Clarifying Psychological Risk Factors for Self-Injury and Suicidal Behaviors: Clinical Applications of Behavioral Measures

Clarifying Psychological Risk Factors for Self-Injury and Suicidal Behaviors: Clinical Applications of Behavioral Measures

A dissertation presented

by

Christine Boram Cha

to

The Department of Psychology

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Clinical Psychology

Harvard University

Cambridge, Massachusetts

May 2015

Advisor: Matthew K. Nock Author: Christine Boram Cha

Clarifying Psychological Risk Factors for Self-Injury and Suicidal Behaviors:

Clinical Applications of Behavioral Measures

Abstract

Self-injurious thoughts and behaviors are life-threatening, prevalent, and challenging clinical outcomes to predict. This dissertation explores the use of behavioral measures to improve prediction of nonsuicidal self-injury (NSSI) and suicidal ideation. This builds on the growing body of literature supporting the clinical application of behavioral measures such as the Implicit Association Test (IAT) and emotional Stroop Task with self-injurious and suicidal individuals. I tackle three questions to inform continued research and application. First, is it safe to administer behavioral measures related to self-injurious thoughts and behaviors? Study 1 reveals that there is minimal change in self-injurious or suicidal urges from before to after completing Suicide and Self-Injury IATs. This was found across three distinct samples. A small to moderate mood decline was consistently detected, which was isolated to female respondents and one type of IAT that presented NSSI-related images. Female participants' negative mood after viewing NSSI-related images appeared to be transient in nature--possibly be alleviated by viewing positive images. Second, can the Self-Injury IAT be used in acute care settings to predict NSSI? Study 2 shows that the Self-Injury IAT can be used for short-term prediction: it predicted NSSI occurring during hospital stays above and beyond other risk factors. Surprisingly, patients' explicit self-report was a more robust predictor of NSSI than the Self-Injury IAT. Admission-to-discharge change in patients' explicit self-report, but not IAT performance, predicted whether they engaged in NSSI after hospital discharge. Third, how does transient mood affect the predictive validity of the Suicide IAT and Stroop task? In Study 3, suicide ideators demonstrated significantly stronger

implicit identification with death after (vs. before) the mood induction, as indicated by post-induction IAT performance. Controlling for history of suicidal ideation, post-induction performance was most predictive of suicidal ideation when assessed categorically (i.e., identification with Death vs. Life). Suicide Stroop performance remained unrelated to suicidal ideation on its own, but enhanced prediction of suicidal ideation when combined with Suicide IAT performance. All baseline suicide ideators who achieved one particular type of IAT/Stroop scoring profile experienced suicidal thoughts six months later. This collection of studies balances clinical application and psychological science, and presents a number of important considerations for future research and practice.

Table of Contents

Acknowledgements

First and foremost, I am grateful for the support, dedication, wisdom, and humor of my graduate advisor and mentor, Matthew Nock. He and his work have been an inspiration to me over the past ten years, and I am incredibly honored to have been a part of his lab. I am also grateful to the other members of my dissertation committee for their invaluable feedback and guidance: Jill Hooley, Richard McNally, and Mahzarin Banaji. I thank my collaborators through Project Implicit Mental Health (Bethany Teachman), Children's Hospital Boston (Eugene D'Angelo), and the University of Glasgow (Rory O'Connor). Completion of this research was facilitated by funding from the Sackler Fellowship in Psychobiology, Biddle Funds and Merit Fellowship from Harvard University, Norlien Foundation, the Linda and Timothy O'Neill Foundation, and the John D. and Catherine T. MacArthur Foundation. Importantly, many thanks to all the members of the Laboratory for Clinical and Developmental Research, especially Charlene Deming, Jeffrey Glenn, Julia Harris, Mark Knepley, and Katarina Slama, and a phenomenal team of research assistants: Britta Ekdahl, Katherine Frost, Katie Gallagher, Amy Harmon, Julia Harris, Gizem Iskenderoglu, Ryn Linthicum, and Rachel Snow. I thank my early mentors, Susan Nolen-Hoeksema and Lori Hilt, for believing in me and strongly encouraging me to pursue this career. I thank my colleagues and dear friends, Nancy Lau and Mei Yi Ng, for help with this dissertation and beyond. Lastly, I could not have completed this project without the constant love, patience, and sacrifice of my family: especially my parents, Bruce and Kim Cha, my brother, David Cha, and my best friend and fiancé, Dustin Saldarriaga.

INTRODUCTION

Self-injurious thoughts and behaviors (SITBs) are serious health risk outcomes that require greater attention and treatment. At the most extreme end, suicidal thoughts and behaviors are directly life-threatening. They consist of suicide attempt, which are deliberate actions to end one's own life, and suicidal ideation, which are serious thoughts or desires to kill oneself. They are relevant precipitants of suicide, which results in nearly one million deaths each year. This statistic is particularly concerning since prevalence rates have remained persistently high despite significant increases in treatment (Kessler, Berglund, Borges, Nock, & Wang, 2005).

Another concern is that nearly 6% of adults in the United States have engaged in nonsuicidal self-injury (NSSI), which is defined as direct and intentional damage to one's body tissue in the absence of any intent to die (Klonsky, 2011). In addition to causing immediate physical harm, NSSI increases the likelihood of making a suicide attempt (Nock, Joiner, Gordon, Lloyd-Richardson, & Prinstein, 2006; Wilkinson, Kelvin, Roberts, Dubicka, & Goodyer, 2011; You & Lin, in press; Zlotnick, Donaldson, Spirito, & Pearlstein, 1997). Trends suggest a recent increase in NSSI, as nearly 20% of adults currently under the age of 30 have engaged in this harmful behavior at least once in their lives (Klonsky, 2011). This pattern may generalize to other self-injurious behaviors, as the Centers for Disease Control and Prevention (CDC, 2015) reports a gradual increase in nonfatal self-harm (suicidal and non-suicidal) injuries over the past decade (112.8 per 100,000, age-adjusted in 2001 to 160.5 per 100,000, age-adjusted in 2013). Given the increasing rates of self-injurious behaviors in this country, it remains imperative that we improve our ability to identify and ultimately treat at risk individuals.

Challenges to Assessing Self-Injurious Thoughts and Behaviors

There are inherent challenges to identifying those at risk of SITB. Historically, the past occurrence and future likelihood of SITBs have been assessed using various self-report

questionnaires and clinical interviews (e.g. Lloyd, Kelley, & Hope, 1997; Nock, Holmberg, Photos, & Michel, 2007; Linehan, Comtois, Brown, Heard, & Wagner, 2006). A critical limitation of explicit self-report is that respondents, if motivated, could readily misreport or underreport the truth. Just as people more actively manage the impression they make on others when discussing socially sensitive topics (Greenwald, Poehlman, Uhlmann, & Banaji, 2009), it has been shown that 78% of suicidal patients deny ideation during their last communication before dying by suicide (Busch, Fawcett, & Jacobs, 2003). A patient may underreport self-injurious urges in a treatment setting, if for instance, they believe it may prevent hospital admission or shorten their stay. Supporting this possibility, prior work reveals weak predictive validity of self-reported self-injury risk (i.e., suicidal or nonsuicidal self-injury; Janis & Nock, 2008), and low self-disclosure rates among suicidal adolescents (Horesh & Apter, 2006; Horesh, Zalsman, & Apter, 2004). An alternative explanation is that self-injurious individuals lack insight on whether they will or will not hurt themselves in the future. This latter possibility is supported by social psychology research demonstrating people's limited insight into their own mental processes (e.g., Nisbett & Wilson, 1977).

In addition to impression management concerns, psychological processes that perpetuate SITB may not be fully captured by explicit self-report. For instance, self-injurers[1] may continue engaging in NSSI because of their implicit identification with this behavior ("*I am a cutter*") over alternative behaviors (e.g., "*I am a runner*;" "*I am a smoker*;" Nock, 2009a). That is, they associate the concept of NSSI with themselves (i.e., *Me=Cutting*). Implicit associations occur automatically, meaning outside of conscious awareness and often times not detectible by explicit self-report. While a person may not be fully aware of them, implicit associations can guide future decisions and/or behaviors in a way that is critical to study. Indeed, they have been shown to

[1] People who engage in NSSI are hereafter referred to as self-injurers.

predict suicide attempt above and beyond explicit self-report (Nock, Park, Finn, Deliberto, Dour, & Banaji, 2010).

Behavioral Measures of SITB Risk

To address the limitations of explicit self-report, clinical science has begun to develop behavioral measures of SITB risk. Implicit identification with cutting, for example, can be objectively assessed using the Self-Injury Implicit Association Test (Self-Injury IAT; Nock & Banaji, 2007), a computerized behavioral measure that evaluates the strength of mental association a self-injurer holds between the constructs of NSSI and themselves. Based on the assumption that people pair related constructs more quickly than unrelated constructs, the Self-Injury IAT captures reaction times of endorsing one association (e.g., *"Cutting=Me"*) versus another (e.g., *"Cutting=Not Me"*). The IAT has tremendous potential as a clinical tool to determine how much a patient associates him/herself with SITB. The adapted Suicide IAT, tested by Nock and colleagues (2010), predicted future suicide attempt above and beyond well-known risk factors (e.g., mood disorder, prior suicide attempt) in addition to clinician and patient prediction.

Another behavioral measure of suicide risk is the Suicide Stroop task. This is an adapted version of the emotional Stroop task (Becker, Strohbach, & Rinck, 1999; Cha, Najmi, Park, Finn, & Nock, 2010; Williams, Mathews, & MacLeod, 1996), which gauges responsiveness to the semantic content of words. Based on the assumption that emotionally salient content interferes with one's ability to attend to other aspects of the word (e.g., font color), the Suicide Stroop task measures reaction times of identifying the color of suicide-related (vs. neutral) words. Earlier studies have demonstrated that *interference* for suicide-related words is associated with recent suicide attempt (Williams & Broadbent, 1986), and that this association is specific to suicide-

related content since no such interference appeared for general negatively-valenced words (Becker et al., 1999). A more recent study prospectively demonstrated that this task can predict future suicide attempt, controlling for well-known risk factors (e.g., mood disorder, prior suicide attempt) and clinician prediction (Cha et al., 2010). The Suicide Stroop task appears to be a promising tool, serving a similar function as the Suicide IAT.

Much effort has been geared toward demonstrating the predictive validity of these behavioral measures. To build on their functional importance, I identify whether these measures are: safe (Study 1), predictive and sensitive to change within a hospital setting (Study 2), and impacted by transient mood (Study 3). Results from these three studies offer critical knowledge to inform future application and dissemination of these clinical tools.

STUDY 1:

Examining Iatrogenic Effects of the Self-Injury and Suicide IAT

Distinct qualities of SITB research have historically raised ethical questions and concerns. For instance, SITB research involves studying vulnerable individuals, most of whom have a current mental disorder and heightened risk of future self-harm (Nock et al., 2006; Guan, Fox, & Prinstein, 2012; Klonsky, May, & Glenn, 2013). SITB research also involves the use of measures that contain sensitive and emotionally provocative content, such as questions about one's history of suicidal thoughts or behaviors. These features of SITB research increase perceived iatrogenic risk, and influence decisions of Human Research Ethics Committees when reviewing research projects addressing SITBs (Lakeman & Fitzgerald, 2009). The impact of these ethical concerns calls for empirical study of iatrogenic effects. Indeed, a recently convened NIMH consensus panel on the protection of vulnerable research participants concluded that evidence, rather than untested assumptions, should guide practices of Human Research Ethics Committees (DuBois et al., 2012). The current study responds to this call by examining the iatrogenic risk of more recently developed behavioral measures assessing SITBs.

Prior research has demonstrated that surveys that include questions about suicidal thoughts and behaviors do not increase individuals' subjective distress or subsequent risk of such behaviors overall (Gould et al., 2005). This result has been replicated in subsequent studies (Gibson, Boden, Benson, & Brand, 2014; Whitlock, Pietrusza, & Purington, 2013) and extended to NSSI-specific survey questions (Muehlenkamp, Swenson, Batejan, & Jarvi, 2015). As another example, there appears to be no difference in emotional impact between reading warning signs for suicide versus reading warning signs for heart attack or diabetes (Rudd et al., 2006).

Studies focusing on research participants with prior history of SITB yield mixed findings. On the one hand, participants may benefit from answering SITB-related questions. Suicidal adolescents and depressed adults responding to such questions have demonstrated an overall decline in suicidal ideation (Gould et al., 2005; Cukrowicz, Smith, & Poindexter, 2010). This extends to young adults engaging in moderate/severe NSSI, who report a relative decrease in subjective distress after responding to SITB-related questions (Muehlenkamp et al., 2015). On the other hand, answering SITB-related questions may be harmful to some individuals. A notable minority of higher-risk youth (e.g., those with a history of suicidal ideation or attempt) reportedly feel more upset after completing survey questions about their history of SITB (Robinson et al., 2011). Similar examination of iatrogenic effects across *all* types of SITB measures is sorely needed.

The design of behavioral measures, such as Self-Injury or Suicide IATs (Nock & Banaji, 2007; Nock et al., 2010), offer a novel set of iatrogenic concerns. This is due to their distinct design, as they repeatedly present SITB-related stimuli consisting of words or images pertaining to NSSI and suicide (e.g., the words *cutting* or *death*, images of bleeding skin or of a person standing on the ledge of a tall building). Although the use of such SITB-related stimuli has become increasingly wide-spread, there have not yet been any investigations testing their potential iatrogenic impact. Repeatedly showing dozens of SITB-related words and images in rapid succession may be more distressing than simply asking about such concepts, and may elicit new SITBs—especially among those with a history of such experiences.

Aims

The purpose of Study 1[2] was to investigate potential iatrogenic effects of viewing SITB-related stimuli. The lack of prior work in this area yields three exploratory aims:

1. Examine potential iatrogenic effects of viewing SITB-related stimuli among all research participants. This was done by measuring changes in people's self-injurious urges (i.e., *desire to self-injure*), suicidal urges (i.e., *desire to die*), and mood before and after IAT administration. Findings addressing this first aim most readily apply to studies implementing a case-control design, which are critical for identifying risk factors for SITB. Relatedly, the potential iatrogenic effect among participants with a history of SITB was also examined.
2. Isolate other characteristics that may heighten iatrogenic risk. For instance, it remains possible that male versus female research participants may be differentially affected. Prior empirical work suggests that male suicide attempters are more likely to believe that talking about suicide would make people more likely to attempt (Shaffer et al., 1990). In contrast, suicidal women may *benefit* from exposure to suicide prevention content through reduced negative mood (Bryan, Dhillon-Davis, & Dhillon-Davis, 2009). It was also tested whether a particular set of stimuli (e.g., images vs. words) heightens iatrogenic risk.
3. Examine changes in participants' mood in response to SITB-related stimuli. It was tested whether viewing SITB-related images, or more specifically NSSI-related images, has an incremental impact on mood over that associated with viewing negatively-valenced images more generally. It was also explored whether any

[2] Since four studies are embedded within Study 1, these studies are referred to as *Studies 1a-d* within this section. They are collectively referred to as *Study 1* throughout the rest of the text (e.g., Introduction, General Discussion).

increase in negative mood is transient, and can possibly be alleviated by showing positively-valenced images. Results from addressing this aim could provide guidelines for conducting SITB research more safely in the future.

These aims were collectively addressed across four studies. This collection of four studies offered the chance to yield robust, generalizable findings about the iatrogenic risk of innovative SITB research. Study 1a featured web-based respondents who completed a SITB-related IAT. This first study helps address the increasing implementation of online data collection methods (Buchanan & Hvizdak, 2009), and yielded a large sample size (*n*=3,304). This sample size provided sufficient statistical power to detect even the smallest iatrogenic effects and to inform subsequent studies. Studies 1b and 1c examined in-person IAT administration in nonclinical and clinical settings: with undergraduate students (*n*=100), and with adolescent psychiatric inpatients (*n*=83), respectively. Finally, Study 1d featured female adult research participants recruited for a neuroimaging study (*n*=30). The unique design of this last study allowed pursuit of the aforementioned third aim, since the research participants rated their mood after they were shown blocks of images varying in emotional valence and content (neutral, NSSI-related, negative, positive).

Method

Samples, measures and data analytic for Studies 1a-c are collectively described due to their similarities. Study 1d is described separately given its distinct aim.

Samples and Measures

Study 1a. Data for Study 1a were from a large-scale web-based IAT study aimed at testing group differences between self-injurious and non-injurious adults. Participants were 3,304 adults (*M*=28.4 years, *SD*=11.3) from 105 countries (64.2% U.S. residents) who voluntarily

visited the *Project Implicit Mental Health* (PIMH) website (www.ImplicitMentalHealth.com) and completed a brief SITB-related IAT (Table 1.1). All participants provided informed consent to completing an IAT related to self-injury, suicide, or death. See Appendix for additional Study 1a sample and IAT description.

In order to test for potential iatrogenic effects, a set of three self-report items was given to each participant before and after each IAT that assessed: (i) desire to self-injure (*How much do you want to hurt yourself right now?*); (ii) desire to die (*How much do you want to die right now?*); and (iii) mood (*How would you rate your mood right now?*). The first two questions were rated on a five-point scale ranging from 0 (*not at all*) to 4 (*extremely*). In the description of results below, reports of *increased* desire to self-injure or desire to die indicate that post-IAT desire to self-injure/die was stronger than pre-IAT desire to self-injure/die (and vice-versa). Further, mood was rated on a seven-point scale ranging from -3 (*extremely positive*) to +3 (*extremely negative*); these items were reverse-coded to maintain consistency of presentation, so that -3 represented negative mood and +3 represented positive mood. Reports of *mood decline* indicate that post-IAT mood was relatively less positive than pre-IAT mood, and reports of *mood improvement* indicate that post-IAT mood was relatively more positive than pre-IAT mood.

