or self-reported depression, hopelessness, psychological flexibility, belief in the afterlife, or death anxiety (see Table 2.5), with one exception: death anxiety and the "others-life" trial-type ( $p < .05$ ). Given the large number of correlations performed and the lack of a systematic pattern, however, this effect should be interpreted with extreme caution.

## Relationship between the death-identity IAT and IRAP

Responses on the death-identity IAT demonstrated a "self-life/others-death" effect. In contrast, the results of the death-identity IRAP suggest that the IAT effect may have been driven by a "self-life" bias specifically, given that the biases on the other three trial-types

were all relatively weak. A dependent *t*-test showed that the magnitude of the IRAP’s “self-life” trial-type did not differ significantly from the IAT *D* score,  $t(38) = 1.4, p = .16$ . On balance, however, a series of correlations failed to find any correlations between the IAT and any of the IRAP trial-types, including the “self-life” trial-type ( $r_s = -.16$  to  $.14$ ).

Table 2.5. *Correlations between the implicit and self-report measures.*

	DI-IAT	DASS-D	BHS	AAQ-II	BIA	TDA	Age
Death–identity IAT	1.0	.19	.23	.27	-.09	.18	.03
<i>Death–identity IRAP</i>							
Self-death	.04	-.09	.00	.05	.14	-.05	-.12
Self-life	-.16	-.07	.03	.15	-.10	.07	-.30
Others-death	.14	-.22	-.14	-.31	.01	-.15	.15
Others-life	.12	-.15	-.06	-.26	.14	-.33*	.17
<i>Death-evaluation IRAP</i>							
Living-positive	-.27	.05	.00	.21	.17	.04	.02
Living-negative	-.38*	-.14	-.15	-.15	-.21	.03	.08
Dying-positive	.01	.05	.03	.06	.12	.09	.08
Dying-negative	-.07	.07	.26	.05	.01	-.01	.06

*Note:* DI-IAT = death-identity IAT, DASS-D = Depression Anxiety Stress Scales’ depression subscale, BHS = Beck Hopelessness Scale, AAQ-II = Acceptance and Action Questionnaire II, TDA = Templer Death Anxiety scale. For ease of interpretation, correlations among the IRAP trial-types are not included here.

\*  $p < .05$

### Death-evaluation IRAP

After latencies of 10,000 ms were removed, mean response time on the death-evaluation IRAP was found to be 1256 ms (SD = 693). Mean scores for the four death-evaluation IRAP trial-types are presented in Figure 2.4. Participants’ demonstrated strong “life-positive” biases (i.e., they were faster to respond to “life” and “positive” with “true” relative to “false”:  $M = .54, SD = .38$ ). Smaller “life-not-negative” ( $M = .16, SD = .46$ ) and “death-positive” biases ( $M = .21, SD = .41$ ) were also found. No biases were found on the “death-negative” trial-type ( $M = .06, SD = .43$ ). A within-subjects ANOVA demonstrated significant differences between the IRAP trial-types,  $F(3, 40) = 11.4, p < .0001, \eta^2 = .34$ .

Follow-up Bonferroni-corrected one sample *t*-tests demonstrated significant “life-positive” and “death-positive” effects ( $ps < .0125$ ). Effects on the “life-negative” and “death-negative” trial-types were not significantly different from zero ( $ps > .0125$ ). As such, effects on the life-related trial-types are relatively intuitive for a normative sample: participants confirm that life is positive. However, the “death-positive” effect is both unexpected and counterintuitive. That is, the sample responded “True” more quickly than “False” on the “death-positive” trial-type, despite reporting normative levels of depression, hopelessness, and death anxiety on the self-report measures. Specifically, a correlation matrix indicated that none of the IRAP trial-types were significantly correlated with age, depression, hopelessness, psychological flexibility, belief in the afterlife, or death anxiety ( $rs = -.21$  to  $.26$ ; see Table 2.5). Finally, it should be noted that a significant correlation was found between the “life-negative” trial-type and the death-identity IAT ( $r = -.38, p = .02$ ). This indicates that a bias towards rejecting that life is negative on the IRAP was associated with a “self-life/others death” bias on the IAT.

### 2.3 Discussion

The current study exposed participants to the death-identity IAT, which has previously been used to study attitudes to death and their relation to suicidality (e.g., Nock et al., 2010; Dickstein et al., 2015). In addition, participants were asked to complete two versions of the IRAP, one of which targeted death-identity and the other of which targeted death-evaluations. The IRAPs were designed to provide information on the strength of specific relational responses rather than overall biases, as is the case with the IAT. The key findings in this largely exploratory study were as follows. The “self-life/others-death” bias found in previous studies using the IAT with normative samples was replicated in the current results (e.g., Dickstein et al., 2015; Harrison et al., 2014; Tang et al., 2013). Intuitively, one might expect that such biases reflect a general positivity to and preference for life over death. Indeed, such attitudes are reflected in many psychological theories (e.g., need for meaning in

life, Greenberg et al., 1986; suicidality, Joiner, 2005). Somewhat unexpectedly, therefore, although both IRAPs employed in the current research produced biases that reflected strong “pro” life responses, they also produced either weak “anti” or “pro” death responses. Indeed, one trial-type (“death-positive” on the death-evaluation IRAP) was in a significant “pro” death direction. How might we explain this curious implicit bias that reflects the lack of an aversive response to death, or even more bizarrely, a bias towards it?

At this stage, it is worth emphasizing that this was a normative sample of participants, and that the IRAP and IAT effects, in general, failed to correlate with, or even produce effects that were consistent with any of the self-report measures (i.e., death anxiety, belief in the afterlife). As such, it would be difficult to explain these implicit biases as reflecting the participants’ self-reported beliefs and attitudes towards death or dying or as indicators of non-normative attitudes to death, such as suicidality (i.e., hopelessness, depression). Indeed, as discussed in the introduction, positive evaluations of death in normative participants run contrary to a wide range of psychological theories of death and suicidality, which assume that death is aversive and/or negatively evaluated (e.g., Joiner, 2005; R. C. O’Connor, 2011).

Given that we cannot appeal readily to something specific or unusual about the sample of participants, perhaps we should consider the impact of the stimuli employed within the IRAPs. Specifically, it is possible that the death-related words employed were not strongly valenced, particularly for a sample of young college students. In other words, death and mortality may be so far removed from the immediate psychological awareness of a young person that these words lack the salience to produce relatively strong IRAP effects (i.e., on the death-identity IRAP). Thus, it might be useful to consider the possible effects of using death-related words that are less abstract to a young person, in the sense that they refer to more proximal adverse events. For example, the word “cancer” would likely have been more evocative of specific functions (e.g., pain, frailty, fear: see Barnes-Holmes, Keane, Barnes-

Holmes, & Smeets, 2000) because these may have been observed directly in a family member who contracted the disease. Of course, while this explanation may account for the absence of death biases on one IRAP (i.e., death-identity), it is harder to explain the presence of a “death-positive” effect on the other IRAP (i.e., death-evaluation).

One alternative explanation for these patterns of bias might be found in the stimuli that were included in the IRAPs. First, it should be noted that the death-related stimuli differed between the death-identity and death-evaluation IRAPs, both in the number of exemplars and the specific stimuli employed (see Tables 2.2 & 2.3). It is possible that this difference may contribute to any differences between them. Possibly more importantly, death can have a variety of valence functions depending on the context. For example, death could involve a lonely, painful decline, or it could involve a romantic (e.g., Romeo and Juliet) or heroic (e.g., Martin Luther King) demise. In retrospect, it is therefore difficult to know exactly what functions of death the stimuli evoked within the IRAPs. Future studies will therefore seek to narrow the range of likely functions that their stimulus sets target. For example, future work might at minimum specify a relevant deictic relation. That is, *whose* death is being referred to (personal death, death of a specified other, death of an unspecified other, etc.)? This question will be addressed in Chapters 4 and 5.

The absence of correlations between the death-identity IAT and IRAP is also worth commenting upon. This result might be seen as somewhat surprising, given that the two measures employed highly similar stimuli. On balance, however, previous research has reported mixed results in this regard, with some studies reporting correlations between implicit measures and others not (see Bosson, Swann, & Pennebaker, 2000; cf. Golijani-Moghaddam, Hart, & Dawson, 2013). Of course, the two tasks do bear important methodological differences. For example, as a relative measure, the IAT presents all four categories on each trial, whereas the IRAP presents only one of the four possible pairings of

these categories per trial. Thus, the lack of a strong and consistent relationship between the measures, which is common in the literature, should not cause excessive concern for the current study. It should also be noted that the lack of correlation between the IAT and IRAP should not necessarily be interpreted as problematic for either measure, given that meta analyses have shown both to have relatively high levels of predictive validity in clinically relevant domains (see Fazio & Olson, 2003; Vahey et al., 2015 for reviews).

In closing, we conclude that the unexpected “death-positive” effects found in the current study ought to be explored further. There are at least two ways in which this might be pursued. First, subsequent studies will seek to better understand the behavioural processes involved in IRAP performances themselves. In so doing, we may be in a better position to explain how unexpected or counter-intuitive patterns of bias emerge, such as the “death-positive” biases observed in the current study. Chapters 3 will therefore examine one as-yet unexplored source of context control within the IRAP: the interactions between the stimulus categories. This strategy will require us to move beyond the notion of the IRAP as a measure of “implicit attitudes” and towards understanding it as a measure of the dynamics of relational responding, as has been argued recently elsewhere (Barnes-Holmes, Barnes-Holmes, & Hussey, in press; Hussey, Barnes-Holmes, & Barnes-Holmes, 2015). Second, future studies will pose questions about these effects in terms of understanding relational response biases in the context of a history of suicidal behaviours, in order to understand whether these effects are consistent across individuals with both normative and non-normative attitudes to death. Specifically, Chapter 5 will compare evaluations of death on the IRAP between normative individuals and those with a history of suicidal behaviours (e.g., ideation and/or attempts). Admittedly, these represent two very different approaches to IRAP research, but both would seem to warrant attention.

### **CHAPTER 3: AN EMPIRICAL INVESTIGATION OF THE ROLE OF THE CONTRAST CATEGORY WITHIN THE IRAP**

Abstract: This chapter presents two proof-of-concept experiments that explore the degree to which effects on the IRAP's four trial-type are independent of one another. In order to do this, a "contrast category manipulation" method is employed, whereby a single category is manipulated between IRAPs and changes in responding to other category pairings are observed. Experiment 1 does this within the context of self-esteem, and Experiment 2 within the objectification of women. Results from Experiment 1 were found to be difficult to interpret. After a number of alterations to the design, results from Experiment 2 demonstrate the viability of the contrast category method as a way to more precisely specify which functions of a stimulus class are being targeted within an IRAP.

The previous chapter explored the strength of specific relational responses around death-identity and death-evaluation. In doing so, it utilised the procedural non-relativity of the IRAP's four trial-types, which can be contrasted with the IAT's single relative bias score. Specifically, whereas the IAT presents all four categories on each trial and assesses the relative bias for one pattern of category pairings over the other (e.g., self-life/others-death vs. self-death/others-life), the IRAP presents individual category pairings separately and can provide bias scores for each (e.g., self-life, self-death, others-life, and others-death). At this point however, it is worth noting that, although the IRAP is indeed *procedurally* non-relative, no research to date has systematically explored the degree to which behaviour on the four trial-types is independent. Behaviour itself is by definition contextually determined; therefore it would be useful to consider how contextual control is exerted by the task itself. Indeed, previous research has noted that "the precision of any particular IRAP is fundamentally

intertwined with the degree of experimental control it is capable of applying to a given analytic question” (Vahey, Boles, & Barnes-Holmes, 2010, p.469).

Let us illustrate the rationale for understanding contextual control within the IRAP using an example. The death-identity IRAP employed within Chapter 2 employed “self” as what we will describe as the *category of interest*, and “others” as the *contrast category* (see Karpinski, 2004 for a similar approach). In this study, “others” was therefore simply used as a category that was deemed to have low salience and would have minimal impact on how participants relate themselves to life and death. Imagine, however, if instead of using a “bland” unspecified “others” category, the category “my mother” was used instead. The question arises, what impact would using a specific other with high emotional valence have on how participants related themselves to life and death. In other words, do we value our own lives less when they are contrasted against the life of a loved one rather than an unspecified other? More technically, this refers to the question of whether responding on one IRAP trial-type is meaningfully influenced by responding on the other three, despite their procedural separation. And, if so, this begs the question of what strategy researchers might employ when selecting the four stimulus categories, given that they cannot be considered in isolation.

The experiments within the current chapter therefore sought to determine whether manipulations of the contrast category affect responding on the category of interest. Subsequently, we attempt to account for such differences in terms of the relational structure of the IRAP. Finally, we discuss how contrast category manipulations might be employed to provide increased specificity in future IRAP research on relational responding to evaluations of life and death.

### **3.1 EXPERIMENT 1**

This experiment represents a first attempt to examine whether responding on the IRAP’s category of interest is influenced by changes to the contrast category. As such, we

elected to adopt a “good candidate” domain here, with the intention of conducting subsequent experiments within the context of implicit attitudes to death specifically. Karpinski (2004: Experiment 2) is one of a small number of empirical studies that has asked similar questions within the context of the IAT (see also Houben & Wiers, 2006; Karpinski, 2004; Robinson, Meier, Zetocha, & McCaul, 2005). As such, the current study employed similar stimulus categories to Karpinski (2004: i.e., self, “Santa”, “Hitler”, “positive”, & “negative”). Specifically, participants completed two self-esteem IRAPs, where one assessed implicit evaluations of self in the context of Santa, and another assessed implicit evaluations of self in the context of Hitler. As such, the two IRAPs were identical other than the contrast category each employed (i.e., Santa vs. Hitler). We hypothesised that “self” would be evaluated more negatively in the context of responding to a positively-valenced other (i.e., Santa) than in the context of responding to a negatively-valenced other (i.e., Hitler).

### **3.2 Method**

#### **Participants**

Forty undergraduate students at the National University of Ireland Maynooth (28 female and 12 male,  $M_{\text{age}} = 21.5$ ,  $SD = 6.9$ ) were recruited for the current study. No incentives were offered for participation. Inclusion criteria were identical to the previous experiment in Chapter 2.

#### **Ethical considerations**

Ethical approval was granted by Maynooth University’s Social Research Ethics Subcommittee. Participants were informed as to the nature and purpose of the study prior to participation. Furthermore, they were made aware that participation was entirely voluntary, that they could cease participation at any time without giving a reason, and that all data would be irrevocably anonymised immediately after participation. Written informed consent was obtained from all participants. Upon completion of the tasks, participants were fully

debriefed on the nature and purpose of the experiment and provided with an opportunity to ask questions, which were then fully answered by the researcher. During this debrief, the researcher made each participant aware of the free student counselling service available on campus, whose contact details were included at the bottom of their copy of the consent form, should they experience any form of post-experimental distress.

## **Measures**

**Recognition and rating scales.** This 4-item measure asked participants if they were aware of the general characteristics associated with “Santa” and “Hitler” (yes/no response format), and then asked them to rate them using the same descriptors that were employed in the IRAP (see Table 3.1). Specifically, participants were asked “How likable or unlikable do you find Santa to be, from 0 to 10, where 0 means “manipulative, dishonest, cruel, horrible, selfish, or heartless” and 10 means “loyal, trustworthy, kind, moral, generous, or friendly”. The same question was asked with regard to Hitler.

**Rosenberg Self-Esteem Scale.** The RSES (see Appendix J) is a widely used 10-item measure of self-esteem (Rosenberg, 1965). Participants respond on a 4-point Likert scale from 1 (strongly disagree) to 4 (strongly agree), with possible scores ranging from 10 to 40. Internal consistency was good in the current sample ( $\alpha = .85$ ).

**Self-Santa and self-Hitler IRAPs.** The procedural details of both IRAPs were identical to those used in Chapter 2 in all respects, other than the stimulus set and responding rules employed. For the current experiment, stimulus categories were drawn from Karpinski (2004), who employed the categories “self”, “Santa”, “Hitler”, “positive” and “negative”. However, in light of the conceptual criticisms raised by Pinter and Greenwald (2005) regarding the appropriateness of Karpinski’s (2004) specific exemplars (e.g., using “Jews” for the category “Hitler”), an entirely new set of exemplars was derived for the purposes of the current study (see Table 3.1). Specifically, all stimuli were identical in both IRAPs other

than the contrast category (i.e., which was set to either “Santa is” or “Hitler is”). Thus, the stimuli contained within the “self-positive” and “self-negative” trial-types were identical in both IRAPs (see Figure 3.1). For the self-Santa IRAP, rule A was “I am good and Santa is bad” and rule B was “I am bad and Santa is good”. For the self-Hitler IRAP, rule A was “I am good and Hitler is bad” and rule B was “I am bad and Hitler is good”.

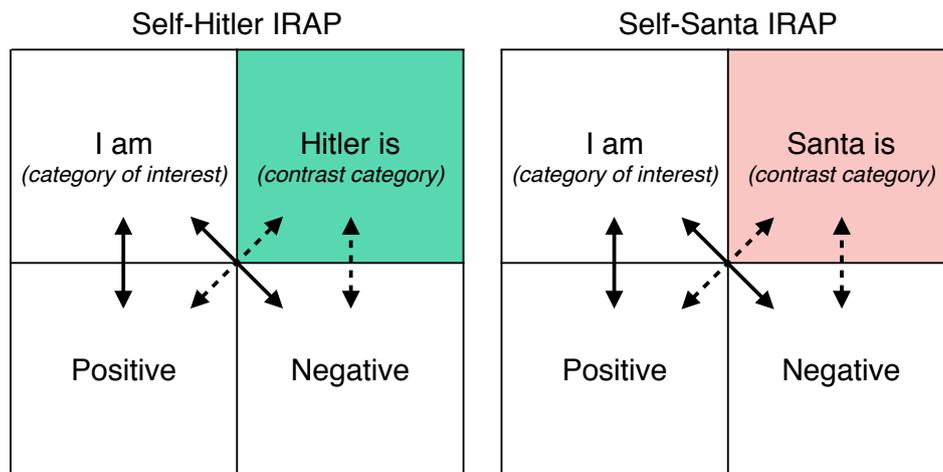


Figure 3.1. The stimulus categories employed in the self-Hitler and self-Santa IRAPs. Solid arrows indicate trial-types that were identical in both IRAPs (i.e., “self-positive” and “self-negative”), whereas dotted arrows indicate trial-types that differed between the IRAPs (e.g., “Hitler is-positive” vs. “Santa is-positive”).

Table 3.1. Stimulus sets for the self-Santa and self-Hitler esteem IRAPs.

Label stimuli				Target stimuli	Response options		
Santa IRAP		Hitler IRAP					
I am	Santa is	I am	Hitler is	Loyal	Manipulative	True	False
				Trustworthy	Dishonest		
				Kind	Cruel		
				Moral	Horrible		
				Generous	Selfish		
				Friendly	Heartless		

### Procedure

Participants completed the recognition and rating scales and the Rosenberg Self-Esteem Scale followed by the self-Santa and self-Hitler IRAPs. The order of presentation of

the two IRAPs and the rule blocks within each IRAP (i.e., rule A first vs. rule B first) were counterbalanced between participants. The format and delivery of the two IRAPs was identical to the previous experiment in all respects.

### **IRAP data processing**

Raw latencies on the IRAP were converted into  $D$  scores and excluded from the analysis if they failed to maintain the mastery criteria in an identical manner to the previous experiment. No participants failed to meet the mastery criteria on the practice blocks, and 3 participants failed more than one test block-pair and thus had their data excluded from the analyses. Twenty-nine individuals therefore remained in the final analysis (this includes exclusions made on the recognition and rating scales, below). In order to create a common axis, trial-types 3 and 4 (i.e., the Santa and Hitler trial-types) were inverted in both IRAPs (similar to Experiment 1). As such, positive  $D$  scores can be interpreted as “positive” or “not-negative” biases, and negative  $D$  scores can be interpreted as “negative” or “not-positive” biases.

## **3.3 Results**

### **Self-reports measures**

**Recognition and rating scales.** All participants affirmed that they were aware of the characteristics associated with both Santa and Hitler. Participants who did not rate Santa as highly positive (i.e.,  $\leq 7$ ) and did not rate Hitler as highly negative (i.e.,  $\geq 3$ ) were excluded from the analysis. Six participants were excluded on this basis. This ensured that the final sample included only those whose explicit ratings of Santa and Hitler differed significantly,  $t(28) = 43.8, p < 0.001, \text{Hedges' } g_{av} = 8.99^5$ . After excluding participants based on their

---

<sup>5</sup> Hedges  $g$  is a variant of Cohen's  $d$  (J Cohen, 1977) and is therefore an estimate of effect size. It is recommended for smaller sample sizes (e.g.,  $n < 20$ : Lakens, 2013). Its interpretation is identical to Cohen's  $d$  (i.e.,  $>.1$  = small,  $>.3$  = medium,  $>.5$  = large,  $>.8$  = very large), and the two scores converge with increasing sample size. As such, in order to provide for direct comparisons of effect size across the different experiments in this body of work, Hedges'  $g$  will be reported throughout (see Lakens, 2013).

responses on the recognition and rating scales and the IRAPs, 29 individuals remained in the final sample.

**Rosenberg self-esteem scale.** Following standard practice, items 3, 5, 8, 9, and 10 were reverse-scored in order to produce a single scale, where higher scores represent higher self-esteem. All participants scored within the “normal” ( $\geq 15$ ) or “high” ( $\geq 25$ ) ranges (see Rosenberg, 1965;  $M = 24.1$ ,  $SD = 4.4$ ).

### **Self-Santa and self-Hitler IRAPs**

Mean  $D$  scores on both IRAPs are depicted in Figure 3.1. Positive  $D$  scores indicated that large “self-positive” biases ( $M_{\text{hitler}} = 0.44$ ,  $M_{\text{santa}} = 0.43$ ) were found on both IRAPs. Similarly, positive  $D$  scores indicated that “Santa-positive” ( $M = 0.18$ ) and, more surprisingly, “Hitler-positive” biases ( $M = 0.20$ ) were found on both IRAPs. No strong “Santa-negative” or “Hitler-negative” effects were found ( $M_{\text{hitler}} = 0.10$ ,  $M_{\text{santa}} = -0.05$ ). Unexpectedly, the only trial-type on which clear differences emerge was the “Self-negative” trial-type. Specifically, participants demonstrated a “self-*not*-negative” bias on the Hitler IRAP, but no clear “self-negative” biases on the Hitler IRAP ( $M = -0.05$ ).

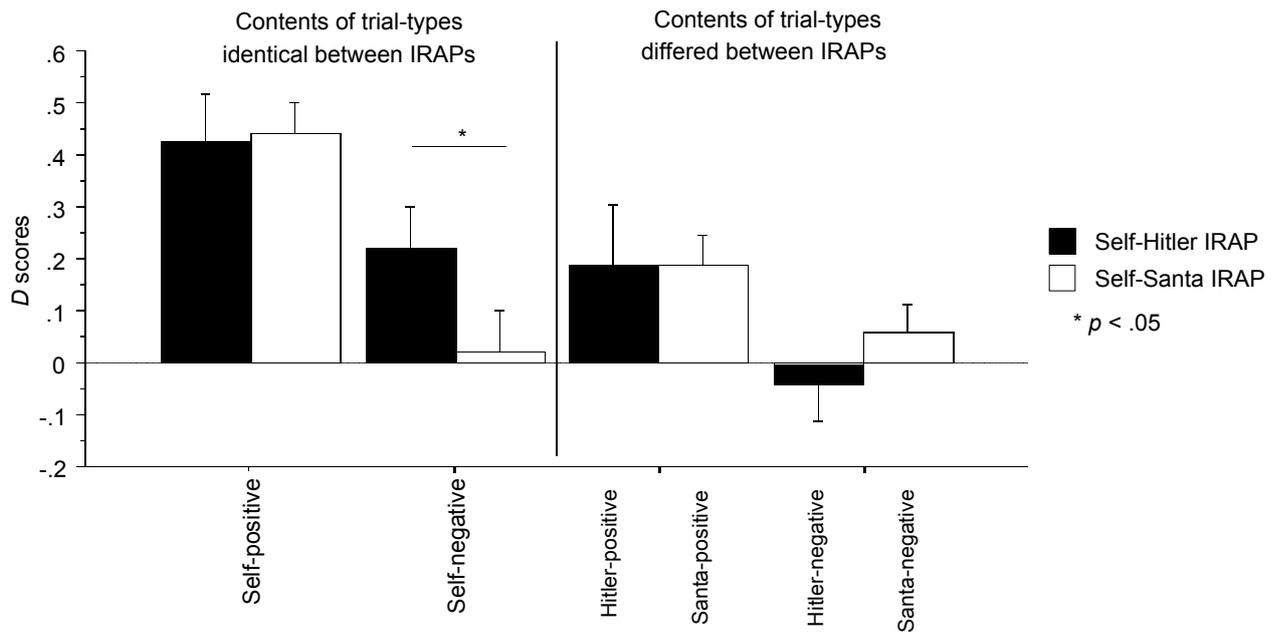


Figure 3.2. Performance on the Self-Santa and Self-Hitler IRAPs. Positive  $D$  scores represent “positive” or “not-negative” biases and negative  $D$  scores represent “negative” or “not-positive” biases. Error bars represent standard errors.

Differences between the self-Hitler and self-Santa IRAPs were assessed using a  $2 \times 4$  within-subjects ANOVA, with IRAP (self-Hitler vs. self-Santa) and IRAP trial-type as the two main variables. No significant main effect for IRAP type was found ( $p = 0.43$ ), but a main effect for trial-type was recorded,  $F(3, 28) = 11.78, p < .0001, \eta^2 = .30$ . Critically, as hypothesized, an interaction effect emerged,  $F(3, 28) = 2.82, p < .05, \eta^2 = .09$ . Follow-up Bonferroni-corrected dependent  $t$ -tests for each trial-type indicated that marginally significant differences were found on the “self-negative” trial-type between conditions,  $t(28) = 2.57, p = .02$  (all other  $ps > .14$ ). These differences were found to be large, Hedges'  $g_{av} = .67$ . Thus, participants did not evaluate Santa and Hitler differentially on the two IRAPs. However, “self-positive” biases did not differ between the two IRAPs.

### Relationship between the implicit and self-report measures

For the purposes of consistency across experiments reported in the current body of work, a correlation matrix was calculated for the trial-type  $D$  scores and the Rosenberg Self-

Esteem Scale. All eight correlations were non-significant,  $r_s = -.21$  to  $.10$ , all  $p_s > .28$  (see Table 3.2).

Table 3.2. *Correlations between the implicit and self-report measures.*

IRAP trial-type	<i>r</i>
<i>Self-Hitler IRAP:</i>	
Self-positive	0.10
Self-negative	0.05
Santa-positive	0.20
Santa-negative	-0.21
<i>Self-Santa IRAP:</i>	
Self-positive	0.09
Self-negative	-0.11
Hitler-positive	0.12
Hitler-negative	-0.12

*Note:* No correlations reached significance.

### 3.4 Discussion

The results of the current study were somewhat counterintuitive. Three main points should be noted. First, marginally significant yet highly unexpected “Hitler-positive” biases were found on the IRAP, despite participants evaluating Hitler negatively on the self-report rating scale. In contrast, participants evaluated Santa as positive on both the self-report rating scale and the IRAP. Second, and perhaps equally puzzlingly, despite Santa and Hitler being evaluated as equally positive on the two IRAPs, differences in the evaluation of the negativity of “self” nonetheless emerged. This pattern was somewhat more intuitive. In the context of Hitler, a positive bias was demonstrated on the “self-negative” trial-type, whereas in the presence of Santa, no bias was found (see Figure 3.1). As one might expect, self was therefore evaluated as comparatively less negative relative to Hitler and comparatively more negative to Santa. Third, and critically, our ability to draw conclusions from the current results about whether the contrast category can influence responding on the category of

interest are limited, given that participants responses to both Santa and Hitler were equally positive on the two IRAPs.

In summary, differences were found in the category of interest (i.e., self) between the two IRAPs, however these differences were not necessarily due to differential evaluations of the contrast categories (i.e., Santa vs. Hitler), which did not differ between conditions. The results therefore provide mixed evidence for whether the contents of the contrast category influences responding on the category of interest.

It should be noted that the results of the current study are therefore somewhat consistent with Karpinski (2004), insofar as differences were found between the two implicit measures, and furthermore none of the implicit measures in either study (i.e., self-Santa IAT, self-Santa IRAP, self-Hitler IAT, or self-Hitler IRAP) correlated with the Rosenberg Self-Esteem Scale. Research on implicit self-esteem using the IRAP has elsewhere sometimes found correlations between self-reported and implicit self-esteem. For example, Vahey, Barnes-Holmes, Barnes-Holmes and Stewart (2009) found significant correlations between a self-esteem IRAP and a self-esteem feeling thermometer. Further mixed evidence comes from a study by Remue, Hughes, De Houwer and De Raedt (2014) who found correlations between “actual” implicit self-esteem (e.g., “I am positive”) and the RSES, but not “ideal” implicit self-esteem (e.g., “I want to be positive”). However, these correlations were found only in the high dysphoric group (i.e., Beck Depression Inventory II  $\geq 14$ ) but not in the low dysphoric group (i.e., BDI-II  $\leq 13$ ). The results of the current study are therefore somewhat difficult to integrate with the broader literature on implicit self-esteem. Unpacking self-esteem will likely require more systematic analyses in future research.

At this point, it is worth considering whether there were an insufficient number of exemplars employed for the categories Santa and Hitler. This has been found to reduce the strength of biases on the IAT (Nosek et al., 2005). However, single category exemplars have

previously been employed in a number of successful IRAP studies (see (Golijani-Moghaddam et al., 2013; Vahey et al., 2015). Furthermore, while the above points might explain the absence of effects on the IRAP (e.g., the lack of a “Hitler-negative” bias), neither can readily explain the presence of unexpected effects, such as the marginally significant “Hitler-positive” effects found in the current experiment. In this regard, there are some similarities to be drawn between the unexpected results found on the Hitler IRAP and the death-evaluation IRAP employed in the previous chapter.

In closing, we concluded that it would perhaps be more fruitful to explicate and revise our strategy for demonstrating and manipulating the contrast category. Our explicit rationale for the current experiment was that the two categories (i.e., Santa and Hitler) would be evaluated differentially, and that evaluations of self may differ as a function of this. On reflection, it could be argued that the contrast categories were therefore selected based on a *single dimension* of comparison (i.e., evaluation), with the assumption that Hitler and Santa would be comparatively different along this dimension (e.g., “Santa is more *positive*”, “Hitler is less *positive*”: both are responded to along the dimension of positivity). It may be the case that, rather than manipulating the comparative strength or specificity of the contrast category along a single dimension (e.g., positive other vs. negative other, or a specified vs unspecified other: see Karpinski, 2004), a more useful contrast category manipulation may involve *two different* functional dimensions of comparison across the conditions. This will be elaborated upon in the next experiment.

### **3.5 EXPERIMENT 2**

This experiment sought to address some of the potential methodological shortcomings of the previous experiment. First, the experimental design was changed from within groups to between groups. While a within groups design arguably would provide more convincing evidence of whether responding on the category of interest can be influenced by the contrast

category, it is possible that training effects between two similar IRAPs may obscure such results. While the counterbalancing of the two IRAPs would appear to make this unlikely, a between-groups design was nonetheless deemed useful to eliminate this possibility.

Second, given the unexpected “Hitler-positive” biases found in the previous experiment, we opted to change the domain in the current study. Specifically, we sought a stimulus category that possessed multiple functions that may be brought to bear differentially, depending on the context. It should be noted that the previous study attempted to influence responding on the category of interest by manipulating the degree to which the contrast categories differed along a *single* dimension of comparison (i.e., their assumed differential valence). In contrast, the current study attempted to use the contrast category to specify *two different* dimensions of comparison. Specifically, we hypothesized that “women” would be objectified and dehumanized when contrasted with “men” (i.e., compared along the dimension of *gender*), whereas they would be humanized when contrasted with “inanimate objects” (i.e., the dimension of *agency*: capable of action, possessing mind and autonomy).

We elected to employ attitudes to women as our target domain based on its apparent suitability to this research question for the following reasons. Previous research has shown that there is a general tendency for women to be evaluated more positively than men (e.g., as the more helpful, kind and empathic gender: see Eagly, Mladinic, & Otto, 1991; Glick et al., 2004). However, research elsewhere has demonstrated that women are, simultaneously, all too often stereotyped as being ill-suited to leadership in occupational settings (see Eagly & Karau, 2002 for review). Importantly, this difference in evaluations of women as either positive (e.g., “empathic”) or negative (i.e., “weak”) has been shown to be highly context dependent (i.e., Glick et al., 2004; Rudman & Glick, 2001 for in depth treatments of these issues). We therefore attempted to utilize these differential, context-dependent evaluations in the present study.

Similar to the previous experiment, stimuli were borrowed from a published study on the implicit de-humanization of women (Rudman & Mescher, 2012) using the Implicit Association Test (Greenwald et al., 1998). Two IRAPs were created that differed only in their contrast category. The Gender IRAP employed stimuli identical to those used by Rudman and Mescher (2012, Experiment 2: i.e., women, men, objects, and human). A second IRAP was created as a variant of the first: the Agency IRAP replaced the category “men” with “inanimate objects” (i.e., women, inanimate objects, objects, and human: see Table 3.3). In so doing, it sought to change the dimension of comparison from the *gender* of women (i.e., male vs. female) to the *agency* of women (i.e., capable of independent action, possessing mind and autonomy).

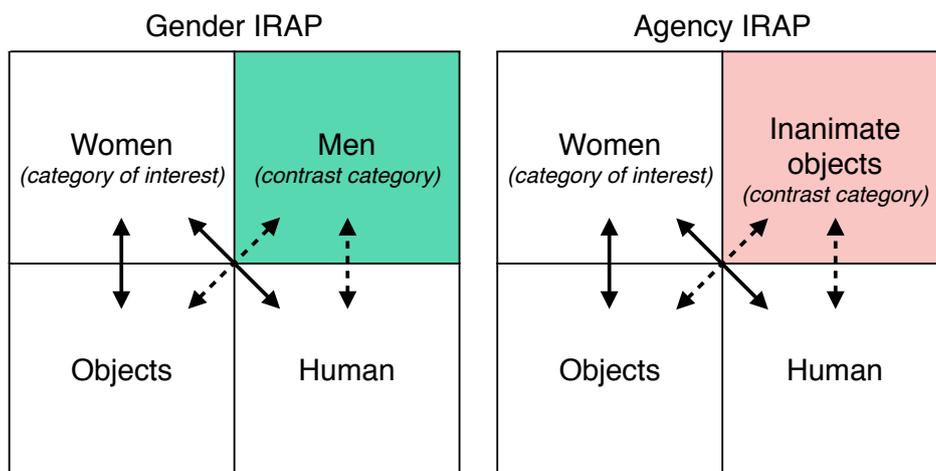


Figure 3.3. The stimulus categories employed in the Gender and Agency IRAPs. Solid arrows indicate trial-types that were identical in both IRAPs (i.e., “women–objects” and “women–human”), whereas dashed arrows indicate trial-types that differed between the IRAPs (e.g., “men–human” vs. “inanimate objects–human”).

Table 3.3. *Stimulus sets for the Gender and Agency IRAPs.*

Label stimuli				Target stimuli		Response options	
Gender IRAP		Agency IRAP					
<i>Women</i>	<i>Men</i>	<i>Women</i>	<i>Inanimate objects</i>	<i>Objects</i>	<i>Human</i>		
Women	Men	Women	Pencil	Thing	Human	Similar	Different
Woman	Man	Woman	Fork	Object	Culture		
Female	Male	Female	Phone	Tool	Logic		
Girl	Guy	Girl	Keys	Device	Rational		

We hypothesized that “women” would be differentially objectified and/or humanized across the two IRAPs depending on the context in which these stimulus classes were presented (i.e., “men” vs. “inanimate items”), despite the fact that the stimuli presented on these trial-type were identical in both cases. Should such differences emerge, results would therefore demonstrate that the contents of one category within the IRAP provide a potentially important source of contextual control over responding to the other categories.