History of SITB was measured using an abbreviated self-report version of the Self-Injurious Thoughts and Behaviors Interview (SITBI; Nock et al., 2007). See Studies 1b and 1c for a more detailed description of the SITBI. This self-report version of the SITBI assessed NSSI, suicide ideation, suicide plan, and suicide attempt. Participants also provided demographic information through a survey on the website. For prevention purposes, any participant indicating extreme desire to self-injure or die when answering questions on the PIMH website received a follow-up message encouraging the participant to seek help and providing mental health

resources to consider. See Werntz et al. (2015) for additional information on Study 1a sample and measures.

Study 1b. Data for Study 1b were from a laboratory-based pilot study testing nine different SITB-related IATs, described in the Appendix. Participants for Study 1b were 100 consenting students enrolled in an undergraduate psychology course at a university in the Northeast.

The wording of pre/post-IAT questions was identical to Study 1a, but the rating scales were different. Participants responded to questions about desire to self-injure and desire to die along a 10-point scale (1=*not at all*, 5=*somewhat*, 10=*very much*), and about mood along a similar 10-point scale (1=*negative*, 5=*neutral*, 10=*positive*).

Respondents completed the IATs in a laboratory room. They then completed self-report questions from the SITBI (Nock et al., 2007) and provided basic demographic information. If the respondent indicated any SITB history, a suicide risk assessment interview took place at the conclusion of the visit.

Study 1c. Data for Study 1c are from a hospital-based study using a battery of seven IATs to examine between-group differences among self-injurious and non-injurious psychiatric hospital inpatients. Study 1c participants were 83 adolescents from a psychiatric inpatient unit at a metropolitan children's hospital. Data from patients' electronic medical records revealed the following current psychiatric diagnoses: mood disorder (78.7%), anxiety disorder (34.7%), eating disorder (30.7%), impulse control disorder (13.3%), alcohol/substance use disorder (13.3%), and psychotic disorder (2.7%). Patients' Global Assessment of Functioning scores at admission were low (M=31.8, SD=6.7) and, on average, patients had at least one previous psychiatric hospitalization (M=1.1, SD=1.7). See Appendix for additional sample information.

Study 1c used similar measures as Studies 1a-b. The pre-/post-IAT questions for Study 1c were identical to those used in Study 1b. Seven distinct IATs were administered to patients. They included the *Suicide* and *Death*-related IATs from Study 1b, and the *Self-Injury IAT* from Study 1a. Self-injury was measured via the SITBI (Nock et al., 2007). Finally, demographic information and psychiatric history were gathered from medical records, with patients' consent.

Table 1.1. Demographic information and SITB history across IAT Studies

	Study 1a	**Study 1b**	**Study 1c**
	Web-based Respondents (*n*=3,304)	College Students (*n*=100)	Adolescent Inpatients (*n*=83)
Age, *M* (*SD*)	28.4 (11.3)	20.1 (2.8)	14.8 (1.5)
Gender (% female)	67.1	63.0	67.5
Race/Ethnicity (%)			
Caucasian/White	77.4	55.0	85.5
Asian	6.7	27.0	1.2
Biracial/Multiple	5.0	7.0	1.2
Hispanic	9.2	6.0	4.8
African American	3.3	4.0	3.6
Other	7.6	1.0	3.6
Lifetime SITB History (%)			
NSSI	56.5	11.0	65.1
Suicidal ideation	81.9	31.0	78.3
Suicide plan	37.3	3.0	56.6
Suicide attempt	28.3	1.0	49.4
Any SITB	86.3	33.0	83.1

Note. IAT=Implicit Association Test; SITB=Self-injurious thoughts and behaviors (NSSI, suicidal ideation, suicide plan, suicide attempt); NSSI=Nonsuicidal self-injury. Valid percentages are reported above. Hispanic/Latino ethnicity was asked separately from race in Study 1a.

Study 1d. Data for Study 1d were drawn from a neuroimaging study intended to examine reward processing in adults who engage in NSSI. Study 1d participants were 30 women, half who were currently engaging in NSSI in the form of cutting (*M*=21.3 years, *SD*=3.7) and half without NSSI history and who did not meet full diagnostic criteria for any current or past psychiatric disorders (*M*=22.8 years, *SD*=3.3).

Measures and procedure for Study 1d were largely distinct from Studies 1a-c. The initial laboratory visit included the SITBI (Nock et al., 2007) and Structured Clinical Interview for DSM-IV-TR Axis I Disorders (SCID I; First, Spitzer, Gibbon, & Williams, 2002). During the subsequent scanning session, participants viewed blocks of positive, negative, or neutral images. These were taken from the International Affective Picture Set (IAPS) (Lang, Bradley, & Cuthbert, 2008). In addition, participants also viewed a block of pictures that depicted NSSI. The NSSI picture set was developed by the Study 1d researchers (Allen & Hooley, 2015). All study participants saw the same pictures in the following order: neutral, NSSI, negative, positive. After viewing each block of pictures, participants rated their current mood using the Positive and Negative Affect Scales (PANAS; Watson, Clark, & Tellegen, 1988). See Appendix for additional information on Study 1d sample and measures.

Data Analyses

Studies 1a-c featured similar analyses to examine potential iatrogenic effects. Paired sample *t* tests were used to first test changes in desire to self-injure, desire to die, and mood among the entire sample, using pooled *SD* to estimate effect size. Alpha was set at .05 for two-tailed significance tests.[3] After examination of overall changes, repeated measures ANOVA were conducted to test whether ratings changed as a function of respondent gender. An additional repeated measures ANOVA was conducted using Study 1a data to determine whether ratings changed as a function of IAT type (cutting/self-injury, death, Suicide IATs). Similar to prior work on iatrogenic effects (Gould et al., 2005), two sets of analyses were conducted overall: among all respondents, to test the overall impact SITB stimuli exposure, and among respondents

[3] This cutoff is less useful when interpreting Study 1 analyses given its exceptional sample size and statistical power. In this case, interpretation of effect sizes may be more informative.

with SITB history, to test the same impact on higher-risk individuals. SITB history was defined as engagement in NSSI, suicidal ideation, suicide plan, or suicide attempt at least once.

For Study 1d, negative mood scores were calculated by summing the rating for the 10 PANAS negative mood items. A repeated measures MANOVA was then used to compare mood in the NSSI and control groups at baseline and after viewing the four different types of images (neutral, NSSI-related, negative, positive). Follow up analyses used paired sample *t* tests to compare participants' negative mood after viewing different types of images. An analysis of covariance was also conducted to specifically test mood change after viewing NSSI images in the NSSI and control participants, controlling for baseline level of negative mood.

Results[4]

Examining potential iatrogenic effects (Studies 1a-c)

Participants in Studies 1a-c participants revealed minimal variation in desire to self-injure and desire to die (Table 1.2). Regarding participants' desire to self-injure, two out of the three studies reported no significant change in desire to self-injure. Participants in Study 1 showed an increase in desire to self-injure from before to after IAT completion; however, this change was small (d=.02) and not replicated in Studies 1b and 1c. See Supplemental Data for related post-hoc analyses. Regarding participants' desire to die, two out of the three studies reported a significant, slight decrease in severity. This pattern emerged in Studies 1a and 1c, and remained unchanged in Study 1b. Ratings for desire to self-injure and die remained low across all three samples, both before and after IAT completion.

[4] Non-parametric tests were run for Studies 1b and 1c due to violation of the assumption of normality. Non-parametric and parametric tests produced identical results. Statistics from parametric tests are reported for consistency/comparison of effect sizes across all three studies (including 1a).

Table 1.2. Mean Pre-/Post-IAT ratings across three samples (Studies 1a-c)

	Desire to self-injure						Desire to die						Mood					
	Pre		Post		*t*	*d*	Pre		Post		*t*	*d*	Pre		Post		*t*	*d*
	M	(*SD*)	*M*	(*SD*)			*M*	(*SD*)	*M*	(*SD*)			*M*	(*SD*)	*M*	(*SD*)		
1a. Web-based respondents[a]	0.35	(0.74)	0.37	(0.76)	2.19*	.02	0.47	(0.87)	0.43	(0.83)	7.48***	.06	0.32	(1.45)	0.32	(1.41)	0.27	.00
1b. College students	1.16	(0.60)	1.14	(0.53)	0.28	.04	1.05	(0.26)	1.05	(0.33)	0.00	.00	6.84	(1.58)	6.23	(1.61)	4.32***	.38
1c. Adolescent inpatients	2.57	(2.49)	2.42	(2.27)	1.23	.06	2.55	(2.64)	2.39	(2.36)	2.11*	.07	6.19	(2.57)	5.97	(2.55)	1.76	.09

Note. IAT=Implicit Association Test; Pre=Pre-IAT rating; Post=Post-IAT rating. *SD*=Standard Deviation. Note that Study 1a ratings for desire to self-injure and die ranged from 0 (*not at all*) to 4 (*extremely*) scale, and *Mood* was reverse coded so the ratings above range from -3 (*extremely negative*) to +3 (*extremely positive*). Studies 1b and 1c involved ratings on a 10-point scale (1=*not at all*, 5=*somewhat*, 10=*very much*) for desire to self-injure and die, and a similar scale for mood (1=*negative*, 5=*neutral*, 10=*positive*).

[a]Given the exceptionally large sample size in Study 1a, interpretation of *p* values at the .05 level is cautioned. If applying a more conservative cutoff of $p<.01$ as adopted by other PIMH papers, there is no change in desire to self-injure among web-based respondents. Greater emphasis on effect size (vs. *p*) values is encouraged.

*$p<.05$, ***$p<.001$

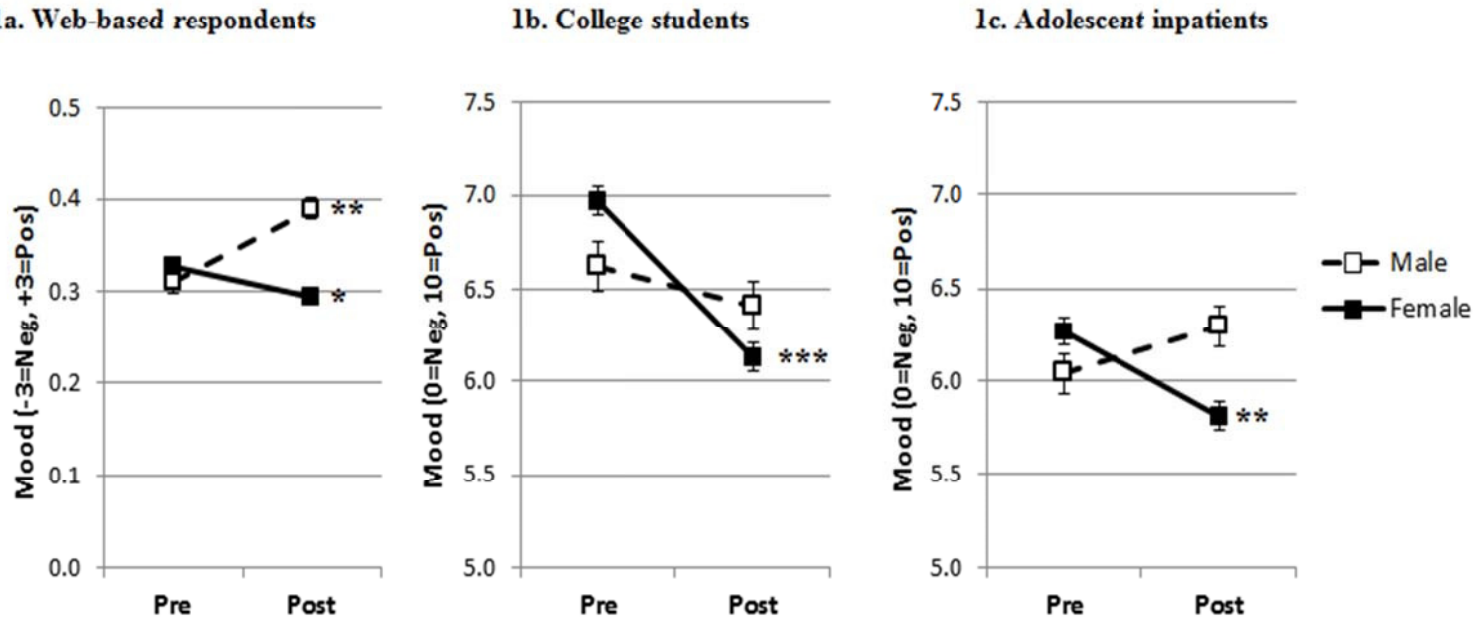
The largest observed pre- to post-IAT change occurred with participants' reported mood. Specifically, Study 1b revealed a significant mood decline producing a medium effect size. Mood remained unchanged among web-based respondents (Study 1a) and adolescent inpatients (Study 1c). All three samples reported positive mood both before and after IAT completion.

When higher-risk participants (i.e., only those with SITB history) were isolated across Studies 1a-c, results were less remarkable than what was found among participants overall. Only higher-risk web-based respondents from Study 1a reported a slight increase in desire to self-injure, $t(2824)=2.38$, $p=.02$, $d=.03$. There was no change in desire to self-injure among any other high-risk participants. Consistent with findings involving all participants, there remained a significant decrease in desire to die in Studies 1a and 1c, ts=2.12-7.45, ps=.00-.04, ds=.06-.08. None of the higher-risk participants across the three studies reported change in mood, ts=-.14-1.30, ps=.21-.89, ds=.001-.26.

Identify characteristics that heighten iatrogenic risk (Studies 1a-c)

Participant gender did not affect changes in desire to hurt oneself or die, Fs=0.00-3.06, ps=.08-1.00, η^2s=.00-.04, but did affect mood change (Figure 1.1.a-c). All three samples produced significant Gender × Time interactions, Fs=4.73-15.38, ps=.00-.03, η^2s=.005-.09. There was a slight mood decline among female web-based participants (Study 1a) and adolescent inpatients (Study 1c), ts=2.06-3.10, ps=.003-.04, ds=.02-.18, and a moderate mood decline among female college students (Study 1b), $t(62)=5.41$, $p<.001$, $d=.54$. Male participants generally did not experience a change in mood, except for a slight mood improvement detected among web-based male respondents, $t(1075)=3.37$, $p=.001$, $d=.05$.

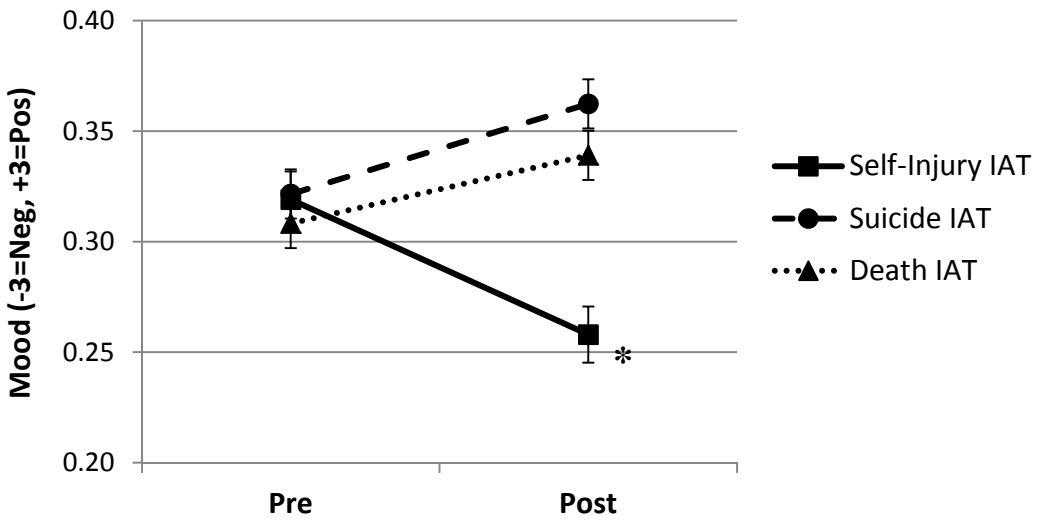
Figure 1.1. Pre-/Post-IAT Mood, by Gender



Note. IAT=Implicit Association Test. There were significant Gender x Time interactions for web-based respondents, $F(1,3266)=15.38$, $p<.001$, $\eta^2=.005$, college students, $F(1,98)=4.73$, $p=.03$, $\eta^2=.05$, and adolescent inpatients, $F(1,81)=7.55$, $p=.01$, $\eta^2=.09$. The breakdown of male and female participants, respectively, are as follows: Study 1a (n=1076, n=2192), Study 1b (n=37, n=63), Study 1c (n=27, n=56). *$p<.05$, **$p<.01$, ***$p<.001$ for within-group differences. Error bars represent ±1 standard error, as calculated for paired sample t tests.

As shown in Study 1a, patterns of mood change also varied across the three types of IAT, $F(2,3301)=5.76$, $p=.003$, $\eta^2=.003$ (Figure 1.2). Specifically, there was a slight but significant mood decline for the Self-Injury IAT, $t(1096)=2.40$, $p=.02$, $d=.04$, but not the suicide or death IATs, ts=1.38-1.84, ps=.07-.17, ds=.02-.03.

Figure 1.2. Pre-/Post-IAT Mood, by IAT Type (Study 1a)



Note. IAT=Implicit Association Test. *p<.05 for within-group differences. Error bars represent ±1 standard error, as calculated for paired sample t tests.

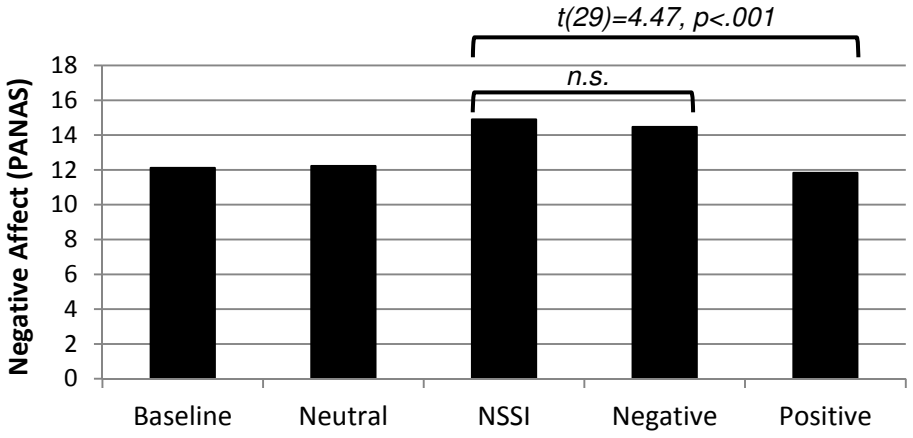
Track mood response to SITB-related stimuli (Study 1d)

The most salient and robust trend across studies thus far was the mood decline observed among female participants. Having identified this subgroup as being prone to mood decline, the female-only sample of Study 1d allows more in-depth examination of mood change in this potentially more vulnerable group. Study 1d also focused on mood response to NSSI-related images, which according to Study 1a may be more emotionally provocative (vs. suicide-related words). A repeated measure MANOVA revealed a main effect of group. At all assessment time points, participants in the NSSI group reported more negative mood than control participants did, $F(1,28)=10.80$, $p=.003$, $\eta^2_p=.28$. There was also a main effect of condition, with negative mood

levels changing according to the type of pictures presented, $F(4,25)=6.23$, $p<.001$, $\eta^2_p=.50$. Importantly, however, there was no Group x Condition interaction, $F(4, 25)=1.50$, $p=.23$, $\eta^2_p=.19$. This indicates that the pattern of negative mood changes reported by the NSSI participants in response to the stimulus pictures was not significantly different from that reported by the controls.

Follow up analyses revealed that, as shown in Figure 1.3, all participants reported an increase in negative mood after viewing the NSSI pictures, $t(29)=3.46$, $p=.002$, $d=.49$. However, when controlling for baseline negative mood, NSSI participants did not show a greater increase in negative mood in response to viewing the NSSI pictures than the control participants did, $F(1, 27)=0.97$, $p=.33$. Moreover, across all participants, negative mood after viewing the NSSI images was comparable to negative mood after viewing the negative IAPS pictures, $t(29)=-0.79$, $p=.44$. Finally, after viewing positive pictures, all participants showed a significant decrease in negative mood, $t(29)=-4.00$, $p< .001$, $d=.53$. Indeed, after viewing the positive images, participants' negative mood levels were comparable to their negative mood levels assessed at baseline, before any images had been viewed, $t(29)=0.89$, $p=.32$.

Figure 1.3. Mood Response to Images (Study 1d)



Note. PANAS=Positive and Negative Affect Scales; NSSI=Nonsuicidal Self-Injury. NSSI images represent SITB-related images.

Discussion

There are several key findings from this series of studies. First, research participants who were repeatedly presented with SITB-related stimuli did not show any reliable increase in their desire to self-injure or desire to die. Second, there was a moderate mood decline, specific to female research participants and the presentation of images (vs. words). Third, women's mood decline from viewing SITB-related images appeared to be both transient and reversible. These findings are discussed in greater detail below.

Importantly, SITB-related stimuli have minimal impact on participants' desire to self-injure and desire to die. The degree of change in these ratings remained small in magnitude across Studies 1a-c. The small effect size is critical to interpret in Study 1a, where *p* values are less informative due to the exceptionally large sample size. Because increased desire to self-injure from Study 1a was not significant at a more conservative *p* value, and not replicated across Studies 1b and 1c, this finding likely emerged as a function of large sample size rather than a meaningful and concerning change. The more robust finding found across Studies 1a thru 1c was participants' decreased desire to die; but this change was small in magnitude as well (*d*s=.00-.07). Given the size of these changes, the most balanced conclusion is that SITB-related stimuli of the type tested here do not increase self-injurious or suicidal urges. This counters the common assumption that the mere mention of suicide or self-injury increases its likelihood of occurrence (Stoppe, Sandholzer, Huppertz, Duwe, & Staedt, 1999).

Another consistent pattern was participants' generally positive mood both before and after the IATs. Even the significant mood decline detected in Study 1b featured students shifting from a markedly positive pre-IAT mood to post-IAT *less positive* (i.e., not negative) mood, and likely reflected regression to the mean. No other systematic change in mood was replicated or

even moderate in magnitude (d<.05) when all research participants were considered. This confirms prior work suggesting no increased distress from SITB-related interview questions or warning signs (Gibson et al., 2014; Gould et al., 2005; Muehlenkamp et al., 2015; Rudd et al., 2006).

At the same time, evidence of any mood decline, even that of minimal concern, warrants further investigation. Across studies, I tested *who* is most vulnerable to mood decline, and which *stimuli type* drives such decline. Regarding *who* is most vulnerable, female (vs. male) participants consistently reported more severe mood decline. Male participants reported no such change. This runs contrary to prior gender-specific findings in the SITB literature suggesting that male participants may react adversely to and female participants may benefit from SITB-related assessment or prevention efforts (Bryan et al., 2009; Shaffer et al., 1990). However, this finding is less surprising in light of basic psychological science, which suggests stronger emotional reactivity of women compared to men. This has been demonstrated with psychological, physiological, and neurobiological responses to emotional images and memories (Canli, Desmond, Zhao, & Gabrieli, 2002; Lang, Greenwald, Bradley, & Hamm, 1993).

Regarding *stimuli type*, Study 1a revealed that the NSSI-related stimuli from the Self-Injury IAT yielded mood decline, whereas stimuli in the Death/Suicide IATs did not. It remains unclear exactly why the Self-Injury IAT would have this effect. One of the unique characteristics of the Self-Injury IAT is its specific focus on NSSI. It remains possible that the topic of NSSI is more emotionally provocative than suicide or death. But recent evidence demonstrating *decreased* distress from answering NSSI-specific questions render this explanation unlikely (Muehlenkamp et al., 2015). The unique use of images (vs. words) as stimuli in the Self-Injury IAT may point toward a more plausible explanation. That is, images related to self-injury and

suicide may be more emotionally provocative than words. The image-driven explanation is also more congruent with cognitive and affective neuroscience, which suggests differential neurobiological processing of emotionally-valenced pictures versus words (Kensinger & Schacter, 2006).

Study 1d offered the chance to examine this 'higher-risk' combination of female research participants and the presentation of NSSI-related images. It was found that women's mood decline in response to NSSI-related images was transient in nature, no more severe among women with NSSI history, and no more severe compared to negative non-SITB stimuli. In fact, the observed mood decline from SITB-related images appeared to be temporary and declined soon after viewing SITB-related images. This may be due to the subsequent presentation of positive images, consistent with prior work showing that asking positive questions after SITB-related questions results in a 'corrective' impact on respondents (Deeley & Love, 2009). It also remains possible that mood improved naturally, rather than in response to positively-valenced images. Future research using SITB-related behavioral measures is advised to exercise heightened awareness of gender- and stimuli-specific mood decline among participants.

STUDY 2:

Using the Self-Injury IAT in the Hospital to Predict NSSI

Having established that adapted IATs have minimal iatrogenic effects (Study 1), Study 2 focuses on the application of the Self-Injury IAT (Self-Injury IAT) in greater detail. In contrast to Study 1, Study 2 focuses on the underlying psychological process that the Self-Injury IAT captures: implicit identification with cutting.

Initial studies examining implicit identification with cutting were cross-sectional in nature. The first study to develop and test the Self-Injury IAT revealed that adolescent self-injurers have a stronger implicit identification with cutting compared to non-injurers (Nock & Banaji, 2007). A second cross-sectional study highlighted the specificity of the Self-Injury IAT, demonstrating that self-injurers have a stronger implicit identification with cutting compared to suicide attempters and healthy adolescents (Dickstein et al., in press).

Importantly, not all correlates of past NSSI predict future NSSI. In fact, the most robust predictor of future NSSI is NSSI history itself. Prior studies testing the incremental predictive validity of other psychosocial factors (e.g., temperament, anxiety symptoms, alcohol use) have shown that they no longer predict NSSI[5] when entering them into a multivariate prediction model with NSSI history (Chapman, Derbidge, Cooney, Hong, & Linehan, 2009; Tuisku et al., 2014). There are some exceptions, such as sleep disturbance predicting NSSI frequency one year later, even after controlling for past NSSI history. However, the problem largely remains: beyond knowing whether adolescents have self-injured in the past, we do not know who will continue self-injuring in the future.

[5] NSSI is often defined by either frequency or by presence (vs. absence). In multivariate models, Chapman et al. (2009) failed to predict NSSI frequency, and Tuisku et al. (2014) failed to predict NSSI presence.

Two studies have prospectively examined implicit identification with cutting by evaluating whether it predicts frequency of future NSSI episodes (Franklin, Puzia, Lee, & Prinstein, 2014; Glenn & Klonsky, 2011).[6] The predictive validity of the Self-Injury IAT and other implicit measures was compared to explicit self-report (i.e., respondents' self-reported likelihood of engaging in NSSI in the future), characteristics of prior NSSI (e.g., frequency, number of methods used, recency), and psychosocial factors (e.g., emotion reactivity, features of borderline personality disorder, impulsive urgency). The time frames examined were six months (Glenn & Klonsky, 2011) and twelve months (Franklin et al., 2014) after initial assessment. Consistent with the aforementioned work, these studies reported that implicit identification with cutting does not predict future NSSI frequency, remission, or recurrence. Instead, robust predictors included characteristics of prior NSSI, specifically frequency and number of NSSI methods used.

There are at least three possible reasons why implicit identification with cutting has not been shown to predict future NSSI. First, a six- or twelve-month follow-up period may be too broad a time frame for some NSSI-specific risk factors[7] to predict NSSI. Indeed, Nock and Cha (2009) distinguish proximal NSSI-specific risk factors (e.g., implicit identification with cutting) from distal risk factors (e.g., childhood maltreatment). The latter set of factors predisposes a

[6] A third study by Randall and colleagues (2013) prospectively tested a battery of NSSI- and suicide-related IATs, including the Self-Injury IAT. It is excluded from the present discussion as the predicted outcome combined NSSI, suicidal ideation, and suicide attempt, and therefore it remains unclear which specific outcome the findings pertain to. Moreover, the majority of Randall et al.'s sample lies outside the range of adolescence or early adulthood (65% were adults over the age of 30 years).

[7] An example of an NSSI-specific risk factor that could be a longer-term predictor is overcoming aversion to NSSI, as measured by the Affect Misattribution Task (Franklin et al., 2014a). Aversion to NSSI, specifically cutting and blood-related images, is conceptualized as an instinctual psychological process. Once someone has overcome the visceral aversion to cutting and blood, this likely does not wax and wane as implicit identification with cutting could. Implicit identification with cutting is a cognitive process that may change depending on other viable behaviors that an adolescent identifies with at that time (e.g., identification with cutting may go down as identification with exercise goes up; Wallenstein & Nock, 2007). Indeed, overcoming aversion to cutting prospectively predicts NSSI one year later, suggesting that it may be a longer-term predictor than implicit identification with cutting (Franklin et al., 2014).

person to long-term maladaptive outcomes (e.g., NSSI, drug use, eating disorders), and could independently or interactively predict NSSI occurring years later. Implicit identification with cutting, in contrast, represents a process that is likely acquired later in life and may only correspond with NSSI in the short-term. For instance, an adolescent who strongly identifies with NSSI when they are 13 may no longer identify as a self-injurer five years or even one year later. Examining shorter-term prediction models could highlight conditions under which implicit identification with cutting predicts future NSSI. Notably, this is aligned with recent calls for short-term prediction models in the related area of suicidal behavior (Glenn & Nock, 2014).

Second, implicit identification with cutting has only been prospectively examined among community-based young adults. Both the clinical severity and age of these samples warrant attention. Regarding clinical severity, implicit identification with cutting may not be as elevated among community-based young adults as it is among adolescent inpatients. For example, community-based young adults with NSSI history have identified more strongly with *Not Cutting* (vs. *Cutting*), on average (Glenn & Klonsky, 2011).[8] Administering the Self-Injury IAT to higher-risk individuals may produce stronger identification with NSSI among self-injurers, and weaker identification with NSSI among non-injurers. Regarding age, young adulthood represents a time where more people stop rather than start or continue engaging in NSSI. Early adolescence (12-14 years) is the most commonly reported age of NSSI onset (Nock, 2009b). Interestingly, studies reporting a slightly older age of onset (15-16 years) report that NSSI typically stops within five years of starting (20-21 years; Whitlock, Eckenrode, & Silverman,

[8] Nock and Banaji (2007) demonstrated that community-based adolescents with NSSI history identified more strongly with Cutting (vs. Not Cutting), on average. This confirms, rather than contradicts, the arguments regarding clinical severity and age. Nock and Banaji's (2007) sample had recently engaged in NSSI (past year history) and were younger (NSSI and no-NSSI group, *M*=17.10 years, *SD*=1.92, range=12-19), whereas Glenn and Klonsky (2011) reported lifetime history of NSSI and involved an older sample (NSSI group, *M*=19.06 years, *SD*=1.89).

2006).[9] Young adulthood is therefore an optimal time to assess remission rather than recurrence (Glenn & Klonsky, 2011), as there are fewer subsequent NSSI episodes to predict. Given the lower base rate of NSSI among young adults and community-based individuals (e.g., 24% past year prevalence vs. >80% among adolescent inpatients; Giletta, Scholte, Engels, Ciairano, & Prinstein, 2012; Nock & Prinstein, 2004), a short follow-up period (e.g., <6 months) would likely not produce sufficient statistical power to predict NSSI. Short-term prediction models of NSSI should ideally be assessed with younger, more clinically severe populations.

Third, implicit identification with cutting may only relate to history of NSSI, with no conditions under which it predicts future NSSI. This is unlikely for several reasons. Implicit associations, as measured by the IAT, are indeed capable of predicting future behavior. This ranges from predicting deliberate voting behavior (Freise, Bluemke, & Wanke, 2007), to clinical outcomes such as drinking (Farris, Ostafin, & Palfai, 2010), major depressive episodes (Elgersma, Glashower, Bockting, Penninx, & de Jong, 2013), and suicide attempt (Nock et al., 2010). IATs have also been shown to predict various clinical outcomes above and beyond history of those outcomes (Elgersma et al., 2013; Farris et al., 2010), and explicit self-report (e.g., Nock et al., 2010). This is likely the case with implicit identification with cutting predicting future NSSI, as well—when optimally assessed.

Addressing these possibilities, the present study used the Self-Injury IAT to prospectively examine whether implicit identification with cutting predicts future NSSI. Short-term prediction models tested whether implicit identification with cutting predicted NSSI during the hospital stay, and within three months of hospital discharge. Multivariate models tested its predictive

[9] Even with the possibility that there may be two trajectories of NSSI, one beginning earlier and another beginning later (see Yates, 2004), those who begin self-injuring earlier in life (i.e., early adolescence) are believed to be a more severe subgroup of self-injurers (Nixon, Cloutier, & Aggarwal, 2002).

validity relative to characteristics of NSSI history (e.g., frequency, number of methods used), as well as patients' explicit self-report.

A final feature of this study examined changes in implicit identification with cutting over time. The Self-Injury IAT was administered at both admission and at discharge to observe changes in implicit identification with cutting over the course of hospitalization. It remains unknown whether such change is possible, or predictive of future NSSI. Prior work suggests that structured treatment can change IAT scores (Clerkin, Fisher, Sherman, & Teachman, 2014; Gamer, Schmuckle, Luka-Krausgrill, & Egloff, 2008; Teachman & Woody, 2003), which in turn predicts subsequent symptom change (Teachman, Marker, & Smith-Janik, 2008). Moreover, Price and colleagues (2009; 2014) used IATs to show that ketamine infusion for the treatment of depression predicts immediate (within 24 hrs) change in implicit identification with *Escape* (i.e., *Me=Escape*). These findings highlight the potential malleability of implicit associations, and their subsequent connection to clinical change. As a comparison, explicit self-report measures were also administered at admission and discharge.

Hypotheses

1. Implicit identification with cutting will be associated with history of NSSI, such that it distinguishes self-injurers and non-injurers at baseline.
2. Self-Injury IAT scores will incrementally improve short-term prediction of future NSSI, occurring during the patients' hospital stay (i.e., hospital-based NSSI) and within three months of hospital discharge (i.e., post-discharge NSSI), above and beyond NSSI history and explicit self-report.
3. Implicit identification with cutting will weaken from hospital admission to discharge. There may be less of a weakening effect from admission to discharge among the

subset of self-injurers who continue engaging in NSSI after hospital discharge. I explored the possibility that admission-to-discharge change will predict post-discharge NSSI above and beyond other risk factors and explicit self-report.