### 3.6 Method

#### Participants

The current study employed only participants who identified as both male and heterosexual in order to limit the number of possible sources of contextual control over participants’ performances. It is therefore useful to reemphasize here that the current study employed the domain of dehumanization of women, but did not seek to explore this domain directly (e.g., by comparing women and men). Instead, we purposefully sought a homogenous sample. Forty-three male undergraduate students at the National University of Ireland Maynooth ( $M_{\text{age}} = 20.2$ ,  $SD = 2.0$ ) were recruited for the current study, and were randomly assigned to one of the two IRAP conditions. Participants reported that they had completed between zero and ten previous IRAPs ( $M = 1.4$ ,  $SD = 2.2$ ). Inclusion criteria were similar to the previous study, in addition to self-identifying as male and heterosexual. No incentives were offered for participation.

## **Ethical considerations**

Ethical approval was granted by Maynooth University's Social Research Ethics Subcommittee. Participants were informed as to the nature and purpose of the study prior to participation. Furthermore, they were made aware that participation was entirely voluntary, that they could cease participation at any time without giving a reason, and that all data would be irrevocably anonymised immediately after participation. Written informed consent was obtained from all participants. Upon completion of the tasks, participants were fully debriefed on the nature and purpose of the experiment and provided with an opportunity to ask questions, which were then fully answered by the researcher. During this debrief, the researcher made each participant aware of the free student counselling service available on campus, whose contact details were included at the bottom of their copy of the consent form, should they experience any form of post-experimental distress.

## **Measures**

**Attitudes towards women scale.** Despite its age, the ATWS (see Appendix K) is a widely used measure of sexist beliefs against women, which was used to compare the two groups on their levels of self-reported sexist attitudes towards women. This 25-item scale asks participants to respond to statements which are either overtly sexist or egalitarian, such as "There should be a strict merit system in job appointment and promotion without regard to sex" and "It is insulting to women to have the 'obey' clause remain in the marriage service." (Spence, Helmreich, & Stapp, 1973). It uses a 1 (strongly agree) to 4 (strongly disagree) response format. Internal consistency was good in the current sample ( $\alpha = .72$ ).

**Likelihood to sexually harass scale. The LSH** (see Appendix L) was used to compare the two groups on their levels of self-reported sexual objectification of women. This scale asks participants to read 10 paragraph-length depictions of specific scenarios and then to respond to three items for each scenario (Pryor, 1987). Each item asks the participant to

imagine that they are working in a specific position of power (e.g., as an editor for a large publisher), and that they then have an interaction with a young, attractive, and/or junior woman. Three questions are then presented that ask whether the participant would be likely to show preferential bias for such a woman. Subscale A does not specify a contingency for this preferential bias (e.g., “Would you agree to read Betsy’s novel?”), subscale B specifies that it is in return for sexual favors (e.g., “Would you agree to reading Betsy’s novel in exchange for sexual favors?”), and subscale C specifies that it is in return for going on a date (e.g., “Would you ask Betsy to have dinner with you the next night to discuss your reading her novel?”). Each item employs a 1 (Not at all likely) to 5 (Very likely) response scale. Internal consistency was excellent in the current sample ( $\alpha = .91$ ).

**Gender and Agency IRAPs.** The procedural details of both IRAPs were similar to those used in previous experiments other than the stimulus set and responding rules employed. As previously stated, stimulus categories for the gender IRAP were drawn from Rudman and Mescher (2012), who used the categories “women”, “men”, “objects”, and “humans”. In order to create the second IRAP through the use of a contrast category manipulation, the category “men” was replaced with “inanimate objects” in the Agency IRAP (see Table 3.3). As such, in an identical manner to the previous experiment, all stimuli were identical in both IRAPs other than the contrast category. For the Gender IRAP, rule A was “Women are objects and men are human” and rule B was “Women are human and men are objects”. For the Agency IRAP, rule A was “Women are human and inanimate objects are objects” and rule B was “Women are objects and inanimate objects are human”.

**IRAP data processing.** Raw latencies on the IRAP were converted into *D* scores and then excluded from the analysis based on mastery criteria in an identical manner to the previous experiments. No participants failed to meet the mastery criteria on the practice blocks. One participant failed more than one test block-pair and therefore had their data

excluded from the analyses. Two participants failed only one test block and therefore the data from this pair were excluded from the calculation of their final *D* scores. Forty-two individuals therefore remained in the final sample, with 21 participants randomly assigned to each of the two groups. Additionally, similar to previous experiments, trial-types 1 and 2 (i.e., the women trial-types) were inverted in both IRAPs in order to create a common axis. As such, positive *D* scores can be interpreted as “human” or “*not*-objects” biases, and negative *D* scores can be interpreted as “objects” or “*not*-human” biases.

### 3.7 Results

#### Demographics and self-reports measures

Two independent t-tests demonstrated that the Gender IRAP and Agency IRAP groups did not differ in terms of their age ( $p = .15$ ) or scores on either the Attitudes Towards Women Scale ( $p = .67$ ) or Likelihood to Sexually Harass Scale ( $p = .75$ ).

Given that no differences were found between the two groups on any of the scales, scores for the two groups were then collapsed in to one group for the following descriptive analysis. Scores on each of the LSH subscales were recoded into two groups: responses rejecting a willingness to abuse power or an uncertainty over willingness to abuse power over women (i.e., 1 [not at all likely] to 3 [unsure]) and responses asserting an explicit willingness to abuse power over women (i.e., 4 [somewhat likely] to 5 [very likely]). After recoding, the overwhelming majority of the sample (97.7%) reported that they would in at least some circumstances show preferential bias for certain women when in a position of power over them. Furthermore, large proportions of the sample reported that they would, in at least some circumstances, sexually harass women by such power in exchange for a date (50.0%) or sexual favours (29.6%). For the purposes of the remaining analyses, scores on the LSH’s three subscales were collapsed in to one (Rudman & Mescher, 2012).

## Gender and Agency IRAPs

Differences between the Gender and Agency IRAPs were assessed using a 2×4 mixed within-between ANOVA, with IRAP type (gender vs. agency) as the between-participant variable and IRAP trial-type as the within-participant variable. No main effect for IRAP type was found ( $p = 0.29$ ). Critically, an interaction effect was found between IRAP type and trial-type, as hypothesized,  $F(40, 3) = 8.96, p < .0001, \eta^2 = .23$ . Tukey’s planned comparison tests demonstrated large and significant differences between the conditions on the “women-human” trial-type, and between both the “men” and “inanimate objects” trial-types between the two IRAPs (all  $ps < .05$ ; Hedges’  $g_s = .70$  to  $1.09$ ). No differences were found between the “women-objects” trial-types ( $p = .24$ ).

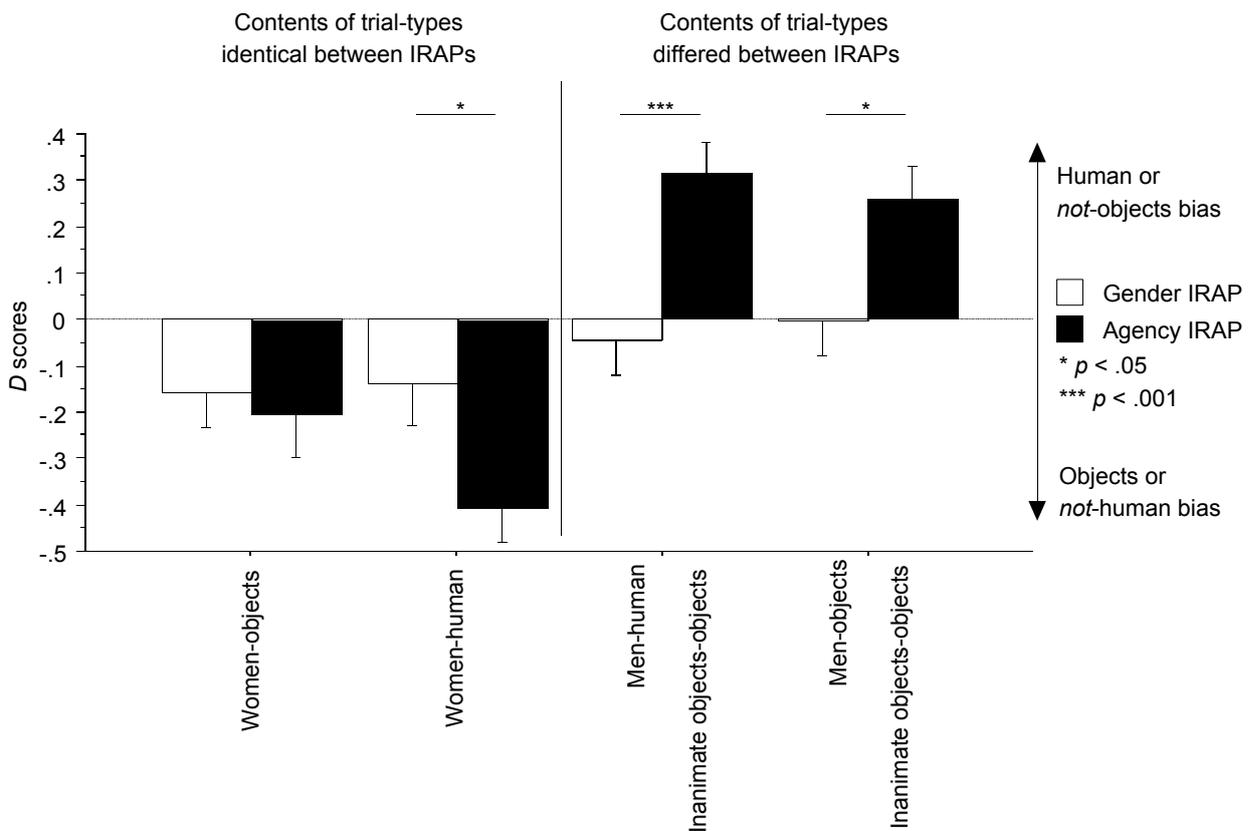


Figure 3.4. Performance on the Gender and Agency IRAPs. Positive  $D$  scores represent “human” or “not-object” biases and negative  $D$  scores represent “object” or “not-human” biases. Error bars represent standard errors.

One-sample *t*-tests were used to explore the significance of the mean bias scores on individual trial-types across the two IRAPs. Significant “women-objects” biases were found on both IRAPs ( $ps < .05$ ). A significant “women-*not*-human” bias was found on the Agency IRAP ( $p < .0001$ ), but not the Gender IRAP ( $p < .14$ ). Significant “inanimate objects-*not*-objects” and “inanimate objects-human” biases were found on the Agency IRAP ( $ps < .01$ ), but no “men-objects” or “men-human” biases were found on the Gender IRAP ( $ps \geq .56$ ). In summary, participants therefore produced differential patterns of responses to “men” and “inanimate objects” across the two IRAPs, as expected. However, the direction of this pattern of bias was counter-intuitive, insofar as participants humanized and de-objectified inanimate objects and did not humanize or de-objectify men. Critically, differences were also found on the category of interest (i.e., the “women-objects” and “women-human” trial-types: see Figure 3.2). Again, however, the pattern of this effect was counterintuitive, as “women” were dehumanized to a greater degree when contrasted with “inanimate objects” rather than “men”. Nonetheless, the current results clearly demonstrate that responding to the contrast category can have a large influence on responding to the category of interest (Hedges’  $g_s \geq .70$ ).

### **Relationship between self-report and implicit measures**

The ATWS and LSH were included primarily as screening measures to compare the two groups. In light of the results of previous research (Rudman & Mescher, 2012), no correlations with the IRAPs were predicted. Nonetheless, correlations between the implicit and explicit measures are reported here for the purpose of consistency across studies.

Table 3.4. *Correlations between the IRAPs and self-report measures.*

	ATWS	LSH
<i>Gender IRAP trial-type:</i>		
Women-objects	-.02	.16
Women-human	-.09	.32
Men-objects	-.36	.41 <sup>†</sup>
Men-human	.00	.05
<i>Agency IRAP trial-type:</i>		
Women-objects	.00	.09
Women-human	.00	.12
Inanimate objects-objects	-.03	-.28
Inanimate objects-human	-.20	-.48*

*Note:* ATWS = attitudes towards women scale, LSH = likelihood to sexually harass scale.

<sup>†</sup>  $p < .1$ , \*  $p < .05$

A correlation matrix demonstrated a significant negative correlation between the “inanimate objects-human” trial-type and the LSH. This indicated that a higher self-reported likelihood to sexually harass women was associated with a stronger tendency to confirm rather than reject that inanimate objects are not human. However, in light of a recent meta-analysis of criterion effects for the IRAP (Vahey et al., 2015), it should be cautioned that this correlation is likely to be underpowered. No other correlations were found to be significant (see Table 3.4). Finally, Fischer’s *r*-to-*z* tests demonstrated that the magnitude of the correlations between either the ATWS or the LSH and the “women” trial-types did not differ significantly between the two IRAPs (all  $ps \geq .5$ ).

### 3.8 Discussion

The two groups were shown not to differ in their age, their sexist attitudes towards women, or their level of self-reported likelihood to sexually harass women. Therefore, we concluded that any differences between the two IRAPs’ “women-human” trial-types were likely due to the contrast category manipulation (i.e., responding to “women” in the context of “men” vs. “inanimate objects”). The results demonstrate that participants differentially

humanized and objectified the contrast categories “men” and “everyday objects”, as assumed. Critically, behaviour on a trial-type of interest (i.e., women-human) was therefore found to differ based on the context provided by the contrast category. As such, while the IRAP’s trial-types are *procedurally non-relative*, behaviour within the task is *not a-contextual*. While this may be relatively unsurprising to functional-analytically-orientated researchers, given that behaviour is by definition contextually determined, this is the first time that this form of contextual control has been demonstrated within the IRAP.

Research elsewhere using the IRAP has sometimes targeted a single trial-type (e.g., Nicholson, McCourt, & Barnes-Holmes, 2013). However, the current results indicate that this must be done in the knowledge that behaviour within that trial-type may be influenced in important ways by the contents of the others. Future research should therefore note that the theoretical reasons for targeting specific trial-types in an analysis should be ideally supported by the contextual control brought to bear by the contrast categories. This support could be either 1) theoretical, for example by selecting optimal contrast categories with considered reference to domain-relevant literature, or 2) empirical, for example by manipulating the contrast category across IRAPs in order to attempt to target specific functions (e.g., the gender vs. agency of women).

It is worth noting that the current research differs in a key way to previous work on the role of the contrast category within other tasks such the IAT, which has pivoted on the questions of a) whether the necessity of a contrast category is inherently problematic, and b) how to overcome this (e.g., De Houwer, 2006; Houben & Wiers, 2006; Huijding, de Jong, Wiers, & Verkooijen, 2005; Karpinski, 2004; Nosek et al., 2005; Ostafin & Palfai, 2006; Palfai & Ostafin, 2003; Pinter & Greenwald, 2005; Robinson et al., 2005; Swanson, Rudman, & Greenwald, 2001). Rather than seeing the contrast category as a procedural “nuisance”, the current results suggests that increased consideration of the choice or manipulation of the

IRAP's contrast category may enhance the precision with which specific relational responses can be targeted, thereby facilitating increasingly fine-grained functional analyses of verbal behaviour using the IRAP.

Specifically, while the majority of research to date using measures such as the IRAP and IAT has operated under a common assumption about the nature of the relation between stimulus categories (i.e., that they should be "obvious opposites", see Robinson et al., 2005, p.208), the current research highlights the fact that relatively less attention has been paid to which specific psychological functions are specified by this relation (e.g., opposite gender vs. opposite in agency), and how this influences behaviour within the task.

While we have focused on the question of whether contrast category manipulations can influence behaviour on other trial-types, it is also worth considering possible reasons for the direction of the specific effect that was found, and what implications this has for the domain of objectification of women. Specifically, the current results indicate that women were more strongly dehumanized on the IRAP in the context of "inanimate objects" relative to "men". Additionally, participants strongly de-objectified and humanized "inanimate objects". Intuitively, one might expect this pattern of effect to be in the reversed direction (i.e., a tendency to respond under accuracy and latency pressure that women are human insofar as inanimate objects are objects). The reasons for this are unclear at present and no explanation readily presents itself, given the absence of previous literature on this specific question. At minimum, the current results may suggest that researchers should give consideration to the interactions between the categories within a task and how these map on the behaviours of interest within a domain. However, in our view, such uncertainty around what current or historic contextual factors gives rise to effects such as this only serves to further underscore the need for a more systematic understanding of the sources of contextual

control within tasks such as the IRAP, and how these can be manipulated to increase the prediction-and-influence of behaviour (Hayes, Barnes-Holmes, & Wilson, 2012).

The current results demonstrate the viability of the contrast category manipulation method. It is therefore useful to consider how subsequent experiments in this thesis will be shaped by these findings. Chapter 2 noted that there was uncertainty about the “type” of death that was being referred to in the IRAP (e.g., personal death vs. death of an other, or heroic death vs. tragic death). This lack of clarity around what contextual control governed participants’ responses might, for example, account for the unexpected death positive effects. The current results suggest that contrast category manipulations may provide a way in which to specify these dimensions of comparison more precisely. For example, similar to the death-evaluation IRAP in Chapter 2, the stimulus categories “My death”, “positive” and “negative” might notionally be used to target implicit evaluations of death. In order to go beyond the typical study design, however, different aspects of the compound stimulus “my death” could be brought to bear within the IRAP using the contrast category manipulation method. For example, two IRAPs could be created, with one targeting the distinction between life and death (e.g., “my death” vs. “my life”), and a second targeting the self-other distinction (e.g., “my death” vs. “others’ deaths”). Although the “my death–positive” or “my death–negative” trial-types would be identical across both tasks, it is possible that differences would emerge across the two tasks (e.g., in mean bias scores and/or predictive validity). Importantly, any differences would be accompanied by greater clarity as to what functional dimension of comparison drives such effects, thereby helping to link such results directly with the domain-specific theories to which they attempt to speak.

The next experiment therefore applies two forms of contrast category manipulations (including the above example) in order to further explore death-evaluations on the IRAP in normative participants. Specifically, we manipulate whose death is being referred to (e.g.,

self vs. other), and what dimension of evaluation is being drawn, such as whether the consequences specified are positive punishers versus positive reinforcers (e.g., “painful” vs. “satisfying”) or positive punishers versus negative reinforcers (e.g., “painful” vs. “escape”).

## **CHAPTER 4: EXPLORING DEATH EVALUATIONS ON THE IRAP USING MORTALITY SALIENCE INDUCTIONS AND CONTRAST CATEGORY MANIPULATIONS**

Abstract: The current chapter details four experiments run in sequence that attempt to address some of the issues with the death-evaluation IRAP raised in Chapter 2. This includes the use of the contrast category manipulation method developed in Chapter 3, in order to manipulate both the target categories (i.e., the positive vs. negative reinforcement associated with death), and the label categories (i.e., whose death was being referred to, self vs. others). Additionally, participants completed a death-evaluation IRAP before and after an induction in order to manipulate the salience of death. The results of each experiment are first discussed individually, and a series of post hoc analyses are subsequently employed to compare results across experiments. Results demonstrate that manipulation of the contrast categories produced differential performance on the “my death-negative” biases across the four IRAPs. However, this impact was limited to the manipulation of the positive versus negative reinforcement associated with death, and not whose death was being referred to. Furthermore, the effects of these the contrast category manipulation was found only after mortality salience induction. The results therefore suggest that the absence of “death-negative” biases found at baseline in Chapter 2 may have been due to inadequate salience of death for normative participants, and not inadequate specification of whose death is being referred to.

In opening, it is worth recalling that Chapter 2 found unexpected “death-positive” biases on the death-evaluation IRAP, and concluded by suggesting several potential avenues by which these unexpected effects might be explored. One suggestion was that the stimulus “dying” might not be particularly salient or meaningful to a young healthy student, whose concept of mortality is likely to be highly abstract. As such, the current chapter’s first research aim was to attempt to manipulate the salience of death and explore the resulting

implicit evaluations of death on the IRAP. This was done through the use of “mortality salience inductions” – a paradigm that is commonly used within the attitudes to death and Terror Management Theory literatures.

Terror Management Theory is a social psychology theory that attempts to account for the ubiquitous need for self-esteem and meaning in life (Greenberg et al., 1990). It postulates that humans’ concurrent desire to live, and awareness of the inevitability of their death, produces terror of death. The theory posits that cultural values function to manage this terror of death by imbuing life with meaning. Tests of the theory commonly manipulate the salience of mortality for participants before assessing changes in cultural values (e.g., harsher sentencing recommendations for hypothetical crimes: Rosenblatt, Greenberg, Solomon, Pyszczynski, & Lyon, 1989). The most commonly-employed mortality salience manipulation is the Mortality Attitudes Personality Survey (MAPS: Rosenblatt et al., 1989), which asks participants to write open-ended responses about the emotions that the thought of their own death arouses in them. A recent meta analysis demonstrated that such inductions are effective in increasing the salience of death (Burke, Martens, & Faucher, 2010), and previous research has demonstrated that responding on the IRAP is sensitive to comparable inductions (e.g., mood inductions: see Hussey & Barnes-Holmes, 2012). In each experiment in the current chapter, participants therefore completed a death-evaluation IRAP before and after a modified mortality salience induction. The standard mortality salience induction (i.e., the MAPS) was extended upon for the purposes of the current studies, in order to emphasise the brevity of life as well as the inevitability of death (see Section 4.2.2).

A second suggestion discussed in Chapter 2 was that there might have been uncertainty about the type of death that was being referred to in the IRAP. For example, did the stimuli refer to personal death, or the death of a stranger, or that of a loved one? Moreover, was this a tragic or a heroic death? In the service of developing methodological and

conceptual tools with which to better address these questions, Chapter 3 therefore explored the role of the contrast category as a potentially useful source of contextual control within the IRAP. The results of Chapter 3 demonstrated that the IRAP's contrast category could be used, at least in some circumstances, to better specify the functional dimension of comparison within the task (e.g., contrasting women with men [dimension of gender] produces different outcomes to contrasting women with everyday items [dimension of agency]). The current chapter's second research aim was therefore to use such contrast category manipulations to explore implicit evaluations of death. For clarity, the results of each study will first be discussed individually and with regard to the effect of the mortality salience induction on IRAP performances. Subsequently, a series of post hoc comparisons will assess how the systematic manipulation of the IRAPs' contrast categories across the four studies influenced performance on the trial-type of interest, which was common across all four experiments (i.e., the "my death-negative" trial-type).

#### **4.1 EXPERIMENT 1**

Recall that the death-evaluation IRAP employed in Chapter 2 employed the categories "living", "dying", "positive" and "negative". The current study sought to improve on this previous IRAP by specifying whose death is referred to, and by manipulating mortality salience between IRAPs. Specifically, the death-evaluation IRAP was modified so that the label stimuli 1) referred to the state of being dead (i.e., life vs. death) rather than the process of dying (i.e., living vs. dying), and 2) included the deictic cue "my" so that these stimuli referred to *personal* life and death rather than leaving this unspecified (i.e., "my death"). The evaluative categories employed were left unchanged from those in Chapter 2 (see Table 4.1).

Table 4.1. *Stimuli used in Experiment 1's death-evaluation IRAP*

Label stimuli		Target stimuli		Response options	
<i>My life</i>	<i>My death</i>	Enjoyable	Awful	True	False
		Exciting	Distressing		
		Great <sup>1</sup>	Hurtful		
		Lovely	Horrible		
		Pleasant	Painful		
		Satisfying	Upsetting		

*Note:* The stimuli in italics were modified relative to the death-evaluation IRAP employed in Chapter 1. <sup>1</sup> Vernacular usage of the word “great” in Ireland is most frequently as a synonym of positively valenced words (e.g., “excellent”) rather than denoting quantity or rank (cf. “substantial” and “prominent”).

## 4.2 Method

### Participants

Twenty-three students at Maynooth University were recruited from the participant pool (18 female and 5 male,  $M_{\text{age}} = 20.35$ ,  $SD = 2.29$ ). No incentives were offered for participation. Participants reported having completed between 0 and 2 IRAPs previous to the current study ( $M = .43$ ,  $SD = .66$ ). Inclusion criteria were similar to previous experiments with one addition: individuals were excluded from participation if they responded with anything other than “No” on a single-item screening questionnaire “Do you currently suffer from a serious, life threatening or terminal illness?” No participants were excluded on this basis.

### Ethical considerations

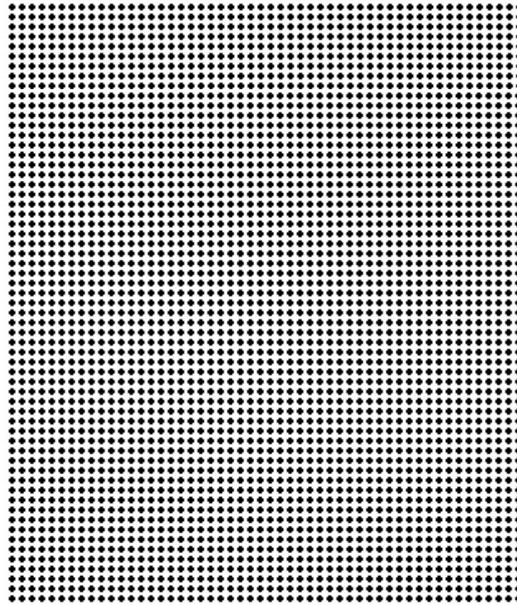
Ethical approval was granted by Maynooth University’s Social Research Ethics Subcommittee. Participants were informed as to the nature and purpose of the study prior to participation. Furthermore, they were made aware that participation was entirely voluntary, that they could cease participation at any time without giving a reason, and that all data would be irrevocably anonymised immediately after participation. Written informed consent was obtained from all participants. It should therefore be noted that exclusions based on responses to the above single-item screening question were completed on the irrevocably

anonymised data. Upon completion of the tasks, participants were fully debriefed on the nature and purpose of the experiment and provided with an opportunity to ask questions, which were then fully answered by the researcher. During this debrief, the researcher made each participant aware of the free student counselling service available on campus, whose contact details were included at the bottom of their copy of the consent form, should they experience any form of post-experimental distress.

## **Measures**

**Self-reports.** Participants completed a number of self-report measures that were previously used in Chapter 2: the Beck Hopelessness Scale ( $\alpha = .71$ ), Depression Anxiety Stress Scale's depression subscale ( $\alpha = .59$ ), and Acceptance and Action Questionnaire-II ( $\alpha = .89$ ).

**Modified mortality salience induction.** The MAPS (Rosenblatt et al., 1989) asks participants to write open ended-responses about the emotions that the thought of their own death arouses in them (see Appendix M). In order to directly emphasise the brevity of life as well as the inevitability of death, this standard mortality salience manipulation was modified in the following manner, which was created specifically for the current experiment. First, the researcher selected the appropriate "Weeks to Live" sheet from a battery based on the participant's age and gender and an assumed lifespan of 79 years for men and 83 years for women (Irish Central Statistics Office, 2014). Each "Weeks to Live" sheet consisted of a pattern of dots arranged in a rectangle roughly 2.5 inches wide, where the number of dots equalled the participant's expected age minus their current age (see Figure 4.1 and Appendix N).



*Figure 4.1.* Example “Weeks to Live” sheet for a 20-year-old female. Each dot represents one week left to live, assuming a life expectancy of 83 years (see Central Statistics Office, 2014).

The researcher then delivered the following script (notes for the researcher are in italics: see Appendix O):

“I think it’s often very easy to forget just how short life is, especially for young healthy students. To help convey this, I’ve put together this diagram for you. Given that I know your age and gender, it’s trivial for me to estimate your expected lifespan.

*[Place Weeks to Live Sheet in front of participant, say very slowly and carefully.]*

Based on that, the number of dots on this piece of paper is equal to the number of weeks you have left to live.

*[Long pause.]*

I promise that I’m not trying to trick or deceive you. It’s a surprisingly small number of dots, isn’t it? And, the thing about dots is that once you spend them you can’t get them back. This is not a rehearsal; you will not get a second shot.

This is your life, right now, ending, one day at a time. The other thing about dots is that they run out, no matter what you do. Make no mistake: death *is* coming. You have a limited number of days left on this planet, and, like all of us, you're faced with the difficult question of what you're going to do with them. How many of these dots will be "well-spent" dots, doing things that you truly value, and how many dots will you feel you have wasted? The dots don't care either way; they just run out. With all of that in mind, I'd like to ask you to write a few lines about what you think about death, and what dying will be like."

Participants were then presented with the MAPS and provided with sufficient time to complete it. For ethical and privacy reasons, participants were informed that they did not have to return their MAPS for content analysis, and that they could take it away with them or ask that it be destroyed.

**Positive and Negative Affect Schedules.** A meta analysis revealed that the effect size of mortality salience inductions are larger when there is a delay between the induction and the critical task (Burke et al., 2010). We therefore followed what this review suggested to be the most common strategy, by employing 1) a single "delay task" between the MAPS and the IRAP, and 2) selecting the Positive and Negative Affect Schedules (PANAS: Watson, Clark, & Tellegen, 1988) for this purpose. The PANAS is a pair of 20-item scales that are commonly used to assess changes in positive and negative affect. The two versions of the scale (PANAS-A and PANAS-B) have been shown to be psychometrically equivalent (Watson et al., 1988). Both versions list 20 emotions (e.g., cheerful, distressed, blue) and ask participants "to what extent you feel this way right now, at this moment" using a 1 (very slightly or not at all) to 5 (extremely) scale.

**Death-evaluation IRAP.** Participants completed a death-evaluation IRAP at baseline, and again after a mortality salience induction (see below). The stimulus set and responding rules were similar to the death-evaluation IRAP employed in Chapter 2, except the label stimuli were set to “My life” and “My death”. The evaluative categories were identical to Chapter 2 (see Table 4.1). Responding rule A was “my life is positive and my death is negative”, and rule B was “my death is positive and my life is negative”.

### **Procedure**

First, participants completed the AAQ-II, BHS, DASS, and PANAS-A self-report measures. Second, they completed the baseline IRAP. The structure and delivery of the IRAP was identical to that described in previous experiments. Third, they were exposed to the modified mortality salience induction, followed by a PANAS-B “delay task”. Finally, participants completed the post induction IRAP. While being exposed to the second IRAP, the researcher left the “Weeks to Live” sheet on the table between the participant and the keyboard, and instructed the participant to look at it between the presentation of blocks on the IRAP and to “remember what we talked about and how it made you feel”. Upon completion of their participation, individuals were fully debriefed on the nature and purposes of the study and any questions they had were answered. This debrief included the following script, which was included in order to alleviate any transient negative mood states remaining after the induction:

“We’ve talked about how death is inevitable, and why that makes a lot of people afraid of it. But, another way to look at death is that it is ‘the mirror in which all meaning in life is reflected’. Death is the ultimate motivator to live your life as fully as possible, and as it happens. After all, if you knew you would live forever, why do anything today, or tomorrow, or ever?”

In addition, participants were provided with the contact details for the on-campus free and confidential student counselling service, and encouraged to make contact with this service if they were in any distress in their lives. It is worth noting that informal polling of individuals after participation suggested that an overwhelming majority reported that they were glad to have participated in the experiment.

**IRAP data processing.** Data from the IRAP were processed in an identical manner to the previous experiments. *D* scores were calculated and accuracy and latency performances were compared against mastery criteria. Participants who failed either the baseline or follow-up IRAP had all their data excluded from the analysis. One participant failed to meet the mastery criteria on one IRAP's practice blocks, and was therefore not presented with the critical test blocks. Three participants failed more than one test block-pair within a single IRAP and therefore had their *D* scores excluded from the analyses. In one case, a participant failed one test block-pair and therefore had that block-pair only excluded from the calculation of their final *D* score. The final sample therefore contained 19 participants.

Similar to previous experiments, trial-types 3 and 4 were inverted in both IRAPs (i.e., the “death” trial-types) so as to create a common vertical axis for all analyses. Positive *D* scores therefore represented “positive” or “*not-negative*” biases, whereas negative *D* scores represented “negative” or “*not-positive*” biases.

### **4.3 Results and Discussion**

#### **Self-report measures**

The sample reported normative levels of hopelessness ( $M = 3.21$ ,  $SD = 2.74$ ), depression ( $M = 2.68$ ,  $SD = 1.7$ ), and psychological flexibility ( $M = 19.68$ ,  $SD = 8.53$ ), and therefore appeared to be representative of normative university students.

### **Mortality salience induction**

A dependent *t*-test demonstrated that participants' moods became significantly more negative as a function of the induction,  $t(16) = 2.87, p < 0.01$ , Hedges'  $g_{av} = 1.18$ . This change in PANAS scores is interpreted as indicating a successful induction of mortality salience within the sample (see Burke et al., 2010).

Sixty-eight per cent of participants included in the final sample chose to return their MAPS to the researcher. The current study took a further, novel step relative to previous research in assessing the efficacy of and adherence to the induction by performing a content analysis on participants' responses within the induction. Participants' responses were transcribed and then analysed using the text parsing and analysis software ("Linguistic Inquiry and Word Count": LIWC: Pennebaker, Booth, & Francis, 2007; see Kahn, Tobin, Massey, & Anderson, 2007). This software searches inputted text for occurrences of words from a dictionary of roughly 4,500 words, each of which are predefined into grammatical and semantic categories. For the purposes of the current experiment, we selected two categories of words to analyse: "positively valenced words" and "negatively valenced words". LIWC then calculates the percentage of the overall input text that was made up of each category as a percentage of total number of words. Analyses using LIWC revealed that participants used an average of 6.3% ( $M = SD = 3.3$ ) "negative emotional" words and an average of 3.4% ( $SD = 2.5$ ) "positive emotional" words. A dependent *t*-test confirmed that participants used significantly more negative emotional words than positive emotional words,  $t(12) = 2.2, p < 0.05$ , Hedges'  $g_{av} = 1.01$ . Results therefore suggested that, when asked to describe their expectations of death in an open-ended manner, participants used predominantly negatively-valenced words (e.g., sad or afraid).

## Death-evaluation IRAP

Comparable to the death-evaluation IRAP in Chapter 2 on which the current IRAP was based, the baseline IRAP revealed strong “my life-positive” biases ( $M = .41$ ,  $SD = .30$ ), as well as moderate “my life-not-negative” ( $M = .26$ ,  $SD = .47$ ) and “my death-positive” biases ( $M = .14$ ,  $SD = .46$ ). No biases were demonstrated on the “my death-negative” trial-type at baseline ( $M = -.03$ ,  $SD = .49$ ), although a “my death-negative” bias was found after mortality salience induction ( $M = -.25$ ,  $SD = .47$ ; see Figure 4.2). Following the mortality salience induction, biases on two of the trial-types changed, whereas the other two did not. Specifically, participants became more positive about life (my life-positive trial-type), and more negative about death (my death-negative trial-type).

A  $2 \times 4$  ANOVA was conducted on the IRAP  $D$  scores, with time-point (baseline vs. post induction) and IRAP trial-type as within-participant variables. A significant main effect was found for trial-type,  $F(3, 54) = 13.77$ ,  $p < .001$ ,  $\eta^2 = .11$ , but not time-point ( $p = .89$ ). A significant interaction effect was also found,  $F(3, 54) = 4.87$ ,  $p < .01$ ,  $\eta^2 = .21$ . Four planned comparisons were made using Bonferroni-correct  $t$ -tests. These revealed a significant change from pre to post of large effect size on the “my life-positive” trial-type,  $t(18) = 3.01$ ,  $p < 0.01$ , Hedges’  $g_{av} = 0.86$ . A marginal effect of medium effect size was found on the “my death-negative” trial-type;  $p < 0.05$ , Hedges’  $g_{av} = 0.46$ . The remaining trial-types did not change significantly ( $ps > .27$ ).

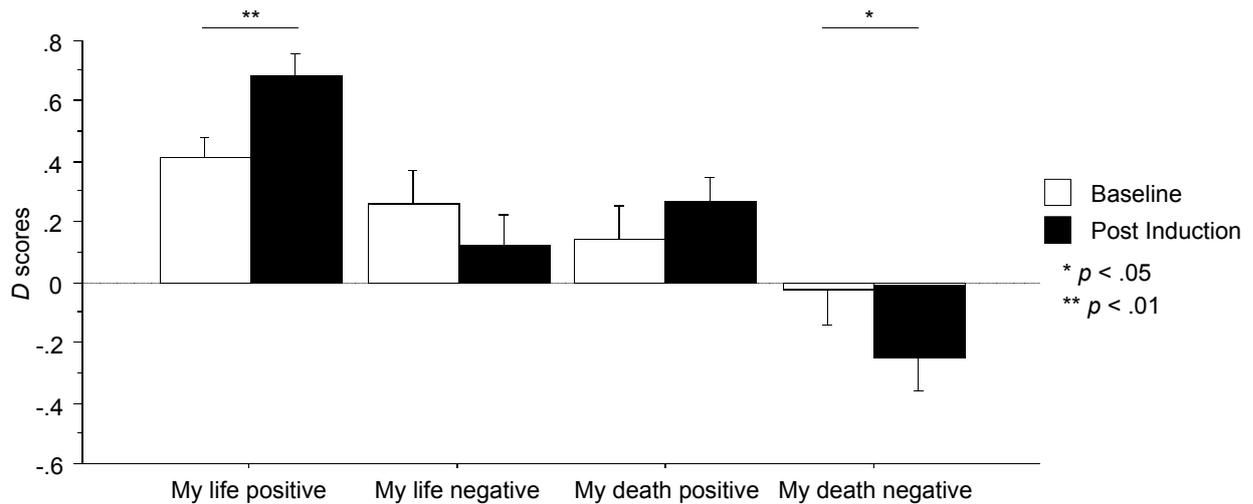


Figure 4.2. Performance on Experiment 1’s death-evaluation IRAP at baseline and after mortality salience induction. Positive *D* scores represent “positive” or “not-negative” biases, and negative *D* scores represent “negative” or “not-positive” biases.

Results therefore suggest that the absence of negative evaluations of death on the IRAP found in Chapter 2 were due, at least in part, to the lack of salience of death in normative students. The current results demonstrated a comparable absence of “my death-negative” bias at baseline, but showed a “my death-negative” bias after the induction. This change was found to be significant and of medium effect size. Participants also became significantly more positive on the “my life-positive” trial-type after mood induction. While no significant “my death-positive” bias was found at baseline, this was possibly due to the relatively low sample size, as the average bias on this trial-type was comparable to that in Chapter 2. Surprisingly, however, participants remained positive about their death after the induction. Specifically, although this change was not significant, the magnitude of the “my death-positive” bias became significant after the induction.

This positivity toward death was not readily explainable via participants’ self reports about their own death. When asked to describe their thoughts about death and dying in an open-ended manner on the MAPS, participants used a significantly greater number of negative than positive words. The open-ended nature of this exercise allowed individuals to

express positive evaluations about death as well as negative (or otherwise) in a way that the death anxiety scale employed in Chapter 2 did not.

In summary, the results of the current experiment suggest that some aspects of the unexpected pattern of bias found in Chapter 2 (i.e., the absence of negativity about death) were explainable by a lack of salience of death for normative students. However, the results also suggest that participants' positivity about death may not be attributable to a lack of clarity about whose death is being referred to or a lack of salience around the gravity and implications of personal death.

On reflection, we considered it possible that these evaluations of death as positive effects may be attributable to some aspect of the "positive" category of stimuli that were employed. Specifically, these stimuli could be argued to refer to positive reinforcement associated with life and death (e.g., exciting, satisfying). The next experiment therefore modified this stimulus category so that, while the stimuli remain positively valenced, they instead referred to the negative reinforcement that is associated with life and death (e.g., calm, relief, or escape).

## 4.4 EXPERIMENT 2

Whereas the previous experiment's IRAP targeted the positive reinforcement associated with life and death (e.g., exciting, satisfying), the current experiment instead targeted negative reinforcement (e.g., calm, relief, or escape: see Table 4.2). Note that this stimulus category nonetheless remained to be positively valenced. The experimental design and measures were otherwise identical to the previous experiment. We specifically sought to explore whether the death-positive biases found in previous studies were resilient to changes in the nature of these positively-valenced stimuli (i.e., from specifying positive reinforcement to negative reinforcement).

Table 4.2. *Stimuli used in Experiment 2's death-evaluation IRAP*

Label stimuli		Target stimuli		Response options	
My life	My death	<i>Relief</i>	Awful	True	False
		<i>Escape</i>	Distressing		
		<i>Calm</i>	Hurtful		
		<i>Soothing</i>	Horrible		
		<i>Peaceful</i>	Painful		
		<i>Comfort</i>	Upsetting		

*Note:* Stimuli in italics differ relative to Experiment 1.

## 4.5 Method

### Participants

Twenty-six students at Maynooth University were recruited from the volunteer pool (18 female and 8 male,  $M_{\text{age}} = 21.04$ ,  $SD = 3.18$ ). No incentives were offered for participation. Participants reported having completed between 0 and 10 IRAPs previous to the current study ( $M = 1.04$ ,  $SD = 2.54$ ). Inclusion criteria were identical to the previous experiment. No participants were excluded on the basis of reporting that they suffered from a serious, life-threatening or terminal illness.

## **Ethical considerations**

Ethical considerations were identical to the previous experiment. In summary, ethical approval was granted by Maynooth University's Social Research Ethics Subcommittee, participants were informed as to the nature prior to participation, its voluntary nature, the right to cease participation at any time, and that all data would be irrevocably anonymised after participation. Written informed consent was obtained from all participants. Finally, all participants were fully debriefed after completing the tasks, and provided with the contact details for the free student counselling service available on campus should they experience any form of post-experimental distress.

## **Measures and procedure**

The order and counterbalancing of the measures was identical to the previous experiment, including the Beck Hopelessness Scale ( $\alpha = .85$ ), DASS-depression subscale ( $\alpha = .92$ ), and Acceptance and Action Questionnaire II ( $\alpha = .92$ ). The structure and delivery of the IRAPs were similar to previous experiments, however the stimulus set within the IRAP differed slightly from the previous experiment. The stimuli categories "My life", "My death", and "negative" were identical to the previous experiment, but the evaluative contrast category was altered to refer to negative reinforcement (e.g., "relief", see Table 4.2). The responding rules were identical to the previous experiment.

**IRAP data processing.** Data from the IRAP were processed in an identical manner to the previous experiment. Four participants failed to meet the mastery criteria on one IRAP's practice blocks, and were therefore not presented with the critical test blocks. Three participants failed more than one test block-pair within a single IRAP and therefore had their *D* scores excluded from the analyses. The final sample therefore contained 17 participants. Identical to the previous experiment, trial-types 3 and 4 were inverted in both IRAPs (i.e., the "death" trial-types) so as to create a common vertical axis for all analyses. Positive *D* scores

therefore represented “positive” or “*not-negative*” biases, whereas negative *D* scores represented “negative” or “*not-positive*” biases.

#### 4.6 Results and Discussion

The analytic strategy was similar to the previous experiment in all respects.

##### **Self-report measures and mortality salience induction.**

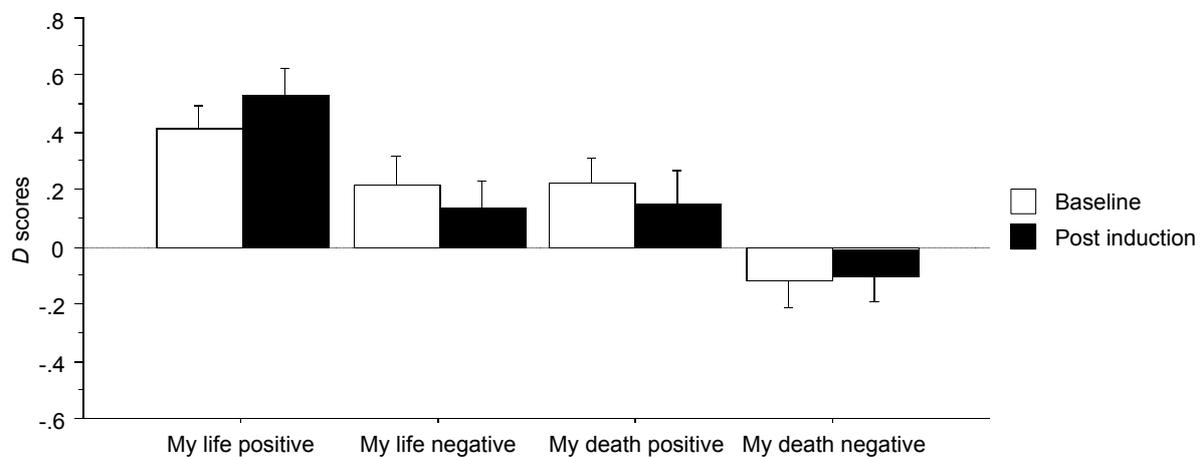
The sample reported normative levels of hopelessness ( $M = 4.25, SD = 3.65$ ), depression ( $M = 4.15, SD = 4.39$ ), and psychological flexibility ( $M = 17.74, SD = 7.86$ ) and therefore appeared to be representative of normative university students. A dependent *t*-test demonstrated that participants’ moods became significantly more negative as a function of the induction,  $t(19) = 4.51, p < 0.001$ , Hedges’  $g_{av} = .90$ . The mortality salience induction was therefore interpreted as being successful for the sample.

Seventy-five per cent of participants included in the final sample chose to return their MAPS to the researcher and were therefore analysed using LIWC. These analyses revealed that participants used an average of 13.0% ( $SD = 10.2$ ) “negative emotional” words and an average of 3.7% ( $SD = 3.5$ ) “positive emotional” words. A dependent *t*-test confirmed that participants used significantly more negative emotional words than positive emotional words,  $t(14) = 3.34, p < 0.01$ , Hedges’  $g_{av} = 1.22$ . Results therefore suggest that participants used predominantly negatively valenced words to describe their feelings around their own death.

##### **Death-evaluation IRAP**

Comparable to the previous experiment, the baseline IRAP revealed strong “my life-positive” biases ( $M = .42, SD = .35$ ), as well as moderate “my life-*not-negative*” ( $M = .21, SD = .48$ ) and “my death-positive” ( $M = .22, SD = .38$ ) biases. Weak “my death-negative” biases were also demonstrated ( $M = -.12, SD = .43$ ). No changes were apparent following the induction.

A 2×4 ANOVA was conducted on the IRAP *D* scores, with time-point (baseline vs. post induction) and IRAP trial-type as within-participant variables. A significant main effect was found for trial-type,  $F(3, 57) = 11.05, p < .001, \eta^2 = .37$ , but not time-point ( $p = .91$ ). In contrast to the previous experiment, no interaction effect was found ( $p = .62$ ). Four planned comparisons were made using Bonferroni-correct *t*-tests. These revealed no significant changes from pre to post on the four trial-types (all  $ps > .29$ ).



*Figure 4.3.* Performance on Experiment 2’s death-evaluation IRAP at baseline and after mortality salience induction. Positive *D* scores represent “positive” or “not-negative” biases, and negative *D* scores represent “negative” or “not-positive” biases.

The current results therefore suggest that the “my death-positive” biases found in previous experiments are not necessarily an artefact of the specific “positive” stimuli employed, because participants continued to demonstrate death-positive biases. The negative reinforcement associated with life and death targeted by the current study’s IRAP also produced small but significant “my death-positive” biases at baseline, and showed no change after the induction. These “death-positive” biases therefore appear to be relatively robust to both changes in mortality salience and changes to the specific dimension of evaluation (i.e., positive vs. negative reinforcement).

Interestingly, the results of the current study differ from the previous study insofar as the mortality salience induction did not impact performance on the IRAP significantly, even though the contents of one of the trial-types affected by the induction in the previous experiment (i.e., my death-negative) did not change in the current study. Whether these differences can be attributed to the contrast category manipulation will be explored below using *post hoc* comparisons (see Section 4.13). For now, it suffices to say that the mortality salience induction did not change results on the IRAP, in contrast to the previous experiment.

It is important to note at this stage that, while Experiments 1 and 2 included the word “my” in their label stimuli (i.e., “my life” and “my death”), this aspect of the stimulus did not necessarily participate in subjects’ responses on the IRAP. Specifically, as the word “my” was present on every trial, participants’ relational responses were not necessarily brought under the control of this deictic cue. As such, it is not possible to definitively conclude that participants were responding to personal death (i.e., relative to the death of someone else). The next experiment therefore employs a different form of contrast category manipulation on the IRAP employed in Experiment 1, this time in order to specifically target this deictic relation (i.e., death of self vs. other).

### **4.7 EXPERIMENT 3**

In an attempt to address whose death was being referred to within the death-evaluation IRAP, Experiments 1 and 2 included the word “my” in their label stimuli (i.e., “my life” and “my death”). That is, both studies specified that personal death of the participant was being referred to. However, it is critical to note that because the word “my” was present on every trial, this stimulus did not discriminate between trials, and therefore participants’ relational responses were not necessarily brought under the control of a deictic cue for “self”. The current experiment therefore employed a different form of contrast category manipulation relative to Experiment 1, in order to specifically target *whose* death is

being referred to (i.e., “*others’* deaths” vs. “*my* death”) rather than the dimension of life versus death (i.e., “*my life*” vs. “*my death*”, as in Experiments 1 & 2: see Table 4.3).

Table 4.3. *Stimuli used in Experiment 3’s death-evaluation IRAP*

Label stimuli		Target stimuli		Response options	
<i>Others’ deaths</i>	My death	Enjoyable	Awful	True	False
		Exciting	Distressing		
		Great	Hurtful		
		Lovely	Horrible		
		Pleasant	Painful		
		Satisfying	Upsetting		

*Note:* Stimuli in italics differ relative to Experiment 1.

## 4.8 Method

### Participants

Twenty-three students at Maynooth University were recruited from the volunteer pool (20 female and 3 male,  $M_{\text{age}} = 21.57$ ,  $SD = 4.72$ ). No incentives were offered for participation. Participants reported having completed between 0 and 2 IRAPs previous to the current study ( $M = .57$ ,  $SD = .79$ ). Inclusion criteria were identical to the previous experiment. No participants were excluded on this basis of reporting that they suffered from a serious, life-threatening or terminal illness.

### Ethical considerations

Ethical considerations were identical to the previous experiment. In summary, ethical approval was granted by Maynooth University’s Social Research Ethics Subcommittee, participants were informed as to the nature prior to participation, its voluntary nature, the right to cease participation at any time, and that all data would be irrevocably anonymised after participation. Written informed consent was obtained from all participants. Finally, all participants were fully debriefed after completing the tasks, and provided with the contact

details for the free student counselling service available on campus should they experience any form of post-experimental distress.

### **Measures and procedure**

The order and counterbalancing of the measures was identical to the previous experiment, including the Beck Hopelessness Scale ( $\alpha = .77$ ), DASS-depression subscale ( $\alpha = .74$ ), and Acceptance and Action Questionnaire II ( $\alpha = .84$ ). The structure and delivery of the IRAPs were similar to previous experiments, however the stimulus set within the IRAP were modified slightly from Experiment 1. The stimuli categories “My death”, “positive” and “negative” were identical to Experiment 1, but the label contrast category was set to “Others’ deaths” (see Table 4.3). Rule A was “other people’s deaths are positive and my death is negative”, and rule B was “my death is positive and other people’s deaths are negative”.

**IRAP data processing.** Data from the IRAP were processed in an identical manner to the previous experiment. Five participants failed to meet the mastery criteria on one IRAP’s practice blocks, and were therefore not presented with the critical test blocks. One participant failed more than one test block-pair within a single IRAP and therefore had their *D* scores excluded from the analyses. The final sample therefore contained 20 participants. Identical to the previous experiment, trial-types 3 and 4 were inverted in both IRAPs (i.e., the “death” trial-types) so as to create a common vertical axis for all analyses. Positive *D* scores therefore represented “positive” or “not-negative” biases, whereas negative *D* scores represented “negative” or “not-positive” biases.

## **4.9 Results and Discussion**

The analytic strategy was similar to the previous experiments in this chapter in all respects.

### **Self-report measures and mortality salience induction.**

The sample reported normative levels of hopelessness ( $M = 4.12, SD = 2.47$ ), depression ( $M = 2.24, SD = 2.39$ ), and psychological flexibility ( $M = 18.35, SD = 8.8$ ) and therefore appeared to be representative of normative university students. Similar to the previous experiment, a dependent  $t$ -test demonstrated that participants' moods became significantly more negative as a function of the induction,  $t(15) = 2.78, p < 0.05$ , Hedges'  $g_{av} = .96$ . The mortality salience induction was therefore interpreted as being successful for the sample.

Eighty-two per cent of participants included in the final sample chose to return their Mortality Attitudes Personality Survey to the researcher and were therefore analysed using LIWC. These analyses revealed that participants used an average of 11.0% ( $SD = 6.0$ ) "negative emotional" words and an average of 2.6% ( $SD = 3.1$ ) "positive emotional" words. A dependent  $t$ -test confirmed that participants used significantly more negative emotional words than positive emotional words,  $t(13) = 5.99, p < .001$ , Hedges'  $g_{av} = 1.76$ . Results therefore suggest that participants used predominantly negatively valenced words to describe their feelings around their own death.

### **Death-evaluation IRAP**

The baseline IRAP revealed weak "others' deaths-positive" ( $M = .21, SD = .45$ ) and strong "my death-positive" biases ( $M = .57, SD = .36$ ); however, no negative biases towards death for others ( $M = .01, SD = .41$ ) or self ( $M = .05, SD = .43$ ) were found. After the induction, participants became more negative about personal death ( $M = -.29, SD = .50$ ), whereas the direction and magnitudes of the biases on the other three trial-types did not appear to change (see Figure 4.4). Following the mortality salience induction, biases on one of the trial-types changed, whereas the other three did not. Specifically, participants became more negative about personal death.

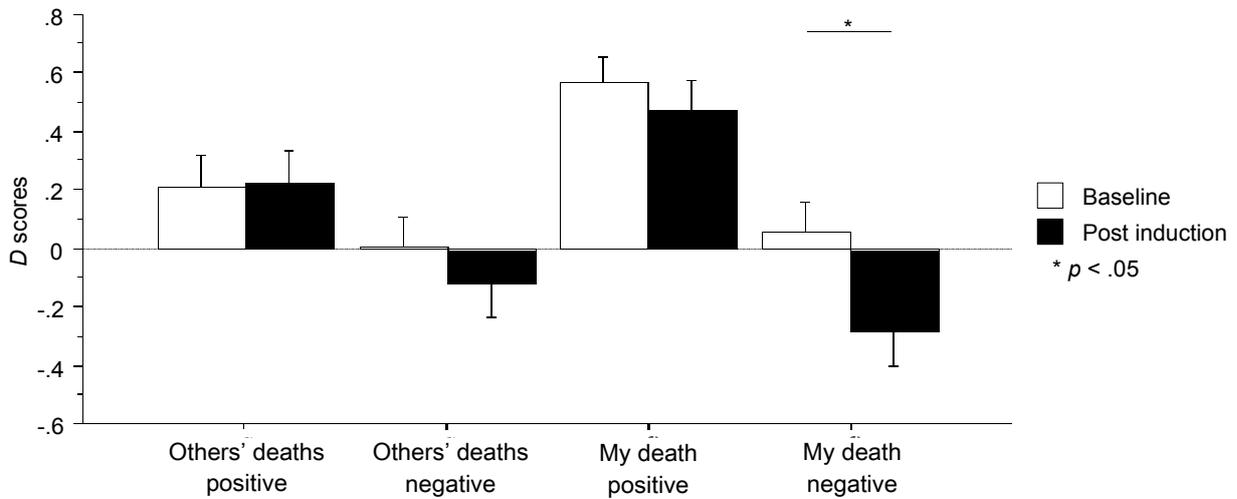


Figure 4.4. Performance on Experiment 3’s death-evaluation IRAP at baseline and after mortality salience induction. Positive *D* scores represent “positive” or “not-negative” biases, and negative *D* scores represent “negative” or “not-positive” biases.

A 2×4 ANOVA was conducted on the IRAP *D* scores, with time-point (baseline vs. post induction) and IRAP trial-type as within-participant variables. A significant main effect was found for trial-type,  $F(3, 48) = 12.24, p < .001, \eta^2 = .43$ , and a marginal effect was found for time-point ( $p = .06$ , Hedges’  $g_{av} = 0.63$ ). Similar to Experiment 1, but in contrast to the previous experiment, no interaction effect was found ( $p = .32$ ). Four planned comparisons were made using Bonferroni-correct *t*-tests. These revealed no significant changes from pre to post on the four trial-types (all  $ps > .3$ ), although a marginal effect of large effect size was found on the “my death–negative” trial-type;  $p < .05$ , Hedges’  $g_{av} = .73$ .

The current study differed from Experiment 1 by requiring participants to respond to the stimulus “my death” along the dimension of the deictic (i.e., self vs. others) rather than life versus death. Emphasising the deictic cue produced large “my death-positive” biases both before and after the induction. Marginal “others’ deaths-positive” biases were also found, suggesting that positive evaluations of death may not be limited to personal death. Similar to Experiment 1, no “my death-negative” bias was found at baseline, but a moderate “my death-

negative” bias was found after induction. The current results, in conjunction, with those from the previous two studies, therefore suggest that the absence of “my death-negative” biases in normative participants at baseline is due to inadequate salience of death. Increasing this salience via an induction produces the expected negative evaluations of death on the IRAP (although not positive evaluations of death). Additionally, the results suggest that positive evaluations of death on the IRAP are not readily explainable through inadequate salience of death, because these biases were not affected by the induction, or through lack of clarity around whose death is being referred to (i.e., self vs. others).

In order to complete a 2 (positive reinforcement vs. negative reinforcement)  $\times$  2 (my life vs. others’ deaths) factorial design for the purposes of conducting post hoc analyses, one further experiment was conducted. Recall that Experiment 2’s contrast category manipulation focused on the target category, whereas Experiment 3’s manipulation focused on the label category. Experiment 4 therefore combined these manipulations by changing both the label and target contrast categories (relative to Experiment 1).

#### **4.10 EXPERIMENT 4**

The current experiment included both the contrast category changes introduced by Experiments 2 and 3 within the same IRAP. Specifically, the label contrast category was set to “others’ deaths” and the target contrast category targeted negative reinforcement (e.g., “relief”: see Table 4.4). The IRAP therefore targeted evaluations of death along the dimension of self versus other, and the evaluative dimension of escape versus pain. All other aspects of the study were identical to the previous three experiments.

Table 4.4. *Stimuli used in Experiment 4's death-evaluation IRAP*

Label stimuli		Target stimuli		Response options	
<i>Others' deaths</i>	My death	<i>Relief</i>	Awful	True	False
		<i>Escape</i>	Distressing		
		<i>Calm</i>	Hurtful		
		<i>Soothing</i>	Horrible		
		<i>Peaceful</i>	Painful		
		<i>Comfort</i>	Upsetting		

*Note:* Stimuli in italics differ relative to Experiment 1.

#### 4.11 Method

##### Participants

Twenty-eight students at Maynooth University were recruited from the volunteer pool (20 female and 8 male,  $M_{\text{age}} = 20.21$ ,  $SD = 1.81$ ). No incentives were offered for participation. Participants reported having completed between 0 and 5 IRAPs previous to the current study ( $M = .43$ ,  $SD = 1.29$ ). Inclusion criteria were identical to the previous experiment. No participants were excluded on the basis of reporting that they suffered from a serious, life-threatening or terminal illness.

##### Ethical considerations

Ethical considerations were identical to the previous experiment. In summary, ethical approval was granted by Maynooth University's Social Research Ethics Subcommittee, participants were informed as to the nature prior to participation, its voluntary nature, the right to cease participation at any time, and that all data would be irrevocably anonymised after participation. Written informed consent was obtained from all participants. Finally, all participants were fully debriefed after completing the tasks, and provided with the contact details for the free student counselling service available on campus should they experience any form of post-experimental distress.

## Measures and procedure

The order and counterbalancing of the measures was identical to the previous experiment, including the Beck Hopelessness Scale ( $\alpha = .71$ ), DASS-depression subscale ( $\alpha = .42$ ), and Acceptance and Action Questionnaire II ( $\alpha = .86$ ). The structure and delivery of the IRAPs were similar to previous experiments, however the stimulus set within the IRAP was modified from Experiment 1 so as to combine the changes to the contrast categories employed within both Experiment 2 and Experiment 3. Specifically, the stimuli categories “My death” and “negative” were identical to Experiment 1, but the target contrast category targeted negative reinforcement (e.g., “relief”, as in Experiment 2), and the label contrast category was set to “Others’ deaths” (as in Experiment 3: see Table 4.4). The responding rules were identical to the previous experiment.

**IRAP data processing.** Data from the IRAP were processed in an identical manner to the previous experiment. Five participants failed to meet the mastery criteria on one IRAP’s practice blocks, and were therefore not presented with the critical test blocks. Four participants failed more than one test block-pair within a single IRAP and therefore had their *D* scores excluded from the analyses. In two cases, participants failed one test block-pair and therefore had that block-pair only excluded from the calculation of their final *D* score. The final sample therefore contained 19 participants. Identical to the previous experiment, trial-types 3 and 4 were inverted in both IRAPs (i.e., the “death” trial-types) so as to create a common vertical axis for all analyses. Positive *D* scores therefore represented “positive” or “not-negative” biases, whereas negative *D* scores represented “negative” or “not-positive” biases.

## 4.12 Results and Discussion

The analytic strategy was similar to the previous experiments in this chapter in all respects.

### **Self-report measures and mortality salience induction.**

The sample reported normative levels of hopelessness ( $M = 2.53$ ,  $SD = 2.01$ ), depression ( $M = 2.95$ ,  $SD = 2.01$ ), and psychological flexibility ( $M = 17.58$ ,  $SD = 6.85$ ), and therefore appeared to be representative of normative university students. Similar to the previous experiment, a dependent  $t$ -test demonstrated that participants' moods became significantly more negative as a function of the induction,  $t(18) = 2.19$ ,  $p < 0.05$ , Hedges'  $g_{av} = .61$ . The mortality salience induction was therefore interpreted as being successful for the sample.

Eighty-nine per cent of participants included in the final sample chose to return their MAPS to the researcher and were therefore analysed using LIWC. These analyses revealed that participants used an average of 9.4% ( $SD = 6.8$ ) "negative emotional" words and an average of 1.6% ( $SD = 2.2$ ) "positive emotional" words. A dependent  $t$ -test confirmed that participants used significantly more negative emotional words than positive emotional words,  $t(16) = 5.11$ ,  $p < .001$ , Hedges'  $g_{av} = 1.53$ . Results therefore suggest that participants used predominantly negatively-valenced words to describe their feelings around their own death.

### **Death-evaluation IRAP**

As illustrated in Figure 4.5, the baseline IRAP revealed weak "my death-positive" biases both before ( $M = .18$ ,  $SD = .38$ ) and after ( $M = .17$ ,  $SD = .40$ ) the induction. No biases were found on the "others' deaths-positive", "others' deaths-negative", or "my death-positive" trial-types ( $M_s = -.07-.09$ ).

A  $2 \times 4$  ANOVA was conducted on the IRAP  $D$  scores, with time-point (baseline vs. post induction) and IRAP trial-type as within-participant variables. No main or interaction

effects were found (all  $ps > .22$ ). Four planned comparisons were made using Bonferroni-correct  $t$ -tests. These revealed no significant changes from pre to post on the four trial-types (all  $ps > .28$ ).

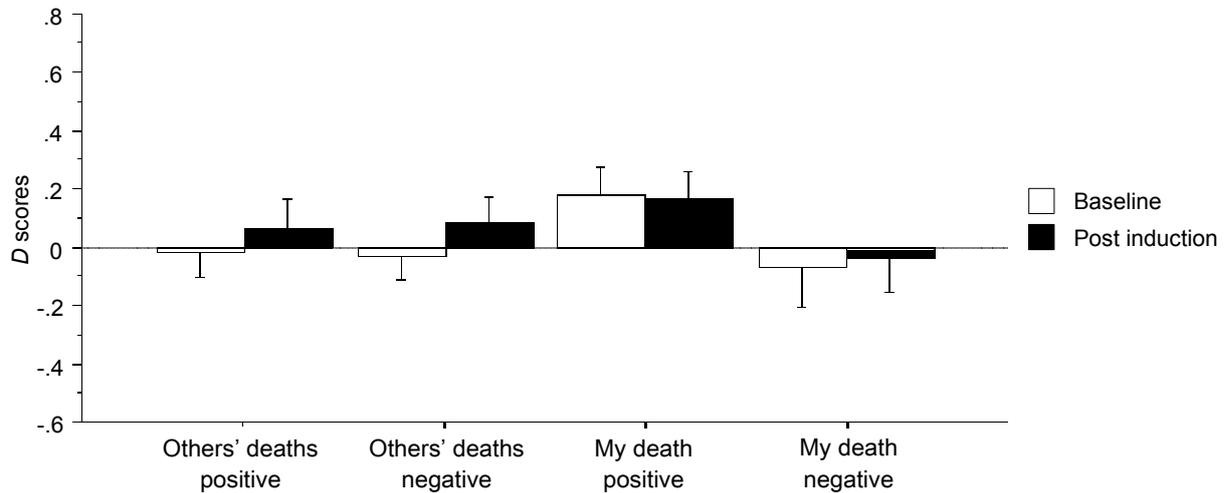


Figure 4.5. Performance on Experiment 4's death-evaluation IRAP at baseline and after mortality salience induction. Positive  $D$  scores represent "positive" or "not-negative" biases, and negative  $D$  scores represent "negative" or "not-positive" biases.

Similar to Experiment 3, which also employed "others' deaths" as the label contrast category, effects on the "others' deaths" trial-types were weak or absent, both at baseline and after the induction. Although the "my death-positive" biases were only marginally significant, their magnitude was comparable to previous studies (i.e.,  $D$  score  $\approx .2$ ). Furthermore, similar to Experiment 2, which also employed the negative reinforcement category, the mortality salience induction did not increase participants' negative evaluations of death on the IRAP.

The results of the current experiment are most meaningful via their similarities and differences to the previous three experiments. A series of post hoc analyses were therefore conducted on the data from all four of the experiments within the current chapter, as will now be discussed.

### 4.13 POST HOC ANALYSES AND DISCUSSION

At this point, it is worth recalling exactly how the IRAPs differed across the four experiments. Specifically, all four IRAPs shared a single common trial-type: “my death-negative”, with the label and target contrast categories being manipulated across the four experiments (see Figure 4.6 and Table 4.5).

In order to appreciate the nature of these contrast category manipulations, three points must be highlighted. First, the choice to keep the “my death-negative” trial-type constant across the IRAPs was based on the following rationale. Although the “death-positive” biases obtained in Chapter 2 were unexpected and therefore interesting, preliminary data from a separate ongoing experiment suggested that normative individuals and suicidal ideators were separated by their performance on the “my death-negative” trial-type specifically, whereas groups did not differ on the “my death-positive” trial-type (see Chapter 2). Given that our ultimate aim was to examine the predictive validity of the IRAP, it therefore seemed more appropriate to keep the “my death-negative” trial-type static across the current chapter’s experiments.

Second, the label contrast category manipulations were designed to specify two different functional dimensions of comparison for the stimulus “my death”. This was done by employing “my life” as the contrast in Experiments 1 and 2, and “others’ deaths” in Experiments 3 and 4. These IRAPs were intended to differentially target the distinction between life and death (i.e., *my life* vs. *my death*) or whose death was being referred to (i.e., *others* deaths vs. *my* death). Third and finally, the target contrast category manipulations were designed to manipulate the nature of evaluation. This was done by employing terms that refer to positive reinforcement (e.g., exciting) as the target contrast category in Experiments 1 and 3, and by employing terms that refer to negative reinforcement (e.g., calming) in Experiments 2 and 4. It is worth noting that the “negative” category, which remained constant across all

four experiments, could therefore be said to refer to the positive punishment associated with personal death (i.e., painful, distressing, awful).

In order to differentiate between the four different death-evaluation IRAPs employed across the four experiments, each IRAP will now be referred to by the contents of their contrast categories, as these are the only way in which they differed. The similarities and differences between the categories employed in the four experiments' IRAPs are illustrated in Figure 4.6.

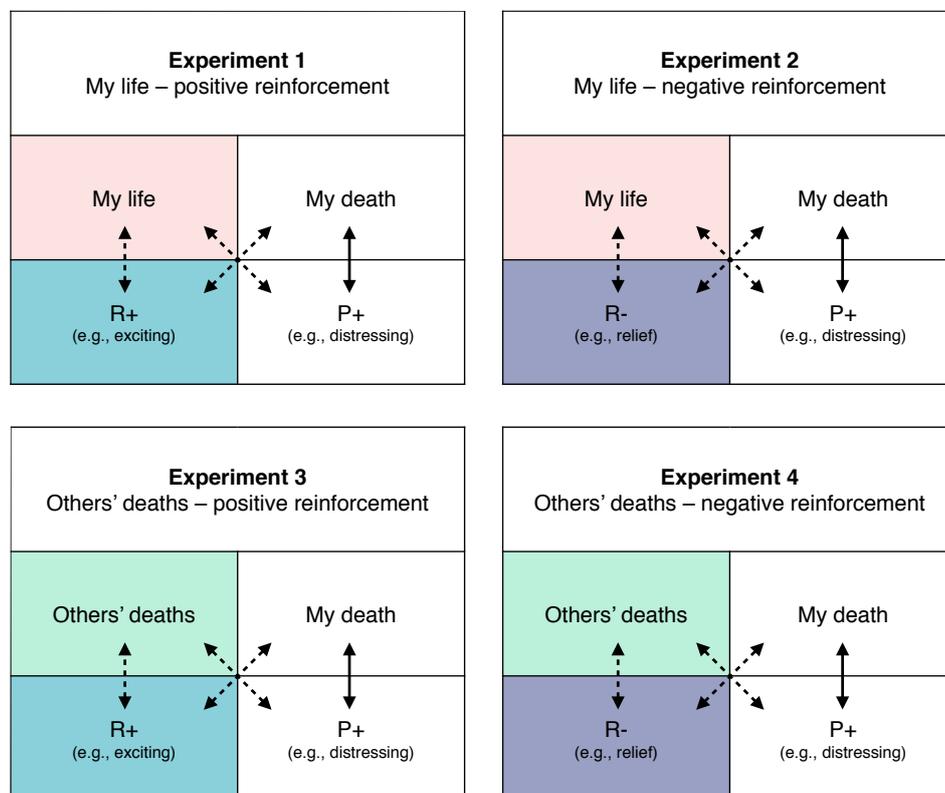


Figure 4.6. The stimulus sets employed in the four IRAPs used in this chapter. The categories “my death” and positive punishers (e.g., “distressing”), and therefore the “my death-negative” trial-type, were common to all four IRAPs. Both the label and target contrast categories were manipulated systematically. Solid arrows denote identical trial-types between IRAPs and dotted arrows denote trial-types that differed.