Method

Sample

Research team members approached 246 adolescents recently admitted to a psychiatric inpatient unit, of which 137 provided parental consent and child assent (response rate=55.7%). Among this sample, 123 patients provided data relevant to study hypotheses. Fourteen patients were withdrawn from the study due to: behavioral, medical, or cognitive limitations to providing reliable self-report responses (*n*=6), invalid IAT data at baseline (*n*=4), or withdrawal from the study (*n*=4). More than half of the final sample reported recent (past year) history of NSSI (*n*=67, 54.5%). 115 participants completed follow-up assessments at discharge (response rate=93.5%), and 100 completed 3-month follow-up assessments (response rate=81.3%). The average duration between admission and discharge assessments was approximately two weeks (*M*=16.2 days, *SD*=21.6).

Measures

Self-Injurious Thoughts and Behaviors Interview (SITBI; Nock et al., 2007)**.** Adolescents' engagement in NSSI was assessed using the SITBI, a structured interview about engagement in self-injurious thoughts and behaviors including NSSI. The SITBI was used to assess history of NSSI at admission, outcomes of hospital-based NSSI and post-discharge NSSI, and adolescents' self-reported future likelihood of NSSI. History of NSSI was captured through questions about whether an adolescent had engaged in NSSI in their lifetime or in the past year, how frequently they engaged in NSSI, and what types of methods they used. Hospital-based

NSSI and post-discharge NSSI were assessed using similar questions but pertaining to different time points, the former referring to NSSI occurring during the hospital stay and the latter referring to NSSI occurring within three months of hospital discharge. Adolescents' self-reported future likelihood of NSSI was assessed through a question asked at admission: *On the scale of 0 to 4, what do you think are the chances that in the future you will purposely hurt yourself without wanting to die?* (0=*Not at all likely;* 4=*Extremely likely*). With regard to the SITBI overall, Nock and colleagues (2007) determined excellent interrater reliability (average κ=.99) and test-retest reliability over a six-month period (average κ=.70). The SITBI also demonstrated excellent construct validity when compared to another established measure of NSSI (Functional *Assessment of Self-Mutilation*; Lloyd, Kelley, & Hope, 1997), as there was perfect agreement on lifetime presence of NSSI (κ=1.0), and excellent agreement on lifetime frequency of NSSI (κ=.99).

Self-Injury Implicit Association Test (Self-Injury IAT; Nock & Banaji, 2007). I used Self-Injury IAT scores to capture the primary predictor: implicit identification with cutting. The Self-Injury IAT is a brief reaction time test during which patients sort images and words into concept categories (*Cutting, Not Cutting*) and attribute categories (*Me, Not Me*) by pressing either a left or right key on a computer keyboard. Forty two trials presented category pairs of *Cutting/Me* on one side, *Not Cutting/Not Me* on the other side. Another forty-two trials presented category pairs of *Cutting/Not Me* on one side, *Not Cutting/Me* on the other side. See www.implicitmentalhealth.com for samples of the test. The order of trials was randomized across adolescents, and order did not significantly impact performance on the Self-Injury IAT at either time point, *t*s=0.62-1.85, *p*s=.07-.54, *d*s=.12-.34. The Self-Injury IAT compares the speed at which a person classifies stimuli when the paired concept-attribute categories match an

individual's implicit associations (e.g., *Cutting/Me*) versus the speed of classification when the paired categories do not match an individual's implicit associations (e.g., *Cutting/Not Me*). Faster response latencies indicate stronger implicit associations between the paired concept-attribute categories. In this case, I would expect self-injurers to categorize pictures of cut skin into the *cutting* category faster when *Cutting* is paired with *Me* (vs. *Not Me*). This is indicated by a higher and more positive *D* score, whose calculation is based on prior IAT work (e.g., Greenwald, Nosek, & Banaji, 2003). This IAT was administered within a larger battery of suicide-related IATs that will be reported separately (Millner, D'Angelo, & Nock, 2015). The IAT typically shows stronger internal consistency estimates than other latency-based measures, and has been conceptualized as assessing both stable and occasion-specific variance (Schnabel, Asendorpf, & Greenwald, 2008). Despite some accounts of strong convergent validity among similar IATs (Cunningham, Preacher, & Banaji, 2001), IATs have generally been shown to have weak to modest convergent validity with other implicit measures (Schnabel et al., 2008). This is also the case between the Self-Injury IAT and other implicit measures (e.g., Affect Misattribution Procedure; Franklin et al., 2014). Finally, IATs typically have moderate associations with explicit measures, which could vary depending on social desirability concerns (Hofmann, Gschwendner, Nosek, & Schmitt, 2005).

Explicit self-report. I also examined patients' explicit self-report of how closely they identify themselves with NSSI. This was examined using a visual analog scale, which presented a horizontal line with *Not Cutting* labeled on the left and *Cutting* labeled on the right. Patients were asked to place a single mark along the line indicating which one they identified with more. The visual analog scale was scored by measuring in millimeters from left to right, such that higher values indicated stronger explicit identification with NSSI. Visual analog scales are

commonly used to assess mood, and have demonstrated high interrater agreement, moderate to strong test-retest reliability, and convergent validity with other measures of mood (Ahearn, 1997). Since visual analog scale scores were strongly correlated with adolescents' self-reported future likelihood of NSSI (r=.74, p<.001), I calculated composite scores by summing up z scores from each explicit measure. Similar to previous comparisons of implicit and explicit measures (Price et al., 2009, 2014), this allowed us to preserve power and reduce risk of Type I error.

Demographic information and psychiatric history. Information on demographic factors and psychiatric diagnoses was collected following baseline assessment. This was extracted from medical charts by research assistants to decrease patients' time and burden.

Procedure

Each patient's admission packet included a recruitment brochure about the current study. Patients and their parents who indicated interest in this study informed a unit or research staff member, and were then approached by a research assistant to potentially provide informed assent/consent. Patients who agreed to participate met with a research assistant to complete all measures at admission, and once more before discharge. They were also contacted by phone three months after discharge for their follow-up interview. Patients were given a $25 gift card for each assessment time point.

Data Analysis

I initially conducted independent samples t tests and chi-square analyses to identify demographic and psychiatric factors that differentiate baseline self-injurers and non-injurers. To test the first hypothesis, I conducted independent samples t tests comparing self-injurers and non-injurers' Self-Injury IAT scores at baseline. Given the clinical severity and diversity of this

sample, I also conducted supplemental analyses to test the specificity of Self-Injury IAT performance to methods of NSSI (i.e., cutting vs. non-cutting).

To test the second hypothesis, I conducted a series of logistic regressions predicting hospital-based NSSI and post-discharge NSSI. This included one set of bivariate models and two sets of multivariate models. The first multivariate model controlled for known risk factors entered into the first step of a hierarchical logistic regression analysis, and IAT scores entered second. The second multivariate model controlled for a composite explicit self-report score entered into the first step of a hierarchical logistic regression analysis, and IAT scores entered second. All prospective analyses were conducted only among adolescents reporting lifetime history of NSSI (n=78),[10] thereby controlling for prior history of NSSI.

To test the third hypothesis, I conducted paired sample t tests among baseline self-injurers to test change in Self-Injury IAT scores from admission to discharge. The same was done for explicit self-report scores. I then examined whether post-discharge self-injurers and non-injurers demonstrated distinct admission-to-discharge trajectories, via repeated measures ANOVA. This would determine whether patients who continued engaging in NSSI after hospital discharge were marked by a significant increase in scores from admission to discharge, and whether patients who did not engage in NSSI after hospital discharge were marked by a significant decrease of change from admission to discharge. Finally, I tested whether change scores predicted post-discharge NSSI. Change scores were computed by subtracting Self-Injury IAT scores at admission from Self-Injury IAT scores at discharge. Since higher Self-Injury IAT scores indicated stronger implicit identification with cutting, I interpreted positive change scores as *increased* identification with cutting throughout the hospital stay, and negative change scores

[10] There is a distinction between n=78 reported here at the $n=68$ self-injurers reported in Table 2.1. This is because NSSI in Table 2.1 was defined based on past year history of NSSI to maintain consistency with the initial Self-Injury IAT study (Nock & Banaji, 2007).

as *decreased* identification with cutting. I also computed admission to discharge change in explicit self-report scores, such that higher scores indicated increased explicit identification with and likelihood of cutting. Hierarchical logistic regressions predicting follow-up NSSI were then conducted by entering known risk factors into the first step, and entering change scores (either Self-Injury IAT or explicit self-report) in the second step.

Results

At baseline, past year self-injurers and non-injurers did not differ across demographic factors, except for race. The only psychiatric factors differentiating self-injurers and non-injurers were mood disorder diagnosis, lifetime frequency of NSSI episodes, and number of NSSI methods (Table 2.1).

Table 2.1. Baseline Comparisons between Recent Self-Injurers vs. Non-Injurers

Variable	NSSI (n = 68)	No NSSI (n = 55)	Test	Effect Size
Age in years ($M \pm SD$)	15.0 ± 1.4	14.6 ± 1.7	$t_{(121)} = 1.39$	$d = 0.25$
Sex (% female)	76.5	65.5	$\chi^2_{(1)} = 1.81$	$\Phi = 0.12$
Race (%)				
White	85.3	89.0		
Hispanic	8.8	0.0		
Black	4.4	0.0	$\chi^2_{(3)} = 12.32$*	$\Phi = 0.32$
Asian	0.0	4.0		
Other	1.5	7.0		
Psychiatric Diagnoses (%)				
Mood disorder	94.1	74.5	$\chi^2_{(1)} = 9.32$**	$\Phi = 0.28$
Anxiety disorder	42.7	34.5	$\chi^2_{(1)} = 0.84$	$\Phi = 0.08$
Eating disorder	23.5	23.6	$\chi^2_{(1)} = 0.00$	$\Phi = 0.00$
Impulse control disorder	16.5	16.6	$\chi^2_{(1)} = 0.00$	$\Phi = 0.00$
Alcohol/substance use disorder	17.6	12.7	$\chi^2_{(1)} = 0.56$	$\Phi = 0.07$
Psychotic disorder	2.9	3.6	$\chi^2_{(1)} = 0.05$	$\Phi = 0.02$
Other disorder	13.2	24.5	$\chi^2_{(1)} = 2.55$	$\Phi = 0.15$
Other Psychiatric Factors				
Total number of disorders	2.1 ± 0.8	1.9 ± 0.9	$t_{(117)} = 0.72$	$d = 0.21$
GAF at admission	29.9 ± 7.8	30.4 ± 6.7	$t_{(120)} = 0.32$	$d = 0.06$
Number of prior hospitalizations	1.3 ± 1.8	0.9 ± 1.5	$t_{(121)} = 1.45$	$d = 0.27$
Lifetime NSSI frequency	94.4 ± 245.6	7.9 ± 47.5	$t_{(72)} = 2.82$**	$d = 0.67$
Number of NSSI methods	3.3 ± 1.6	0.5 ± 1.3	$t_{(117)} = 10.27$***	$d = 0.90$

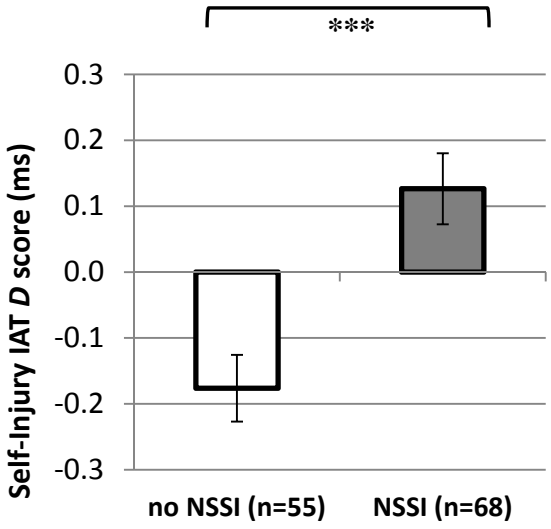
Note. NSSI=Nonsuicidal self-injury; GAF=Global Assessment of Functioning. *p<.05 **p<.01***p<.001

Hypothesis 1: Association with past NSSI

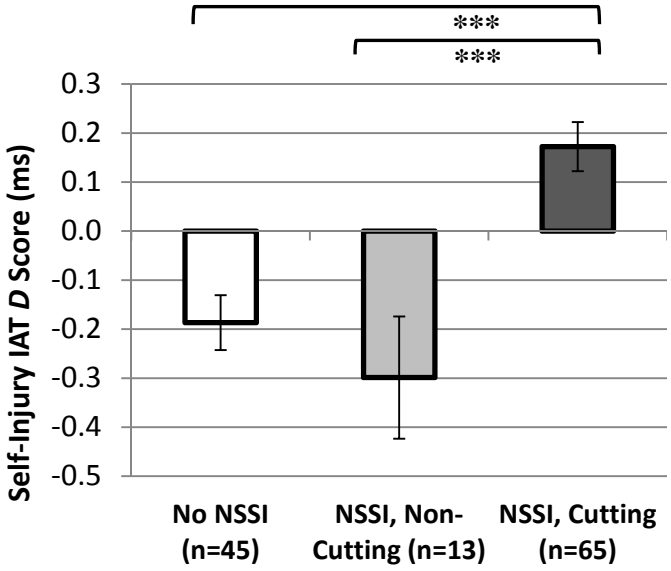
Consistent with hypotheses, recent self-injurers had significantly higher Self-Injury IAT scores than non-injurers, revealing stronger implicit identification with cutting, $t(121)=4.02$, $p<.001$, $d=0.73$ (Fig 2.1a). Self-Injury IAT scores continued to differentiate groups controlling for race of NSSI, $OR=5.75$, $CI=2.22\text{-}14.92$, $p<.001$.[11]

Fig 2.1. Self-Injury IAT Scores at Baseline

a. NSSI engagement (past year)



b. NSSI method used (lifetime)



Note. Self-Injury IAT=Self-Injury Implicit Association Test; NSSI=Nonsuicidal self-injury; ms=milliseconds. Error bars represent ±1 standard error of the mean. ***$p<.001$

Self-Injury IAT results at admission appeared to be driven by self-injurers who specifically used cutting as their NSSI method, which matched the stimuli featured in the Self-

[11] Logistic regression models controlling for psychiatric risk factors did not hold up to assumptions of linearity of the logit or multicollinearity. Entering mood disorder suppressed criterion-irrelevant variance of Self-Injury IAT scores, thereby *inflating* its odds ratio, $OR=6.07$, $CI=2.27\text{-}16.26$, $p<.001$. Moreover, entering either lifetime frequency of NSSI or total number of NSSI methods violated assumptions of linearity of the logit, as the interaction terms between each predictor variable and its log transformation significantly predicted past year history of NSSI, $ps<.001$. For these reasons, multivariate analyses predicting past year NSSI history *only* controlled for race. This assumption was not violated for the other outcome variables of hospital-based NSSI and post-discharge NSSI.

Injury IAT (e.g., images of cut skin). Self-injurers who cut themselves differed from non-injurers, and from self-injurers who used alternative methods of NSSI (e.g., hitting, scratching), $F(2,120)=14.67$, $p<.001$ (Fig 2.1b). Post-hoc analyses revealed that self-injurers who cut themselves had significantly higher scores than other groups, *p*s<.001, and that there was no significant difference between non-injurers and self-injurers who did not cut, $p=1.00$.

Hypothesis 2: Short-term prediction of future NSSI

Only patients with lifetime history of NSSI were included in prospective analyses, providing a more conservative test of the study hypotheses. Bivariate analyses revealed that Self-Injury IAT scores predicted both hospital-based and post-discharge NSSI (Table 2.2). Self-Injury IAT scores also predicted hospital-based NSSI above and beyond other risk factors (i.e., mood disorder, lifetime frequency of NSSI, number of NSSI methods). However, Self-Injury IAT scores did not predict NSSI above and beyond explicit self-report measures. Explicit self-report scores continued predicting hospital-based and post-discharge NSSI controlling for Self-Injury IAT scores, *OR*s=1.88-2.37, *CI*s=1.29-3.80, *ps*<.01.