### **Self-report measures and mortality salience induction**

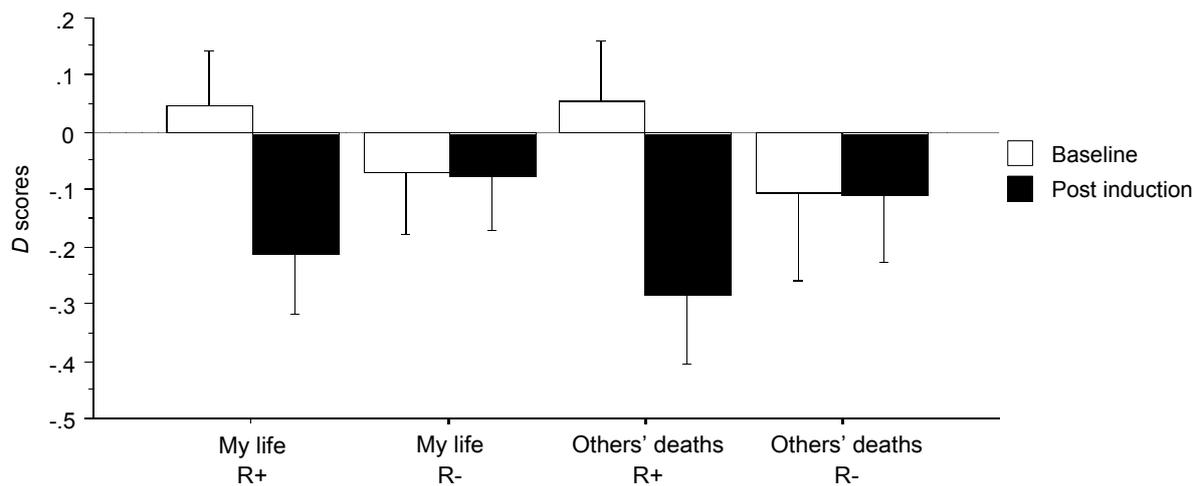
A series of ANOVAs demonstrated that participant variables across the four experiments did not differ in their age or levels of hopelessness, depression, or psychological flexibility (all  $p$ s > .08). Significant decreases in mood on the PANAS from baseline to post induction indicated that the induction was successful for the sample as a whole,  $t(71) = 5.93$ ,  $p < .001$ , Hedges'  $g_{av} = 0.62$ .

On the whole, seventy-nine per cent of participants returned their MAPS, which were analysed using LIWC. This revealed that participants used an average of 10.5% ( $SD = 7.3$ ) “negative emotional” words and an average of 2.8% ( $SD = 2.9$ ) “positive emotional” words. A dependent  $t$ -test confirmed that participants used significantly more negative emotional words than positive emotional words,  $t(58) = 7.06$ ,  $p < .0001$ , Hedges'  $g_{av} = 1.30$ . Results therefore suggest that participants used predominantly negatively valenced words to describe their feelings around their own death.

### **Death-evaluation IRAPs**

Performance on the “my death–negative” trial-type only (i.e., the only trial-type common to all four trial-types) was compared between the four IRAPs and across the mortality salience induction using a 4×2 mixed within-between ANOVA, with IRAP type (“my life–R+”, “my life–R-”, “others’ deaths–R+”, and “others’ deaths–R-”) as the between-participant variable and time point (baseline vs. post mortality salience induction) as the within-participant variable. A main effect was found for time point,  $F(3, 71) = 4.56$ ,  $p < .05$ , Hedges'  $g_{av} = .27$ , but not for IRAP type ( $p = .93$ ). A trend towards an interaction effect was found ( $p = .12$ ,  $\eta^2 = .08$ ). Given this trend for an interaction effect with a relatively low  $n$ , four planned comparisons were made using Bonferroni-corrected  $t$ -tests to assess the significance of changes from baseline to post induction for each IRAP condition. These yielded a marginally significant effect on the “others’ deaths–R+” IRAP,  $t(16) = 2.28$ ,  $p =$

0.017, Hedges'  $g_{av} = .72$ , and a trend on the “my life–R+” IRAP,  $t(18) = 2.58, p = .04$ , Hedges'  $g_{av} = .45$ . No significant changes were found for the “my life–R-” ( $p = .94$ ) or “others’ deaths–R-” IRAP ( $p = .82$ ). Given the effect sizes of the marginally significant changes, and given the small samples sizes involved, these were therefore deemed relevant.



*Figure 4.7.* Differences on the “my death-negative” trial-type between the four death-evaluation IRAPs, at baseline and after mortality salience induction. Positive  $D$  scores represent “my death-positive” or “my death-not-negative” biases, and negative  $D$  scores represent “my death-negative” or “my death-not-positive” biases.

Together, the results of the four experiments suggest that the contrast category manipulations had an impact on the “my death-negative” biases within the IRAPs. However, this impact was limited to the manipulation of the target categories (i.e., differential targeting of the positive vs. negative reinforcement associated with death), and not the label categories (i.e., whose death was being referred to). Furthermore, the effects of these the contrast category manipulation was found only after mortality salience induction. The results therefore suggest that the absence of “death-negative” biases found at baseline in Chapter 2 may have been due to inadequate salience of death for normative participants, and not inadequate specification of whose death is being referred to. Interestingly, however, the reactivity of these death-negative biases to mortality salience induction appear to be reliant

on whether this negativity was contrasted with positive or negative reinforcement. Specifically, despite the contents of the death-negative trial-type being identical across IRAPs, reactivity to the induction was found for the two “positive reinforcement” IRAPs, but not the two “negative reinforcement” IRAPs. These results therefore serve to emphasise that although the IRAP’s four trial-types are procedurally non-relative, performance within them is not a-contextual.

It is also worth noting the lack of difference on the “death–negative” trial-type across the deictic contrast category manipulations (i.e., *my life/my death* vs. *others’ deaths/my death* IRAPs) suggests that the addition of “my” to the stimulus set does not impact significantly on the performance. While somewhat speculative, this could suggest that responding within the IRAP is assumed to be self-referential even when this is not manipulated, and not that the deictic is simply ignored even when the task structure does not require its participation in individuals’ relational responses.

Finally, the “death-positive” effects found in Chapter 2 were also found in several of the current experiments. In contrast to the “death–negative” biases, these “death-positive” effects proved to be resilient (i.e., unaffected by) to the mortality salience inductions across the four experiments. As such, this unexpected positivity towards death cannot be easily explained as being due to inadequate salience of death. This will be dealt with in some detail in the general discussion.

### **Implications for the study of implicit death-evaluations in suicidal individuals**

The current results also have implications for future research within the study of suicidal behaviours. While implicit measures have been shown to have encouraging predictive utility within suicidal behaviours (e.g., Nock et al., 2010; Randall et al., 2013), research to date has been largely a-theoretical, and no work has attempted to directly connect the content or results of implicit measures with theories of suicide. It is therefore worth

noting that leading cotemporary theories of suicidal behaviour, such as the Interpersonal Theory (Joiner, 2005) and the Integrated Motivational-Volitional model (R. C. O'Connor, 2011), posit that normative individuals evaluate death negatively (e.g., fear it), and that suicidality is characterized in part by a loss of this fear of death. In order to maximize a measure's power to reveal differences between individuals with and without suicidal behaviours, future research should therefore employ a stimulus set that produces sensitivity to one's own mortality over one that does not. The current results suggest that stimulus sets that *contrasted* the punishment associated with death with positive reinforcement (e.g., the R+ IRAPs employed in Experiments 1 & 3) produce greater sensitivity than those that specify negative reinforcement (e.g., the R- IRAPs employed in Experiments 2 & 4).

At this point, having now developed and employed the contrast category manipulation method in order to select a "good candidate" IRAP stimulus set, it therefore seems important to explore differences between normative and suicidal individuals. The next study therefore aimed to explore whether individuals with suicidal ideation would demonstrate "death-negative" effects, as they are by definition high in mortality salience, or whether they would demonstrate "death-*not*-negative" effects, due to the acquired fearlessness of death that has been observed in suicidal individuals (see Joiner, 2005; R. C. O'Connor, 2011). For the purposes of the final experiment, we selected Experiment 1's IRAP (i.e., my life–my death/positive reinforcement–positive punishment) from among the four IRAPs employed in the current chapter. This IRAP was selected on the basis that 1) it was found to be sensitive to changes in the salience of mortality, and 2) a focus on whose death was being referred to did not appear to influence results. As such, it seemed most useful to also capture responses to life rather than others' deaths within the suicidal sample.

## **CHAPTER 5: COMPARING DEATH EVALUATIONS ON THE IRAP BETWEEN NORMATIVE AND SUICIDAL INDIVIDUALS**

Abstract: In this experiment, implicit death-evaluations are compared between normative and suicidal individuals. Participants completed two versions of the death-identity IRAP, one of which specified personal death (i.e., was self-focused) and the other of which targeted death in the abstract. Self-focused evaluations of death reliably distinguished between the two groups, correctly classifying 74% of cases, but evaluations of death in the abstract did not. The suicidal group produced specific biases indicating a rejection of the negativity of death. Results are consistent with the definition of suicidality as involving a self-focused wish to die. Furthermore, results are consistent with the assertion that suicidal ideation is specifically associated with a decreased aversion to the negative aspects of death. Results are therefore consistent with two key points that are common to leading contemporary theories of suicidal behaviour. The utility of implicit death-evaluations should therefore be considered alongside self-evaluations and self-death associations in the future.

It is worth reiterating at this point that while research on suicide to date has explored “self-evaluation” and “death-identity” biases (Franck et al., 2007; Nock et al., 2010), no work has examined the third possible combination of these categories (“death-evaluation” biases) within suicidal populations. This is somewhat surprising, given the central role that evaluations of death (and life) play in both of the leading contemporary theories of suicidal behaviour: the Interpersonal Theory of Suicide (IPT: Joiner, 2005; Van Orden et al., 2010) and the Integrated Motivational-Volitional model of suicide (IMV: R. C. O’Connor, 2011). As discussed in Chapter 1, both theories posit that unbearable psychological pain associated with living provides a motivation for the development of suicidal ideation. Similarly, both theories argue that the movement from ideation to attempts is moderated by an acquired

capacity for suicide, including a lack of fear of death. Elsewhere, empirical research has also focused on evaluations of life and death across normative and suicidal individuals, for example using the Multi-Attitude Suicide Tendency scale (Ferrara et al., 2012; Muehlenkamp & Gutierrez, 2004; Orbach et al., 1991; Osman et al., 2000); by examining the comfort some individuals derive from suicidal ideation (Crane et al., 2014); and by comparing the desirability of life versus death (Brown et al., 2005; Kovacs & Beck, 1977). The current study therefore seeks to fill this gap in the literature by examining the utility of implicit evaluations of death across normative and suicidal individuals. In doing so, we therefore hoped to link the results of the current study more closely with existing theory on evaluations of death within suicidality, both of which have implicated a fearlessness of death (Joiner, 2005; Van Orden et al., 2010).

A second way in which the current study sought to better link implicit measures with existing theory is through an exploration of the role of self in generating effects on the IRAP. Given that suicide is defined in part as involving a *self-focused* wish to die (Silverman, 2006; Silverman, Berman, Sanddal, O'Carroll, & Joiner, 2007b), the self is therefore thought to play a key role in how death and the future are verbally constructed for individuals contemplating suicide. In order to assess this theoretical supposition, we created two separate death-evaluation IRAPs and manipulated the role of self across them. Given that that manipulation of the relative emphasis on self was not shown to influence results in the previous chapter, the current study instead elected to manipulate whether the IRAP contained a reference to self or not (rather than merely the relative emphasis of self). The "personal IRAP" therefore included a reference to self (e.g., "my death"), whereas the "abstract IRAP" did not (e.g., "death"). The tasks were otherwise identical, and as such any differential effects between them may be attributed to the presence or absence of a reference to the self.

Based on the results and recommendations of the previous chapter, we elected to employ the positive reinforcement stimulus category over the negative reinforcement category, given that the former was shown to be sensitive to mortality salience. However, it should be noted that no mortality salience induction procedure was employed in the current study, due to the vulnerable nature of the clinical population.

We made two hypotheses, one specific and one exploratory. First, given that suicidality is characterized in part by a self-focused wish to die, we hypothesized that the self-focused measure of implicit evaluations of death (i.e., personal IRAP) would be a superior predictor of group membership than the measure of abstract implicit evaluations of death (i.e., abstract IRAP). Second, we hypothesized that the normative and suicidal groups would be separated by performance on one or more specific biases (e.g., performance on the IRAPs' life-positive, life-negative, death-positive, or death-negative trial-types). However, given the mixed results obtain by previous research using self-report measures, no specific predictions were made regarding which trial-type(s) would separate the groups (e.g., negative evaluations of life, or negative evaluations of death).

## **5.1 Method**

### **Participants**

Participants were recruited from two populations: the normative sample consisted of undergraduate students at Maynooth University, Ireland; and the suicidal ideation sample consisted of psychiatric patients attending St. Patrick's University Hospital, a large private psychiatric hospital in Dublin. Participants at the St. Patrick's site were recruited in an ad hoc manner from an evidence-based treatment group for self-harm based on the skill's training portion of Dialectic Behavior Therapy (see Booth, Keogh, Doyle, & Owens, 2014). Twenty-five university students and twenty-four service users at St. Patrick's were recruited (see Table 5.1 for demographic information). No incentives were offered for participation.

Participants reported that they had completed a maximum of one IRAP previous to the current study ( $M = 0.15$ ,  $SD = 0.36$ ). Written informed consent was obtained prior to participation and no remuneration was offered.

Inclusion criteria were similar to the experiments in previous chapters. Participants in the suicidal ideation group were additionally required to report current suicidal ideation (i.e., Beck Scale for Suicidal Ideation [BSSI]  $\geq 2$ : see (Beck, Steer, & Ranieri, 1988; Brown et al., 2000). For ethical reasons, participants at the Maynooth University site were not screened for suicidal ideation. Instead, participants at this site were recruited if they scored in the normative range of a proxy measure of suicidality that has also been shown to be predictive of suicide risk: the Beck Hopelessness Scale (BHS: Beck et al., 1974; Brown et al., 2000).

### **Ethical considerations**

It is important to note the current research was embarked upon only after careful consideration of the literature on iatrogenic effects (i.e., causing that which one seeks to assess) associated with assessments within vulnerable populations, and suicidality specifically. This literature strongly suggests that the assessment of suicidality has been shown not to be associated with significant distress during the assessment or at follow-up several days or weeks later, or with an increase in future suicidal behaviour (see Cha & Nock, 2009; Gould et al., 2005; Muehlenkamp, Swenson, Batejan, & Jarvi, 2015; Reynolds, Lindenboim, Comtois, Murray, & Linehan, 2006; see also Ceci & Bruck, 2009; Hussey, McEnteggart, Nicholson, & Thompson, 2015; Jacomb et al., 1999; Jorm, Kelly, & Morgan, 2007; Yeater, Miller, Rinehart, & Nason, 2012). Indeed, such research has demonstrates that participation in research is frequently associated with perceived benefits to the individual, and that the higher the degree of (e.g., Jacomb et al., 1999). Nonetheless, careful adherence to ethical guidelines was followed throughout, as will now be discussed.

Ethical approval was granted by Maynooth University's Social Research Ethics Subcommittee for data collected at the Maynooth site, and by St. Patrick's University Hospital's Research Ethics Committee for data collected at the St. Patrick's site. Furthermore, the researcher was made an honorary researcher at St Patrick's University Hospital.

At both sites, participants were informed as to the nature and purpose of the study prior to participation. Additionally, at the St. Patrick's site, information regarding the study was provided at least one day prior to participation. In the majority of cases participation occurred several days after being informed about the study. Furthermore, participants at both sites were made aware that participation was entirely voluntary, that they could cease participation at any time without giving a reason, and that all data would be irrevocably anonymised immediately after participation. Written informed consent was obtained from all participants. Upon completion of the tasks, participants were fully debriefed on the nature and purpose of the experiment and provided with an opportunity to ask questions, which were then fully answered by the researcher. At the Maynooth site, this debrief included the provision of the contact details for the free student counselling service available on campus, should they experience any form of post-experimental distress.

At the St. Patrick's site, where suicidality was assessed, several procedural choices were made based on ethical considerations. First, all assessments via the self-report measures were administered by an assistant psychologist who was a member of the participant's clinical care team at the hospital and was already known to the participant. The assistant psychologist provided a debriefing for the self-report measures section of participation and followed the hospital's established internal reporting procedures in the instance of a participant being distressed during the assessment. Immediately afterwards, participants then met the current author to complete the implicit measures. In this way, the researcher was blinded to the participant's suicidal history. After the completion of the implicit measures, the

researcher debriefed the participant. This included answering any questions the participant might have about the nature or purpose of the implicit measures. Appropriate reporting procedures in the instance of a participant becoming distressed during the research. Specifically, the researcher would contact the assistant psychologist assigned to the participant's care team (who previously administered the self-report measures), and further action would be taken based on the hospital's established internal procedures.

## **Measures**

The Beck Hopelessness Scale ( $\alpha = .93$ ) and Depression Anxiety Stress Scale's depression subscale ( $\alpha = .92$ ) were identical to previous chapters. In addition, participants at the St Patrick's site completed two self-report measures of suicidal behaviours, as follows.

**Beck Scale for Suicidal Ideation.** The BSSI is a 21-item Thurstone scale that assesses an individual's motivation and preparedness for a suicide attempt (Beck et al., 1988). Due to ethical constraints, this scale was used to assess the severity of past week suicidal ideation in the suicidal ideation group only. The BSSI has been shown to have high internal consistency and test-retest reliability in psychiatric samples, and to be predictive of subsequent suicide attempts (Beck, Brown, Steer, Dahlsgaard, & Grisham, 1999). The "current" (i.e., past week) version of the scale was employed in the present study. Internal consistency was found to be good in the current sample ( $\alpha = .88$ ).

**Self-Injurious Thoughts and Behaviors Interview.** Although originally designed as a structured clinical interview (SITBI: Nock, Holmberg, Photos, & Michel, 2007, see Appendix P) this measure has since been converted to a self-report format (Latimer, Meade, & Tennant, 2013), which was used here. Both formats have demonstrated good reliability and validity. The SITBI defines a variety of behaviours for the participant and asks the individual to report their presence, severity, recency, and frequency, including suicide attempt (i.e., "an actual attempt to kill yourself in which you had at least *some* intent to die"), non-suicidal self-

injury (i.e., “purposely hurting yourself *without* wanting to die”) and suicidal ideation (i.e., “*thoughts* about killing yourself”). The SITBI was used to establish the lifetime prevalence of thoughts about non-suicidal self-injury, non-suicidal self-injury, suicidal ideation, suicidal planning, and suicide attempts.

**Death-evaluation IRAPs.** The Personal Death IRAP’s stimulus set and responding rules were identical to that employed in Chapter 4 Experiment 1. The Abstract Death IRAP’s stimulus set was identical to the Personal Death IRAP except that the word “my” was removed from both label stimuli (see Table 5.1).

Table 5.1. *Stimuli used in the Personal and Abstract IRAPs.*

Label stimuli				Target stimuli		Response options	
Personal IRAP		Abstract IRAP					
My life	My death	Life	Death	Enjoyable	Awful	True	False
				Exciting	Distressing		
				Great	Hurtful		
				Lovely	Horrible		
				Pleasant	Painful		
				Satisfying	Upsetting		

## Procedure

Participants completed the self-report measures first, followed by the two death-evaluation IRAPs. The order of presentation of the two IRAPs, and the block order within them (i.e., rule A block first vs. rule B block first), was fully counterbalanced between participants. The structure and delivery of both IRAPs was identical to that in previous experiments in all respects.

## IRAP data processing

Raw latencies on the IRAP were converted into *D* scores and excluded based on failure to meet mastery criteria in an identical manner to the previous experiments. One

participant failed to meet the mastery criteria on one IRAP's practice blocks and was therefore were not presented with the critical test blocks. In eight cases, participants failed one test block-pair and therefore had that block-pair only excluded from their final *D* score. In four cases, participants failed more than one test block-pair and thus had their *D* scores for that IRAP excluded from the analyses. The final sample therefore contained 25 individuals in the normative group and 23 individuals in the suicidal ideation group who had data for at least one IRAP. Identical to the previous experiment, trial-types 3 and 4 were inverted in both IRAPs (i.e., the "death" trial-types) so as to create a common vertical axis for all analyses. Positive *D* scores therefore represented "positive" or "not-negative" biases, whereas negative *D* scores represented "negative" or "not-positive" biases.

## 5.2 Results

### Self-reports measures

A series of analyses explored differences between the normative and suicidal ideation groups on the demographic and self-report measures. A Pearson's Chi-squared test revealed no significant differences in gender distribution between the normative and suicidal ideation groups ( $p = .94$ ). Independent *t*-tests revealed that the suicidal ideation group reported significantly higher hopelessness and depression than the normative group, as expected based on the use of these measures as screening tools (all  $ps < 0.001$ ). The suicidal ideation group also reported a higher average age ( $p < .001$ ). All effects were very large (Hedges'  $g_s \geq 2.22$ ). As per the inclusion criteria, all participants in the suicidal ideation group reported experiencing suicidal ideation in the last week, as measured by the BSSI. Furthermore, 52% of the sample reported a lifetime history of non-suicidal self-injury on the SITBI, 65% reported a history of suicidal planning, and 57% reported one or more suicide attempts (see Table 5.2).

Table 5.2. Demographic and self-report data for the normative and suicidal ideation groups.

Variable	Normative group	Suicidal ideation group	Statistical test	Effect size
Gender	14 female, 11 male	13 female, 10 male	$\chi^2(1) = .75$	$\Phi = .13$
SITBI lifetime prevalence:				
Non-suicidal self-injury	-	52%	-	-
Suicidal ideation	-	100%	-	-
Suicidal planning	-	65%	-	-
Suicide attempt	-	57%	-	-
Mean:				
Age	18.7 (2.7)	38.6 (12.4)	$t(46) = -7.81^{***}$	$g_s = 2.22$
BHS	3.4 (2.0)	13.6 (4.5)	$t(46) = -10.10^{***}$	$g_s = 2.87$
DASS depression	4.1 (3.6)	13.6 (4.5)	$t(46) = -8.20^{***}$	$g_s = 2.33$
BSSI	-	16.0 (9.4)	-	-

Note: For means, standard deviations are given in parentheses. SITBI = Self-Injurious Thoughts and Behaviors Interview, BHS = Beck Hopelessness Scale; DASS depression = Depression Anxiety Stress Scales depression subscale; BSSI = Beck Scale for Suicidal Ideation;  $g_s$  = Hedges'  $g_s$ .  
\*\*\* $p < .001$ .

### Group differences on the IRAPs

Mean  $D$  scores on both IRAPs are depicted in Figure 5.1. Both groups produced comparable “life-positive” and “death-positive” biases on the two IRAPs. The pattern of effect appeared to diverge on the negative trial-types, however. The normative group produced moderate “life-not-negative” biases on both IRAPs, whereas the suicidal ideation group produced strong “life-not-negative” biases on the abstract IRAP and no bias on the personal IRAP. Finally, the suicidal ideation group produced moderate “death-not-negative” biases on both IRAPs, whereas the normative group showed no such biases.

A  $2 \times 4 \times 2$  mixed within-between ANOVA was employed to assess the relationship between performances on the two IRAPs across the two groups, with IRAP (abstract vs. personal) and trial-type as within group variables, and group (normative vs. suicidal ideation) as a between groups variable. No main effects were found for either group ( $p = .08$ ) or IRAP

type ( $p = .36$ ). In other words, performance did not differ significantly based on group membership or whether the content of the IRAP referred evaluations of personal death or abstract death. Critically, however, a three-way interaction effect was found between group, IRAP type, and trial-type,  $F(3, 135) = 3.49, p < .05$ . A series of Bonferroni-corrected post hoc comparisons demonstrated that this three-way interaction was driven by differential performance on a single trial-type: only  $D$  scores on the personal IRAP's "my death-negative" trial-type differed between the two groups. Specifically, the suicidal ideation group produced a moderate "my death-not-negative" bias ( $M = 0.29, SD = .41$ ), whereas the normative group produced a weak "my death-negative" bias ( $M = -0.12, SD = .38, p < .01$ ). This differential performance was of a very large effect size (Hedges'  $g_{av} = 1.02$ ).

The suicidal ideation group was therefore characterized by a specific rejection of negative emotions related to their own death, but not death in the abstract. In contrast, the normative group had no strong biases regarding the negativity of death, whether personal or in the abstract. It is worth noting that the presence or absence of a reference to the self (i.e., the word "my") therefore produced a significantly different pattern of effect across the two IRAPs and between the two groups.

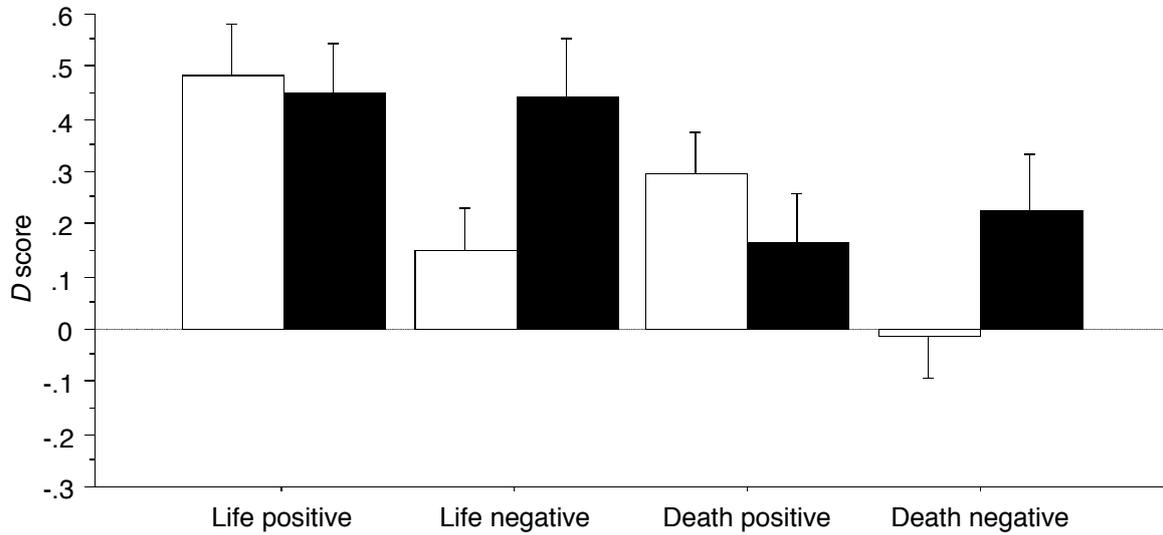


Figure 5.1. Performance on the IRAPs between the normative and suicidal ideation groups. Upper panel: abstract death-evaluation IRAP. Lower panel: personal death-evaluation IRAP. Positive *D* scores represent “positive” or “not-negative” biases and negative *D* scores represent “negative” or “not-positive” biases. Error bars represent standard errors. \*\*  $p < .01$

### Concurrent predictive validity of the IRAP

A series of Receiver Operating Characteristic (ROC) curves were then calculated to assess the degree to which individuals could be classified into their known groups (normative vs. suicidal ideation) based on the results of the abstract or personal IRAPs alone. ROC curves are used to assess the accuracy of a test provide a cutoff-independent method for the evaluation of the accuracy of a test by assessing the proportionate of correctly ranked cases (i.e., the "Area Under the Curve" [AUC]; see Gardner & Greiner, 2006). These demonstrated that only one IRAP trial-type was found to correctly rank a significant number of cases: the personal IRAP's "My death-negative" trial-type,  $AUC = .76, p < 0.01$  (all other  $ps \geq .13$ ; see Table 5.3 and Figure 5.2). In effect, this trial-type correctly ranked 76% of cases by their known groups. It is worth noting, therefore, that the "death-negative" trial-type on the abstract IRAP did not significantly rank individuals ( $AUC = .64, p = .13$ ). The IRAP's ability to correctly rank individuals by their known groups was therefore dependent on the presence of a reference to self (i.e., the word "My") in the stimulus set.

Table 5.3. *The abstract and personal death-evaluation IRAPs as predictors of group membership (normative vs. suicidal ideation).*

IRAP trial-type	AUC	95% CI
<i>Abstract death IRAP</i>		
Life positive	.46	(.28 to .65)
Life negative	.62	(.44 to .80)
Death positive	.40	(.22 to .58)
Death negative	.64	(.47 to .82)
<i>Personal death IRAP</i>		
My life positive	.49	(.30 to .68)
My life negative	.36	(.19 to .54)
My death positive	.60	(.42 to .78)
My death negative	.76**	(.60 to .91)

*Note:* AUC = Area under the curve, 95% CI = 95% Confidence interval.

\*\*  $p < 0.01$

Whereas ROC curves allow us to assess the proportion of correctly-ranked individuals, contingency tables and classification statistics based on a cut-off value allow us to make predictions about specific individual cases; for example, to estimate the probability that an individual who produced a given  $D$  score is a member of the suicidal ideation group rather than the normative group. A cut-off  $D$  score was therefore derived from the ROC curve for the “my death-negative” trial-type (on the grounds that only it yielded a significant AUC). The choice of an “optimum” cut-off value is necessarily subjective (Rutter & Miglioretti, 2003), therefore we elected to follow the approach employed by relevant previous research (Beck et al., 1999; Harriss & Hawton, 2005; Niméus, Alsén, & Träskman-Bendz, 2002). Specifically, the optimal cut-off value was considered to be the one that maximized the correct classification of both groups (i.e., maximized the sum of sensitivity and specificity, at the point of furthest displacement of the ROC curve). This was found to be at  $D$  score = 0.03. Interestingly, this value closely corresponds to the more procedurally meaningful value at which participants responded to “my death-negative” with both response options (True and False) at equal speed (i.e.,  $D$  score = 0). Previous research using the IAT to explore implicit death-identity effects has also found optimal cut-off score to correspond closely with a  $D$  score of zero (Nock et al., 2010). We therefore selected 0 as a cut-off point.

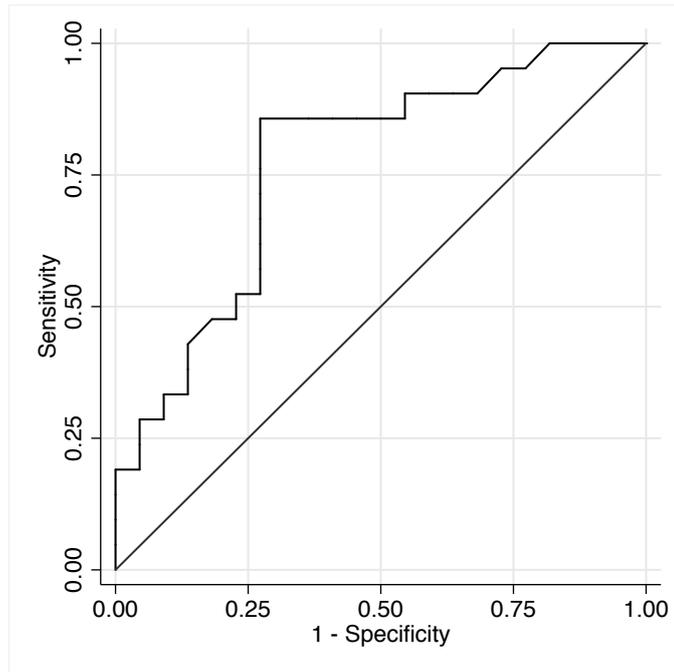


Figure 5.2. Receiver Operating Characteristic curve for the personal death-evaluation IRAP’s “my death-negative” trial-type as a predictor of group membership (normative vs. suicidal ideation).

Scores on the “my death-negative” trial-type were dichotomized using this cut-off. A Fisher’s exact test demonstrated that the IRAP was a significant predictor of group membership (OR = 10.50, 95% CI [2.34 to 47.03],  $p < 0.01$ ), correctly classifying 76% of individual cases as being a member of the suicidal ideation or normative group<sup>6</sup>. A  $D$  score  $> 0$  on the “my death-negative” trial-type therefore increased the probability of being in the suicidal ideation group by approximately 10 times. This cut-off also yielded good sensitivity and adequate specificity (see Table 5.4). Specifically, the proportion of true positives to false negatives was high (.86), and the ratio of true negatives to false positives was moderate (.64). The positive likelihood ratio implies that 2.36 individuals were correctly identified as suicidal ideators for every 1 normative individual who was misidentified as an ideator. The negative

---

<sup>6</sup> It should be noted that when the exact cut-off value derived from the ROC curve (i.e.,  $D$  score  $> 0.03$ ) was employed in an identical analysis, concurrent predictive validity was slightly improved (OR = 12.86). However, this was at the expense of face validity and consistency with comparable previous research.

likelihood ratio implies that .22 ideators were misidentified as normative for every 1 normative individual who was correctly identified as normative.

Table 5.4. *Classification table for the personal death-evaluation IRAP's "my death-negative" trial-type in predicting group membership (cut-off D score  $\geq 0$ ; N = 43).*

D score	Group		Sensitivity	Specificity	+LR	-LR	Diagnostic Odds Ratio [95% CI]
	Suicidal ideation	Normative					
> 0	18	8	.86	.64	2.36	.22	10.50
$\leq 0$	3	14					[2.34 to 47.03]**

Notes: Scores on the IRAP were dichotomized to indicate either "My death negative" ( $D$  score  $\leq 0$ ) or "My death *not*-negative" ( $D$  score  $> 0$ ) effects. +LR = positive likelihood ratio, -LR = negative likelihood ratio, OR = odds ratio, CI = confidence interval.

\*\* Fischer's exact test:  $p < .001$

### 5.3 Discussion

The current study examined differences in death-evaluation on the IRAP between normative students and psychiatric patients with past-week suicidal ideation, roughly half of who also had a lifetime history of suicide attempts. A within-groups comparison was also included: as discussed in the opening chapter, suicide is defined in part by a self-focused desire to die, rather than a general endorsement of death or suicide generally (O'Carroll et al., 1996; Silverman, 2006). However, little empirical research has directly assessed this assertion. This study manipulated whether the stimulus set contained a reference to self or not between two IRAPs (e.g., required participants to respond to "death" or "my death") in order to observe whether this influenced the IRAPs' ability to discriminate between the two groups. The stimulus sets were otherwise identical to that selected by the previous chapter.

Results demonstrated that both the normative and suicidal groups demonstrated "death-positive" biases, providing further evidence that these biases are relatively robust across studies and populations. In contrast to the normative group (in both the current study

and the studies contained in previous chapters), the suicidal ideation group demonstrated “life–*not*–negative” effects.

Importantly, the presence or absence of a reference to self within the IRAP stimulus sets significantly influenced results: large and significant differences were found between the normative and suicidal ideation groups on the “my death–negative” trial-type in the personal death IRAP (Hedges’  $g_s = 1.02$ ). In contrast, no significant differences were found on the “death–negative” trial-type on the abstract death IRAP. This finding supports the assertion that suicidality involves a self-focused wish to die; in doing so, results also emphasise the utility of appealing to domain specific theorizing when designing stimulus sets. In summary, results from the ANOVAs demonstrated that the normative and suicidal ideation groups demonstrated large and significant differences between their mean bias score (i.e.,  $D$  scores) on the personal death IRAP’s “my death–negative” trial-type.

In order to also assess the IRAPs’ inferential utility, we then assessed whether scores on each trial-type were predictive of group membership using a series of Receiver Operating Characteristic curves. Analyses demonstrated that performance on the “my death–negative” trial-type (and no other trial-types in the two IRAPs) could correctly ordinaly rank a significant proportion of individuals (76% AUC,  $p < .01$ ) by their respective group memberships. This compares favourably with previous research that has attempted to discriminate between normative individuals and those with suicidal ideation. For example, the self-cutting IAT demonstrated an AUC of 74% in one previous study (Nock & Banaji, 2007b).

Finally, in order to assess whether individual performances on the IRAP were predictive of a history of suicidal behaviours at an individual level, a cut-off value was derived from the ROC curve at the point of maximum combined sensitivity and specificity. Interestingly, this cut-off value was found to correspond with a procedurally meaningful

value (i.e.,  $D$  scores  $\geq 0$ ) that indicates whether participants' performance reflects either "my death–negative" or "my death–*not*-negative" effects. Classification statistics indicated that, at this cut-off point, the personal death IRAP's "my death–negative" trial-type possessed good sensitivity and adequate specificity. On the basis of an individual's performance on this trial-type only, the IRAP could distinguish between individuals with and without a history of suicidal ideation with 76% accuracy. More specifically, demonstrating  $D$  score of greater than 0 (i.e., a "death–*not*-negative" effect on the IRAP) made that individual roughly ten times more likely to have a history of suicidal behaviours (OR = 10.50, 95% CI [2.34 to 47.03],  $p < 0.01$ ). Results therefore support the concurrent validity of death-evaluations on the IRAP, and demonstrate the ability to make inferences about specific individuals with a relatively high degree of confidence.