Table 2.2. Predictive Validity of Self-Injury IAT Scores

	Bivariate		Multivariate $_a$		Multivariate $_b$	
Outcome	**OR**	**CI**	**OR**	**CI**	**OR**	**CI**
Hospital-based NSSI	6.65**	1.79-24.71	5.46*	1.30-22.97	2.53	0.56-11.41
Post-discharge NSSI	3.55*	1.09-11.57	2.61	0.75-9.01	1.69	0.44-6.48

Note. NSSI=Nonsuicidal self-injury; Self-Injury IAT=Self-Injury Implicit Association Test, assessed at admission. Multivariate$_a$ tested whether Self-Injury IAT scores predict NSSI above and beyond mood disorder, lifetime frequency of NSSI, and number of NSSI methods. Multivariate$_b$ tested whether Self-Injury IAT scores predict NSSI above and beyond Explicit Measure Composite scores, which draw from the Cutting/Me visual analog scale & patient prediction of future NSSI. Explicit Measure Composite scores predicted Hospital-based NSSI, *OR*=2.37, *CI*=1.48-3.80, *p*<.001, & post-discharge NSSI, *OR*=1.88, *CI*=1.29-2.76, *p*=.001, above and beyond Self-Injury IAT scores. *p <.05, **p <.01

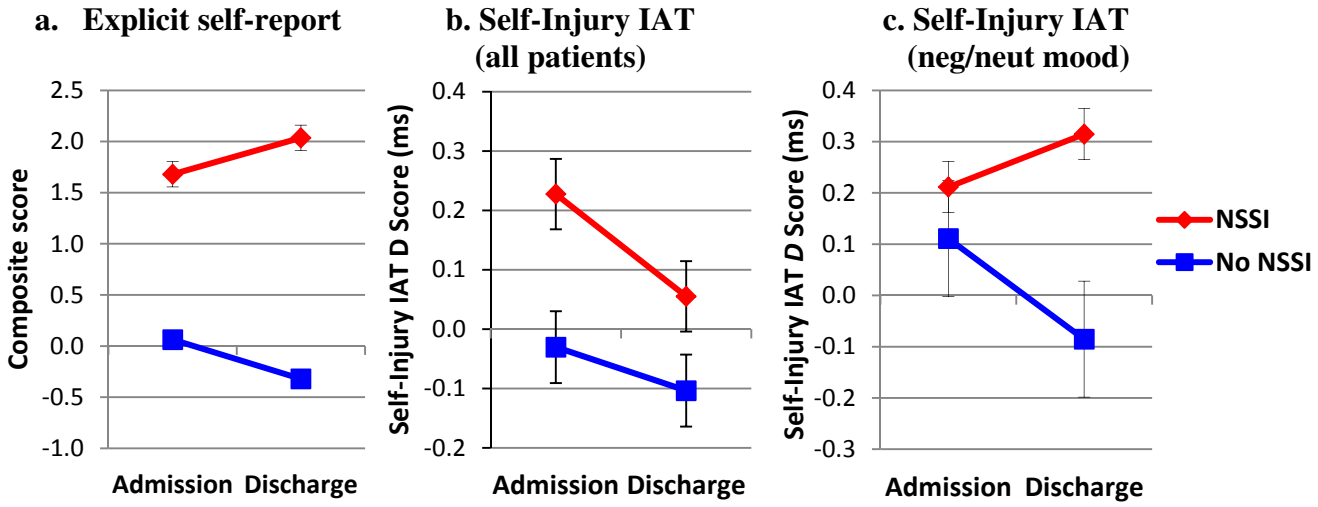
Given that admission results revealed that Self-Injury IAT scores may be specific to method of NSSI used, I also examined whether Self-Injury IAT scores more strongly predicted hospital-based cutting behavior (vs. non-cutting methods of NSSI). Among the 26 patients reporting hospital-based NSSI, 23 specifically engaged in cutting and 3 engaged in other forms of NSSI. Indeed, when *only* cutting behavior was counted as hospital-based NSSI, bivariate analyses revealed stronger predictive validity of the Self-Injury IAT, *OR*=13.34, *CI*=2.84-64.67, *p*=.001. Self-Injury IAT also predicted hospital-based cutting behavior above and beyond other risk factors, *OR*=12.14, *CI*=2.38-61.90, *p*=.003, and nearly predicted above and beyond explicit self-report scores, *OR*=5.77, *CI*=0.995-33.50, *p*=.051. Self-Injury IAT scores did not predict post-discharge cutting behavior in either bivariate or multivariate analyses, *OR*s=1.48-3.10, *CI*=0.39-9.94, *p*s=.06-.56.

One limitation of this study is that patients had varying lengths of stay in the hospital, ranging from 30 hours to as many as 116 days. On the one hand, assessing hospital-based NSSI at the naturally occurring time point of discharge strengthens the ecological validity of these findings. On the other hand, it introduces variability between the initial Self-Injury IAT and hospital-based NSSI, including exceptional cases that far exceed a two-week stay. I therefore conducted post-hoc analyses, replicating bivariate and multivariate models only among baseline self-injurers who stayed two weeks or less in the hospital (*n*=57). Results were identical, except that Self-Injury IAT scores appeared to more strongly predict hospital-based NSSI. There were larger odds ratios for the Self-Injury IAT in the bivariate model, *OR*=9.53, *CI*=1.60-56.89, *p*=.01, and in the multivariate model controlling for risk factors, *OR*=8.59, *CI*=1.27-58.09, *p*=.02. Self-Injury IAT scores, albeit showing a stronger effect than before, still did not significantly predict hospital-based NSSI above and beyond explicit self-report, *OR*=4.22, *CI*=0.59-30.05, *p*=.15.

Hypothesis 3: Change from admission to discharge

Self-Injury IAT and explicit self-report scores did not change from admission to discharge when examining all baseline self-injurers. Admission-to-discharge score changes were small and nonsignificant, *t*s=0.80-1.46, *p*=.15-.43, *d*=0.07-0.11. Prospective analyses revealed some predictive validity of change scores, as patients who did versus did not go on to engage in post-discharge NSSI displayed distinct trajectories of explicit self-report scores. This was captured by a significant interaction between Group (post-discharge NSSI vs. no NSSI) and Time (Admission vs. Discharge), $F(1,61)=6.00$, $p=.02$ (Fig 2.2a). Patients who did not engage in post-discharge NSSI reported a significant decrease in explicit self-report score from admission to discharge, $t(30)=0.24$, $p=.03$, $d=0.17$. Patients who did engage in post-discharge NSSI reported a small, nonsignificant increase in their explicit self-report scores, $t(31)=1.44$, $p=.16$, $d=0.21$. Admission-to-discharge change in explicit self-report predicted post-discharge NSSI, *OR*=1.74, *CI*=1.07-2.84, *p*=.03.

Fig 2.2. Change in Explicit Self-Report and Self-Injury IAT



Note. NSSI=Nonsuicidal self-injury, as assessed during the three-month follow-up interview. Admission-to-discharge change in explicit self-report (Fig 2a) and Self-Injury IAT scores (Fig 2b) was examined across all patients. Self-Injury IAT scores were then examined only among patients reporting neutral or negative mood at both time points (Fig 2c). Error bars represent ±1 standard error, as calculated for paired sample *t* tests.

Results were less remarkable for admission-to-discharge change in Self-Injury IAT scores. Trajectories of Self-Injury IAT scores did not significantly vary between patients who did versus did not engage in post-discharge NSSI, $F(1,57)=0.34$, $p=.56$ (Fig 2b). There were nonsignificant decreases in Self-Injury IAT scores among both groups of patients (post-discharge NSSI vs. no post-discharge NSSI), *t*s=0.61-1.45, *p*s=.16-.61, *d*s=0.22-0.56. Change in Self-Injury IAT scores did not predict post-discharge NSSI, *OR*=0.79, *CI*=0.35-1.75, $p=.56$, and multivariate models were therefore not tested. These Self-Injury IAT findings counter my hypothesis and warrant further investigation.

One important consideration is the context under which patients completed the Self-Injury IAT. Consider the expected trajectory of a hospital stay: patients were likely admitted following a recent incident of acute safety risk, and at discharge were presumably in an improved emotional state. Indeed, patients reported being in a more positive mood at discharge than at admission, $t(51)=3.23$, $p=.002$, $d=0.51$. While this pattern is desirable from a clinical perspective, it introduces systematic variation in study procedures. Changes in contextual factors, such as transient mood, can impact performance on clinically-relevant IATs (Gemar, Segal, Sagrati, & Kennedy, 2001). Specifically, IAT scores make more meaningful differentiations between clinical groups when respondents are in a more negative state. Because most patients in the current study reported recovered mood at discharge, directly subtracting discharge from admission Self-Injury IAT scores may not fully capture meaningful clinical change over time. In an effort to standardize assessment time points, I excluded patients reporting positive mood at both admission and discharge, and reexamined Self-Injury IAT score change. The Group x Time interaction, albeit still not significant, produced a larger effect, $F(1,7)=1.72$, $p=.23$ (Fig 2c). This was also the case with the observed decrease in Self-Injury IAT scores among patients who did

not engage in post-discharge NSSI, $t(3)=0.87$, $p=.45$, $d=1.00$, and with the observed increase in Self-Injury IAT scores among patients who did not engage in post-discharge NSSI, $t(4)=1.04$, $p=.46$, $d=1.04$. A similar pattern emerged when reexamining the predictive validity of Self-Injury IAT change scores among patients in a relatively more negative mood, *OR*=22.58, *CI*=0.16-3244.26, $p=.22$.

Discussion

There are several key findings from this study. First, implicit identification with cutting was related to NSSI history among adolescent psychiatric patients, and predicted short-term future NSSI above and beyond other clinical risk factors (e.g., frequency of NSSI, number of methods used, mood disorder). Second, implicit identification with cutting more strongly related to past and future NSSI when NSSI was strictly defined by cutting behavior (vs. hitting, scratching). Third, adolescents' explicit self-report predicted NSSI above and beyond implicit identification with cutting. Fourth, change in explicit self-report from admission to discharge predicted which patients went on to engage in NSSI after hospitalization. I discuss each of these findings in more detail below.

The first finding that implicit identification with cutting relates to NSSI history replicates and extends prior work. I specifically compared self-injurers and non-injurers based on past year history of NSSI, since this was the time frame used in the original Nock and Banaji (2007) study. Baseline comparisons produced significant but more modest group differences than ($d=0.73$) than what Nock and Banaji found among community-based adolescents ($d=1.20$). This is not surprising, since clinical samples tend to produce more moderate between-group effects due to greater clinical severity of groups (Franklin, Glenn, & Nock, 2015). Conceptually speaking, a self-injurer may be more likely to identify with NSSI simply because they have engaged in it

before. Similarly, a person who begins smoking may see themselves as more of a smoker than one who has never smoked before. Of note, the two prior Self-Injury IAT studies did not report a significant relationship between NSSI history and implicit identification with cutting (Franklin et al., 2014; Glenn & Klonsky, 2011). The two notable differences between the present study and prior studies are that the latter studies were conducted among young adults (vs. adolescents) and predicted NSSI frequency (vs. past year presence of NSSI). Future work is encouraged to continue exploring these inconsistencies.

Relatedly, implicit identification with cutting prospectively predicted hospital-based NSSI above and beyond prior NSSI frequency, number of NSSI methods, and mood disorder. This finding exceeded my expectations that implicit identification with cutting can predict future NSSI in the short-term[12], as it only predicted hospital-based NSSI. In fact, when patients who happened to have longer than a two-week stay in the hospital were excluded, implicit identification with cutting was even more predictive. Since implicit identification with cutting did not predict NSSI occurring three-months later, it is less surprising that it did not predict NSSI six- or twelve-months later in prior studies (Franklin et al., 2014; Glenn & Klonsky, 2011). This finding prospectively confirms the *implicit identification hypothesis* (Nock, 2009). How a person implicitly identifies him or herself (self-injurer vs. non-injurer) matters, and relates to whether they will engage in NSSI in the near future. Implicit identification with cutting appeared to only have a short-term impact of NSSI, thereby confirming its status as a more proximal, NSSI-specific risk factor.

[12] There is admittedly no consistent definition of 'short-term' prediction (Glenn & Nock, 2014). However, a two-week assessment window represents an incrementally narrower window of time than before in NSSI research. I urge the field of both NSSI and suicide research to shorten assessment windows to predict self-injurious thoughts and behaviors over the next hours, days, or weeks.

The second finding was that there were stronger and more specific associations between Self-Injury IAT performance and cutting behavior in the present study. That is, patients who used NSSI methods other than cutting (e.g., burning, inserting objects under their skin) did not differ from controls in their Self-Injury IAT scores. Moreover, only including patients who cut themselves in prospective analyses strengthened the predictive validity of the Self-Injury IAT. Implications of these findings are discussed under Future Directions.

The third finding is that although implicit identification with cutting prospectively predicted NSSI above and beyond other risk factors, it did not predict above and beyond patients' explicit self-report. In fact, patient self-report predicted short- and long-term NSSI above and beyond implicit identification with cutting. This counters initial expectations that patients may be reluctant to openly discuss sensitive topics (e.g., NSSI) due to fear of embarrassment or consequences (Tourangeau & Yan, 2007). There are several reasons why NSSI may not be as sensitive a topic as previously expected. First and foremost, it may be an issue related to the outcome assessed. Adolescent inpatients may not perceive disclosure of NSSI as affecting their inpatient stay as much as suicidal urges. An important caveat of studies reporting weak predictive validity of self-report is that they all involved suicide-related outcomes (Horesh & Apter, 2006; Horesh et al., 2004; Janis & Nock, 2008). The few studies focusing *only* on NSSI risk reported significant prospective prediction of long-term NSSI (e.g., Glenn & Klonsky, 2011). Adolescent inpatients may therefore be more forthright about how much they identify with cutting behavior, or how likely they are to self-injure again. Similar patterns of more open self-reporting have been observed among violent inpatients, whose self-reported perceptions of future risk have been shown to predict post-discharge instances of self-injurious and violent behavior (Peterson, Skeem, & Manchak, 2011; Skeem, Manchak, Lidz, & Mulvey,

2013). Other possible explanations for why and when to rely explicit self-report are included in the General Discussion.

The fourth finding is that admission-to-discharge change in explicit self-report, but not implicit identification with cutting, predicted long-term (i.e., post-discharge) NSSI. On the one hand, this brings forward a useful set of self-report questions that could be used to track progress across treatment. On the other hand, change in implicit identification with cutting may be more temperamental and require a more standardized repeated assessment approach. According to post-hoc findings, consistent mood across time points produced more clinically meaningful changes in Self-Injury IAT scores. Future work is encouraged to more systematically examine the role of mood on Self-Injury IAT performance over time.

Notably, neither implicit identification with cutting nor explicit self-report changed among the *overall* group of patients over the course of inpatient stabilization. A more structured and standardized course of treatment, as used previously (Clerkin et al., 2014; Gamer et al., 2008; Price et al., 2009, 2014; Teachman et al., 2008; Teachman & Woody, 2003), may be required to change automatic cognitions that the Self-Injury IAT captures. Indeed, the only case of SITB-related implicit associations changing over time was in response to short-term, highly controlled, hospital-based pharmacological intervention (Price et al., 2009, 2014).

STUDY 3:

Assessing the Role of Negative Mood on the Suicide IAT and Stroop Tasks

Study 2 post-hoc findings suggest that contextual factors such as negative mood may impact the predictive validity of applied IATs. In Study 3, I systematically examine the role of negative mood on the psychological processes underlying Suicide IAT and Stroop Task performance. This marks the first examination to test the mood-dependent nature of implicit identification with death, as well as suicide-specific attentional bias.

We currently do not know how a suicidal person's mood influences his or her performance on these behavioral measures. Relevant literature on depression suggests that negative mood may be what triggers maladaptive cognitive processes, such as implicit identification with death. For example, Gemar and colleagues (2001) found that formerly depressed adults demonstrated stronger implicit associations between the concepts of *Me* and *Negative* (vs. *Me* and *Positive*) following a negative mood induction, whereas non-depressed adults did not exhibit such a change. This finding goes beyond identification of cognitive biases and suggests that it may be equally if not more important to understand the contexts in which these biases arise. Teasdale's *Differential Activation Hypothesis* (1988) proposes that people vary in how their maladaptive cognitive biases respond to emotional states; those low in *cognitive reactivity* have more mood-resistant biases and are less likely to endure persistent depression, and those high in cognitive reactivity have mood-dependent biases and are more likely to be persistently depressed.

Extending Teasdale's theory to suicidal individuals, latent suicide-specific cognitive vulnerabilities may be observed in the context of negative mood. Evidence supporting this claim would demonstrate more pronounced deficits that emerge after (vs. before) a negative mood

induction (see Lau, Segal, & Williams, 2004). Indeed, there is an observed deterioration in the ability to problem solve and generate positive future events among suicidal individuals after negative mood inductions (Williams, Barnhofer, Crane, & Beck, 2005; Williams, Van der Does, Barnhofer, Crane, & Segal, 2008). Such mood-dependent decline in cognition is not observed among nonsuicidal individuals, nor among nonsuicidal depressed controls (Williams et al., 2005, 2008). Negative mood therefore appears to reinstate (i.e., reactivate) suicidal individuals' thinking patterns. William and colleagues (2005) suggest that this cognitive reactivity among prior suicide ideators[13] may reveal how suicidal thoughts and behaviors perpetuate over time. To date, however, this question has not been prospectively assessed, nor applied to implicit identification with death or suicide-specific attentional bias. Especially when these cognitive processes are examined in a neutral laboratory setting, a mood induction may be required to 'activate' the precise automatic associations and attentional biases that have been captured in previous studies conducted under more inherently distressing circumstances (e.g., psychiatric hospitalization immediately following a suicide attempt).

The aim of the current study is to test mood-dependent cognitive risk factors of suicide. Through a laboratory-based study, I addressed this aim by exploring the effects of a temporary negative mood induction on two aspects of suicide-related cognition: implicit identification with death, and suicide-specific attentional bias.

Hypotheses

1. Suicidal adults will demonstrate stronger implicit identification with death (i.e., *Death=Me*) and stronger suicide-specific attentional bias after (vs. before) the negative mood induction. I expect no changes in pre- to post-induction cognition among nonsuicidal adults.

[13] People who engage in suicidal ideation are hereafter referred to as suicide ideators.

2. Post-induction cognition will more strongly differentiate groups based on history of suicidal ideation, compared to pre-induction cognition.
3. Post-induction cognition will more strongly predict future suicidal ideation than pre-induction performance. Post-induction cognition will also be predictive above and beyond people's explicit self-report.

Method

Sample

Research team members recruited 150 community-based suicidal and nonsuicidal adults (*M*=32.0, *SD*=13.6 years; 56.0% Female; 74.7% White, 9.3% Black, 6.7% Hispanic, 4.7% Asian, 4.0% Multiracial, and 0.7% Other). Inclusion criteria included adult status (≥18 years-old), English fluency, and availability to complete the full two-hour interview and two follow-up calls. Suicide ideators were recruited based on recent history of active suicidal ideation (e.g., those who "did not want to live" but were uncertain about death/suicide were not recruited). Nonsuicidal adults were recruited based on having no history of suicidal thoughts or behaviors, and self-reported high desire to live. In all cases, I did not recruit individuals who indicated the presence of any factor impairing the individual to effectively participate in the study, including the presence of gross cognitive impairment, agitated/violent behavior, or high/imminent risk of suicide. Among the 150 recruited and consenting adults, one participant withdrew due to a self-reported learning disability that prevented them from being able to complete the computer tasks. The sample of 149 participants were young adults (*M*=32.02 years, *SD*=13.68), with the majority being Female (55.7%) and White (74.5% White, 9.4% Black, 6.7% Hispanic, 4.7% Asian, 2.7% Multi-Racial, 2.0% Other).