### **Links with theories of suicide**

As noted in the general introduction in Chapter 1, we elected to employ the IRAP over other more commonly employed measures such as the IAT due to the greater conceptual clarity over the interpretation of its results. Specifically, the IAT effect can be interpreted as the strength of one pattern of categorisation relative to another (e.g., a "death-negative/life-positive faster relative to death-positive/life-negative" effect), whereas trial-type effects on the IRAP can be interpreted as the strength of asserting versus rejecting a specific proposition (e.g., a "death-positive-true relative to death-positive-false" effect). As stated in Section 1.4, this ability to tie effects on the IRAP back to specific propositions rather than overall patterns of response bias affords it greater ability to connect with and test theories of suicidality, as will now be discussed.

**Suicide as a self-focused wish to die.** Results demonstrated that self-focused evaluations of death on the IRAP reliably distinguished between individuals with suicidal ideation and normative individuals, whereas evaluations of abstract death did not. The simple inclusion of the word “my” thus produced key differences in the concurrent predictive validity of the IRAP, underscoring the need for theoretically well-informed stimulus selection strategies in the service of maximizing predictive validity (Nosek & Greenwald, 2009). Critically, results are therefore consistent with definitions of suicidal behaviour as involving a self-focused wish to die (Silverman, 2006; Silverman et al., 2007b), as the IRAP that contained a reference to the self successfully discriminated between individuals in the suicidal ideation and control groups.

**Suicide as involving a loss of “fearlessness of death”.** Additionally, suicidal ideation was associated with a specific implicit bias: the personal death IRAP’s “my death-negative” trial-type. Participants who rejected the negativity of their own death (i.e., produced positive *D* scores) were roughly 10 times more likely to be in the suicidal ideation group. This cut-off score was also shown to have good sensitivity and adequate specificity. In contrast, positive evaluations of death, and both positive and negative evaluations of life, showed no concurrent predictive validity. Results therefore indicate that evaluations of death may be useful in predicting risk of suicide among low and high-risk individuals. This adds to previous research on suicidal behaviours which has focused predominantly on evaluations of self (e.g., Franck et al., 2007; Glashouwer et al., 2010; Price et al., 2014) and associations between death and self (e.g., Dickstein et al., 2015; Ellis et al., 2015; Harrison et al., 2014; Nock et al., 2010; Randall et al., 2011).

Furthermore, such results are consistent with the “acquired capacity for suicide” postulated by both leading theories of suicidal behaviour (Joiner, 2005; R. C. O’Connor, 2011). Such theories argue that individuals’ innate fear of death is undermined or eroded by

either direct or indirect experience of physically dangerous and life threatening behaviours (e.g., ideation, self-harm, suicide attempts). The suicidal ideation group's rejection of the negativity of death on the IRAP is therefore supportive of the concept of fearlessness of death within suicidality. Indeed, it is interesting to note the conceptual overlap between Joiner's use of the word fearlessness (i.e., loss of pre-existing fear: Joiner, 2005, p. 147) and the specific rejection of negativity of death found in the current results. It should be noted that, to the best of the authors' knowledge, this is the first time that fearlessness of death in suicidality has been observed using an "objective" behavioural measure (Glenn & Nock, 2014a) rather than self-reports (Ribeiro et al., 2014a).

In summary, evaluations of personal death on the IRAP were found to successfully discriminate between individuals with current suicidality and controls. The IRAP's inferential utility was comparable with that demonstrated by previous research using the IAT (and targeting a variety of stimuli). However, in contrast with previous research, the nature of effects on the IRAP allowed for two specific links to be made between the data generated from the implicit measure and existing theories of suicidality (i.e., suicide as a self-focused wish to die, and suicide as involving a loss of "fear of death").

Future research might therefore follow the pattern set in previous research using the IAT (e.g., Nock & Banaji, 2007b; Nock et al., 2010) by examining the IRAP's criterion validity in the prospective prediction of self-harmful behaviours. The relative utility of multiple implicit measures could be compared (*à la* Randall et al., 2013). However, the use of measures such as the IRAP may allow such future research to go beyond a merely outcomes-based evaluation of the utility of such measures to also include a greater number of theory-based hypotheses. This could serve to constrain experimentation in a useful way in order to conserve resources in this resource-intensive area of research.

One limitation of the current study's sampling method should be acknowledged at this point. For ethical reasons, suicidal ideation was not assessed in the student population, thus the groups were not necessarily mutually exclusive. As such, it is possible that some individuals in this sample may have met criteria for current suicidal ideation. Should this have been the case, it may have artificially inflated the false positive rate (and therefore suppressed the specificity and negative likelihood ratio) of responses on the IRAP. Future research might employ mutually exclusive groups in order to explore whether this influenced the results.

## CHAPTER 6: GENERAL DISCUSSION

### 6.1 Summary of findings

The following chapter provides an overview of the current research findings, and then considers the conceptual and theoretical issues that were raised in the course of this work. Throughout these discussions, recommendations for future research on the use of implicit measures within the study of suicidality (and more generally) are provided.

The systematic review conducted in Chapter 1 concluded that implicit measures appear to have utility within the study of suicidal behaviours, having demonstrated a degree of both construct and prospective criterion validity across at least 13 published articles. This review also noted several directions for future research that formed the basis of the current body of research. To recall, these included (a) a need to link the effects generated by implicit measures to theories of suicide; (b) the fact that no research to date had examined implicit evaluations of death, despite the centrality of such evaluations to most theories of suicide; and (c) closer consideration of the role of individuals' immediate psychological context (e.g., current mood, mortality salience) while completing implicit measures related to suicide, for example within analogue studies.

In order to address these points, several variations of death-evaluation IRAP were employed across six empirical studies, employing both experimental analogue (i.e., within normative participants) and cohort comparison designs (between normative students and psychiatric patients with past-week suicidal ideation). Two further experiments provided a “proof of concept” of the contrast category manipulation method within other amenable domains (i.e., implicit self-esteem and objectification of women). The aims and results of each study will be briefly summarised, and their implications highlighted.

The study contained in Chapter 2 explored implicit attitudes to death in a normative population using the widely used death-identity IAT originally employed by Nock and

colleagues (2010), as well as a death-identity IRAP. The results on the IRAP indicated that the normative participants demonstrated strong “self-life” biases specifically (i.e., not “self-not-death”, “others-life” or “others-death” biases). Existing research using the IAT has assumed that normative participants’ responses were driven by self-life biases (Harrison et al., 2014). However, this is difficult to directly assess due to the relative nature of biases on the IAT (Nosek et al., 2005), as discussed in Chapter 1. The results on the death-identity IRAP therefore support the assumption made within several previous studies that normative participants demonstrate specific implicit self-life attitudes. Results therefore underscore the IRAP’s utility in examining specific classes of relational responding through its four procedurally-independent trial-types.

In addition, this study took the novel step of examining implicit evaluations of death via a second IRAP. While the expected “life-positive” and “life-not-negative” effects were found, performances on the remaining trial-types were less intuitive. Specifically, the normative sample demonstrated a “death-positive” effect, while also demonstrating no significant “death-negative” effect. A number of *post hoc* explanations were proffered in order to attempt to account for such findings, including the unspecified nature of whose death was being referred to within the death-evaluation IRAP (e.g., self, a loved one, or a stranger); and the questionable salience of mortality within a sample of young, healthy students.

In order to attempt to develop a method by which such fine-grained comparisons might be manipulated within the IRAP (e.g., whose death is being referred to), the two experiments contained in Chapter 3 represent attempts to provide “proof of concept” that the contents of a given trial-type on the IRAP influences behaviour on others. In doing so, this chapter attempts to develop such “contrast category manipulations” as a generic experimental methodology to aid the refinement of IRAP stimulus sets in order to meet our analytic goals. This approach therefore built on comparable manipulations of the IAT’s stimulus categories

(e.g., Ostafin & Palfai, 2006; Robinson et al., 2005; Swanson et al., 2001), albeit with different analytic goals. Specifically, previous research on the IAT's contrast category has commonly been conducted from the perspective of problematizing the need for a contrast category, whereas the current research sought to clarify whether the contrast functions as a source of contextual control within the IRAP.

Based on the results of a conceptually similar study using the IAT by Karpinski (2004), Experiment 1 did this within the domain of self-evaluation. Participants completed two IRAPs that both contained the categories "self", "positive", and "negative", but differed by whether "self" was contrasted with a positively or negatively valenced other (i.e., "Hitler" or "Santa"). Results were inconclusive: a difference was found between the "self-negative" biases between the conditions; however equivalent "Santa-positive" and "Hitler-positive" biases were also found. As the evaluations of the contrast category did not differ between IRAPs, the contrast category effect on the "self-negative" trial-type could not be attributed to this manipulation with any confidence.

Experiment 2 attempted to address the methodological shortcomings of the previous study. First, in contrast to Experiment 1, a between-groups design was employed so as to eliminate the possibility that performance on one IRAP influenced performance on the other. Second, the target domain was changed to the implicit dehumanisation of women, which was again selected as a "good candidate" domain based on previously published empirical and theoretical research that suggested women are differentially humanised or objectified based on context (Glick et al., 2004; Rudman & Glick, 2001). Two IRAPs were created based on the stimulus categories employed by Rudman and Mescher (2012), which both contained the categories "women", "objects", and "human", but which differed by whether "women" were contrasted with "men" (i.e., compared along the dimension of gender) or "everyday objects" (i.e., compared along the dimension of agency). In this case, large and significant differences

were found between both the “men” versus “everyday items” trial-types, which differed between IRAPs, and also the “women” trial-types, which were held constant. Women were therefore differentially humanised or objectified based on the context in which they were presented, suggesting that, in some cases at least, the contrast category serves to specify the functional dimension of comparison within the IRAP.

Having demonstrated the viability of such contrast category manipulations in the previous chapter, Chapter 4 then applied the contrast category manipulation method to the interrogation of implicit death-evaluations in normative participants. In doing so, we sought to explore some of the explanations proffered for the unexpected effects found in Chapter 2 by differentially emphasising (a) the self and (b) the nature of evaluation between IRAP. Specifically, across four experiments, the “my death-negative” trial-type was kept static, and the contents of the other three were systematically manipulated. A first manipulation of the label stimuli involved contrasting “my death” with either “my life” (i.e., emphasising the life/death distinction) or “other peoples’ deaths” (i.e., emphasising the self/others distinction). A second manipulation of the target stimuli involved contrasting the “negative” category (which specified positive punishment such as “distressing” and “painful”) with positive reinforcement (e.g., “exciting”) or negative reinforcement (e.g., “calming”). A 2×2 post hoc comparisons design was therefore employed across the four experiments.

In addition to the contrast category manipulations, these studies also attempted to address other possible explanation of the unexpected effects found in Chapter 2. Specifically, it was noted that young healthy students may not be particularly aware of or threatened by their mortality, and this may have contributed to the presence of death–positive effects and absence of death–negative effects on the IRAP. In order to examine this, all four experiments in Chapter 4 also manipulated the salience of death across time points. Specifically,

participants completed a variant of the death-evaluation IRAP before and after a mortality salience induction.

Results demonstrated that “my death–positive” effects were generally found across the four experiments, and furthermore these effects were unaffected by the mortality salience induction. Participants also demonstrated relatively weak effects regarding the positivity or negativity when responding to other peoples’ deaths, but evaluated their own life positively. Of central importance were the *post hoc* analyses, which focused on the “my death–negative” trial-type. Critically, despite the fact that the stimuli presented in this trial-type were identical across all four tasks, conditions were differentially sensitive to the mortality salience induction. Specifically, when the “negative” category (i.e., positive punishment) was contrasted with positive reinforcement, significant changes were found between the baseline and follow-up IRAPs. In contrast, no differences were found between time points when the contrast category instead specified negative reinforcement. Furthermore, the manipulation of the role of self (i.e., “my life” vs. “others’ deaths” contrasts) was also found to not influence results. In summary, a combination of contrast category manipulations and mortality salience inductions were employed in order to select, by systematic comparison, the stimulus set that both produced strong effects in normative participants and showed greatest sensitivity to mortality salience (i.e., that employed in Chapter 4 Experiment 1, which included the categories “my life”, “my death”, positive reinforcement, and negative reinforcement). This IRAP stimulus set was therefore selected for use in the final study, on the basis that it was most likely to be able to uncover differences in implicit evaluations of death between individuals with and without recent suicidal ideation, should they exist.

Finally, Chapter 5 examined whether implicit evaluations of death differed between normative university students and psychiatric patients reporting suicidal behaviours (e.g., current ideation, history of planning, and/or attempts). Two implicit measures were compared,

one of which referred to life and death without a reference to self (i.e., “life” and “death”), and the other of which contained a reference to self (i.e., “my life” and “my death”). Results demonstrated that only self-focused evaluations of death on the IRAP reliably distinguished the groups. Such results are consistent with two aspects of existing theories of suicidality. First, only results on the IRAP that contained a reference to self were predictive of group membership. This is consistent with definitions of suicidal behaviour as involving a self-focused wish to die (Silverman, 2006; Silverman et al., 2007b). Second, suicidal ideation was associated with a specific implicit bias: the personal death IRAP’s “my death-negative” trial-type. Participants who rejected the negativity of their own death (i.e., produced positive *D* scores) were roughly 10 times more likely to be in the suicidal ideation group, and this cut-off score was also shown to have good sensitivity and adequate specificity. This result is consistent with the concept of suicide involving an acquired capacity for suicide through the loss of “fear of death” (Joiner, 2005; R. C. O’Connor, 2011). It should be noted that the nature of the effects that distinguish between groups also have theoretical importance, which will be discussed in detail in a separate section below.

It is also worth broadly comparing the results found between Chapter 4 and 5. Specifically, an identical IRAP stimulus set was used in both Chapter 4 (Experiment 1) and Chapter 5. In both cases, the normative group demonstrated no significant “my death–negative” effects. However, after mortality salience induction, the normative participants in Chapter 4 demonstrated no significant effects at baseline (i.e., *D* scores close to 0, similar to Chapter 2), but demonstrated the expected “my death–negative” effects (i.e., negative *D* scores) after mortality salience induction. In contrast, the suicidal ideation group in Chapter 5 demonstrated “my death–*not*-negative” effects (i.e., positive *D* scores). While the two groups are not directly comparable, given that one completed a mortality salience induction and one did not, this may suggest that the differences found between normative

participants and suicidal ideators are due to their history of suicidal behaviours specifically, rather than merely a greater salience of death generally within suicidality. That is, the IRAP may be sensitive to a history of contemplating suicide (where death is appetitive) rather than merely being aware of ones own mortality (which is typically aversive). Of course, the forgoing is somewhat speculative. Future work might therefore directly examine the effect of mortality salience inductions within suicidal samples in order to explore this more directly.

Two limitations of the study conducted in Chapter 5 must also be noted. First, the two groups were not necessarily comprised of mutually exclusive categories, as individuals in the normative groups were not assessed for a history of suicidal behaviours. This was due to an insurmountable limitation of the conditions for ethical approval of this research. However, should some member of this group have, unknown to us, possessed a history of suicidality, this is more likely to have artificially inflated the false positive rate than suppress it (and therefore suppress the specificity and negative likelihood ratio). Nonetheless, future research could draw stronger conclusions through the use of mutually exclusive categories. Second, it should be recognised that self-report measures of suicidal history were employed as the “gold standard” against which to compare the utility of the death-evaluation IRAPs, despite the fact that this body of research is predicated on the flaws of such self-report measures. Of course, this limitation is somewhat difficult to avoid when employing a cohort comparison design, unless one has access to independently verifiable reports, such as past or future records of hospital admissions, or coroner’s reports. As will be discussed below, the ultimate test of a measure’s utility within suicidality necessarily comes from the use of prospective designs.

The results contained in the current body of work provide initial evidence for the construct validity of implicit death-evaluations on the IRAP, via the analogue studies contained in Chapter 4, and more directly from the cohort comparison study in Chapter 5. As such, death-evaluations and the IRAP may represent a “good candidate” for future

prospective studies that attempt to predict suicidal behaviours prospectively. Specifically, given the low incidence rate of actual suicide attempts and the difficulty in differentiating ideators from those who go on to make an attempt, the most persuasive evidence for a given measure's utility must come from prospective studies (see Jacob Cohen, 1986; Glenn & Nock, 2014b; Klonsky & May, 2014). As such, future work might examine, either together or separately, the prospective criterion validity of (a) the IRAP, given the demonstrated utility of its four procedurally independent trial-types, and (b) death-evaluations on implicit measures.

Finally, in recommending that future studies examine the predictive utility of the IRAP within prospective designs, it is important that we note that there is no singular set of task parameters that defines the IRAP. That is, although all the IRAPs employed in the current body of research employed identical task parameters, we are keen to see that the IRAP's formal properties are not ossified across time, but instead are adapted to meet researchers' specified analytic goals. As such, future research that is concerned with the prediction of suicidal behaviours should examine the properties of the IRAP that increase or decrease its predictive validity within such forms of behaviour. Immediately pressing concerns for such research include the use of particular response options (i.e., true vs. false or same vs. different, which may represent different forms of relational cues); the presence, absence or format of the responding rules (e.g., whether these serve to orientate participants towards certain trial-types, and whether this increases or decreases the utility of such performances); the rate of switching between blocks (e.g., whether predictive validity is driven by behavioural momentum within blocks or the ability to switch responding patterns between in a flexible manner), and the overall task length (e.g., whether the utility of effects on the task is driven by either the rate of acquisition of the patterns of responding, or increasingly stable patterns of responding across blocks). In any case, future research into these intriguing possibilities certainly seems warranted.

## 6.2 Issues raised

### Theories of suicide and theories of learning

As discussed in the introduction, leading contemporary theories of suicidal behaviour such as the IPT and IMV emerged from the diathesis-stress paradigm. However, both theories self-admittedly place stronger emphasis on environmental stressors than on biological diathesis. This is particularly the case where each attempts to account for an individual's progression from suicidal ideation to actual attempts, which is proposed to be moderated by an acquired capability for suicide. In their descriptions of this acquired capacity, both theories acknowledge that suicidal behaviours are an acquired behavioural repertoire that is learned within an individual's lifespan, and seek to understand what environmental events give rise to these behaviours. As such, both theories have a key learning psychology component to them.

Specifically, the IPT argues that the acquired capability for suicide involves respondent-type learning. Specifically, that individuals habituate to the initially aversive functions of death and bodily harm through repeated exposure to painful and life-threatening situations. The IMV additionally emphasises the role of observational learning, insofar as individuals can acquire the capability for suicide through exposure to others' self harm. In the broad sense, then, both theories appeal to learning processes in order to explain suicidal behaviours.

However, neither theory provides an in-depth application of learning principles to suicidal behaviour. Many behaviourally orientated researchers would argue that any lack of precision here arguably decreases our capacity for intervention (e.g., Hayes & Brownstein, 1986). However, this explanatory gap between theories of suicide and basic learning psychology is likely for good reason: traditional respondent and operant learning principles have difficulty accounting for suicidal behaviour due to its definition. As noted in the introduction (Section 1.1), the majority of definitions of suicide require that it be an

“intentional” or “instrumental” behaviour, where its explanation lies in the future (Silverman, 2006). Hjelmeland and Knizek refer to this intentionality as “the subject trying to achieve something by the act” (1999, p.278). In contrast, theories of learning by definition place the explanation of behaviour in the past (i.e., interactions with the environment). Loosely speaking, an individual may be *motivated* to make a suicide attempt due to unbearable psychological distress, but the act of a suicide attempt is initiated with the *intention* escaping that aversive state (Hjelmeland & Knizek, 1999). A related issue for the application of a learning account to theories of suicide is the fact that the intended outcome (i.e., death, or loss of life) cannot be directly contacted by the individual, and therefore cannot serve to reinforce future behaviour. As such, any learning-based account of suicidal behaviour arguably must first account for the ability to learn via indirect experience and to engage in intentional or instrumental behaviour, and therefore go beyond operant and respondent learning (Hayes, 1992).

As discussed in the introduction (Section 1.4), Relational Frame Theory attempts to provide a general account of such indirect learning through the concept of derived relations (Hayes et al., 2001). To briefly reiterate, RFT’s foundational concept of derived relational responding refers to the arguably uniquely human ability to learn to relate stimulus A to stimulus B, and stimulus B to stimulus C, and then derive the relation between stimulus A and stimulus C (although see Hughes & Barnes-Holmes, 2014 for some important qualifiers). In this respect, RFT can therefore be said to be an account of the learning history that gives rise to phenomena that are commonly referred to at the cognitive level of analysis as inference, intention, and perspective taking (Hughes & Barnes-Holmes, 2015).

In principle, this ability to derive relations among events that have not been directly contacted by the individual can allow, in principle, for the application of a learning account to suicidal behaviour (Hayes, 1992). Broadly, this could be done in terms of more complex

instances of derived relational responding such as the construction of future (Y. Barnes-Holmes, McHugh, & Barnes-Holmes, 2004) and transfer of escape or avoidance functions to events that have not been directly experienced (Dymond, Roche, Forsyth, Whelan, & Rhoden, 2007). Loosely speaking, the concept of derived relations may be the “missing link” that theories of suicide require to ground themselves in historic environmental causes. Of course, such theoretical work is not the primary purpose of the current body of work. However, it’s perceived feasibility did provide an important motivation. Future theoretical and empirical work will be required to assess the utility of appealing to derived relations within theories of suicide.

It should be noted, however, that more recent developments in the RFT literature argue that existing work may have created a false dichotomy between trained versus derived relations (Hughes et al., 2012). This refers to the idea that the first time that an individual derives a relation between given stimuli may be importantly different from the  $n^{\text{th}}$  time this is done, and that there may be one or more properties of relational responding that may be useful in parsing this behaviour (D. Barnes-Holmes, Barnes-Holmes, Hussey, & Luciano, 2015). Broadly speaking, there may for example be important differences between an individual who has a shallow history of emitting a given response (e.g., a passing thought that “I’d rather die than tolerate this”) versus an individual who has a lengthy history (e.g., has ruminated on this same thought for years). This notion of a continuum from relatively unpracticed, “extended and elaborated” relational responses versus highly practiced, “brief and immediate” relational responses forms the basis of interpretation of effects on the IRAP. Indeed, several notional properties of relational responding have been proposed but remain as yet untested (D. Barnes-Holmes et al., 2015; Hughes et al., 2012). Indeed, this difference between more brief and immediate over more extended and elaborated relational responding forms the basis of the RFT account of effects on the IRAP. Within the current context, the

key finding in Chapter 5 that individuals with suicidal ideation demonstrate stronger “my death-*not*-negative” effects than normative individuals suggests that individuals with suicidal ideation have a more well-established history of rejecting the negativity of their own death than have normative individuals. This learned rejection of negativity is consistent with the concept of an acquired capacity for suicide (and specifically the loss of fear of death). However, the appeal to derived relations here provides a far more precise account of the learning history that gives rise to these effects on the IRAP, whereas mere habitation or observational learning cannot (Hayes, 1992).

Of course, it is important to acknowledge, even if only briefly, that such a conceptualization of the historic environmental determinants of these behaviours runs parallel to, and not against, cognitive-level theorising. Although the cognitively orientated researchers might attribute behaviours within the IRAP to causal cognitive mechanisms (e.g., “the strength of associations in memory”: Greenwald et al., 2002), it is important to note that this represents an alternative and separate level of explanation to the one adopted in the current body of research; one functional-analytic and one cognitive (D. Barnes-Holmes & Hussey, 2016; De Houwer, 2011). Fruitful comparisons between these two levels of analysis are possible of course, but this was not the goal of the current body of work (see Hughes et al., 2012).

### **Linking the effects found on implicit measures with theories of suicide**

The results found within the current body of work underscore the notion that the IRAP’s four separate bias scores (i.e., its assessment of four individual classes of relational responding) provide greater conceptual clarity around its results, and allow for more precise conclusions to be drawn. Indeed, this appears to facilitate greater contact between IRAP effects and domain specific theories than is possible when using “relative” measures, such as the IAT (see Robinson et al., 2005). Specifically, in contrast with previous research using

implicit measures within suicidality, which has been largely a-theoretical, the current results can be linked with existing theory in two ways. First, the finding that suicidal ideation is associated with a specific rejection of the negativity of death is consistent with suicidality through the loss of fear of death, as postulated by both the Interpersonal Theory of suicide (IPT: Joiner, 2005; Van Orden et al., 2010) and the Integrated Motivational-Volitional Model (IMV: R. C. O'Connor, 2011). Second, performances on the “death-negative” trial-type were predictive of suicidality only when the stimuli contained a reference to the self (i.e., the personal death IRAP was predictive, whereas the abstract death IRAP was not). These results are consistent with the assertion that suicide involves a *self-focused* wish to die specifically rather than a general endorsement of death or suicide (O'Carroll et al., 1996; Silverman, 2006).

At this point, we would like to reemphasize that the current body of research is the first to examine such implicit evaluations of death. Furthermore, it should be noted that there are relatively few measures available that have attempted to assess the acquired capability for suicide (e.g., the Acquired Capability for Suicide Scale: Ribeiro et al., 2014b). Of those that do exist all rely on self-reports, to the best of the current author's knowledge. As such, the current results regarding death-evaluations on the IRAP appear to represent a novel contribution to the literature on suicidality, both in the use of objective behavioural measures and in drawing links between such measures and existing research.

Future research could attempt to elaborate these links in a number of other ways. For example, the IPT and IMV both posit specific “pathways” towards suicide. For example, a perceived inability to relieve intolerable psychological pain is seen as leading to suicidal ideation, whereas the development of suicidal intent and attempts are thought to be moderated by an acquired capability for suicide. As such, these theories would predict that specific implicit attitudes should be predictive of specific stages (i.e., ideation, intent,

attempts) along these pathways. Research on these pathways has been difficult, however, because existing work has frequently focused on the results of implicit measures in isolation. Future research could therefore examine the *interactions* among several measures of implicit attitudes (e.g., death-evaluative, death-identity, and self-escape) across these various stages. For example, the IPT and IMV appear to predict that self-escape effects should be predictive of ideation alone, whereas self-death and death-evaluation effects should be predictive of intent and attempts.

On a separate note, while the current research has largely focused on the results generated by implicit measures, as noted previously, the work has been conducted through the lens of Relational Frame Theory. As a result, both suicidal behaviours and the results generated by the implicit measures can be conceptualised within this single theoretical framework; for example, the relative strength of brief and immediate relational responses and derived avoidance behaviours would be seen as playing a critical role. This ability to encompass both types of behaviours under a single theoretical umbrella likely has the potential for broad benefits that should be explored in greater depth by future research. Specifically, as discussed in Chapter 1, existing theories of suicide assume the presence of behavioural repertoires (e.g., language, intentional behaviour, and the ability to construct death), but from a functional-analytic perspective one must account for these repertoires rather than simply assume them. In doing so, such accounts may then feed back into domain level theorising. For example, at a conceptual level, RFT allows for the specification of the learning pathway for the postulated acquired capacity for suicide, as well as the conceptualisation of suicidal behaviour more generally (i.e., through derived relations), both of which are difficult to explain using traditional operant principles alone (see Hayes, 1992). Future research should therefore attempt to elaborate the role of derived relational responding within suicidality at both conceptual and empirical levels, for example, through a more

detailed interrogation of rule-governed (e.g., “intentional”) behaviour in the emergence and maintenance of suicidality, and the acquisition of a fearlessness of death through derived relations (e.g., indirect experience through language).

### **Unexpected positivity effects on the IRAP**

As noted in the results summary, somewhat unexpected positive implicit evaluations were found on the IRAP in several experiments (i.e., “death–positive” effects in Chapters 3, 5, and 6; a “Hitler–positive” effect in Chapter 4 Experiment 1). Furthermore, positivity biases towards death were shown to be unaffected by mortality salience inductions (in contrast to “death–negative” effects), and did not appear to be related to self-reported evaluations of death on either psychometric measures (e.g., the Death Anxiety Scale or Belief in the Afterlife Scale) or from content analysis of responses on open-ended questions (e.g., within the Mortality Attitudes Personality Survey). Given their somewhat counterintuitive nature, it is worth considering the possible determinants of these effects, especially in light of a recent paper which has suggested that they represent a flaw in the IRAP itself (i.e., O’Shea, Watson, & Brown, 2015).

It is worth noting that similar positivity effects have been observed and noted in previous research using the IRAP (e.g., D. Barnes-Holmes, Murphy, Barnes-Holmes, & Stewart, 2010; Hughes & Barnes-Holmes, 2011; Power, Barnes-Holmes, Barnes-Holmes, & Stewart, 2009). In addition, comparable valence asymmetries have also been found on other implicit measures, such as the Brief IAT (Nosek, Bar-Anan, Sriram, Axt, & Greenwald, 2014). In line with the conclusions of previous IRAP research, one explanation might appeal to a general positivity bias in natural language to which the IRAP may be sensitive (see D. Barnes-Holmes, Murphy, et al., 2010, pp. 75–76). That is, all things being equal, in natural language interactions speakers tend to confirm rather than deny, and to emphasize the positive over the negative. Evidence for this suggestion can be found, for example, in the fact

that many languages possess a larger number of positively-valenced words than negatively-valenced ones (Dodds et al., 2015). Given that the IRAP was specifically designed to capture differential probabilities (i.e., biases) in patterns of relational responding that are found in natural language (D. Barnes-Holmes, Hayden, Barnes-Holmes, & Stewart, 2008), it seems reasonable to assume that such biases may also be reflected in IRAP performances.

Of course, the forgoing point is not an attempt to merely dismiss positivity effects as irrelevant. Rather, they may serve to highlight unresolved questions and guide future research. Specifically, if the IRAP is to be considered a measure of the response patterns found in natural language, it might even be important to capture such positivity biases, especially if the variance they create helps to optimize IRAP's predictive validity. Recall, for example, that the results from Chapters 3 and 4 indicate that performance on one trial-type is influenced by the stimuli presented in the other trial-types. As such, it is possible (even likely) that the interaction or response competition between general positivity biases and specific response biases towards the positive and negative aspects of life and death were jointly responsible for producing the observed level of predictive validity on the "death-negative" trial-type in Chapter 5.

In conclusion, it is important to note that, while such positivity effects appear to affect the mean bias scores on particular trial-types, they do not appear to undermine the IRAP's predictive validity (see Vahey et al., 2015 for meta analysis of clinically relevant criterion effects). Indeed, this is the case in the current results: while "death-positive" effects were demonstrated, performance on the "death-negative" trial-type were nonetheless predictive of a history of suicidal behaviours. On a related point, the current results therefore reemphasize the importance of conducting analyses at the trial-type level (rather than merely overall *D* scores) in order to better understand the psychological processes that may be operating through the measure.

### **Broader utility of contrast category manipulations**

Chapters 3 and 4, respectively, developed and employed the contrast category manipulation method. To recap, this method involves holding the trial-types of interest static across IRAPs and manipulating the contents of the remaining trial-types in a theoretically meaningful manner. For example, holding “my death” constant across IRAPs, and manipulating whether the contrast category refers to “my life” (i.e., to emphasize the life/death distinction) or “others’ deaths” (i.e., to emphasise the role of self). The current results indicated that the contrast category manipulation was an effective method. As such, it may have general utility within research using the IRAP in other basic or applied domains in order to elucidate the extent to which the comparison category serves to guide effects on the task in a theoretically meaningful way.

It must be recognised that this method was directly inspired by comparable research using the IAT (e.g., Houben & Wiers, 2006; Karpinski, 2004; Robinson et al., 2005; Swanson et al., 2001). However, a key difference exists: research using the IAT has – either tacitly or explicitly – treated the requirement of a contrast category as a procedural nuisance that serves to limit the ability to interpret results (e.g., Robinson et al., 2005). In contrast, the procedural independence of the IRAP’s four trial-types allows us to fundamentally alter this conceptual narrative. Rather than a procedural irritation, the selection (or manipulation) of the contrast category can be seen as a source of contextual control within the task, thereby facilitating increasingly fine-grained functional analyses.

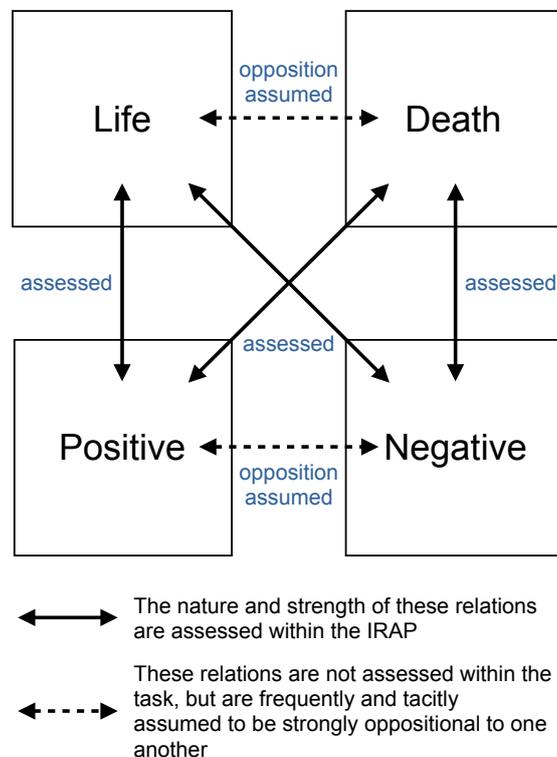
At this point, it is worth recalling that the above section on unexpected positivity biases emphasised the need to analyse the results of the four trial-types separately, given that they reflect discrete classes of relational responding. Considering the results of the contrast category manipulation studies, this point should be elaborated as follows: while it appears to be both useful and important to analyse effects on the four trial-types separately, it should be

noted that the contents of a given trial-type appears to influence responding on the others. As such, as a cautionary note, future research should therefore ensure that information regarding the contents of, and performance on, all trial-types are reported, even when only one trial-type is considered to be of theoretical importance (see Nicholson et al., 2013).

In summary, the results generated through the use of contrast category manipulations within the current body of research serve to emphasise that, although the IRAP's four trial-types are procedurally non-relative (i.e., separated), behaviour within the task is not a-contextual. On a related point, it is worth mentioning that, having observed the influence of the contrast category, one could seriously question the logic of attempting to develop a procedure (e.g., to measure "attitudes", "associative strengths", or "relational responding") that is a-contextual or "absolute" in some sense (e.g., free of or unaffected by a contrast category: see O'Shea, Watson, & Brown, 2015). That is, we would argue that all measures of implicit attitudes (i.e., the relative strength of brief and immediate relational responding) are moderated by contextual variables. As such, it may be more useful to embrace and harness (rather than problematise) such sources of contextual control in the service of meeting analytic goals.

Whereas the current research explored whether the contrast category influences responding on the trial-type(s) of interest, future research might go further by examining what moderates this influence. Furthermore, such work could attempt to describe in functional terms the desirable properties of a stimulus set in the service of a given analytic goal (e.g., the maximisation of predictive validity). In order to elaborate on what is meant here, recall that the IRAP assesses both the *nature* and *strength* of relational responding between the two label categories (e.g., life and death) and two target categories (e.g., positive and negative). That is, for each trial-type (e.g., life-positive, life-negative, death-positive, and death-negative), the IRAP provides a metric that describes whether participants were faster to

respond using one response option or the other (e.g., positive vs. negative *D* scores represent faster responding with “True” relative to “False”) as well as quantifying the relative strength of the bias (e.g., the absolute strength of the *D* score). In contrast, the two label and the two target categories are each commonly assumed by researchers to be “obvious opposites” (e.g., life vs. death, positive vs. negative; Robinson et al., 2005, p. 208). However, the nature and strength of the relations among the “label–label” and “target–target” stimulus categories is not formally assessed by the IRAP<sup>7</sup>. This distinction between the relations that are assessed versus assumed by the IRAP is illustrated in Figure 6.1.