Measures

Suicide Implicit Association Test (IAT). I used the Suicide IAT (Nock et al., 2010) to measure implicit identification with death. The Suicide IAT is similar to the Self-Injury IAT described in Study 2, except that half paired the constructs of *Death* and *Me* together, and the other half presented *Death* and *Not Me* together. The format and presentation of the Suicide IAT was identical to that used in Nock et al. (2010). I would expect suicide ideators to categorize death-related words into the *Death* category faster when *Death* is paired with *Me* (vs. *Not Me*).

Suicide Stroop Task. I used a Suicide Stroop task (Becker et al., 1999; Cha et al., 2010; Williams & Broadbent, 1986), to measure implicit identification with death. The Suicide Stroop task is a computerized, adapted emotional Stroop task which presents a series of words on the screen that are printed in either red or blue. Participants identified the color of the word by pressing one of two different keys. Following eight practice trials involving neutral words (*cat, ball, desk, radio, basket, table, tape, bag*), critical trials presented four categories of words in new random order: suicide-related (*dead, funeral, suicide*), negatively-valenced (*alone, rejected, stupid*), positively-valenced (*happy, success, pleasure*) and neutral (*paper, engine, museum*). Each category was represented 12 times throughout the task, such that participants completed a total of 48 critical trials. I measured response latencies from identifying font color: greater response latencies were interpreted as indicating greater interference due to the semantic content of the presented words. For example, I would expect an individual with an attentional bias toward suicide-related stimuli to demonstrate greater response latencies due to fixation on the semantic content of suicide-related words.

Mood induction paradigms. Both the negative and positive mood inductions, in that order, were administered to participants during the laboratory visit. Each mood induction lasted ten minutes.

Negative Mood Induction Paradigm. The Velten mood induction was used to temporarily induce negative mood among participants prior to retaking the Suicide IAT and Stroop tasks. Similar to previous studies with suicidal adults (Williams et al., 2005, 2008), the negative mood induction paradigm consisted of participants reading negative Velten statements and being asked to reflect on how those statements apply to them (e.g., *There are things about me that I do not like; The way I feel now, the future looks boring and hopeless*). While reading these statements, participants also listened to "Russia under the Mongolian Yoke" by Prokofiev played at half speed. This procedure is identical to the one used in prior studies involving adult suicide ideators (Williams et al., 2005, 2008), and in studies producing the strongest evidence for cognitive reactivity in depression (see Clark, Beck, Alford, 1999). Participants underwent this mood induction immediately following the first round of computer tasks (IAT, Stroop). The Velten mood induction has been shown to induce negative mood which peaks for more than twelve minutes, and naturally returns to baseline within 30 minutes of the mood induction (Chartier & Ranieri, 1989).

Positive Mood Induction Paradigm. A positive mood induction paradigm was used to remove any residual negative mood following negative mood inductions, similar to previous studies (e.g., Clark & Teasdale, 1985; Frost & Green, 1982). Participants watched a novel compilation of videos, which pilot testing revealed ranged from amusing to inspiring. Participants underwent the positive mood induction immediately before debriefing and the suicide risk assessment.[14]

[14] The positive mood induction procedure that typically complements the Velten negative mood induction involves participants reading positive statements aloud (e.g., *I have complete confidence in myself*; *I know that in the future I won't let so-called "problems" get me down*). I initially proposed this mood induction procedure, but later modified this since pilot testing revealed that people perceived such positive statements as invalidating (i.e., forcing oneself to actively endorse positive statements while in a negative mood).

Repeated assessments of transient mood. Transient mood between measures and mood inductions was assessed using a two-dimensional visual analog scale (VAS) aligned with the Circumplex Model of emotions (Russell, 1980). This is adapted from Russell and colleagues' (1989) Affect Grid, which is a single-item scale indicating degree of valence along the horizontal axis (negative-positive) and arousal along the vertical axis (high-low). Participants were asked to mark one coordinate indicating both valence and arousal in that moment. Examples were provided for the participant prior to measure completion (e.g., anger=negative valence, high arousal; calm=positive valence, low arousal). Mood valence was measured by determining how many horizontal millimeters away from the middle (i.e., neutral valence) participants marked. Larger positive values indicated more positive mood, and larger negative values indicated more negative mood. Mood arousal was measured by determining how many vertical millimeters away from the middle (i.e., neutral arousal) participants marked. Larger positive values indicated higher arousal, and larger negative values indicated lower arousal.

Assessment of demographic factors. Participants completed a demographic information sheet to provide information about their e.g., age, race, and gender.

Assessment of psychiatric factors. The Mini International Neuropsychiatric Interview (MINI; Sheehan et al., 1998) was used to assess current presence of psychiatric symptoms, so that additional differences between suicide ideators and non-ideators could be detected. The MINI is a brief, structured interview assessing various categories of current DSM-IV Axis I diagnoses. Information on psychiatric symptoms was collected to compare across participants. Following the MINI was a brief interview asking participants to indicate lifetime history of psychiatric diagnoses.

Assessment of self-injurious thoughts and behaviors. Suicidal ideation was assessed using the SITBI (Nock et al., 2007), identical to what was used in Study 2. The SITBI was used in the present study to assess suicidal ideation prior to the baseline laboratory visit, as well as suicidal ideation one- and six-months after the laboratory visit. In each module, if the initial screening question is endorsed, then the corresponding follow-up questions are included in the interview (e.g., suicidal ideation in the past week vs. past year). Participants' self-reported likelihood of future suicidal ideation (i.e., explicit self-report) was assessed at the end of the baseline SITBI suicidal ideation module, to compare its predictive validity with the predictive validity of the behavioral measures. Participants were asked to predict their future likelihood of suicidal ideation on a 0 to 4 scale (*0=Not at all likely; 4=Extremely likely*). The suicidal ideation module had good agreement (κ=.59) with another established measure of recent suicidal ideation, the Beck Sale for Suicidal Ideation (BSI; Beck, Steer, & Ranieri, 1988). See Study 2 for additional validity and reliability information regarding the SITBI.

Procedure

During a two-hour lab visit, participants first completed a series of pre-mood induction measures. This included an assessment of baseline mood, along with administration of the Suicide IAT and Stroop task. An alternative measure of attentional bias, the probe discrimination task, was also administered.[15] All three tasks were presented in random order, and there were no order effects between groups or on task performance, *p*s=.16-.99. Participants next completed a brief negative mood induction procedure, during which they listened to sad music while reading negative statements (e.g., "*Today feels like a gloomy day"*). This is identical procedure has changed behavioral measure performance of suicidal individuals in prior studies (see Martin,

[15] The probe discrimination task lies outside the scope of the present study, hence its description and results are presented elsewhere.

1990; Williams et al., 2005, 2008). Participants then completed the post-induction Suicide IAT and Stroop task, and reported demographic information, psychiatric history, and prior self-injurious thoughts and behaviors. Transient mood was measured throughout the lab visit in order to check for adequate manipulation. Because of potential ethical concerns associated with inducing a negative mood among suicidal individuals, at the end of the visit participants completed a positive mood induction (e.g., *Candid Camera* clips). Finally, participants were contacted by phone one- and six-months later to assess whether they engaged in subsequent suicidal thoughts or behaviors.

Data Analysis

I initially conducted independent samples *t* tests and chi-square analyses to test for demographic and psychiatric differences between participants with versus without suicidal ideation history at baseline (i.e., suicide ideators vs. non-ideators). I also conducted paired sample *t* tests to determine whether the mood inductions had their intended effect (i.e., negative mood induction resulting in mood decline, positive mood induction resulting in mood improvement).

To test the first hypothesis, I conducted repeated measures ANOVAs with Suicide IAT and Stroop scores to test for Time (pre vs. post) by Group (suicidal vs. nonsuicidal) interactions. Group was determined by baseline suicide ideator status. Simple effects analyses were conducted via paired sample *t* tests and one-way ANOVAs.

To address the second hypothesis, I tested whether pre- versus post-mood induction task performance differentiated suicidal from nonsuicidal participants at baseline. Bivariate analyses via logistic regression tested whether each set of scores predict suicide ideator status at baseline. Multivariate analyses via logistic regression then directly compared pre- and post-induction

scores as they predict suicide ideator status at baseline. Multivariate analyses were only conducted for IAT scores because bivariate analyses involving Stroop scores were not significant.[16] Identical analyses were conducted with a more easily interpretable version of task scores, consisting of either a positive or negative value, which has been shown to also predict suicide-related outcomes (Nock et al., 2010). This value was determined by whether the original IAT or Stroop task score was a numerical value above or below zero. A score above zero (i.e., positive) meant that, for instance, the participant responded more quickly in IAT trials where *Death* and *Me* were presented together. A score below zero (i.e., negative) meant that the participant responded more quickly in IAT trials where *Life* and *Me* were presented together. This version of task scores is hereafter referred to as *categorical* scores.

Analyses to test the third hypothesis were similar to those testing the second hypothesis, except predicting future (vs. past) suicidal ideation. All prospective analyses were conducted among baseline ideators, thereby controlling for history of suicidal ideation. Both continuous and categorical task scores were tested as predictors. More promising predictors (i.e., post-induction scores that predicted above and beyond pre-induction scores) were subsequently entered into a multivariate model to test whether they predicted 1- or 6-month ideation above and beyond explicit self-report.

Post-hoc analyses were also conducted using IAT and Stroop categorical scores, to test whether positive scores across more than one behavioral task could more strongly predict future suicidal ideation. These *dual-task* scoring profiles, described below, were also entered into a multivariate model to compare with explicit self-report.

[16] Despite a significant correlation detected across scores (pre- and post-induction IAT scores, $r=.55$, $p<.001$), the assumption of multicollinearity was met as indicated by Tolerance values (>0.1) and VIF values (<10).

Results

There were several demographic and psychiatric differences between participants reporting suicidal ideation history versus no history at baseline (Table 3.1). Suicidal participants were younger, more likely to identify as White, more often met criteria for a current mood or anxiety disorder, and more likely to report lifetime history of alcohol/substance use disorders. These demographic and psychiatric factors were largely controlled for in subsequent analyses predicting suicidal ideation at baseline.

Table 3.1. Comparison of demographic and psychiatric factors

Variable	Suicidal (n = 103)	Nonsuicidal (n = 46)	Test	Effect Size
Age in years ($M \pm SD$)	30.3 ± 11.8	35.9 ± 16.6	$t_{(66)} = 2.09$*	$d = 0.51$
Sex (% female)	50.5	67.4	$\chi^2_{(1)} = 3.68$	$\Phi = 0.16$
Race (%)				
White	78.6	65.2		
Hispanic	5.8	8.7		
Black	4.0	21.7	$\chi^2_{(5)} = 14.26$*	$\Phi = 0.31$
Asian	5.8	2.2		
Multiracial	2.9	2.2		
Other	2.9	0.0		
Psychiatric Diagnoses (%)				
Mood disorder	64.1	4.3	$\chi^2_{(1)} = 45.73$***	$\Phi = 0.55$
Anxiety disorder	67.0	4.3	$\chi^2_{(1)} = 50.02$***	$\Phi = 0.58$
Eating disorder	12.6	2.2	$\chi^2_{(1)} = 4.08$	$\Phi = 0.17$
Alcohol/substance use disorder	22.3	2.2	$\chi^2_{(1)} = 9.56$**	$\Phi = 0.25$

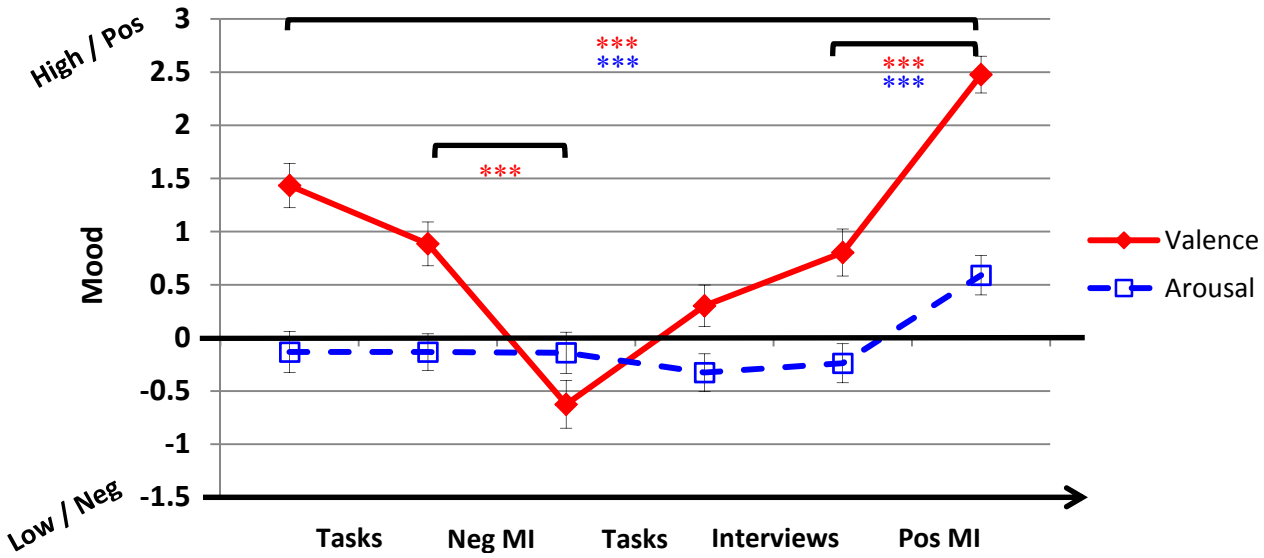
Note. Mood, anxiety, and eating disorder categories represent current diagnoses determined by the MINI. Participants self-reported lifetime history of alcohol/substance use disorder. *p <.05 **p <.01 ***p <.001

Mood Change over Time

The mood inductions had their intended effect on participants' mood valence (Fig 3.1). Participants reported relatively positive mood immediately before the negative mood induction (M=0.89, SD=2.50), and r negative mood immediately after (M=-0.61, SD=2.75), t(146)=10.69, p<.001, d=0.57. Arousal remained relatively low both before (M=-0.13, SD=2.09) and after (M=-

0.17, *SD*=2.36) the negative mood induction, *t*(146)=0.17, *p*=.86, *d*=-0.02). Of note, the positive mood induction at the end of the laboratory visit effectively improved mood, *t*(147)=-12.30, *p*<.001, *d*=-0.69, and heightened arousal, *t*(147)=-4.85, *p*<.001, *d*=-0.37. Participants reported overall improvement in mood across the lab visit, and reported more positive mood at the end than at the beginning, *t*(147)=-6.34, *p*<.001, *d*=-0.44. They also reported higher arousal at the end than at the beginning of the lab visit, *t*(147)=-3.71, *p*<.001, *d*=-0.31.

Figure 3.1. Mood change during laboratory visit



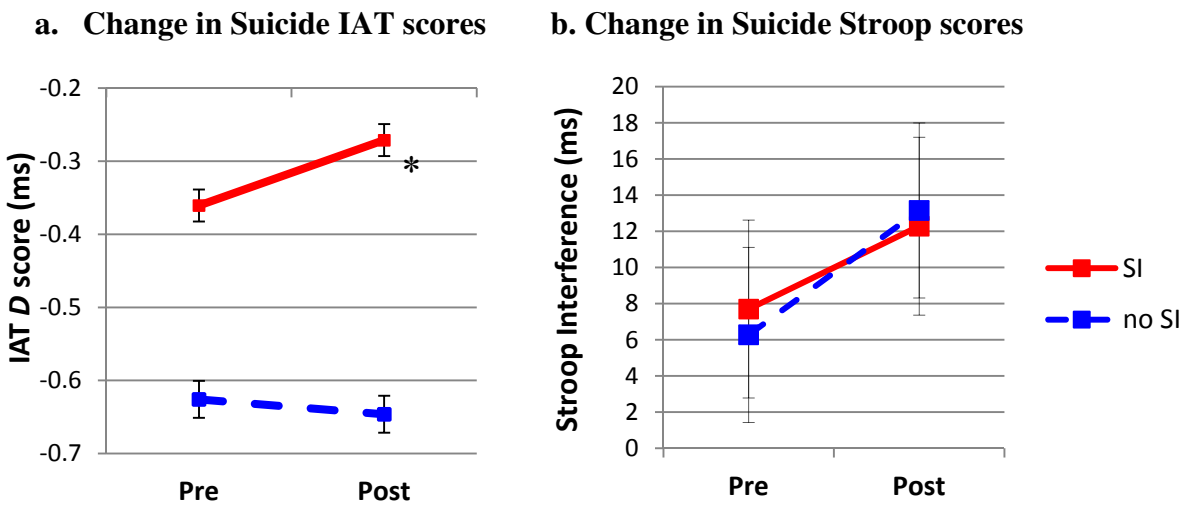
Note. Neg MI=Negative Mood Induction; Pos MI=Positive Mood Induction. Error bars represent ±1 standard error, calculated at each time point.***p <.001

Hypothesis 1: Change in score among suicide ideators

When examined all together, participants' task performance did not change before and after the negative mood induction, *t*s=.74-1.63, *p*s=.11-.46, *d*s=.09-.13. When baseline suicide ideators and non-ideators were examined separately, distinct patterns in IAT scores began to emerge (Fig 3.2a). Whereas suicide ideators demonstrated a significant *increase* in pre- to post-induction IAT scores, *t*(98)=-2.04, *p*=.04, *d*=-0.21, non-ideators' scores remained unchanged,

$t(43)=0.40$, $p=.69$, $d=.06$. The Group x Time interaction, however, was not significant, $F(1,141)=2.21$, $p=.14$, $\eta_p^2=.02$. There were no Stroop score changes among ideators or non-ideators over time, *t*s=.47-.71, *p*s=.48-.64, *d*s=.07-.14, nor was there a significant Group x Time interaction, $F(1,126)=0.02$, $p=.88$, $\eta_p^2=.00$ (Fig 3.2b).