*Figure 6.1.* Stimulus relations within the IRAP that are assessed versus those that are commonly assumed by researchers, but not assessed by the task.

In light of unexpected results on the IRAP, future researchers might therefore wish to consider whether the relations between such stimulus classes are indeed strongly oppositional, as may have been assumed. Furthermore, such researchers, in line with their analytic goals,

might examine whether relational responding on the trial-types is influenced by the nature or strength of the “label-label” and “target-target” relations (e.g., by the broader relational coherence within the task). There are some indications in the current results that this may be the case. For example, it should be noted that strong effects were demonstrated on IRAPs whose stimulus categories, when taken at face value, appeared to reflect “obvious opposites” (e.g., Chapter 4, Experiment 1: “my life” vs. “my death” and “enjoyable” vs. “painful”). Whereas, in contrast, no significant effects were found on any trial-type on other IRAPs whose stimulus categories were less “obviously opposite” (e.g., Chapter 4, Experiment 3: “others’ deaths” vs. “my death” and “calming” vs. “painful”). Of course, the foregoing is somewhat speculative. However it may nonetheless serve to orientate our analyses to the broader behavioural processes that serve to generate effects on the IRAP (and indeed other implicit measures).

### **6.3 Broader implications and considerations**

The previous section on “linking the effects found on implicit measures with theories of suicide” (above) made a number of suggestions for future empirical studies on the prospective prediction of behaviour in the context of evaluating theories of suicide. Beyond these, it is also useful to consider the broader real-world utility of implicit measures within the prediction of self-harmful behaviours.

As discussed in the introduction, suicidal behaviour has proved to be very difficult to predict prospectively, and both self-report measures and clinician judgement are extremely limited in this regard (see Section 1.2). For this reason, the encouraging prospective prediction rates being generated by the self-death IAT (i.e., Nock et al., 2010; Randall et al., 2013, see Section 1.3 for systematic review) are undeniably interesting for researchers, clinicians, and policy makers whose goal is predict and influence suicidal behaviours within applied contexts (Glenn & Nock, 2014b). At a practical level, the simplicity of completing

the IAT appears to make it suitable even in a demanding applied context such as an Accident and Emergency waiting room, and therefore worthy of study along with more invasive and intensive assessments (e.g., cerebrospinal fluid biomarkers Lindqvist et al., 2011).

While the current body of work did not add to this prospective prediction literature, it hopes to provide a number of contributions to the broader literature that might serve to guide future research. These include contributions at the level of (1) theory (i.e., discussion around the need for closer scrutiny of the adequacy of the learning psychology underpinning of theories of suicide; potential for closer consideration of derived relations; and how such an approach can serve to link the results generated by implicit measures with the theories of suicide), (2) review (i.e., the first systematic review of implicit measures of within suicidal behaviour), (3) methodology (i.e., propositionally based measures, and use of contrast category manipulations to better target the relations of interest), and (4) empirical (i.e., demonstration of a implicit rejection of negativity around personal death within suicidal individuals).

It is equally important to acknowledge several caveats to their real world potential. First, reaction-times have inherently low reliability (Ratcliff, 1993). As such, all reaction-time based implicit measures and their scoring procedures have relatively low reliability, which limits the ability to make individual-level predictions. Of course, efforts have been made to improve this situation. If we consider scoring metrics in isolation, it is useful to note that several researchers argue that the  $D$  algorithm and its variants that are used in the vast majority of IAT and IRAP research (including the current body of work) represents an inappropriate treatment of reaction time data, given  $D$ 's parametric assumptions and the highly non-normal distribution of reaction times (see De Schryver, 2013; Erceg-Hurn & Mirosevich, 2008). In this respect, recent advances in probabilistic and semi-parametric scoring algorithms appear to represent superior alternatives to  $D$ , insofar as they are more

resistant to outliers and scaling effects (De Schryver, 2013; Sriram, Greenwald, & Nosek, 2010; Thas, Neve, Clement, & Ottoy, 2012). Nonetheless, even these alternatives appear to provide only an incremental improvement over existing practices, and it is possible that real world utility of implicit measures may be limited by the reliability of reaction times within procedures such as the IAT and IRAP.

Second, it would be appropriate to be wary of the fallacy of affirming the consequent in advocating for the wider use of implicit measures within applied contexts. Certainly, existing risk assessment methods are of limited utility, and alternatives such as implicit measures exist, but it does not necessarily follow that these alternatives should therefore be adopted. One point made in the conclusion of the systematic review conducted in Section 1.3 appears to be particularly relevant here: the existing research on the predictive utility of the self-death IAT has always been on subsequent attempts made after a hospital admission for a given attempt. As such, there may be important contextual factors that determine effects on the task that may not generalise to wider use as a predictive tool. More specifically, existing research may have been tacitly assessing implicit attitudes to death and suicide in the context of having *survived* an attempt, rather than generally. Given the malleability of implicit measures to current distress state demonstrated in Chapter 4 (see also Tang et al., 2013) this possibility cannot be ignored. As such, future research may need to explore the predictive utility of implicit measures when assessment is done in a screening context (i.e., prior to any known potential attempts) before their utility as screening tools for attempts, rather than *repeat* attempts, can be known. More generally, any policy decision that would ask or require clinicians to consider the results of a new test when making decisions would require strong evidence of the superior utility of such tasks and policies over the existing state of affairs.

Third and finally, it is important to note that, despite the ubiquity of short-term suicide risk-management assessments in psychiatric practice, recent evidence suggests that such risk

classification systems are not particularly useful in predicting or preventing suicide (Ryan et al., 2015; Ryan & Large, 2013). In addition to this, it should be recognised that that involuntary psychiatric admission is itself a strong risk factor for subsequent suicide attempts (Large, Sharma, Cannon, Ryan, & Nielssen, 2011). As such, the question of the potential predictive utility of implicit measures within suicidal behaviour should not be divorced from the question of what decisions are to be made on the basis of such an assessment (Glenn & Nock, 2014b). To put this another way, it may be tempting to assume that psychological or psychiatric intervention of some form can lower the likelihood of suicide attempts, and that only the decision making process around when to intervene is in question. However, the evidence base for psychosocial interventions for suicidal behaviours specifically, and especially short-term interventions, is relatively lower than the efficacy for broader diagnostic categories more generally (see Tarrier et al., 2008 for systematic review and meta analysis).

#### **6.4 Conclusions**

The current thesis set out to explore the IRAP as a way to target implicit evaluations of death (i.e., the relative strength of brief and immediate relational responses) within suicidality, and in doing so investigate a number of moderating variables (e.g., the role of the IRAP contrast category; the salience of mortality). It culminated with the demonstration of initial evidence for the construct validity of implicit death-evaluations on the IRAP. Results showed that a history of recent suicidal ideation is characterized by a specific rejection of the negativity of death; a finding that is consistent in several respects with the concept of an acquired capacity for suicide, as postulated by contemporary theories of suicide (e.g., IPT, IMV). Future research could therefore examine the prospective validity of such IRAP effects. However, there is also large scope to examine, in a more theoretically fine-grained way, the degree to which the behavioural processes that constitute language (i.e., arbitrarily applicable relational responding) are involved in the development and maintenance of suicidal

behaviours. The work presented in the current thesis is therefore presented as the first step in developing a functional-analytic program of research on suicidal behaviours that is rooted strongly in Relational Frame Theory.

## References

- Baker, T. B., McFall, R. M., & Shoham, V. (2008). Current status and future prospects of clinical psychology toward a scientifically principled approach to mental and behavioral health care. *Psychological Science in the Public Interest*, *9*(2), 67–103.
- Bar-Anan, Y., Nosek, B. A., & Vianello, M. (2009). The Sorting Paired Features Task. *Experimental Psychology*, *56*(5), 329–343. <http://doi.org/10.1027/1618-3169.56.5.329>
- Barnes-Holmes, D., Barnes-Holmes, Y., & Hussey, I. (in press). Relational Frame Theory: Finding its historical and intellectual roots and reflecting upon its future development. In S. C. Hayes, D. Barnes-Holmes, R. D. Zettle, & A. Biglan (Eds.), *Handbook of contextual behavioral science*. New York, NY: Wiley-Blackwell.
- Barnes-Holmes, D., Barnes-Holmes, Y., Hussey, I., & Luciano, C. (2015). Relational Frame Theory: Finding its historical and intellectual roots and reflecting upon its future development. In R. D. Zettle, S. C. Hayes, D. Barnes-Holmes, & A. Biglan (Eds.), *The Wiley Handbook of Contextual Behavioral Science* (pp. 115–128). John Wiley & Sons, Ltd. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1002/9781118489857.ch8/summary>
- Barnes-Holmes, D., Barnes-Holmes, Y., Power, P., Hayden, E., Milne, R., & Stewart, I. (2006). Do you really know what you believe? Developing the Implicit Relational Assessment Procedure (IRAP) as a direct measure of implicit beliefs. *The Irish Psychologist*, *32*(7), 169–177.
- Barnes-Holmes, D., Barnes-Holmes, Y., Stewart, I., & Boles, S. (2010). A sketch of the Implicit Relational Assessment Procedure (IRAP) and the Relational Elaboration and Coherence (REC) model. *The Psychological Record*, *60*, 527–542.
- 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(4), 497–516.
- Barnes-Holmes, D., & Hussey, I. (2012). Implicit Relational Assessment Procedure (Version 2012). Maynooth, Ireland. Retrieved from IRAPresearch.org
- Barnes-Holmes, D., & Hussey, I. (2015). The functional-cognitive meta-theoretical framework: Reflections, possible clarifications and how to move forward. *International Journal of Psychology*, Advance online publication. <http://doi.org/10.1002/ijop.12166>
- Barnes-Holmes, D., & Hussey, I. (2016). The functional-cognitive meta-theoretical framework: Reflections, possible clarifications and how to move forward. *International Journal of Psychology*, 51(1), 50–57. <http://doi.org/10.1002/ijop.12166>
- Barnes-Holmes, D., Keane, J., Barnes-Holmes, Y., & Smeets, P. (2000). A derived transfer of emotive functions as a means of establishing differential preferences for soft drinks. *The Psychological Record*, 50(3), 493–511.
- Barnes-Holmes, D., Murphy, A., Barnes-Holmes, Y., & Stewart, I. (2010). 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, 57–66.
- Barnes-Holmes, D., Regan, D., Barnes-Holmes, Y., Commins, S., Walsh, D., Stewart, I., ... Dymond, S. (2005). Relating Derived Relations as a Model of Analogical Reasoning: Reaction Times and Event-Related Potentials. *Journal of the Experimental Analysis of Behavior*, 84(3), 435–451. <http://doi.org/10.1901/jeab.2005.79-04>
- Barnes-Holmes, Y., McHugh, L., & Barnes-Holmes, D. (2004). A Relational Frame Account of the Development of Complex Cognitive Phenomena: Perspective-taking False

- Belief Understanding and Deception. *International Journal of Psychology and Psychological Therapy*, 4(2), 303–324.
- Barzilay, S., & Apter, A. (2014). Psychological Models of Suicide. *Archives of Suicide Research*, 18(4), 295–312. <http://doi.org/10.1080/13811118.2013.824825>
- Bassett, J. F., & Dabbs, J. M. J. (2003). Evaluating explicit and implicit death attitudes in funeral and university students. *Mortality*, 8(4), 352–371.  
<http://doi.org/10.1080/13576270310001604022>
- Batterham, P. J., Ftanou, M., Pirkis, J., Brewer, J. L., Mackinnon, A. J., Beautrais, A., ... Christensen, H. (2014). A systematic review and evaluation of measures for suicidal ideation and behaviors in population-based research. *Psychological Assessment*, Advance online publication. <http://doi.org/Advance online publication>. <http://dx.doi.org/10.1037/pas0000053>
- Baumeister, R. F. (1990). Suicide as escape from self. *Psychological Review*, 97(1), 90–113.  
<http://doi.org/10.1037/0033-295X.97.1.90>
- Beck, A. T., Brown, G. K., Steer, R. A., Dahlsgaard, K. K., & Grisham, J. R. (1999). Suicide ideation at its worst point: a predictor of eventual suicide in psychiatric outpatients. *Suicide & Life-Threatening Behavior*, 29(1), 1–9.
- Beck, A. T., Steer, R. A., Kovacs, M., & Garrison, B. (1985). Hopelessness and eventual suicide: a 10-year prospective study of patients hospitalized with suicidal ideation. *American Journal of Psychiatry*, 142(5), 559–563.  
<http://doi.org/10.1176/ajp.142.5.559>
- Beck, A. T., Steer, R. A., & Ranieri, W. F. (1988). Scale for Suicide Ideation: psychometric properties of a self-report version. *Journal of Clinical Psychology*, 44(4), 499–505.

- Beck, A. T., Weissman, A., Lester, D., & Trexler, L. (1974). The measurement of pessimism: the hopelessness scale. *Journal of Consulting and Clinical Psychology, 42*(6), 861–865.
- Bentley, K. H., Nock, M. K., & Barlow, D. H. (2014). The Four-Function Model of nonsuicidal self-injury: Key directions for future research. *Clinical Psychological Science, 2*(5), 638–656.
- Berman, A. L., & Silverman, M. M. (2014). Suicide risk assessment and risk formulation part II: Suicide risk formulation and the determination of levels of risk. *Suicide and Life-Threatening Behavior, 44*(4), 432–443. <http://doi.org/10.1111/sltb.12067>
- Bhaskaran, J., Wang, Y., Roos, L., Sareen, J., Skakum, K., & Bolton, J. M. (2014). Method of suicide attempt and reaction to survival as predictors of repeat suicide attempts: a longitudinal analysis. *The Journal of Clinical Psychiatry, 75*(8), e802–808. <http://doi.org/10.4088/JCP.13m08879>
- Bond, F. W., Hayes, S. C., Baer, R. A., Carpenter, K. M., Guenole, N., Orcutt, H. K., ... Zettle, R. D. (2011). Preliminary psychometric properties of the Acceptance and Action Questionnaire-II: a revised measure of psychological inflexibility and experiential avoidance. *Behavior Therapy, 42*(4), 676–688. <http://doi.org/10.1016/j.beth.2011.03.007>
- Booth, R., Keogh, K., Doyle, J., & Owens, T. (2014). Living Through Distress: A skills training group for reducing deliberate self-harm. *Behavioural and Cognitive Psychotherapy, 42*(02), 156–165. <http://doi.org/10.1017/S1352465812001002>
- Bosson, J. K., Swann, W. B., & Pennebaker, J. W. (2000). Stalking the perfect measure of implicit self-esteem: the blind men and the elephant revisited? *Journal of Personality and Social Psychology, 79*(4), 631–643.

- Bradley, M. M., & Lang, P. J. (1999). *Affective norms for English words (ANEW): Stimuli, instruction manual and affective ratings. Technical report C-1*. Gainesville, FL: The Center for Research in Psychophysiology, University of Florida.
- Brown, G. K., Beck, A. T., Steer, R. A., & Grisham, J. R. (2000). Risk factors for suicide in psychiatric outpatients: a 20-year prospective study. *Journal of Consulting and Clinical Psychology, 68*(3), 371.
- Brown, G. K., Steer, R. A., Henriques, G. R., & Beck, A. T. (2005). The internal struggle between the wish to die and the wish to live: A risk factor for suicide. *American Journal of Psychiatry, 162*(10), 1977–1979.  
<http://doi.org/10.1176/appi.ajp.162.10.1977>
- Burke, B. L., Martens, A., & Faucher, E. H. (2010). Two decades of Terror Management Theory: A meta-analysis of mortality salience research. *Personality and Social Psychology Review, 14*(2), 155–195. <http://doi.org/10.1177/1088868309352321>
- Busch, K. A., Fawcett, J., & Jacobs, D. G. (2003). Clinical correlates of inpatient suicide. *Journal of Clinical Psychiatry, 64*(1), 14–19. <http://doi.org/10.4088/JCP.v64n0105>
- Catania, A. C. (1973). The concept of the operant in the analysis of behavior. *Behaviorism, 1*, 103–116.
- Catania, A. C. (1998). *Learning* (4th edition). Upper Saddle River, NJ: Prentice-Hall.
- Ceci, S. J., & Bruck, M. (2009). Do IRBs pass the minimal harm test? *Perspectives on Psychological Science, 4*(1), 28–29.
- Central Statistics Office. (2014). *Measuring Ireland's Progress 2013*. Dublin, Ireland: Central Statistics Office. Retrieved from  
<http://www.cso.ie/en/releasesandpublications/ep/p-mip/measuringirelandsprogress2013/#.VTwY1a3BzGc>

- Cha, C. B., & Nock, M. K. (2009). *Assessing Suicidal Thoughts Without Asking: Development of a Suicide Implicit Association Test*. Presented at the Association for Psychological Science annual conference, Chicago, IL, USA.
- Chen, M., & Bargh, J. A. (1999). Consequences of automatic evaluation: Immediate behavioral predispositions to approach or avoid the stimulus. *Personality and Social Psychology Bulletin*, 25(2), 215–224. <http://doi.org/10.1177/0146167299025002007>
- Chiesa, M. (1994). *Radical behaviorism: The philosophy and the science* (Vol. xiii). Boston, MA, US: Authors Cooperative.
- Claassen, C. A., Harvilchuck-Laurenson, J. D., & Fawcett, J. (2014). Prognostic models to detect and monitor the near-term risk of suicide. *American Journal of Preventive Medicine*, 47(3), S181–S185. <http://doi.org/10.1016/j.amepre.2014.06.003>
- Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.
- Cohen, J. (1986). Statistical approaches to suicidal risk factor analysis. *Annals of the New York Academy of Sciences*, 487(1), 34–41.
- Crane, C., Barnhofer, T., Duggan, D. S., Eames, C., Hepburn, S., Shah, D., & Williams, J. M. G. (2014). Comfort from suicidal cognition in recurrently depressed patients. *Journal of Affective Disorders*, 155, 241–246. <http://doi.org/10.1016/j.jad.2013.11.006>
- Creemers, D. H. M., Scholte, R. H. J., Engels, R. C. M. E., Prinstein, M. J., & Wiers, R. W. (2013). Damaged self-esteem is associated with internalizing problems. *Frontiers in Psychology*, 4. <http://doi.org/10.3389/fpsyg.2013.00152>
- Crowell, S. E., Beauchaine, T. P., & Linehan, M. M. (2009). A Biosocial Developmental Model of Borderline Personality: Elaborating and Extending Linehan's Theory. *Psychological Bulletin*, 135(3), 495–510. <http://doi.org/10.1037/a0015616>

- Dawson, D. L., Barnes-Holmes, D., Gresswell, D. M., Hart, A. J., & Gore, N. J. (2009). Assessing the implicit beliefs of sexual offenders using the Implicit Relational Assessment Procedure: A first study. *Sexual Abuse: A Journal of Research and Treatment, 21*(1), 57–75. <http://doi.org/10.1177/1079063208326928>
- De Houwer, J. (2002). The Implicit Association Test as a tool for studying dysfunctional associations in psychopathology: strengths and limitations. *Journal of Behavior Therapy and Experimental Psychiatry, 33*(2), 115–133. [http://doi.org/10.1016/S0005-7916\(02\)00024-1](http://doi.org/10.1016/S0005-7916(02)00024-1)
- De Houwer, J. (2003). The extrinsic affective Simon task. *Experimental Psychology, 50*(2), 77–85. <http://doi.org/10.1026//1618-3169.50.2.77>
- De Houwer, J. (2006). What are implicit measures and why are we using them? In R. W. Wiers & A. W. Stacy (Eds.), *The handbook of implicit cognition and addiction* (pp. 11–28). Thousand Oaks, CA: Sage Publishers.
- De Houwer, J. (2011). Why the cognitive approach in psychology would profit from a functional approach and vice versa. *Perspectives on Psychological Science, 6*(2), 202–209. <http://doi.org/10.1177/1745691611400238>
- De Houwer, J., Barnes-Holmes, Y., & Barnes-Holmes, D. (2015). Riding the Waves: A Functional-Cognitive Perspective on the Relations Among Behavior Therapy, Cognitive Behavior Therapy, and Acceptance and Commitment Therapy. *International Journal of Psychology*, Advance online publication.
- De Houwer, J., Heider, N., Spruyt, A., Roets, A., & Hughes, S. (2015). The relational responding task: toward a new implicit measure of beliefs. *Cognition, 6*, 319. <http://doi.org/10.3389/fpsyg.2015.00319>

- De Houwer, J., & Moors, A. (2007). How to define and examine the implicitness of implicit measures. In N. Schwarz & B. Wittenbrink (Eds.), *Implicit measures of attitudes: Procedures and controversies* (pp. 179–194). London, UK: Guilford Press.
- De Houwer, J., Teige-Mocigemba, S., Spruyt, A., & Moors, A. (2009). Implicit measures: A normative analysis and review. *Psychological Bulletin*, *135*(3), 347–368.  
<http://doi.org/10.1037/a0014211>
- De Leo, D. (2002). Why are we not getting any closer to preventing suicide? *The British Journal of Psychiatry*, *181*(5), 372–374. <http://doi.org/10.1192/bjp.181.5.372>
- De Schryver, M. (2013). *The Probabilistic Index: A semiparametric scoring framework for the Implicit Association Test*. Retrieved from  
<http://psycnet.apa.org/journals/ccp/56/6/893/>
- DeVellis, R. F. (2012). *Scale Development: Theory and Applications*. SAGE Publications.
- Dickstein, D. P., Puzia, M. E., Cushman, G. K., Weissman, A. B., Wegbreit, E., Kim, K. L., ... Spirito, A. (2015). Self-injurious implicit attitudes among adolescent suicide attempters versus those engaged in nonsuicidal self-injury. *Journal of Child Psychology and Psychiatry*, Advance online publication.  
<http://doi.org/10.1111/jcpp.12385>
- Dodds, P. S., Clark, E. M., Desu, S., Frank, M. R., Reagan, A. J., Williams, J. R., ... Danforth, C. M. (2015). Human language reveals a universal positivity bias. *Proceedings of the National Academy of Sciences*, *112*(8), 2389–2394.  
<http://doi.org/10.1073/pnas.1411678112>
- Doessel, D. P., & Williams, R. F. (2010). The economic argument for a policy of suicide prevention. *Suicidology Online*, *1*, 66–75.

- Dougher, M. J., Augustson, E., Markham, M. R., Greenway, D. E., & Wulfert, E. (1994). The transfer of respondent eliciting and extinction functions through stimulus equivalence classes. *Journal of the Experimental Analysis of Behavior*, *62*(3), 331–351.
- Dougher, M. J., Hamilton, D. A., Fink, B. C., & Harrington, J. (2007). Transformation of the Discriminative and Eliciting Functions of Generalized Relational Stimuli. *Journal of the Experimental Analysis of Behavior*, *88*(2), 179–197.  
<http://doi.org/10.1901/jeab.2007.45-05>
- Dymond, S., & Barnes, D. (1995). A transformation of self-discrimination response functions in accordance with the arbitrarily applicable relations of sameness, more than, and less than. *Journal of the Experimental Analysis of Behavior*, *64*(2), 163–184.
- Dymond, S., & Barnes, D. (1997). Behavior-analytic approaches to self-awareness. *The Psychological Record*, *47*(2), 181.
- Dymond, S., Roche, B., Forsyth, J. P., Whelan, R., & Rhoden, J. (2007). Transformation of Avoidance Response Functions in Accordance with Same and Opposite Relational Frames. *Journal of the Experimental Analysis of Behavior*, *88*(2), 249–262.  
<http://doi.org/10.1901/jeab.2007.22-07>
- Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological Review*, *109*(3), 573–598. <http://doi.org/10.1037//0033-295X.109.3.573>
- Eagly, A. H., Mladinic, A., & Otto, S. (1991). Are women evaluated more favorably than men? *Psychology of Women Quarterly*, *15*(2), 203–216. <http://doi.org/10.1111/j.1471-6402.1991.tb00792.x>
- Eikelenboom, M., Smit, J. H., Beekman, A. T. F., Kerkhof, A. J. F. M., & Penninx, B. W. J. H. (2014). Reporting suicide attempts: consistency and its determinants in a large mental health study. *International Journal of Methods in Psychiatric Research*, *23*(2), 257–266. <http://doi.org/10.1002/mpr.1423>

- Ellis, T. E., & Rufino, K. A. (2015). A psychometric study of the Suicide Cognitions Scale with psychiatric inpatients. *Psychological Assessment, 27*(1), 82–89.  
<http://doi.org/10.1037/pas0000028>
- Ellis, T. E., Rufino, K. A., & Green, K. L. (2015). Implicit measure of life/death orientation predicts response of suicidal ideation to treatment in psychiatric inpatients. *Archives of Suicide Research*, Advance online publication.  
<http://doi.org/10.1080/13811118.2015.1004483>
- Erceg-Hurn, D. M., & Mirosevich, V. M. (2008). Modern robust statistical methods: An easy way to maximize the accuracy and power of your research. *American Psychologist, 63*(7), 591–601. <http://doi.org/10.1037/0003-066X.63.7.591>
- Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology, 54*(1), 297–327.  
<http://doi.org/10.1146/annurev.psych.54.101601.145225>
- Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology, 50*(2), 229–238.  
<http://doi.org/10.1037/0022-3514.50.2.229>
- Ferrara, M., Terrinoni, A., & Williams, R. (2012). Non-suicidal self-injury (NSSI) in adolescent inpatients: assessing personality features and attitude toward death. *Child and Adolescent Psychiatry and Mental Health, 6*(1), 1–8.
- Fiedler, K. (2014). From intrapsychic to ecological theories in social psychology: outlines of a functional theory approach: from intrapsychic to ecological theories. *European Journal of Social Psychology*. <http://doi.org/10.1002/ejsp.2069>
- Foody, M., Barnes-Holmes, Y., & Barnes-Holmes, D. (2012). The role of self in Acceptance and Commitment Therapy. In L. McHugh & I. Stewart (Eds.), *The self and*

- perspective taking: research and applications* (pp. 125–142). Oakland, CA: New Harbinger.
- Franck, E., De Raedt, R., Dereu, M., & Van den Abbeele, D. (2007). Implicit and explicit self-esteem in currently depressed individuals with and without suicidal ideation. *Journal of Behavior Therapy and Experimental Psychiatry*, *38*(1), 75–85.  
<http://doi.org/10.1016/j.jbtep.2006.05.003>
- Gardner, I. A., & Greiner, M. (2006). Receiver-operating characteristic curves and likelihood ratios: improvements over traditional methods for the evaluation and application of veterinary clinical pathology tests. *Veterinary Clinical Pathology / American Society for Veterinary Clinical Pathology*, *35*(1), 8–17.
- Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin*, *132*(5), 692–731. <http://doi.org/10.1037/0033-2909.132.5.692>
- Gawronski, B., & Bodenhausen, G. V. (2011). The Associative-Propositional Evaluation Model: Theory, evidence, and open questions. *Advances in Experimental Social Psychology*, *44*, 59–118.
- Gawronski, B., & De Houwer, J. (2011). Implicit measures in social and personality psychology. In *Handbook of research methods in social and personality psychology* (Vol. 2). New York, NY: Cambridge University Press.
- Gemar, M. C., Segal, Z. V., Sagrati, S., & Kennedy, S. J. (2001). Mood-induced changes on the Implicit Association Test in recovered depressed patients. *Journal of Abnormal Psychology*, *110*(2), 282–289. <http://doi.org/10.1037//0021-843X.110.2.282>
- Glashouwer, K. A., de Jong, P. J., Penninx, B. W. J. H., Kerkhof, A. J. F. M., van Dyck, R., & Ormel, J. (2010). Do automatic self-associations relate to suicidal ideation? *Journal*

*of Psychopathology and Behavioral Assessment*, 32(3), 428–437.

<http://doi.org/10.1007/s10862-009-9156-y>

Glenn, C. R., & Nock, M. K. (2014a). Improving the prediction of suicidal behavior in youth.

*International Journal Of Behavioral Consultation And Therapy*, 9(3). Retrieved from

[http://baojournal.com/IJBCT/IJBCT-9\\_3/A03.pdf](http://baojournal.com/IJBCT/IJBCT-9_3/A03.pdf)

Glenn, C. R., & Nock, M. K. (2014b). Improving the short-term prediction of suicidal

behavior. *American Journal of Preventive Medicine*, 47(3), S176–S180.

<http://doi.org/10.1016/j.amepre.2014.06.004>

Glick, P., Lameiras, M., Fiske, S. T., Eckes, T., Masser, B., Volpato, C., ... Wells, R. (2004).

Bad but bold: Ambivalent attitudes toward men predict gender inequality in 16 nations. *Journal of Personality and Social Psychology*, 86(5), 713–728.

<http://doi.org/10.1037/0022-3514.86.5.713>

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(3-4), 105–119. <http://doi.org/10.1016/j.jcbs.2013.05.002>

Gould, M. S., Marrocco, F. A., Kleinman, M., Thomas, J. G., Mostkoff, K., Cote, J., &

Davies, M. (2005). Evaluating iatrogenic risk of youth suicide screening programs.

*JAMA: The Journal of the American Medical Association*, 293(13), 1635–1643.

Greenberg, J., Pyszczynski, T., & Solomon, S. (1986). The causes and consequences of a

need for self-esteem: A terror management theory. In R. F. Baumeister (Ed.), *Public self and private self* (pp. 189–212). Springer New York.

Greenberg, J., Pyszczynski, T., Solomon, S., Rosenblatt, A., Veeder, M., Kirkland, S., &

Lyon, D. (1990). Evidence for terror management theory II: The effects of mortality salience on reactions to those who threaten or bolster the cultural worldview. *Journal*

*of Personality and Social Psychology*, 58(2), 308–318. <http://doi.org/10.1037/0022-3514.58.2.308>

Greenwald, A. G., Banaji, M. R., Rudman, L. A., Farnham, S. D., Nosek, B. A., & Mellott, D. S. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological Review*, 109(1), 3–25. <http://doi.org/10.1037//0033-295X.109.1.3>

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the Implicit Association Test. *Journal of Personality and Social Psychology*, 74(6), 1464–1480.

Greenwald, A. G., & Nosek, B. A. (2001). Health of the Implicit Association Test at age 3. *Zeitschrift Für Experimentelle Psychologie*, 48(2), 85–93.

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(2), 197–216. <http://doi.org/10.1037/0022-3514.85.2.197>

Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology*, 97(1), 17–41. <http://doi.org/10.1037/a0015575>

Harris, E. C., & Barraclough, B. (1997). Suicide as an outcome for mental disorders: A meta-analysis. *The British Journal of Psychiatry*, 170(3), 205–228.

Harrison, D. P., Stritzke, W. G. K., Fay, N., Ellison, T. M., & Hudaib, A.-R. (2014). Probing the implicit suicidal mind: Does the death/suicide Implicit Association Test reveal a desire to die, or a diminished desire to live? *Psychological Assessment*, 26(3), 831–840. <http://doi.org/10.1037/pas0000001>

- Harriss, L., & Hawton, K. (2005). Suicidal intent in deliberate self-harm and the risk of suicide: The predictive power of the Suicide Intent Scale. *Journal of Affective Disorders, 86*(2-3), 225–233. <http://doi.org/10.1016/j.jad.2005.02.009>
- Hayes, S. C. (1992). Verbal relations, time, and suicide. In S. C. Hayes & L. J. Hayes (Eds.), *Understanding verbal relations* (pp. 109–118). NV: Context Press.
- Hayes, S. C., Barnes-Holmes, D., & Roche, B. (2001). *Relational frame theory: A post-Skinnerian account of human language and cognition*. New York: Kluwer Academic/Plenum Press.
- 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*(1-2), 1–16. <http://doi.org/10.1016/j.jcbs.2012.09.004>
- Hayes, S. C., & Brownstein, A. J. (1986). Mentalism, behavior-behavior relations and a behavior-analytic view of the purposes of science. *The Behavior Analyst, 9*, 175–190.
- Hayes, S. C., Brownstein, A. J., Haas, J. R., & Greenway, D. E. (1986). Instructions, multiple schedules, and extinction: Distinguishing rule-governed from schedule-controlled behavior. *Journal of the Experimental Analysis of Behavior, 46*(2), 137–147.
- Hayes, S. C., Brownstein, A. J., Zettle, R. D., Rosenfarb, I., & Korn, Z. (1986). Rule-governed behavior and sensitivity to changing consequences of responding. *Journal of the Experimental Analysis of Behavior, 45*(3), 237–256. <http://doi.org/10.1901/jeab.1986.45-237>
- Hayes, S. C., Strosahl, K., & Wilson, K. G. (1999). *Acceptance and Commitment Therapy: An experiential approach to behavior change*. Guilford Press.
- Heidegger, M. (1962). *Being and time* (trans. J. Macquarrie and E. Robinson). New York, NY: Harper & Row.

- Hjelmeland, H., & Knizek, B. L. (1999). Conceptual confusion about intentions and motives of nonfatal suicidal behavior: A discussion of terms employed in the literature of suicidology. *Archives of Suicide Research*, 5(4), 277–283.  
<http://doi.org/10.1023/A:1009685700347>
- Houben, K., & Wiers, R. W. (2006). Assessing implicit alcohol associations with the Implicit Association Test: Fact or artifact? *Addictive Behaviors*, 31(8), 1346–1362.  
<http://doi.org/10.1016/j.addbeh.2005.10.009>
- Hughes, S. (2012). *Why we like what we like: A functional approach to the study of human evaluative responding* (unpublished doctoral thesis). National University of Ireland, Maynooth, Maynooth, Ireland. Retrieved from  
<http://eprints.maynoothuniversity.ie/4329/>
- Hughes, S., & Barnes-Holmes, D. (2011). On the formation and persistence of implicit attitudes: New evidence from the Implicit Relational Assessment Procedure (IRAP). *The Psychological Record*, 61, 391–410.
- Hughes, S., & Barnes-Holmes, D. (2014). Associative concept learning, stimulus equivalence, and relational frame theory: working out the similarities and differences between human and nonhuman behavior. *Journal of the Experimental Analysis of Behavior*, 101(1), 156–160. <http://doi.org/10.1002/jeab.60>
- Hughes, S., & Barnes-Holmes, D. (2015). Relational Frame Theory: Implications for the Study of Human Language and Cognition. In R. D. Zettle, S. C. Hayes, D. Barnes-Holmes, & A. Biglan (Eds.), *Handbook of contextual behavioral science*. New York, NY: Wiley-Blackwell.
- Hughes, S., & Barnes-Holmes, D. (2016). Relational Frame Theory: The basic account. In R. D. Zettle, S. C. Hayes, D. Barnes-Holmes, & A. Biglan (Eds.), *Handbook of contextual behavioral science*. New York, NY: Wiley-Blackwell.

- 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), 465–498.
- Hughes, S., Barnes-Holmes, D., & Vahey, N. A. (2012). Holding on to our functional roots when exploring new intellectual islands: A voyage through implicit cognition research. *Journal of Contextual Behavioral Science, 1*(1–2), 17–38.  
<http://doi.org/10.1016/j.jcbs.2012.09.003>
- Hughes, S., Hussey, I., Corrigan, B., Jolie, K., Murphy, C., & Barnes-Holmes, D. (2015). Faking Revisited: Exerting Strategic Control over Performance on the Implicit Relational Assessment Procedure. *Manuscript Submitted for Publication*.
- Huijding, J., de Jong, P. J., Wiers, R. W., & Verkoijen, K. (2005). Implicit and explicit attitudes toward smoking in a smoking and a nonsmoking setting. *Addictive Behaviors, 30*(5), 949–961. <http://doi.org/10.1016/j.addbeh.2004.09.014>
- 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. <http://doi.org/10.1016/j.cbpra.2012.03.002>
- Hussey, I., Barnes-Holmes, D., & Barnes-Holmes, Y. (2015). From Relational Frame Theory to implicit attitudes and back again: clarifying the link between RFT and IRAP research. *Current Opinion in Psychology, 2*, 11–15.  
<http://doi.org/10.1016/j.copsyc.2014.12.009>
- Hussey, I., McEnteggart, C., Nicholson, E., & Thompson, M. (2015). *Level of pre-experimental depression and anxiety is not predictive of experimentally induced distress: implications for ethical review boards*. Manuscript in preparation.