Figure 3.2. Change in task performance between baseline SI and no SI groups



Note. IAT=Implicit Association Test; SI=Suicidal Ideation based on lifetime history. *Pre* and *Post* time points represent ratings immediately before and after each indicated portion of the visit. Error bars represent ±1 standard error, as calculated for paired sample *t* tests. *p <.05

Hypothesis 2: Post- (vs. pre-) induction scores differentiating ideators and non-ideators

Results addressing the second hypothesis directly compared pre- and post-induction IAT scores in their ability to differentiate individuals based on history of suicidal ideation (Table 3.2). Continuous score results are discussed first, followed by categorical score results. Bivariate analyses revealed that both pre- and post-induction scores differentiated baseline ideators from non-ideators, with post-induction scores producing a larger effect size. When entered in a multivariate model, post-induction scores statistically predicted baseline ideator status above and beyond pre-induction scores, and continued to predict baseline ideator status controlling for age,

Table 3.2. Predictive Validity of Pre- vs. Post-Mood Induction IAT Scores

		Baseline SI				1-Month SI[a]				6-Month SI[a]			
		Bivariate		Multivariate		Bivariate		Multivariate		Bivariate		Multivariate	
		OR	CI	OR	CI	OR	CI	OR	CI	OR	CI	OR	CI
Cont.	Pre	6.82***	2.29-20.28	2.61	0.72-9.46	1.95	0.65-5.90	1.38	0.38-4.98	2.16	0.68-6.87	1.18	0.30-4.70
	Post	11.45***	3.84-34.17	6.70**	1.97-22.80	2.20	0.81-5.94	1.83	0.58-5.77	3.46*	1.01-11.83	3.06	0.74-12.73
Cat.	Pre	4.61*	1.02-20.81	3.50	0.74-16.53	4.29*	1.23-14.97	3.57	0.96-13.30	2.41	0.61-9.47	1.73	0.40-7.40
	Post	15.60**	2.05-118.92	13.31**	1.73-102.52	6.09**	1.79-20.70	5.49**	1.56-19.28	11.00*	1.37-88.66	10.62*	1.30-86.84

Note. Cont.=Continuous IAT scores; Cat.=Categorical IAT scores; SI=Suicidal Ideation. Bivariate analyses examine the independent relations between IAT scores and SI. Multivariate analyses directly compare Pre- and Post-Mood Induction IAT scores and their relation to SI.

[a] These are prospective analyses conducted only among baseline suicide ideators, thereby representing persistence of SI over time.

*p <.05, **p <.01, ***p <.001

mood disorder, and anxiety disorder, *OR*=6.58, *CI*=1.59-27.16, *p*=.01.[17] Results for categorical IAT scores were similar, but produced larger effect sizes. Post-induction categorical IAT scores did not predict baseline ideator status above and beyond the aforementioned risk factors, *OR*=8.84, *CI*=0.96-81.75, *p*=.06.

Hypothesis 3: Post- (vs. pre-) induction scores predicting future suicidal ideation

Results addressing the third hypothesis tested whether pre- or post-induction IAT scores prospectively predicted suicidal ideation. All prospective analyses were conducted among baseline suicide ideators, to control for history of suicidal ideation. Again, both continuous and categorical scores were tested.

Categorical scores appeared to be more promising predictors than continuous scores. Categorical post-induction scores predicted future suicidal ideation above and beyond pre-induction scores, and again produced larger odds ratios than continuous scores. This promising set of scores was also compared to explicit self-report (i.e., participants' predicted likelihood of having suicidal thoughts). Such comparisons revealed that categorical IAT scores still predicted suicidal ideation occurring one month later, *OR*=4.99, *CI*=1.30-19.12, *p*=.02. That is, statistically controlling for explicit self-report, a participant who had a positive (vs. negative) score on the post-induction IAT had nearly a five-fold increase in the odds that they would have suicidal thoughts in the near future. Controlling for explicit self-report, categorical post-induction IAT scores did not predict suicidal ideation six months later, *OR*=6.46, *CI*=0.72-57.84, *p*=.09. Explicit self-report continued to predict one-month and six-month ideation above and beyond categorical post-induction IAT scores, *OR*s=2.51-3.21, *CI*s=1.50-5.80, *p*s<.001.

[17] Entering race or history of alcohol/substance use disorder suppressed criterion-irrelevant variance of Self-Injury IAT scores, thereby *inflating* its odds ratio, *OR*=9.86, *CI*=2.02-48.08, *p*<.01. For this reason, multivariate analyses predicting history of suicidal ideation controlled for age, mood disorder, and anxiety disorder.

Exploring novel applications of categorical scoring systems

In light of more promising results with categorical scores, I explored other useful applications of a categorical scoring system. Specifically, I tested whether there is value in examining categorical scores not only within a single behavioral task, but also across tasks. Since both the IAT and the Stroop task can produce categorical scores where positive scores indicate greater suicide risk, participants could take these tasks and produce one of the following *Dual-Task Profiles*: negative score on both the IAT and Stroop (Negative/Negative), negative score on one task and positive score on the other task (Negative/Positive),[18] or positive score on both tasks (Positive/Positive). Theoretically, positive scores on both the IAT and Stroop would indicate that multiple cognitive processes perpetuate suicidal thinking: individuals with this profile would not only be quick to pair together the idea of "death" with themselves (vs. "life" with themselves), but also have difficulty disengaging from the idea of suicide. The following set of post-hoc analyses tested this theory by categorizing baseline ideators into the aforementioned profiles, based on post-induction IAT and Stroop scores. Could Dual-Task Profiles help predict future suicidal ideation?

Indeed, these profiles marked the future likelihood of suicidal ideation (Table 3.3). There was a systematic increase in the proportion of baseline ideators who continued thinking about suicide both one month and six months later, with the lowest proportion among negative/negative profiles and the highest among positive/positive profiles. For example, a little more than half of participants with a negative/negative profile continued thinking about suicide six months later, compared to *all* participants with a positive/positive profile who continued thinking about suicide during that same time frame. Compared to explicit self-report, Dual-Task

[18] This could be defined as having a positive score on either task (Suicide IAT or Stroop). Efforts were made to tease apart performance on these measures (e.g., IAT Positive/Stroop Negative; IAT Negative/Stroop Positive), but the cell count was too low and therefore the two tasks were combined.

Profiles predicted one-month suicidal ideation, *OR*=1.94, *CI*=1.05-.3.57. As an example, a negative/positive profile, compared to a negative/negative profile, marked nearly a two-fold increase in the odds of experiencing suicidal ideation in the near future. Type of profile did not predict six-month suicidal ideation above and beyond explicit self-report, *OR*=2.73, *CI*=0.92-8.17, *p*=.07.

Table 3.3. Proportion of baseline ideators who continued to engage in suicidal ideation (SI)

Dual-Task Profile	1-Mo SI[a] % endorsed	6-Mo SI[b] % endorsed
Negative / Negative	33.3	57.9
Negative / Positive	46.2	73.0
Positive / Positive	83.3	100.0

Note. Negative / Negative=Negative score on both post-induction IAT and Stroop tasks; Negative / Positive=Positive score on either post-induction IAT or Stroop task, Negative score on the other task; Positive / Positive=Positive score on both post-induction IAT and Stroop tasks.
[a] There was a significant distribution across profiles, $\chi^2_{(2)}$=8.12, *p*=.02, Φ=.33
[b] Chi-square analysis was not run on 6-month suicidal ideation since one of the cell counts was < 1 (i.e., there were no Positive / Positive profiles who did *not* experience suicidal ideation six months later)

Discussion

This study yields five key findings. First, implicit identification with death assessed after (vs. before) a negative mood induction was stronger among suicide ideators, and better differentiated past ideators and non-ideators. Second, negative mood inductions appear to have minimal impact on suicide-specific attentional bias. Third, categorically-defined implicit identification with death more strongly predicted future suicidal ideation when assessed after (vs. before) a negative mood induction. Fourth, explicit self-report predicted suicidal ideation above and beyond implicit identification with death. Finally, accounting for *both* types of suicide-related cognitions (implicit identification with death, suicide-specific attentional bias) predicted suicidal ideation above and beyond explicit self-report. I elaborate on each key finding below.

The initial finding confirms the first and second hypothesis, suggesting that implicit identification with death is a mood-dependent cognitive process. Based on present findings, I can conclude that suicide ideators' implicit identification with death significantly strengthens after a negative mood induction. Stronger implicit identification with negative thinking (*Negative-Me*) has been detected among formerly depressed adults, only after (vs. before) a negative mood induction (Gemar et al., 2001). This extends prior work involving suicidal individuals by demonstrating that, similar to interpersonal problem solving and positive future fluency (Williams et al., 2005, 2008), implicit identification with death can be activated by negative mood. Drawing from the extensive cognitive theory of depression literature, there are several ways to interpret this finding (Lau et al., 2004). Applying Teasdale's (1988) *Differential Activation Hypothesis*, for example, an individual may begin associating the concept of death and themselves (i.e., *Death-Me*) at the onset of suicidal ideation. This association may then strengthen over time and become a latent cognitive vulnerability that interacts with, or is retriggered by, stress or negative mood. It is also described as *associative priming,* where dysphoria could prime or enhance the cognitive vulnerability (See Beevers, 2005; Segal & Ingram, 1994). Implicit identification with death would therefore be a variable risk factor for the persistence (vs. onset) of suicidal ideation over time.

Second, suicide-specific attentional bias did not appear to be as mood-dependent a cognitive process as implicit identification with death. To explain this distinction, one possibility is that suicide-specific attentional bias is a trait-like characteristic undeterred by mood or contextual factors. This is unlikely since attentional bias, as measured by the emotional Stroop task, has been shown to change as a function of many contextual factors (e.g., negative mood, positive mood, social stress, social rejection, sleep deprivation; Amir, McNally, Riemann, Burns,

Lorenz, & Mullen, 1996; Gilboa-Schechtman, Revelle, & Gotlib, 2000; Gotlib & Cane, 1987; Jamieson, Koslov, Nock, & Mendes, 2013; Sagaspe et al., 2006). Moreover, suicide-specific attentional bias has been shown to be specific to recency of suicide attempt, as past week suicide attempters demonstrate stronger attentional bias than less recent attempters (Cha et al., 2010). A more likely possibility is that suicide-specific attentional bias is a cognitive process that can be activated using techniques other than the present mood induction. Indeed, it has been argued that negative mood inductions may not be enough to capture cognitive reactivity, and that combining this with a cognitive load manipulation may be most optimal (Phillips, Hine, & Thorsteinsson, 2010).

Third, categorical implicit identification with death predicts future suicidal ideation when it is assessed after a negative mood induction. This finding builds on prior cross-sectional work with suicidal adults (Williams et al., 2005, 2008), and offers prospective evidence that cognitive reactivity predicts suicidal ideation. It also illuminates reasons why implicit identification with death may have predicted suicide-related outcomes when administered in a more acute treatment setting (Nock et al., 2010). Specifically, the context of having recently attempted suicide and being evaluated at a psychiatric emergency department may naturally create the negative mood-inducing context in which implicit identification with death could be more readily assessed.

In addition to explaining prior work, present findings have methodological implication. This study introduces an administration technique that would allow the Suicide IAT to be used outside the hospital setting while still effectively predicting suicidal ideation. This expands the application of promising behavioral risk assessments beyond the hospital setting, which serves only a minority of suicidal individuals. In fact, only 23% of community-based suicidal adults seek care from mental healthcare specialists (Bruffaerts et al., 2011). It is thereby possible for

implicit identification with death to be assessed among individuals who seek care through more accessible and less acute means, such as their primary care physicians or web-based support groups.

Critically, implicit identification with death only predicts suicidal ideation when assessed categorically rather than continuously. That is, *whether* a person identified with death (vs. life) predicted future suicidal ideation, but *the degree* of identification did not. This is practically beneficial, as categorical scores from the Suicide IAT simplifies the interpretation of reaction time tasks.

Fourth, even though categorical implicit identification with death predicted suicidal ideation six months later, it did not do so above and beyond explicit self-report. In fact, people's ability to predict whether they were going to think of suicide in the future was the most predictive risk factor, even after controlling for history of suicidal ideation. While this may initially be surprising, it is important to consider what outcome is being assessed and how. Regarding the outcome, people may be more able and willing to report future likelihood of suicidal ideation rather than suicide attempt. They may be more able because suicidal ideation occurs more frequently than suicide attempt (Kessler et al., 2005), and can likely be more easily monitored and anticipated than brief and intense urges to kill oneself. People may be more willing to report future likelihood of suicidal ideation because there is no imminent risk of thoughts (vs. actions). Related to this last point, the fact that people are providing explicit self-report in a neutral laboratory setting (vs. psychiatric emergency department, inpatient unit) may increase willingness to report to others. Indeed, suicidal adults have been shown to underreport their intent to die in psychiatric hospital settings (Busch et al., 2003). In general, the strength of explicit measure performance should not be underestimated, as it has been shown to be

predictive above and beyond mood-induced implicit measure performance in related areas of depression and dysphoria (Haeffel et al., 2007).

Finally, combining implicit identification with death and suicide-specific attentional bias produced the strongest and most unique predictor of future suicidal ideation. This introduces a parsimonious way to link performance on multiple behavioral measures with suicide risk. Conceptually, a person may be at risk of persistent suicidal ideation if they identify with death. But if they *both* identify with death and attend toward suicide-related content, it is nearly certain that they will continue thinking of suicide. Even though suicide-specific attentional bias may not predict suicidal ideation on its own, it carries explanatory value on top of implicit identification with death.

GENERAL DISCUSSION

This dissertation elucidates underlying psychological mechanisms and practical applications of SITB-related behavioral measures. Below, I elaborate on the overall pattern of findings across the three studies and implications for future research. This is followed by a discussion of limitations and overall conclusions.

Safe to Proceed. Importantly, it is possible to objectively observe risk factors for SITB in an ethical and responsible way. Study 1 findings provide replicated evidence that SITB-related stimuli used in newer behavioral and neurobiological studies do not increase people's self-injurious or suicidal urges, and have a minimal impact on mood. This pattern of findings is remarkably consistent across web-based, laboratory-based, and clinical settings. While Study 1 only tested the impact of IATs, findings likely apply to other behavioral measures such as the suicide Stroop task (Cha et al., 2010), Affect Misattribution Procedure (Franklin et al., 2014), and neuroimaging study paradigms used with self-injurious and suicidal individauls (Plenar, Bubalo, Fladung, Ludolph, & Lulé, 2012; Jollant et al., 2008). Here, I have explored the question of not only *whether* there is iatrogenic risk to novel SITB assessments, but also *how* instances of heightened risk may be handled. In light of these findings, I encourage both researchers and practitioners to take on a balanced perspective of SITB assessments. The stakes are undoubtedly high when approaching topics such as suicide or self-injury, but it is precisely their life-threatening nature that calls for more efforts toward effective and responsible assessment.

Study 1 findings are especially timely given the future direction of clinical science. The National Institute of Mental Health recently introduced the Research Domain Criteria (RDoC) project, which is intended to utilize genetic, neurobiological, and behavioral science to redefine the taxonomy of psychopathology (Insel et al., 2010; Sanislow et al., 2010). The success of the

RDoC project relies on studying mental illness not only through self-report, but also through neural circuits, physiology, and behavior. These latter *units of analyses* often times involve the presentation of rapid visual images or words through experimental paradigms described above, and not unlike those tested in Study 1. RDoC-aligned projects such as Studies 2 and 3 can indeed advance the understanding of psychopathology, while maintaining safety and practicing clinical sensitivity to vulnerable research participants.

Beat the Past. As measured by the self-injury and suicide IATs, implicit associations can predict SITBs above and beyond SITB history. This was the case regarding the prediction of NSSI (Study 2) and suicidal ideation (Study 3). The prediction of NSSI above and beyond NSSI history is particularly notable given prior prospective evidence showing that frequency, recency, and number of methods of NSSI are all more predictive than IATs and other risk factors (e.g., emotional reactivity; impulsivity, temperament, anxiety, alcohol use; Chapman et al., 2009; Franklin et al., 2014; Glenn & Klonsky, 2011; Tuiksu et al., 2014). Conceptually speaking, people's implicit identification with cutting or death does not simply reflect whether they have self-injured or had suicidal thoughts in the past. There is something unique about how closely people pair the idea of themselves with cutting and death that informs their short-term decisions and actions.