- Jacomb, P. A., Jorm, A. F., Rodgers, B., Korten, A. E., Henderson, A. S., & Christensen, H. (1999). Emotional response of participants to a mental health survey. *Social Psychiatry and Psychiatric Epidemiology*, *34*(2), 80–84.
- Janis, I. B., & Nock, M. K. (2008). Behavioral forecasts do not improve the prediction of future behavior: a prospective study of self-injury. *Journal of Clinical Psychology*, *64*(10), 1164–1174. <http://doi.org/10.1002/jclp.20509>
- Joiner, T. E. J. (2005). *Why people die by suicide*. Harvard University Press.
- Jorm, A. F., Kelly, C. M., & Morgan, A. J. (2007). Participant distress in psychiatric research: a systematic review. *Psychological Medicine*, *37*(07), 917. <http://doi.org/10.1017/S0033291706009779>
- Kahn, J. H., Tobin, R. M., Massey, A. E., & Anderson, J. A. (2007). Measuring emotional expression with the Linguistic Inquiry and Word Count. *The American Journal of Psychology*, *120*(2), 263–286. <http://doi.org/10.2307/20445398>
- Karpinski, A. (2004). Measuring self-esteem using the Implicit Association Test: The role of the other. *Personality and Social Psychology Bulletin*, *30*(1), 22–34. <http://doi.org/10.1027/1618-3169.52.1.74>
- Karpinski, A., & Steinman, R. B. (2006). The single category implicit association test as a measure of implicit social cognition. *Journal of Personality and Social Psychology*, *91*(1), 16–32. <http://doi.org/10.1037/0022-3514.91.1.16>
- Kastenbaum, R. (2000). *The psychology of death*. Springer Publishing Company.
- Kazdin, A. E. (2008). Evidence-based treatment and practice: New opportunities to bridge clinical research and practice, enhance the knowledge base, and improve patient care. *American Psychologist*, *63*(3), 146–159. <http://doi.org/10.1037/0003-066X.63.3.146>

- Kennelly, B. (2007). The economic cost of suicide in Ireland. *Crisis: The Journal of Crisis Intervention and Suicide Prevention*, 28(2), 89–94. <http://doi.org/10.1027/0227-5910.28.2.89>
- Kishita, N., Muto, T., Ohtsuki, T., & Barnes-Holmes, D. (2014). Measuring the effect of cognitive defusion using the Implicit Relational Assessment Procedure: An experimental analysis with a highly socially anxious sample. *Journal of Contextual Behavioral Science*, 3(1), 8–15. <http://doi.org/10.1016/j.jcbs.2013.12.001>
- Klonsky, E. D. (2009). The functions of self-injury in young adults who cut themselves: Clarifying the evidence for affect-regulation. *Psychiatry Research*, 166(2-3), 260–268. <http://doi.org/10.1016/j.psychres.2008.02.008>
- Klonsky, E. D., & May, A. M. (2014). Differentiating suicide attempters from suicide ideators: A critical frontier for suicidology research. *Suicide and Life-Threatening Behavior*, 44(1), 1–5. <http://doi.org/10.1111/sltb.12068>
- Knowles, S. E., & Townsend, E. (2012). Implicit and explicit attitudes toward self harm: Support for a functional model. *Journal of Behavior Therapy and Experimental Psychiatry*, 43(2), 730–736. <http://doi.org/10.1016/j.jbtep.2011.10.007>
- Kovacs, M., & Beck, A. T. (1977). The wish to die and the wish to live in attempted suicides. *Journal of Clinical Psychology*, 33(2), 361–365. [http://doi.org/10.1002/1097-4679\(197704\)33:2<361::AID-JCLP2270330207>3.0.CO;2-H](http://doi.org/10.1002/1097-4679(197704)33:2<361::AID-JCLP2270330207>3.0.CO;2-H)
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, 4. <http://doi.org/10.3389/fpsyg.2013.00863>
- Large, M., Sharma, S., Cannon, E., Ryan, C. J., & Nielssen, O. (2011). Risk factors for suicide within a year of discharge from psychiatric hospital: a systematic meta-

- analysis. *The Australian and New Zealand Journal of Psychiatry*, 45(8), 619–628.  
<http://doi.org/10.3109/00048674.2011.590465>
- Larkin, C., Di Blasi, Z., & Arensman, E. (2014). Risk factors for repetition of self-harm: A systematic review of prospective hospital-based studies. *PLoS ONE*, 9(1), e84282.  
<http://doi.org/10.1371/journal.pone.0084282>
- Latimer, S., Meade, T., & Tennant, A. (2013). Measuring engagement in deliberate self-harm behaviours: psychometric evaluation of six scales. *BMC Psychiatry*, 13(1), 4.
- Lindqvist, D., Janelidze, S., Erhardt, S., Träskman-Bendz, L., Engström, G., & Brundin, L. (2011). CSF biomarkers in suicide attempters – a principal component analysis. *Acta Psychiatrica Scandinavica*, 124(1), 52–61. <http://doi.org/10.1111/j.1600-0447.2010.01655.x>
- Linehan, M. M. (1993). *Cognitive-behavioral Treatment of Borderline Personality Disorder*. Guilford Press.
- 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. [http://doi.org/10.1016/0005-7967\(94\)00075-U](http://doi.org/10.1016/0005-7967(94)00075-U)
- MacLeod, C. (1993). Cognition in clinical psychology: measures, methods or models? *Behaviour Change*, 10(03), 169–195. <http://doi.org/10.1017/S0813483900005519>
- McHugh, L., Barnes-Holmes, Y., & Barnes-Holmes, D. (2004). Perspective-taking as relational responding: A developmental profile. *Psychological Record*, 54(1), 115–144.
- McMahon, E. M., Keeley, H., Cannon, M., Arensman, E., Perry, I. J., Clarke, M., ... Corcoran, P. (2014). The iceberg of suicide and self-harm in Irish adolescents: a

- population-based study. *Social Psychiatry and Psychiatric Epidemiology*, 49(12), 1929–1935. <http://doi.org/10.1007/s00127-014-0907-z>
- Meites, T. M., Deveney, C. M., Steele, K. T., Holmes, A. J., & Pizzagalli, D. A. (2008). Implicit depression and hopelessness in remitted depressed individuals. *Behaviour Research and Therapy*, 46(9), 1078–1084. <http://doi.org/10.1016/j.brat.2008.05.008>
- Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ*, 339, b2535. <http://doi.org/10.1136/bmj.b2535>
- Muehlenkamp, J. J., & Gutierrez, P. M. (2004). An investigation of differences between self-injurious behavior and suicide attempts in a sample of adolescents. *Suicide and Life-Threatening Behavior*, 34(1), 12–23. <http://doi.org/10.1521/suli.34.1.12.27769>
- Muehlenkamp, J. J., Swenson, L. P., Batejan, K. L., & Jarvi, S. M. (2015). Emotional and Behavioral Effects of Participating in an Online Study of Nonsuicidal Self-Injury An Experimental Analysis. *Clinical Psychological Science*, 3(1), 26–37. <http://doi.org/10.1177/2167702614531579>
- National Office of Suicide Prevention. (2013). *Annual report 2013*. National Office of Suicide Prevention.
- Neimeyer, R. A., Wittkowski, J., & Moser, R. P. (2004). Psychological research on death attitudes: an overview and evaluation. *Death Studies*, 28(4), 309–340. <http://doi.org/10.1080/07481180490432324>
- 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. <http://doi.org/10.1016/j.jbtep.2012.02.001>

- Nicholson, E., McCourt, A., & Barnes-Holmes, D. (2013). The Implicit Relational Assessment Procedure (IRAP) as a measure of obsessive beliefs in relation to disgust. *Journal of Contextual Behavioral Science*, 2(1–2), 23–30.  
<http://doi.org/10.1016/j.jcbs.2013.02.002>
- Niméus, A., Alsén, M., & Träskman-Bendz, L. (2002). High suicidal intent scores indicate future suicide. *Archives of Suicide Research*, 6(3), 211–219.  
<http://doi.org/10.1080/13811110214142>
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 84(3), 231–259.  
<http://doi.org/10.1037/0033-295X.84.3.231>
- Nock, M. K. (2009a). Suicidal behavior among adolescents: Correlates, confounds, and (the search for) causal mechanisms. *Journal of the American Academy of Child & Adolescent Psychiatry*, 48(3), 237–239.  
<http://doi.org/10.1097/CHI.0b013e318196b944>
- Nock, M. K. (2009b). Why do people hurt themselves? New insights into the nature and functions of self-injury. *Current Directions in Psychological Science*, 18(2), 78–83.
- Nock, M. K. (2012). Future directions for the study of suicide and self-injury. *Journal of Clinical Child & Adolescent Psychology*, 41(2), 255–259.  
<http://doi.org/10.1080/15374416.2012.652001>
- Nock, M. K. (2013). *Updated Suicide IAT Task 2013*. Retrieved from  
<http://www.wjh.harvard.edu/~nock/nocklab/Updated%20Suicide%20IAT%202103.zip>
- Nock, M. K., & Banaji, M. (2007a). Assessment of self-injurious thoughts using a behavioral test. *American Journal of Psychiatry*, 164(5), 820–823.

- Nock, M. K., & Banaji, M. R. (2007b). Prediction of suicide ideation and attempts among adolescents using a brief performance-based test. *Journal of Consulting and Clinical Psychology, 75*(5), 707–715. <http://doi.org/10.1037/0022-006X.75.5.707>
- Nock, M. K., Holmberg, E. B., Photos, V. I., & Michel, B. D. (2007). Self-Injurious Thoughts and Behaviors Interview: Development, reliability, and validity in an adolescent sample. *Psychological Assessment, 19*(3), 309–317. <http://doi.org/10.1037/1040-3590.19.3.309>
- Nock, M. K., Park, J. M., Finn, C. T., Deliberto, T. L., Dour, H. J., & Banaji, M. R. (2010). Measuring the suicidal mind: Implicit cognition predicts suicidal behavior. *Psychological Science, 21*(4), 511–517. <http://doi.org/10.1177/0956797610364762>
- Nock, M. K., Prinstein, M. J., & Sterba, S. K. (2009). Revealing the form and function of self-injurious thoughts and behaviors: A real-time ecological assessment study among adolescents and young adults. *Journal of Abnormal Psychology, 118*(4), 816–827. <http://doi.org/10.1037/a0016948>
- Nosek, B. A., & Banaji, M. R. (2001). The Go/No-Go Association Task. *Social Cognition, 19*(6), 625–666. <http://doi.org/10.1521/soco.19.6.625.20886>
- Nosek, B. A., Bar-Anan, Y., Sriram, N., Axt, J., & Greenwald, A. G. (2014). Understanding and using the Brief Implicit Association Test: Recommended scoring procedures. *PLoS ONE, 9*(12), e110938. <http://doi.org/10.1371/journal.pone.0110938>
- Nosek, B. A., & Greenwald, A. G. (2009). (Part of) the case for a pragmatic approach to validity: Comment on De Houwer, Teige-Mocigemba, Spruyt, and Moors (2009). *Psychological Bulletin, 135*(3), 373–376. <http://doi.org/10.1037/a0015047>
- Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2005). Understanding and using the Implicit Association Test: II. Method variables and construct validity. *Personality &*

*Social Psychology Bulletin*, 31(2), 166–180.

<http://doi.org/10.1177/0146167204271418>

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2007). The Implicit Association Test at age 7: A methodological and conceptual review. In J. Bargh (Ed.), *Automatic processes in social thinking and behavior* (pp. 265–292). New York, NY: Psychology Press.

Nosek, B. A., Hawkins, C. B., & Frazier, R. S. (2011). Implicit social cognition: From measures to mechanisms. *Trends in Cognitive Sciences*, 15(4), 152–159.

<http://doi.org/10.1016/j.tics.2011.01.005>

O’Carroll, P. W., Berman, A. L., Maris, R. W., Moscicki, E. K., Tanney, B. L., & Silverman, M. M. (1996). Beyond the tower of Babel: A nomenclature for suicidology. *Suicide and Life-Threatening Behavior*, 26(3), 237–252. <http://doi.org/10.1111/j.1943-278X.1996.tb00609.x>

O’Connor, E., Gaynes, B., Burda, B. U., Williams, C., & Whitlock, E. P. (2013). *Screening for Suicide Risk in Primary Care: A Systematic Evidence Review for the US Preventive Services Task Force*. (AHRQ Publication No. 13-05188-EF-1). Rockville, MD: Agency for Healthcare Research and Quality. Retrieved from

<http://europepmc.org/abstract/med/23678511>

O’Connor, R. C. (2011). The Integrated Motivational-Volitional model of suicidal behavior. *Crisis: The Journal of Crisis Intervention and Suicide Prevention*, 32(6), 295–298.

<http://doi.org/10.1027/0227-5910/a000120>

O’Connor, R. C., & Nock, M. K. (2014). The psychology of suicidal behaviour. *The Lancet Psychiatry*, 1(1), 73–85. [http://doi.org/10.1016/S2215-0366\(14\)70222-6](http://doi.org/10.1016/S2215-0366(14)70222-6)

- O'Hora, D., & Barnes-Holmes, D. (2004). Instructional Control: Developing a Relational Frame Analysis. *International Journal of Psychology & Psychological Therapy*, 4(2), 263–284.
- Orbach, I., Milstein, I., Har-Even, D., Apter, A., Tiano, S., & Elizur, A. (1991). A multi-attitude suicide tendency scale for adolescents. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, 3(3), 398–404. <http://doi.org/10.1037/1040-3590.3.3.398>
- Osarchuk, M., & Tatz, S. J. (1973). Effect of induced fear of death on belief in afterlife. *Journal of Personality and Social Psychology*, 27(2), 256–260. <http://doi.org/10.1037/h0034769>
- O'Shea, B., Watson, D. G., & Brown, G. D. (2015). Measuring implicit attitudes: a positive framing bias flaw in the Implicit Relational Assessment Procedure (IRAP). *Psychological Assessment*, Advance online publication.
- Osman, A., Gilpin, A. R., Panak, W. F., Kopper, B. A., Barrios, F. X., Gutierrez, P. M., & Chiros, C. E. (2000). The Multi-Attitude Suicide Tendency Scale: Further validation with adolescent psychiatric inpatients. *Suicide and Life-Threatening Behavior*, 30(4), 377–385. <http://doi.org/10.1111/j.1943-278X.2000.tb01104.x>
- Ostafin, B. D., & Palfai, T. P. (2006). Compelled to consume: The Implicit Association Test and automatic alcohol motivation. *Psychology of Addictive Behaviors*, 20(3), 322–327. <http://doi.org/10.1037/0893-164X.20.3.322>
- Palfai, T. P., & Ostafin, B. D. (2003). Alcohol-related motivational tendencies in hazardous drinkers: assessing implicit response tendencies using the modified-IAT. *Behaviour Research and Therapy*, 41(10), 1149–1162. [http://doi.org/10.1016/S0005-7967\(03\)00018-4](http://doi.org/10.1016/S0005-7967(03)00018-4)

- Parling, T., Cernvall, M., Stewart, I., Barnes-Holmes, D., & Ghaderi, A. (2012). Using the Implicit Relational Assessment Procedure to Compare Implicit Pro-Thin/Anti-Fat Attitudes of Patients With Anorexia Nervosa and Non-Clinical Controls. *Eating Disorders, 20*(2), 127–143. <http://doi.org/10.1080/10640266.2012.654056>
- Patterson, W. M., Dohn, H. H., Bird, J., & Patterson, G. A. (1983). Evaluation of suicidal patients: the SAD PERSONS scale. *Psychosomatics, 24*(4), 343–345, 348–349. [http://doi.org/10.1016/S0033-3182\(83\)73213-5](http://doi.org/10.1016/S0033-3182(83)73213-5)
- Payne, K. B., Cheng, C. M., Govorun, O., & Stewart, B. D. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality and Social Psychology, 89*(3), 277–293. <http://doi.org/10.1037/0022-3514.89.3.277>
- Pennebaker, J. W., Booth, R. J., & Francis, M. E. (2007). Linguistic inquiry and word count: LIWC (Version 2007). Austin, TX.
- Phillips, W. J., Hine, D. W., & Thorsteinsson, E. B. (2010). Implicit cognition and depression: A meta-analysis. *Clinical Psychology Review, 30*(6), 691–709. <http://doi.org/10.1016/j.cpr.2010.05.002>
- Pinter, B., & Greenwald, A. G. (2005). Clarifying the role of the “other” category in the self-esteem IAT. *Experimental Psychology, 52*(1), 74–79. <http://doi.org/10.1027/1618-3169.52.1.74>
- Platt, S., Bille-Brahe, U., Kerkhof, A., Schmidtke, A., Bjerke, T., Crepet, P., ... Michel, K. (1992). Parasuicide in Europe: the WHO/EURO multicentre study on parasuicide. I. Introduction and preliminary analysis for 1989. *Acta Psychiatrica Scandinavica, 85*(2), 97–104.
- Pompili, M., Girardi, P., Ruberto, A., Kotzalidis, G. D., & Tatarelli, R. (2005). Emergency staff reactions to suicidal and self-harming patients. *European Journal of Emergency*

- Medicine: Official Journal of the European Society for Emergency Medicine*, 12(4), 169–178.
- Power, P., 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(4), 621–640.
- Preti, A. (2011). Do Animals Commit Suicide? Does It Matter? *Crisis: The Journal of Crisis Intervention and Suicide Prevention*, 32(1), 1–4. <http://doi.org/10.1027/0227-5910/a000077>
- Price, R. B., Iosifescu, D. V., Murrough, J. W., Chang, L. C., Al Jurdi, R. K., Iqbal, S. Z., ... Mathew, S. J. (2014). Effects of ketamine on explicit and implicit suicidal cognition: a randomized controlled trial in treatment-resistant depression. *Depression and Anxiety*, 31(4), 335–343. <http://doi.org/10.1002/da.22253>
- Price, R. B., Nock, M. K., Charney, D. S., & Mathew, S. J. (2009). Effects of intravenous ketamine on explicit and implicit measures of suicidality in treatment-resistant depression. *Biological Psychiatry*, 66(5), 522–526. <http://doi.org/10.1016/j.biopsych.2009.04.029>
- Prinstein, M. J. (2008). Introduction to the special section on suicide and nonsuicidal self-injury: A review of unique challenges and important directions for self-injury science. *Journal of Consulting and Clinical Psychology*, 76(1), 1–8. <http://doi.org/10.1037/0022-006X.76.1.1>
- Pryor, J. B. (1987). Sexual harassment proclivities in men. *Sex Roles*, 17(5-6), 269–290. <http://doi.org/10.1007/BF00288453>
- Quinlivan, L., Cooper, J., Steeg, S., Davies, L., Hawton, K., Gunnell, D., & Kapur, N. (2014). Scales for predicting risk following self-harm: an observational study in 32 hospitals

- in England. *BMJ Open*, 4(5). Retrieved from  
<http://bmjopen.bmj.com/content/4/5/e004732.short>
- Randall, J. R., Colman, I., & Rowe, B. H. (2011). A systematic review of psychometric assessment of self-harm risk in the emergency department. *Journal of Affective Disorders*, 134(1–3), 348–355. <http://doi.org/10.1016/j.jad.2011.05.032>
- Randall, J. R., Rowe, B. H., Dong, K. A., Nock, M. K., & Colman, I. (2013). Assessment of self-harm risk using implicit thoughts. *Psychological Assessment*, 25(3), 714–721. <http://doi.org/10.1037/a0032391>
- Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, 114(4), 510–532.
- Remue, J., De Houwer, J., Barnes-Holmes, D., Vanderhasselt, M. A., & De Raedt, R. (2013). Self-esteem revisited: Performance on the implicit relational assessment procedure as a measure of self-versus ideal self-related cognitions in dysphoria. *Cognition & Emotion*, 27(8), 1441–1449. <http://doi.org/10.1080/02699931.2013.786681>
- Remue, J., Hughes, S., De Houwer, J., & De Raedt, R. (2014). To Be or Want to Be: Disentangling the Role of Actual versus Ideal Self in Implicit Self-Esteem. *PLoS ONE*, 9(9), e108837. <http://doi.org/10.1371/journal.pone.0108837>
- Reynolds, S. K., Lindenboim, N., Comtois, K. A., Murray, A., & Linehan, M. M. (2006). Risky assessments: participant suicidality and distress associated with research assessments in a treatment study of suicidal behavior. *Suicide and Life-Threatening Behavior*, 36(1), 19–34.
- Ribeiro, J. D., Witte, T. K., Van Orden, K. A., Selby, E. A., Gordon, K. H., Bender, T. W., & Joiner, T. E. (2014a). Fearlessness about death: The psychometric properties and construct validity of the revision to the Acquired Capability for Suicide Scale. *Psychological Assessment*, 26(1), 115–126. <http://doi.org/10.1037/a0034858>

- Ribeiro, J. D., Witte, T. K., Van Orden, K. A., Selby, E. A., Gordon, K. H., Bender, T. W., & Joiner, T. E. (2014b). Fearlessness about death: The psychometric properties and construct validity of the revision to the Acquired Capability for Suicide Scale. *Psychological Assessment, 26*(1), 115–126. <http://doi.org/10.1037/a0034858>
- Robinson, M. D., Meier, B. P., Zetocha, K. J., & McCaul, K. D. (2005). Smoking and the Implicit Association Test: When the contrast category determines the theoretical conclusions. *Basic and Applied Social Psychology, 27*(3), 201–212. [http://doi.org/10.1207/s15324834basp2703\\_2](http://doi.org/10.1207/s15324834basp2703_2)
- Roefs, A., Huijding, J., Smulders, F. T. Y., MacLeod, C. M., de Jong, P. J., Wiers, R. W., & Jansen, A. T. M. (2011). Implicit measures of association in psychopathology research. *Psychological Bulletin, 137*(1), 149–193. <http://doi.org/10.1037/a0021729>
- Rönspies, J., Schmidt, A. F., Melnikova, A., Krumova, R., Zolfagari, A., & Banse, R. (2015). Indirect measurement of sexual orientation: Comparison of the Implicit Relational Assessment Procedure, viewing time, and choice reaction time tasks. *Archives of Sexual Behavior*, Advance online publication. <http://doi.org/10.1007/s10508-014-0473-1>
- Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.
- Rosenblatt, A., Greenberg, J., Solomon, S., Pyszczynski, T., & Lyon, D. (1989). Evidence for terror management theory: I. The effects of mortality salience on reactions to those who violate or uphold cultural values. *Journal of Personality and Social Psychology, 57*(4), 681–690. <http://doi.org/10.1037/0022-3514.57.4.681>
- Rudd, M. D., Berman, A. L., Joiner, T. E., Nock, M. K., Silverman, M. M., Mandrusiak, M., ... Witte, T. (2006). Warning signs for suicide: theory, research, and clinical

- applications. *Suicide & Life-Threatening Behavior*, 36(3), 255–262.  
<http://doi.org/10.1521/suli.2006.36.3.255>
- Rudman, L. A., & Glick, P. (2001). Prescriptive gender stereotypes and backlash toward agentic women. *Journal of Social Issues*, 57(4), 743–762.  
<http://doi.org/10.1111/0022-4537.00239>
- Rudman, L. A., & Mescher, K. (2012). Of animals and objects: Men's implicit dehumanization of women and likelihood of sexual aggression. *Personality and Social Psychology Bulletin*, 38(6), 734–746.  
<http://doi.org/10.1177/0146167212436401>
- Rüsch, N., Lieb, K., Göttler, I., Hermann, C., Schramm, E., Richter, H., ... Bohus, M. (2007). Shame and implicit self-concept in women with borderline personality disorder. *American Journal of Psychiatry*, 164(3), 500–508.
- Rutter, C. M., & Miglioretti, D. L. (2003). Estimating the accuracy of psychological scales using longitudinal data. *Biostatistics*, 4(1), 97–107.  
<http://doi.org/10.1093/biostatistics/4.1.97>
- Ryan, C. J., Large, M., Gribble, R., Macfarlane, M., Ilchef, R., & Tietze, T. (2015). Assessing and managing suicidal patients in the emergency department. *Australasian Psychiatry*, 23(5), 513–516. <http://doi.org/10.1177/1039856215597536>
- Ryan, C. J., & Large, M. M. (2013). Suicide risk assessment: where are we now? *Medical Journal of Australia*, 198(9). Retrieved from <https://www.mja.com.au/journal/2013/198/9/suicide-risk-assessment-where-are-we-now>
- Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology*, 4, 1–32.

- Shneidman, E. S. (1993). Suicide as Psychache. *The Journal of Nervous and Mental Disease*, 181(3), 145–147.
- Sidman, M. (1994). *Equivalence relations and behavior: A research story* (Vol. xvii). Boston, MA, US: Authors Cooperative.
- Silverman, M. M. (2006). The language of suicidology. *Suicide and Life-Threatening Behavior*, 36(5), 519–532.
- Silverman, M. M., & Berman, A. L. (2013). Suicide Risk Assessment and Risk Formulation Part I: A Focus on Suicide Ideation in Assessing Suicide Risk. *Suicide and Life-Threatening Behavior*, n/a–n/a. <http://doi.org/10.1111/sltb.12065>
- Silverman, M. M., Berman, A. L., Sanddal, N. D., O'Carroll, P. W., & Joiner, T. E. (2007a). Rebuilding the tower of babel: A revised nomenclature for the study of suicide and suicidal behaviors Part 1: Background, rationale, and methodology. *Suicide and Life-Threatening Behavior*, 37(3), 248–263.
- Silverman, M. M., Berman, A. L., Sanddal, N. D., O'Carroll, P. W., & Joiner, T. E. (2007b). Rebuilding the tower of Babel: a revised nomenclature for the study of suicide and suicidal behaviors part 2: Suicide-related ideations, communications, and behaviors. *Suicide and Life-Threatening Behavior*, 37(3), 264–277.
- Smyth, S., Barnes-Holmes, D., & Forsyth, J. P. (2006). A Derived Transfer of Simple Discrimination and Self-Reported Arousal Functions in Spider Fearful and Non-Spider-Fearful Participants. *Journal of the Experimental Analysis of Behavior*, 85(2), 223–246. <http://doi.org/10.1901/jeab.2006.02-05>
- Spence, J. T., Helmreich, R., & Stapp, J. (1973). A short version of the Attitudes toward Women Scale (AWS). *Bulletin of the Psychonomic Society*, 2(4), 219–220. <http://doi.org/10.3758/BF03329252>

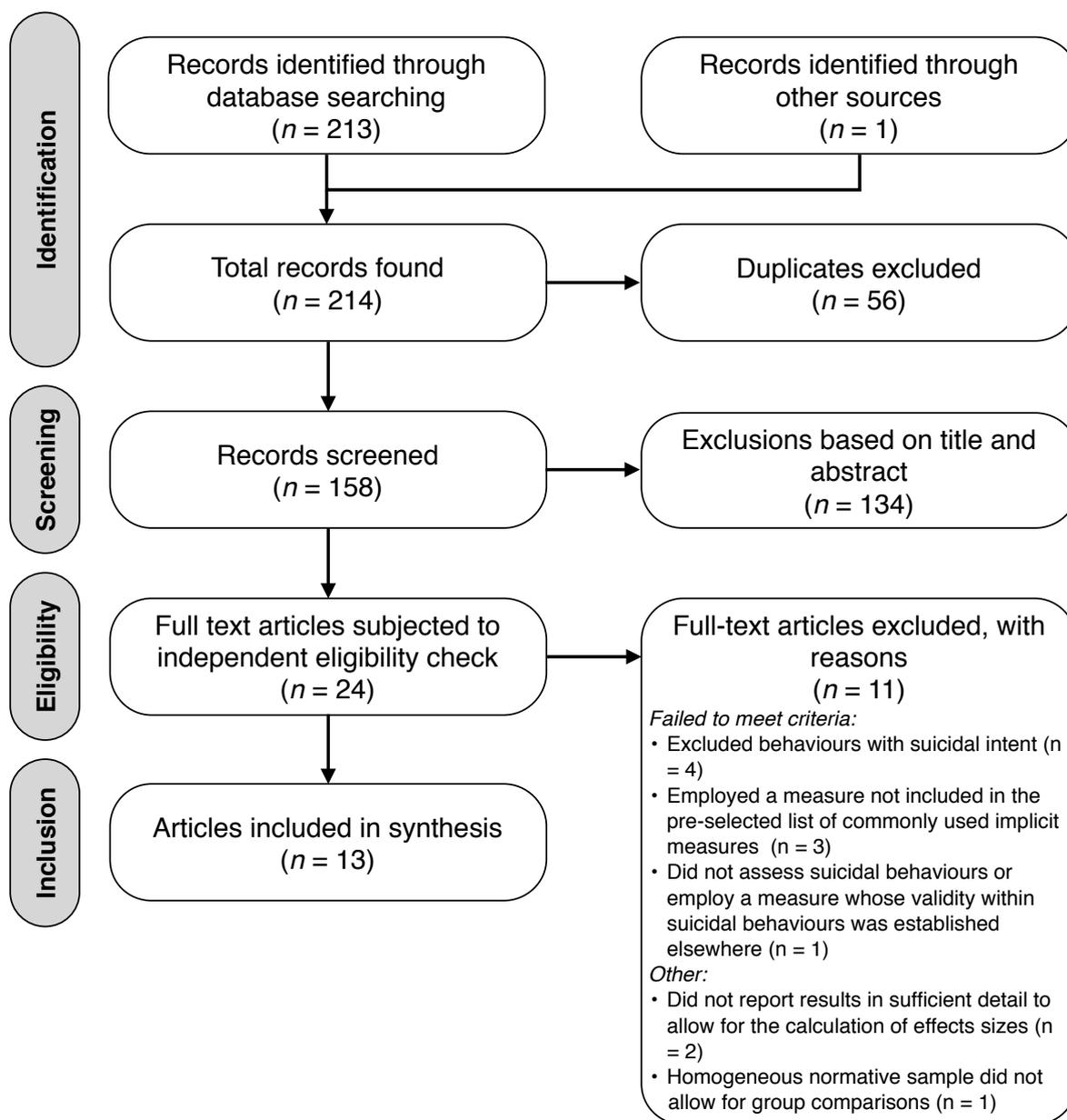
- Sriram, N., & Greenwald, A. G. (2009). The Brief Implicit Association Test. *Experimental Psychology (formerly Zeitschrift Für Experimentelle Psychologie)*, 56(4), 283–294.  
<http://doi.org/10.1027/1618-3169.56.4.283>
- Sriram, N., Greenwald, A. G., & Nosek, B. A. (2010). Correlational biases in mean response latency differences. *Statistical Methodology*, 7(3), 277–291.  
<http://doi.org/10.1016/j.stamet.2009.10.004>
- Stewart, I., Barnes-Holmes, D., & Roche, B. (2004). A functional-analytic model of analogy using the relational evaluation procedure. *The Psychological Record*, 54(4), 531–552.
- Suominen, K., Isometsä, E., Ostamo, A., & Lönnqvist, J. (2004). Level of suicidal intent predicts overall mortality and suicide after attempted suicide: a 12-year follow-up study. *BMC Psychiatry*, 4(1), 11.
- Swanson, J. E., Rudman, L. A., & Greenwald, A. G. (2001). Using the Implicit Association Test to investigate attitude-behaviour consistency for stigmatised behaviour. *Cognition and Emotion*, 15(2), 207–230. <http://doi.org/10.1080/02699930125706>
- Tang, J., Wu, S., & Miao, D. (2013). Experimental test of escape theory: Accessibility to implicit suicidal mind. *Suicide and Life-Threatening Behavior*, 43(4), 347–355.  
<http://doi.org/10.1111/sltb.12021>
- Tarrier, N., Taylor, K., & Gooding, P. (2008). Cognitive-behavioral interventions to reduce suicide behavior: A systematic review and meta-analysis. *Behavior Modification*, 32(1), 77–108. <http://doi.org/10.1177/0145445507304728>
- Templer, D. I. (1970). The construction and validation of a death anxiety scale. *The Journal of General Psychology*, 82(2), 165–177.  
<http://doi.org/10.1080/00221309.1970.9920634>

- Thas, O., Neve, J. D., Clement, L., & Ottoy, J.-P. (2012). Probabilistic index models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *74*(4), 623–671. <http://doi.org/10.1111/j.1467-9868.2011.01020.x>
- 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.
- Vahey, N. A., 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*(3), 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. <http://doi.org/10.1016/j.jbtep.2015.01.004>
- Van Orden, K. A., Witte, T. K., Cukrowicz, K. C., Braithwaite, S. R., Selby, E. A., & Joiner, T. E. (2010). The interpersonal theory of suicide. *Psychological Review*, *117*(2), 575–600. <http://doi.org/10.1037/a0018697>
- Vervliet, B., & Raes, F. (2013). Criteria of validity in experimental psychopathology: application to models of anxiety and depression. *Psychological Medicine*, *43*(11), 2241–2244. <http://doi.org/10.1017/S0033291712002267>
- Warden, S., Spiwak, R., Sareen, J., & Bolton, J. M. (2014). The SAD PERSONS scale for suicide risk assessment: A systematic review. *Archives of Suicide Research*, *18*(4), 313–326. <http://doi.org/10.1080/13811118.2013.824829>
- Wasserman, D., Hoven, C. W., Wasserman, C., Wall, M., Eisenberg, R., Hadlaczky, G., ... Carli, V. (2015). School-based suicide prevention programmes: the SEYLE cluster-

- randomised, controlled trial. *The Lancet*, 385(9977), 1536–1544.  
[http://doi.org/10.1016/S0140-6736\(14\)61213-7](http://doi.org/10.1016/S0140-6736(14)61213-7)
- Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, 54(6), 1063–1070. <http://doi.org/10.1037/0022-3514.54.6.1063>
- Wenzel, A., & Beck, A. T. (2008). A cognitive model of suicidal behavior: Theory and treatment. *Applied and Preventive Psychology*, 12(4), 189–201.  
<http://doi.org/10.1016/j.appsy.2008.05.001>
- Whelan, R. (2008). Effective analysis of reaction time data. *The Psychological Record*, 58(3), 475–482.
- WHO. (2014). *Preventing suicide: A global imperative*. World Health Organisation.
- Williams, J. M. G. (1997). *Suicide and Attempted Suicide: Understanding the Cry of Pain*. London, UK: Penguin.
- Williams, J. M. G., van der Does, W., Barnhofer, T., Crane, C., & Segal, Z. S. (2008). Cognitive Reactivity, Suicidal Ideation and Future Fluency: Preliminary Investigation of a Differential Activation Theory of Hopelessness/Suicidality. *Cognitive Therapy and Research*, 32(1), 83–104. <http://doi.org/10.1007/s10608-006-9105-y>
- Wilson, T. D. (2009). Know Thyself. *Perspectives on Psychological Science*, 4(4), 384–389.  
<http://doi.org/10.1111/j.1745-6924.2009.01143.x>
- Wittenbrink, B., Judd, C. M., & Park, B. (1997). Evidence for racial prejudice at the implicit level and its relationship with questionnaire measures. *Journal of Personality and Social Psychology*, 72(2), 262–274. <http://doi.org/10.1037/0022-3514.72.2.262>

Yeater, E., Miller, G., Rinehart, J., & Nason, E. (2012). Trauma and Sex Surveys Meet Minimal Risk Standards: Implications for Institutional Review Boards. *Psychological Science*, 23(7), 780–787. <http://doi.org/10.1177/0956797611435131>

## Appendix A: Flowchart of Included studies



## Appendix B: Table of excluded articles

Article	Reason for exclusion
1. Bassett, J. F., & Dabbs, J. M. J. (2003). Evaluating explicit and implicit death attitudes in funeral and university students. <i>Mortality</i> , 8(4), 352–371. <a href="http://doi.org/10.1080/13576270310001604022">http://doi.org/10.1080/13576270310001604022</a>	Did not assess suicidal behaviors or employ a measure whose validity within suicidal behaviors was established elsewhere
2. Creemers, D. H. M., Scholte, R. H. J., Engels, R. C. M. E., Prinstein, M. J., & Wiers, R. W. (2012). Implicit and explicit self-esteem as concurrent predictors of suicidal ideation, depressive symptoms, and loneliness. <i>Journal of Behavior Therapy and Experimental Psychiatry</i> , 43(1), 638–646. <a href="http://doi.org/10.1016/j.jbtep.2011.09.006">http://doi.org/10.1016/j.jbtep.2011.09.006</a>	Employed a measure not included in the pre-selected list
3. Creemers, D. H. M., Scholte, R. H. J., Engels, R. C. M. E., Prinstein, M. J., & Wiers, R. W. (2013). Damaged self-esteem is associated with internalizing problems. <i>Frontiers in Psychology</i> , 4. <a href="http://doi.org/10.3389/fpsyg.2013.00152">http://doi.org/10.3389/fpsyg.2013.00152</a>	Did not report results from the implicit measure in sufficient detail to allow for the calculation of an effect size
4. Franklin, J. C., Lee, K. M., Puzia, M. E., & Prinstein, M. J. (2014). Recent and Frequent Nonsuicidal Self-Injury Is Associated With Diminished Implicit and Explicit Aversion Toward Self-Cutting Stimuli. <i>Clinical Psychological Science</i> , 2(3), 306–318. <a href="http://doi.org/10.1177/2167702613503140">http://doi.org/10.1177/2167702613503140</a>	Excluded behaviors with suicidal intent
5. Franklin, J. C., Puzia, M. E., Lee, K. M., & Prinstein, M. J. (2014). Low implicit and explicit aversion toward self-cutting stimuli longitudinally predict nonsuicidal self-injury. <i>Journal of Abnormal Psychology</i> , 123(2), 463–469. <a href="http://doi.org/10.1037/a0036436">http://doi.org/10.1037/a0036436</a>	Excluded behaviors with suicidal intent
6. Galynker, I., Yaseen, Z. S., Briggs, J., & Hayashi, F. (2015). Attitudes of acceptability and lack of condemnation toward suicide may be predictive of post-discharge suicide attempts. <i>BMC Psychiatry</i> , 15(1). <a href="http://doi.org/10.1186/s12888-015-0462-5">http://doi.org/10.1186/s12888-015-0462-5</a>	Employed a measure not included in the pre-selected list
7. Glenn, C. R., & Klonsky, E. D. (2011). Prospective prediction of nonsuicidal self-injury: a 1-year longitudinal study in young adults. <i>Behavior Therapy</i> , 42(4), 751–762.	Excluded behaviors with suicidal intent
8. Hanson, M. D., Niec, A., Pietrantonio, A. M., Johnson, S., Young, M., High, B., ... Eva, K. W. (2007). Effects associated with adolescent standardized patient simulation of depression and suicidal ideation. <i>Academic Medicine</i> , 82(10), S61–S64.	Did not report results from the implicit measure in sufficient detail to allow for the calculation of an effect size
9. Jollant, F., Guillaume, S., Jaussent, I., Bechara, A., & Courtet, P. (2013). When knowing what to do is not sufficient to make good decisions: Deficient use of explicit understanding in remitted patients with histories of suicidal acts. <i>Psychiatry Research</i> , 210(2), 485–490. <a href="http://doi.org/10.1016/j.psychres.2013.07.011">http://doi.org/10.1016/j.psychres.2013.07.011</a>	Employed a measure not included in the pre-selected list
9. Nock, M. K., & Banaji, M. (2007). Assessment of self-injurious thoughts using a behavioral test. <i>American Journal of Psychiatry</i> , 164(5), 820–823.	Excluded behaviors with suicidal intent
10. Violanti, J. M., Mnatsakanova, A., & Andrew, M. E. (2013). Suicidal ideation in police officers: Exploring an additional measure. <i>Suicidology Online</i> , 4, 33–41.	Homogeneous normative sample and absence of an experimental manipulation did not allow for group comparisons

## Appendix C: Acceptance and Action Questionnaire II

Below you will find a list of statements. Please rate how true each statement is for you by circling a number next to it. Use the scale below to make your choice.

1	2	3	4	5	6	7
never true	very seldom true	seldom true	sometimes true	frequently true	almost always true	always true

My painful experiences and memories make it difficult for me to live a life that I would value.	1	2	3	4	5	6	7
---	---	---	---	---	---	---	---

I'm afraid of my feelings.	1	2	3	4	5	6	7
----------------------------	---	---	---	---	---	---	---

I worry about not being able to control my worries and feelings.	1	2	3	4	5	6	7
--	---	---	---	---	---	---	---

My painful memories prevent me from having a fulfilling life.	1	2	3	4	5	6	7
---	---	---	---	---	---	---	---

Emotions cause problems in my life.	1	2	3	4	5	6	7
-------------------------------------	---	---	---	---	---	---	---

It seems like most people are handling their lives better than I am.	1	2	3	4	5	6	7
--	---	---	---	---	---	---	---

Worries get in the way of my success.	1	2	3	4	5	6	7
---------------------------------------	---	---	---	---	---	---	---

## Appendix D: Belief in the Afterlife Scale

Below you will find a list of statements. Please rate how true each statement is for you by circling a number next to it. Use the scale below to make your choice.

<b>Total disagreement</b>											<b>Total agreement</b>
<b>0</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>	<b>10</b>	
Earthly existence is the only existence we have.	0	1	2	3	4	5	6	7	8	9	10
There must be an afterlife of some sort.	0	1	2	3	4	5	6	7	8	9	10
The idea of there exists somewhere some sort of afterlife is beyond my comprehension.	0	1	2	3	4	5	6	7	8	9	10
Many scientists believe in a life after death: they are right, there is one.	0	1	2	3	4	5	6	7	8	9	10
A belief in an afterlife may be useful for some, but I don't believe in one at all.	0	1	2	3	4	5	6	7	8	9	10
There is supportive evidence for the existence of an afterlife.	0	1	2	3	4	5	6	7	8	9	10
Some existentialists claim that when man dies he ceases to exist: I agree.	0	1	2	3	4	5	6	7	8	9	10

## Appendix E: Death Anxiety Scale

Please read the statements carefully one by one. If the statement describes your attitudes for the, please **circle True**. If the statement does not describe your attitude, please **circle False**.

1.	I am very much afraid to die	True / False
2.	The thought of death seldom enters my mind	True / False
3.	It doesn't make me nervous when people talk about death	True / False
4.	I dread to think about having an operation	True / False
5.	I am not at all afraid to die	True / False
6.	I am not particularly afraid of getting cancer	True / False
7.	The thought of death never bothers me	True / False
8.	I am often distressed by the way time flies so very rapidly	True / False
9.	I fear dying a painful death	True / False
10.	The subject of life after death troubles me greatly	True / False
11.	I am really scared of having a heart attack	True / False
12.	I often think about how short life really is	True / False
13.	I shudder when I hear thinking about a World War III	True / False
14.	The sight of a dead body is horrifying to me	True / False
15.	I feel that the future holds nothing for me to fear	True / False

## Appendix F: Depression Anxiety Stress Scales

Please read each statement and circle a number 0, 1, 2 or 3 which indicates how much the statement applied to you **over the past week**

0	1	2	3
Did not apply to me at all	Applied to me to some degree, or some of the time	Applied to me to a considerable degree, or a good part of time	Applied to me very much, or most of the time
1	I found it hard to wind down	0	1 2 3
2	I was aware of dryness of my mouth	0	1 2 3
3	I couldn't seem to experience any positive feeling at all	0	1 2 3
4	I experienced breathing difficulty (eg, excessively rapid breathing, breathlessness in the absence of physical exertion)	0	1 2 3
5	I found it difficult to work up the initiative to do things	0	1 2 3
6	I tended to over-react to situations	0	1 2 3
7	I experienced trembling (eg, in the hands)	0	1 2 3
8	I felt that I was using a lot of nervous energy	0	1 2 3
9	I was worried about situations in which I might panic and make a fool of myself	0	1 2 3
10	I felt that I had nothing to look forward to	0	1 2 3
11	I found myself getting agitated	0	1 2 3
12	I found it difficult to relax	0	1 2 3
13	I felt down-hearted and blue	0	1 2 3
14	I was intolerant of anything that kept me from getting on with what I was doing	0	1 2 3
15	I felt I was close to panic	0	1 2 3
16	I was unable to become enthusiastic about anything	0	1 2 3
17	I felt I wasn't worth much as a person	0	1 2 3
18	I felt that I was rather touchy	0	1 2 3
19	I was aware of the action of my heart in the absence of physical exertion (eg, sense of heart rate increase, heart missing a beat)	0	1 2 3
20	I felt scared without any good reason	0	1 2 3
21	I felt that life was meaningless	0	1 2 3

## Appendix G: Informed Consent Form: Maynooth University Site

### INFORMED CONSENT FORM

In agreeing to participate in this research I understand the following:

- This research is being conducted by Ian Hussey, a postgraduate student at the Department of Psychology, National University of Ireland, Maynooth.
- The method proposed for this research project has been approved by the University Ethics Committee, which means that the Committee does not have concerns about the procedure itself as detailed by the student.
- If I have any concerns about participation I understand that I may refuse to participate or withdraw at any stage.
- I have been informed as to the general nature of the study and agree voluntarily to participate.
- I will complete a number of questionnaires, some of which will ask about symptoms of psychological suffering such as depression and anxiety, as well as a number of computer based pairing tasks where I will be asked to pair words related to death.
- There are no known expected risks associated with any aspect of participation, including the distress task.
- All data from the study will be treated confidentially. The data from all participants will be irrevocably anonymised, compiled, analysed, and submitted as part of a PhD thesis to the Department of Psychology. It may also be submitted for publication in an academic journal. The data will be retained for approximately 5 years before being destroyed. No participant's data will be identified by name at any stage of the data analysis or in the write up.
- At the conclusion of my participation, any questions or concerns I have will be fully addressed.
- I may withdraw from this study at any time without giving a reason, and may withdraw my data at the conclusion of my participation if I still have concerns.

---

Participant's signature

Researcher's signature

---

Participant name (please print)

Date

Should you be in distress or experiencing any form of mental health complaint we encourage you to contact the NUIM student counseling service. This professional, free, and confidential counseling service can be contacted on 01-7083554 or [counselling.nuim.ie](http://counselling.nuim.ie) to schedule an appointment.

**Appendix H: Consent Form & Researcher Checklist: St. Patrick's University Hospital Site**



**CONSENT FORM**

The participant must complete this form herself/himself

Title of Project: **Exploring the role of implicit cognition in suicidal ideation**

Protocol Number: 20/11

**PLEASE TICK YOUR RESPONSE IN THE APPROPRIATE BOX**

- I have read and understood the attached Participant information leaflet Yes  No
- I have had the opportunity to ask questions and discuss the study Yes  No
- I have received satisfactory answers to all my questions Yes  No
- I understand that I am free to withdraw from the study at any time without giving a reason and without this affecting my future medical care Yes  No
- I agree to take part in this study without prejudice to my legal or ethical rights Yes  No

**Patient ID Number:** \_\_\_\_\_

Participant's Signature: \_\_\_\_\_ Date: \_\_\_\_\_

Participant's Name in Print: \_\_\_\_\_

Witness Signature:\* \_\_\_\_\_ Date: \_\_\_\_\_

Witness' Name in Print: \_\_\_\_\_

Investigator's Signature: \_\_\_\_\_ Date: \_\_\_\_\_

Investigator's Name in Print: \_\_\_\_\_

**Please attach the Participant Information Sheet to this Consent Form, ask the participant to sign and date it and, where appropriate, place a copy of both in the participant's case notes.**

\*Witness must be somebody other than the Investigator



---

## Researcher checklist for consent form

---

Date: \_\_\_\_\_ Participant name: \_\_\_\_\_ Sex: M/F

DOB: \_\_\_\_\_ No.: \_\_\_\_\_ Patient ID Number: \_\_\_\_\_

Name of investigator obtaining consent: Ian Hussey

1. Have you given the participant an oral explanation of the proposed research project? Yes/No
  - 1.2. Have you given the information sheet to the participant? Yes/No
  - 1.3. Have you told the participant that he/she will be kept informed of all relevant information that becomes available during the course of the study? Yes/No
2. Did your oral explanation to the participant include:
  - 2.1. that this is a research project? Yes/No
  - 2.2. that participation is voluntary? Yes/No
  - 2.3. the aims of the project? Yes/No
  - 2.4. the likely duration of the participant's involvement? Yes/No
  - 2.5. the expected benefits to the participant and/or others? Yes/No
  - 2.6. the procedures which will be involved in participation? Yes/No
  - 2.7. what risks, inconvenience, discomfort or distress may reasonably be anticipated for this participant: the level and the likelihood? Yes/No
  - 2.8. that a refusal to participate may be given without reasons and will not affect the participant's rights or their right to care? Yes/No
  - 2.9. that the participant may be withdrawn from the study if the study investigator considers this is necessary in the best interests of the participant? Yes/No
  - 2.10. that personal information may be scrutinised during audit by competent authorities and properly authorised people, but all personal information will be treated as strictly confidential and will not be made publicly available? Yes/No
  - 2.11. that information generated by the study may be published but that no details will be divulged from which the participant could be identified? Yes/No
  - 2.12. that some such information will be retained for a period after the end of the trial? Yes/No
  - 2.13. what compensation arrangements are available? Yes/No
  - 2.14. whom to contact in an emergency and how? Yes/No
3. Have you allowed the participant sufficient time to consider the matter on his/her own, to discuss with others if wished, or ask you questions? Yes/No
4. In your opinion, has the participant understood and consented to take part in this research? Yes/No



# Experimenter Script

for the Implicit Relational Assessment Procedure

**Version**

2.0

**Author**

Ian Hussey  
Department of Psychology  
John Hume building  
National University of Ireland Maynooth  
Kildare  
Ireland  
[ian.Hussey@nuim.ie](mailto:ian.Hussey@nuim.ie)

**License**

This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.



Due to the fact that the IRAP requires participants to remember and follow a rule rather than simply categorise stimuli, it is objectively harder than tasks such as an IAT. As such, the large majority of published IRAP work has been conducted one-to-one with an experienced researcher. Attempting to deliver the IRAP in groups, or remotely, typically results in very large attrition rates or unusable data, although this has not been systematically explored. As such, we recommend that the researchers deliver the IRAP, as it is currently constructed, in a one-to-one manner with verbal instruction from a trained researcher.

We have found that for participants to complete the IRAP successfully they need to learn the following steps in this specific order. If, for example, participants try to go quickly before first learning to go accurately, or if they try to be accurate without first understanding that they must be accurate to the rule rather than their own opinion, they are unlikely to complete the IRAP successfully. If a participant is responding very quickly at the sacrifice of accuracy, the experimenter needs to undermine this quickly and effectively. The experiment may stop them mid-practice block and emphasise that they need to concentrate on being accurate first before attempting to go quickly.

1. Conceptualization – Does the participant understand that they're not being asked for their personal opinion, but rather to follow a rule?
2. Accuracy - Can they do this accurately?
3. Speed - Are they getting faster with practice?

The following is a list of what we think are the key 'ingredients' that need to be included in the experimenter's verbal frontend to get participants through the IRAP successfully. They aren't always delivered in the order they're presented here, but are used as needed (**text in red is dependent on the IRAP stimulus set**).

- Unlike a questionnaire, which asks you for your personal opinion, this task just asks you to follow a rule. For the moment, that rule is "**flowers are positive and insects are negative**".
- You will see words related to either **flowers** or **insects** at the top of the screen, and positive and negative words such as "**health**" or "**murder**" in the middle of the screen. You can respond with either "**True**" or "**False**". According to the rule – and not necessarily what you believe – is this first trial "**True**" or "**False**"? What was the rule, again?
- This is just a pairing task. Go as slowly as you need to get them all right according to the rule. You'll naturally start to go faster when you've learned to be accurate.
- Unlike a questionnaire, where you can give whatever answer you want to, in this task if you get one incorrect according to the rule you'll see a red X. Simply give it the correct answer to continue.
- After every block the rule swaps, there are only two rules. As you can see, now the rule is "**flowers are negative and insects are positive**".
- Well done. Keep going until the task is complete. Keep being as accurate as you can, and when you're accurate you'll naturally go quickly. You can take breaks during the feedback screens if you need to.

## Appendix J: Rosenberg Self-Esteem Scale

Below is a list of statements dealing with your general feelings about yourself. Please rate how much you agree with each statement by circling a number next to it. Use the scale below to make your choice.

<b>Strongly Agree</b>	<b>Agree</b>	<b>Disagree</b>	<b>Strongly Disagree</b>
<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>

1.	I feel that I am a person of worth, at least on an equal plane with others.	1	2	3	4
2.	I feel that I have a number of good qualities.	1	2	3	4
3.	All in all, I am inclined to feel that I am a failure.	1	2	3	4
4.	I am able to do things as well as most other people.	1	2	3	4
5.	I feel I do not have much to be proud of.	1	2	3	4
6.	I take a positive attitude toward myself.	1	2	3	4
7.	On the whole, I am satisfied with myself.	1	2	3	4
8.	I wish I could have more respect for myself.	1	2	3	4
9.	I certainly feel useless at times.	1	2	3	4
10.	At times I think I am no good at all.	1	2	3	4

## Appendix K: Attitudes Towards Women Scale

Below you will find a list of statements. Please rate how true each statement is for you by circling a number next to it. Use the scale below to make your choice.

1	2	3	4
Agree strongly	Agree mildly	Disagree mildly	Disagree strongly
1. Swearing and obscenity are more repulsive in the speech of a woman than of a man.	1	2	3 4
2. Women should take increasing responsibility for leadership in solving the intellectual and social problems of the day.	1	2	3 4
3. Both husband and wife should be allowed the same grounds for divorce.	1	2	3 4
4. Telling dirty jokes should be mostly a masculine prerogative.	1	2	3 4
5. Intoxication among women is worse than intoxication among men.	1	2	3 4
6. Under modern economic conditions with women being active outside the home, men should share in household tasks such as washing dishes and doing the laundry.	1	2	3 4
7. It is insulting to women to have the "obey" clause remain in the marriage service.	1	2	3 4
8. There should be a strict merit system in job appointment and promotion without regard to sex.	1	2	3 4
9. A woman should be free as a man to propose marriage.	1	2	3 4
10. Women should worry less about their rights and more about becoming good wives and mothers.	1	2	3 4
11. Women earning as much as their dates should bear equally the expense when they go out together.	1	2	3 4
12. Women should assume their rightful place in business and all the professions along with men.	1	2	3 4
13. A woman should not expect to go to exactly the same places or to have quite the same freedom of action as a man.	1	2	3 4
14. Sons in a family should be given more encouragement to go to college than daughters.	1	2	3 4
15. It is ridiculous for a woman to run a locomotive and for a man to darn socks.	1	2	3 4
16. In general, the father should have greater authority than the mother in the bringing up of children.	1	2	3 4
17. Women should be encouraged not to become sexually intimate with anyone before marriage, even their fiancés.	1	2	3 4
18. The husband should not be favored by law over the wife in the disposal of family property or income.	1	2	3 4
19. Women should be concerned with their duties of childbearing and house tending rather than with desires for professional or business careers.	1	2	3 4

20. The intellectual leadership of a community should be largely in the hands of men.	1	2	3	4
21. Economic and social freedom is worth far more to women than acceptance of the ideal of femininity which has been set up by men.	1	2	3	4
22. On the average, women should be regarded as less capable of contributing to economic production than are men.	1	2	3	4
23. There are many jobs in which men should be given preference over women in being hired or promoted.	1	2	3	4
24. Women should be given equal opportunity with men for apprenticeship in the various trades.	1	2	3	4
25. The modern girl is entitled to the same freedom from regulation and control that is given to the modern boy.	1	2	3	4

## Appendix L: Likelihood to Sexually Harass Scale

### Instructions

On the sheets that follow you will find 10 brief scenarios that describe 10 different interactions between males and females. In each case you will be asked to imagine that you are the main male character in the scenario. Then you will be asked to rate how likely it is that you would perform each of several different behaviors in the described social context. Assume in each scenario that no matter what you choose to do, nothing bad would be likely to happen to you as a result of your action. Try to answer each question as honestly as you can. Your answers will be completely anonymous. No one will ever try to discover your identity, no matter what you say on the questionnaire.

### Scenario #1

Imagine that you are an executive in a large corporation. You are 42 years old. Your income is above average for people at your job level. You have had numerous job offers from other companies. You feel very secure in your job. One day your personal secretary decides to quit her job and you have the task of replacing her. The personnel department sends several applicants over for you to interview. All seem to be equally qualified for the job. One of the applicants, Michelle S., explains during her interview that she desperately needs the job. She is 23 years old, single and has been job hunting for about a month. You find yourself very attracted to her. She looks at you in a way that possibly conveys she is also attracted to you. How likely are you to do the following things in this situation?

*a. Would you give her the job over the other applicants? (Circle a number to indicate your response.)*

1	2	3	4	5
Not at all likely				Very likely

*b. Assuming that you are secure enough in your job that no possible reprisals could happen to you, would you offer her the job in exchange for sexual favors? (Circle a number to indicate your response.)*

1	2	3	4	5
Not at all likely				Very likely

c. Assuming that you fear no reprisals on your job, would you ask her to meet you later for dinner to discuss her possible employment?

1	2	3	4	5
Not at all likely				Very likely

**Scenario #2**

Imagine that you are the owner and manager of an expensive restaurant. One day, while going over the receipts, you discover that one of the waitresses has made some errors in her checks. She has undercharged several customers. The mistake costs you \$100. In talking to some of the other employees, you find that the particular customers involved were friends of the waitress. You call her into your office and ask her to explain her behavior. The waitress confesses to having intentionally undercharged her friends. She promises that she will never repeat this dishonest act and tells you that she will do anything to keep her job. The waitress is someone you have always found particularly attractive. She is a divorcee and about 25 years old. How likely are you to do the following things in this situation?

a. Would you let her keep her job?

1	2	3	4	5
Not at all likely				Very likely

b. Would you let her keep her job in exchange for sexual favors?

1	2	3	4	5
Not at all likely				Very likely

c. Would you ask her to meet you for dinner after work to discuss the problem?

1	2	3	4	5
Not at all likely				Very likely

**Scenario #3**

Imagine that you are the manager of a shipping company. One day your supervisor asks you to study the possibility of buying several computers for the office. You call up several competing companies that sell computers. Each company sends a sales representative over to your office who describes the company's products. A salesperson from company "A" calls you and asks to come to your office. You agree and the next day a very attractive woman shows up. She can offer no real reason for buying her company's products over those of the other companies. However, she seems very sexy. How likely are you to do the following things in this situation?

a. *Would you recommend her line of computers?*

<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
<b>Not at all likely</b>				<b>Very likely</b>

b. *Assuming that you are secure enough in your job that no possible reprisals could happen to you, would you agree to recommend her line of computers in exchange for sexual favors?*

<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
<b>Not at all likely</b>				<b>Very likely</b>

c. *Given the same assumptions as the last question above, would you ask her to meet you later for dinner to discuss the choice of computers?*

<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
<b>Not at all likely</b>				<b>Very likely</b>

#### **Scenario #4**

Imagine that you are a Hollywood film director. You are casting for a minor role in a film you are planning. The role calls for a particularly stunning actress, one with a lot of sex appeal. How likely are you to do the following things in this situation?

a. *Would you give the role to the actress whom you personally found sexiest?*

<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
<b>Not at all likely</b>				<b>Very likely</b>

b. *Would give the role to an actress who agreed to have sex with you?*

<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
<b>Not at all likely</b>				<b>Very likely</b>

c. Would ask the actress to whom you were most personally attracted to talk with you about the role over dinner?

1	2	3	4	5
Not at all likely				Very likely

### Scenario #5

Imagine that you are the owner of a modeling agency. Your agency specializes in sexy female models used in television commercials. One of your models, Amy T., is a particularly ravishing brunette. You stop her after work one day and ask her to have dinner with you. She coldly declines your offer and tells you that she would like to keep your relationship with her "strictly business." A few months later you find that business is slack and you have to lay off some of your employees. You can choose to lay off Amy or one of four other women. All are good models, but someone has to go. How likely are you to do the following things in this situation?

a. Would you fire Amy?

1	2	3	4	5
Not at all likely				Very likely

b. Assuming that you are unafraid of possible reprisals, would you offer to let Amy keep her job in return for sexual favors?

1	2	3	4	5
Not at all likely				Very likely

c. Would you ask Amy to dinner so that you could talk over her future employment?

1	2	3	4	5
Not at all likely				Very likely

### Scenario #6

Imagine that you are a college professor. You are 38 years old. You teach in a large midwestern university. You are a full professor with tenure. You are renowned in your field (Abnormal Psychology) and have numerous offers for other jobs. One day following the return of an examination to a class, a female student stops in your office. She tells you that her score is one point away from an "A" and asks you if she can do some extra credit project to raise her score. She tells you that she may not have a sufficient grade to get into graduate school without the "A." Several other students have asked you to do extra credit assignments and you have declined to let

them. This particular woman is a stunning blonde. She sits in the front row of the class every day and always wears short skirts. You find her extremely sexy. How likely are you to do the following things in this situation?

a. *Would you let her carry out a project for extra credit (e.g. write a paper)?*

1	2	3	4	5
Not at all likely				Very likely

b. *Assuming that you are very secure in your job and the university has always tolerated professors who make passes at students, would you offer the student a chance to earn extra credit in return for sexual favors?*

1	2	3	4	5
Not at all likely				Very likely

c. *Given the same assumptions as in the question above, would you ask her to join you for dinner to discuss the possible extra credit assignments?*

1	2	3	4	5
Not at all likely				Very likely

### Scenario #7

Imagine that you are a college student at a large Midwestern university. You are a junior who just transferred from another school on the East coast. One night at a bar you meet an attractive female student named Rhonda. Rhonda laments to you that she is failing a course in English Poetry. She tells you that she has a paper due next week on the poet, Shelley, and fears that she will fail since she has not begun to write it. You remark that you wrote a paper last year on Shelley at your former school. Your paper was given an A+. She asks you if you will let her use your paper in her course. She wants to just retype it and put her name on it. How likely are you to do the following things in this situation?

a. *Would you let Rhonda use your paper?*

1	2	3	4	5
Not at all likely				Very likely

b. *Would you let Rhonda use your paper in exchange for sexual favors?*

1	2	3	4	5
Not at all likely				Very likely

c. *Would you ask Rhonda to come to your apartment to discuss the matter?*

1	2	3	4	5
Not at all likely				Very likely

### Scenario #8

Imagine that you are the editor for a major publishing company. It is your job to read new manuscripts of novels and decide whether they are worthy of publication. You receive literally hundreds of manuscripts per week from aspiring novelists. Most of them are screened by your subordinates and thrown in the trash. You end up accepting about one in a thousand for publication. One night you go to a party. There you meet a very attractive woman named Betsy. Betsy tells you that she has written a novel and would like to check into getting it published. This is her first novel. She is a dental assistant. She asks you to read her novel. How likely are you to do the following things in this situation.

a. *Would you agree to read Betsy's novel?*

1	2	3	4	5
Not at all likely				Very likely

b. *Would you agree to reading Betsy's novel in exchange for sexual favors?*

1	2	3	4	5
Not at all likely				Very likely

c. *Would you ask Betsy to have dinner with you the next night to discuss your reading her novel?*

1	2	3	4	5
Not at all likely				Very likely

### Scenario #9

Imagine that you are a physician. You go over to the hospital one day to make your rounds visiting your patients. In looking over the records of one of your patients, you discover that one of the attending nurses on the previous night shift made an error in administering drugs to your patient. She gave the wrong dosage of a drug. You examine the patient and discover that no harm was actually done. He seems fine. However, you realize that the ramifications of the error could have been catastrophic under other circumstances. You pull the files and find out who made the error. It

turns out that a new young nurse named Wendy H. was responsible. You have noticed Wendy in some of your visits to the hospital and have thought of asking her out to dinner. You realize that she could lose her job if you report this incident. How likely are you to do each of the following things?

a. *Would you report Wendy to the hospital administration?*

1	2	3	4	5
Not at all likely				Very likely

b. *Assuming that you fear no reprisals, would you tell Wendy in private that you will not report her if she will have sex with you?*

1	2	3	4	5
Not at all likely				Very likely

c. *Assuming that you fear no reprisals, would you ask Wendy to join you for dinner to discuss the incident?*

1	2	3	4	5
Not at all likely				Very likely

## Scenario #10

Imagine that you are the news director for a local television station. Due to some personnel changes you have to replace the anchor woman for the evening news. Your policy has always been to promote reporters from within your organization when an anchor woman vacancy occurs. There are several female reporters from which to choose. All are young, attractive, and apparently qualified for the job. One reporter, Loretta W., is someone whom you personally find very sexy. You initially hired her, giving her a first break in the TV news business. How likely are you to do the following things in this situation?

a. *Would give Loretta the job?*

1	2	3	4	5
Not at all likely				Very likely

b. *Assuming that you fear no reprisals in your job, would you offer Loretta the job in exchange for sexual favors?*

1	2	3	4	5
Not at all likely				Very likely

*c. Assuming that you fear no reprisals in your job, would you ask her to meet you after work for dinner to discuss the job?*

<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
<b>Not at all likely</b>				<b>Very likely</b>

## **Appendix M: Mortality Attitudes Personality Survey**

*Please take a few minutes to answer the following questions.*

**Q1. What emotions does the thought of your own death arouse in you?**

**Q2. Jot down, as specifically as you can, what you think will happen to you physically as you die and once you are physically dead.**

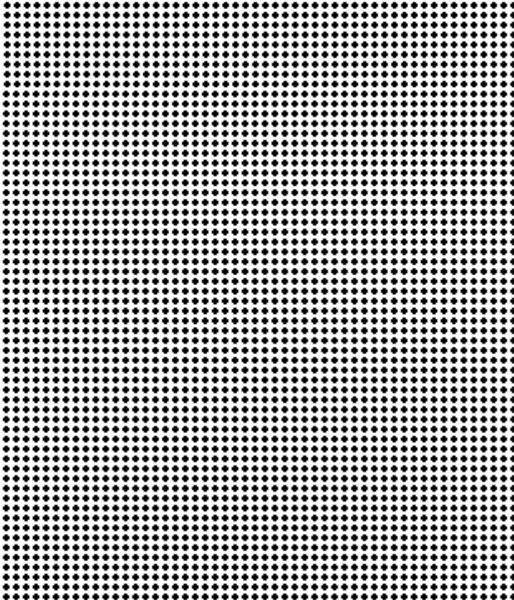
**Q3. “The one thing I fear most about my death is...”**

**Q4. “My scariest thoughts about death are...”**

**Appendix N: Mortality Salience Induction: Example “Weeks to Live” Sheet**

Female

Age 20



## Appendix O: Mortality Salience Induction Script

(Notes to experimenter in italics)

I think it's often very easy to forget just how short life is, especially for young healthy students. To help convey this, I've put together this diagram for you. Given that I know your age and gender, it's trivial for me to estimate your expected lifespan.

*[Place weeks to live sheet in front of participant, say very slowly and carefully.]*

Based on that, the number of dots on this piece of paper is equal to the number of weeks you have left to live.

*[Long pause.]*

I promise that I'm not trying to trick or deceive you. It's a surprisingly small number of dots, isn't it? And, the thing about dots is that once you spend them you can't get them back. This is not a rehearsal; you will not get a second shot. This is your life, right now, ending, one day at a time. The other thing about dots is that they run out, no matter what you do. Make no mistake: death *is* coming. You have a limited number of days left on this planet, and, like all of us, you're faced with the difficult question of what you're going to do with them. How many of these dots will be "well-spent" dots, doing things that you truly value, and how many dots will you feel you have wasted? The dots don't care either way; they just run out. With all of that in mind, I'd like to ask you to write a few lines about what you think about death, and what dying will be like.

## Appendix P: Self-Injurious Thoughts and Behaviours Interview (modified self-report format)

### Non-Suicidal Self-Injury

Have you ever hurt yourself without wanting to die? (for example, cutting or burning) *We will refer to this as non-suicidal self-injury (NSSI).*

- No / Yes
- More than a year ago
- In the last year
- In the last 6 months
- In the last 1 month
- In the last 1 week

When was the last time?

### Ideation

Have you ever had thoughts about killing yourself?

- No / Yes
- More than a year ago
- In the last year
- In the last 6 months
- In the last 1 month
- In the last 1 week

When was the most recent time?

### Planning

Have you ever made a plan to kill yourself?

- No / Yes
- More than a year ago
- In the last year
- In the last 6 months
- In the last 1 month
- In the last 1 week

When was the **most recent** time you made a plan to kill yourself?

### Attempts

Have you ever made an actual attempt to kill yourself? *We will refer to this as an attempt.*

- No / Yes
- More than a year ago
- In the last year
- In the last 6 months
- In the last 1 month
- In the last 1 week

When was the **most recent** attempt?

## Appendix Q: Letter of ethical approval: Maynooth University

NATIONAL UNIVERSITY OF IRELAND, MAYNOOTH  
MAYNOOTH, CO. KILDARE, IRELAND



NUI MAYNOOTH  
Ollscoil na hÉireann Mú Nuad

Dr Carol Barrett  
Secretary to NUI Maynooth Ethics Committee

23 April 2012

Ian Hussey  
Psychology Department  
NUI Maynooth

RE: *Application for Ethical Approval for a project entitled:*  
"Exploring the role of implicit cognition in suicidal ideation"

Dear Ian,

The Ethics Committee evaluated the above project and we would like to inform you that ethical approval has been granted.

Kind Regards,



---

Dr Carol Barrett  
Secretary, NUI Maynooth Ethics Committee

cc. Dermot Barnes-Holmes, Psychology Department

## Appendix R: Letter of ethical approval: St. Patrick's University Hospital



Please reply to:  
St. Patrick's University Hospital, P.O. Box 136, James's St., Dublin 8.  
Tel: 01 249 3200 Fax: 01 249 3417  
www.stpatrickshosp.com

PRIVATE & CONFIDENTIAL

Professor James V. Lucey  
M.D., Ph.D., FRCPI, FRCPsych,  
Medical Director  
Research Council 00646  
15<sup>th</sup> February 2012

Mr Ian Hussey  
Dept of Psychology  
John Hume Building  
NUI Maynooth  
Co. Kildare  
([ian.hussey@nuim.ie](mailto:ian.hussey@nuim.ie))

**Re: Exploring the Role of Implicit Cognition in Suicidal Ideation  
(Protocol No. 20/11)**

Dear Mr Hussey,

We are pleased to inform you that your revised documentation has been granted approval and research can commence immediately.

Approval is granted subject to the following standard conditions:

1. You must adhere fully to the terms and conditions set out in your research protocol.
2. Any persons involved in this research who are not employees of St. Patrick's University Hospital are required to obtain an honorary contract from the hospital. This can be sought through Ms. Mairead Flanagan, HR Department (01 2493619; [mairreadflanagan@stpatmail.com](mailto:mairreadflanagan@stpatmail.com)).
3. If there are any material changes to be made to Protocol 20/11 in the next 12 months, you must contact the Research Ethics Committee for approval.
4. You must report back to the Research Ethics Committee no later than 12 months subsequent to this approval letter (February 15<sup>th</sup>, 2013), with a summary report on the progress of this research. A template for this report can be downloaded from the hospital website. Failure to complete this report will result in ethical approval being withdrawn for your research.

We thank you for your cooperation. With very best wishes.

Yours sincerely,

Prof. James V. Lucey MD., Ph.D., FRCPI., FRCPsych.,  
Secretary to the Research Ethics Committee  
Medical Council 00646

cc. Prof. Dermot Barnes-Holmes, Academic Supervisor  
Dr. Richard Booth, Clinical Supervisor