Of note, Studies 2 and 3 controlled for SITB history by excluding non-injurers and non-ideators in prospective analyses, thereby producing prediction models for SITB persistence (vs. onset). Although impact of implicit associations on SITB onset remains unknown, I speculate that it plays a mediating role. For instance, adolescents may begin engaging in NSSI because peers who they closely identify with are engaging in NSSI. Indeed, adolescents have been shown to share similar attributes and identify with close friends and peer groups (e.g., Tolson & Urberg,

1993), and often learn about NSSI through their friends (Deliberto & Nock, 2008). Perhaps identification with others who engage in NSSI (i.e., *Me=Friend*) may initiate an indirect association between self-concept and NSSI (i.e., *Me=Friend=NSSI*) that would have otherwise not occurred. The influence of peer identity on other maladaptive outcomes has indeed been shown in the area of alcohol use and smoking (e.g., Duan, Chou, Andreeva, & Penz, 2009; Mercken, Candel, Williams, & de Vries, 2007). Using the Self-Injury IAT to capture mechansims of peer influence on NSSI represents one potential area for future study.

Tailor to Improve. Study 2 demonstrated that implicit identification with cutting is more specific to using cutting as a method of NSSI, relative to other NSSI methods. Indeed, the use of idiographic (vs. nomothetic) IAT stimuli in has been shown to effectively differentiate groups based on aggressiveness (Bluemke & Friese, 2012), self-esteem (Greenwald & Farnham, 2000), and drinker status (Lindgren, Westgate, Kilmer, Kaysen, & Teachman, 2012). This finding from Study 2 presents concrete and feasible ways for future IATs to target other outcomes of interest. For example, images of bruised skin may be a preferred set of stimuli to predict NSSI if a person has a reported NSSI history of hitting themselves. Beyond tailoring IAT stimuli to the outcome (e.g., cutting vs. hitting behavior), tailoring "Self"-related stimuli (e.g., respondents' first name, nickname, birthday) may further enhance predictive validity of the IAT as shown in prior studies (Bluemke & Friese, 2012).

Sometimes, Just Ask. In addition to the use of behavioral measures, we can rely on explicit self-report more than previously assumed. Beyond the aforementioned possibility that explicit self-report may be more accurate for predicting certain clinical outcomes (e.g., NSSI) over others (e.g., suicide attempt), here I expand on alternative reasons why people may more openly self-report SITBs. People may be especially accurate in their explicit self-report when

they possess greater awareness of their self-injurious urges. This may be the case for the adolescent inpatients (e.g., those from Study 2) who have received psychotherapeutic interventions that improve their awareness and willingness to seek help (e.g., mindfulness skills through Dialectical Behavior Therapy). This may also be the case for adults who have been experiencing suicidal thoughts for years (e.g., those from Study 3), and have likely encountered opportunities to engage in treatment of some kind.

Alternatively, it may make a difference exactly *when* adolescents are asked about their self-injurious urges. Perhaps patients are less motivated to conceal self-injurious urges once they have already been hospitalized (i.e., upon admission to an inpatient unit), compared to when they are presenting at a psychiatric emergency room (i.e., prior to admission to an inpatient unit).

A final possibility is that there may be a recent cultural shift deeming it acceptable to discuss SITB. This is especially true for NSSI, which is increasingly mentioned in the media (Whitlock, Purington, & Gershkovich, 2009). Adolescents may not perceive the topic of NSSI as sensitive in *either* a hospital or community setting. Future work is encouraged to clarify whether the predictive validity of self-report varies as a function of the predicted outcome, the treatment setting, or broader changes in how SITBs are perceived.

Mood Matters. Study 3 shows that how a person feels when completing behavioral measures may be just as important as the measures themselves, since implicit identification with death predicted suicidal ideation stronger when assessed in a negative (vs. neutral) mood state. This builds on prior work pertaining to SITB risk assessments, which had not addressed the role of respondents' mood. Study 3 experimentally demonstrates the importance of diathesis-stress framework of thinking. The mood induction used within Study 3 could be conceptualized as a proxy for more chronic life stressors or co-occurring depression, which has been proposed to

interact with psychological risk factors to explain depression and suicidality and NSSI (Lau et al., 2004; Nock & Cha, 2009; Teasdale et al., 1988).

Importantly, the specific negative mood induction procedure was selected not only for its effectiveness at prompting a negative emotional state, but also for its relatively short-lasting impact. That is, it lasted long enough to show the potential impact of negative mood on cognitive processes, but short enough not to permanently exacerbate the processes themselves. Functionally, the negative mood induction is similar to diagnostic tests used in other areas of healthcare. For example, dentists conduct pulp sensitivity tests by briefly exposing a troublesome tooth (i.e., pulp) to cold or heat, which helps determine its reactivity. The 'stress' exposure itself does not exacerbate the toothache, but creates the appropriate conditions under which the cause of the toothache is detected. Similarly, briefly inducing a negative mood state may amplify the read of cognitive reactivity related to SITBs.

The mood-dependent nature of IAT performance may help explain Study 2 results. One of the reasons why implicit identification with cutting may have not significantly changed over time in Study 2 was because mood had naturally improved by the second IAT administration at discharge, relative to the initial IAT administration at admission. Without consistent mood at both assessment time points, observing admission-to-discharge change patterns in implicit identification with death may understandably not produce meaningful results. Identifying the expected contextual norms of treatment settings would help determine which risk assessments to implement, and how (i.e., with vs. without a mood induction).

Piecing Together the Puzzle. While much of the dissertation focused on implicit associations related to SITBs, they alone do not predict outcomes. Study 3 results revealed that suicide-specific attentional bias enhances prediction of suicidal ideation. While *many* people who

identified with death continued thinking about suicide, *all* people who identified with death *and* attended specifically to suicide continued thinking about suicide. Together, this finding not only enhances our understanding of cognitive risk factors, but also offers a more accessible and easily interpretable and disseminable scoring method. Future studies are encouraged to continue exploring this technique of dual-task (ideally, multiple-task) profiles.

Limitations

Findings should be interpreted in light of several limitations. Study-specific limitations are reviewed first, followed by general limitations of this work.

Specific Limitations of Study 1. Studies 1a-c did not include a non-SITB-related stimuli control condition, and so it is not possible to draw causal inferences about whether SITB-related stimuli exposure caused such observed changes. This omission is less critical for these series of studies—given the implausible assumption that self-harm and suicidal urges would increase spontaneously for those exposed to no stimuli or non-SITB-related stimuli—and so there is no strong justification for running such a condition. Nevertheless, without such a condition one could argue that the observed patterns represent natural change over time. Present data still suggest this is an unlikely possibility. For instance, mood declined specifically following images rather than words (Study 1a), and mood decreased following negative stimuli and improved following positively-valenced images (Study 1d). These systematic changes would be difficult to explain if this first possible limitation held true.

Additionally, the longer-term impact of SITB-related behavioral measures remains unknown. I can conclude that such measures do not have immediate iatrogenic effects, and merely speculate that these would generalize to a longer-term time frame. Prior work suggests that asking SITB-related questions does not increase future likelihood of SITBs, with effects

ranging from three weeks to two years (Muehlenkamp et al., 2015). Future research to prospectively explore longer-term iatrogenic effects of behavioral measures is strongly encouraged

Specific Limitations of Study 2. Importantly, patients in Study 2 experienced variable lengths of stay at the inpatient unit. Although patients on average stayed for approximately two weeks, this ranged from as few as 30 hours to as many as 116 days. The varying durations between admission and discharge represents another barrier to comparing scores over time, and to predicting NSSI occurring within a hospital stay. According to post-hoc results, narrowing the range to a more standardized time frame indeed enhances the predictive validity of the IAT. Of note, the presence of this limitation highlights the robustness of self-report change over time, since it continued to predict post-discharge NSSI in spite of such contextual variability.

Specific Limitations of Study 3. In Study 3, demand characteristics were inherent to the mood induction design. Mood inductions absent of demand characteristics (e.g, unsolvable anagram tasks, manipulated achievement tests; Chartier & Ranieri, 1989) offer an alternative mood induction procedure. They have recently been used among self-injurious individuals, who experience greater physiological reactivity to a stressful task than nonsuicidal individuals (Nock & Mendes, 2008). Applying alternative mood inductions would help tease apart mechanisms driving the observed mood decline. Another limitation to Study 3 is that nonsuicidal adults were largely absent of psychopathology. While efforts were made to statistically control for prominent differentiating factors such as mood disorder, it was not possible control for all contextual differences between people who had versus had never experienced suicidal thoughts in their lifetime. Future work is encouraged to include a more conservative control group of depressed nonsuicidal adults, similar to prior work (e.g., Williams et al., 2005, 2008).

Limitations across Studies. The measurement of mood across Studies 1-3 relied on self-report. This was the case when examining mood response before and after IATs (Study 1), at the beginning and end of hospitalization (Study 2), and before and after mood induction procedures (Study 3). While Study 3 distinguished emotional valence from arousal following Russell's (1980) Circumplex Model of emotions, both sets of ratings still relied on self-report. Self-report measures were used out of feasibility and convenience, but the cost of this assessment method is potentially inaccurate self-report (Tourangeau & Yan, 2007). Reliance on more objective outcomes, such as behavioral performance, physiological arousal, or actual SITB outcomes may yield more accurate findings.

Additionally, Studies 1-3 included a restricted range of measures besides the IAT. This affects interpretation of Study 1 results, for example, since it remains unknown whether the lack of iatrogenic effects of IATs generalize to alternative implicit measures – some of which have been examined among self-injurers (e.g., Affect Misattribution Procedure; Franklin et al., 2014) and some of which have not (e.g., Affect Priming Paradigm; Fazio, Sanbonmatsu, Powell, & Kardes, 1986). The restricted range of measures and additional risk factors also affects interpretations of Studies 2 and 3, since the incremental predictive validity of the IAT is only relative to comparison risk factors that were controlled for. There remain strongly predictive risk factors that were not accounted for (e.g., childhood adversities, borderline personality disorder features; Bruffaerts et al., 2010; Glenn & Klonsky, 2011). Of note, past SITB has been shown to be one of the most robust predictors of future SITB (Borges, Angst, Nock, Ruscio, & Kessler, 2008; Chapman et al., 2009; Tuisku et al., 2014), so additional predictive variance may have been largely eliminated by controlling for SITB history in Studies 2 and 3. Future work is

encouraged to test shorter-term predictive models involving multiple behavioral measures and risk factors of SITB.

Finally, with the exception of Study 1, there was limited statistical power to detect small effects. Although prior community-based research suggests large group differences in implicit associations about NSSI (Nock & Banaji, 2007), clinical samples are likely to produce more moderate effects (Franklin et al., 2015). An additional consideration is that the predictive validity of IATs may weaken (i.e., yields smaller effect sizes) as the duration of follow-up period lengthens. Larger sample sizes, whether community-based or clinical, may be required to detect the smaller effect sizes of longer-term prediction models.

Conclusion

Behavioral measures assessing SITB risk are safe, predictive, and mood-dependent. By applying behavioral measures, this dissertation has asserted that people who identify with cutting and death continue hurting themselves and thinking of suicide in the short-term future. This applies to hospital settings where it is helpful to determine who will deliberately hurt themselves during their hospital stay, and to more neutral laboratory settings to predict who will continue experiencing suicidal thoughts. In the latter case, a person's negative mood state may enhance the ability to detect implicit associations. Consideration of implicit identification with death with other cognitive risk factors, namely suicide-specific attentional bias, further explains which suicidal people will continue thinking of suicide. Combining cognitive factors can be more predictive of suicidal ideation than explicit self-report. Whether present findings generalize to more severe outcomes such as suicide attempt remains unknown, and represents a promising area for future work. Continuation of this research will inform thoughtful and impactful application of behavioral measures, and ultimately serve a population in need of greater care and attention.

REFERENCES

Ahearn, E.P. (1997). The use of visual analog scales in mood disorders: A critical review. *Journal of Psychiatry Research, 31,* 569-579.

Allen, K. J. D., & Hooley, J. M. (2015). Inhibitory control in people who self-injure: Evidence for impairment and enhancement. *Psychiatry Research, 225*, 631-637.

Amir, N., McNally, R.J., Riemann, B.C., Burns, J., Lorenz, M., & Mullen, J.T. (1996). Suppression of the emotional Stroop effect by increased anxiety in patients with social phobia. *Behavior Research and Therapy, 34,* 945-948.

Beck, A.T., Steer, R.A., & Ranieri, W.F. (1988). Scale for Suicidal Ideation: Psychometric properties of a self-report version, *Journal of Clinical Psychology, 44,* 499-505.

Becker, E.S., Strohbach, D., & Rinck, M. (1999). A specific attentional bias in suicide attempters. *Journal of Nervous and Mental Disease, 187,* 730-735.

Beevers, C.G. (2005). Cognitive vulnerability to depression: A dual process model. *Clinical Psychology Reviews, 25,* 975-1002.

Bluemke, M., & Friese, M. (2012). On the validity of idiographic and generic self-concept implicit association tests: A core-concept model. *European Journal of Psychiatry, 26,* 515-528.

Borges, G., Angst, J., Nock, M.K., Ruscio, A.M., & Kessler, R.C. (2008). Risk factors for the incidence and persistence of suicide-related outcomes: A 10-year follow-up study using the National Comorbidity Surveys. *Journal of Affective Disorders, 105,* 25-33.

Bruffaerts, R., Demyttenaere, K., Borges, G., Haro, J.M., Hwang, I., Karam, E.G.,…Nock, M.K. (2010). Childhood adversities as risk factors for the onset and persistence of suicidal behavior. *British Journal of Psychiatry, 197,* 20-27.

Bruffaerts, R., Demyttenaere, K., Hwang, I., Chiu, W.T., Sampson, N., Kessler, R.C.,…Nock, M.K. (2011). Treatment of suicidal persons around the world. *British Journal of Psychiatry, 199,* 64-70.

Bryan, C.J., Dhillon-Davis, L.E., Dhillon-Davis, K.K. (2009). Emotional impact of a video-based suicide prevention program on suicidal viewers and suicide survivors. *Suicide and Life-Threatening Behavior, 39,* 623-632.

Buchanan, E.A., & Hvizdak, E.F. (2009). Online survey tools: Ethical and methodological concerns of Human Research Ethics Committees. *Journal of Empricial Research on Human Research Ethics, 4,* 37-48.

Busch, K.A., Fawcett, J., & Jacobs, D.G. (2003). Clinical correlates of inpatient suicide. *Journal of Clinical Psychiatry, 64,* 14-19.

Canli, T., Desmond, J.E., Zhao, Z., & Gabrieli, J.D.E. (2002). Sex differences in the neural basis of emotional memories. *Proceedings of the National Academy of Sciences, 99,* 10789-10794.

Cha, C.B., Najmi, S., Park, J.M., Finn, C.T., & Nock, M.K. (2010). Attentional bias toward suicide-related stimuli predicts suicidal behavior. *Journal of Abnormal Psychology, 119,* 616-622.

Center for Disease Control and Prevention (CDC) (2015). Web-based Injury Statistics Query and Reporting System. www.cdc.gov/injury/wisqars/fatal.html.

Chapman, A.L., Derbidge, C.M., Cooney, E., Hong, P.Y., & Linehan, M.M. (2009). Temperament as a prospective predictor of self-injury among patients with borderline personality disorder. *Journal of Personality Disorders, 23,* 122-140.

Chartier, G.M., & Ranieri, D.J. (1989). Comparison of two mood induction procedures.

Cognitive Therapy and Research, 13, 275-282.

Clark, D.A., Beck, A.T., & Alford, B.A. (1999). *Scientific foundations of cognitive theory and therapy for depression.* New York: John Wiley & Sons.

Clark, D.M., & Teasdale, J.D. (1985). Constraints on the effects of mood on memory. *Journal of Personality and Social Psychology, 48,* 1595-1608.

Clerkin, E.M., Fisher, C.R., Sherman, J.W., & Teachman, B.A. (2014). Applying the Quadruple Process model to evaluate change in implicit attitudinal responses during therapy for panic disorder. *Behavior Research and Therapy, 52,* 17-25.

Cukrowicz, K., Smith, P., & Poindexter, E. (2010). The effect of participating in suicide research: Does participating in a research protocol on suicide and psychiatric symptoms increase suicide ideation and attempts? *Suicide and Life-Threatening Behavior, 40,* 535-543.

Cullen, K.R., Westlund, M.K., LaReviere, L.L., & Klimes-Dougan, B. (2013). An Adolescent with nonsuicidal self-injury: A case and discussion of neurobiological research on emotion regulation. *American Journal of Psychiatry, 170,* 828-831.

Cunningham, W.A., Preacher, K.J., & Banaji, M.R. (2001). Implicit attitude measures: Consistency, stability, and convergent validity. *Psychological Science, 12,* 163-170.

Deeley, S.T., & Love, A.W. (2010). Does asking adolescents about suicidal ideation induce negative mood state? *Violence and Victims, 25,* 677-688.

Dickstein, D.P., Puzia, M.E., Cushman, G.K., Weissman, A.B., Wegbreit, E., Kim, K.L., Nock, M.K., & Spirito, A. (in press). Self-injurious implicit attitudes among adolescent suicide attempters versus those engaged in nonsuicidal self-injury. *Journal of Child Psychology and Psychiatry*.

DuBois, J.M., Beskow, L., Campbell, J., Dugosh, K., Festinger, D., Hartz, S.,…Lidz, C. (2012). Restoring balance: A consensus statement on the protection of vulnerable research participants. *Public Health Ethics, 102,* 2220-2225.

Elgersma, H.J., Glashouwer, K.A., Bockting, C.L.H., & Penninx, B.W.J.H. (2013). Hidden scars in depression? Implicit and explicit self-associations following recurring depressive episodes. *Journal of Abnormal Psychology, 122,* 951-960.

Farris, S.R., Ostafin, B.D., & Palfai, T.P. (2010). Distractibility moderates the relation between automatic alcohol motivation and drinking behavior. *Psychology of Addictive Behaviors, 24,* 151-156.

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,* 229-238.

First, M.B., Spitzer, R.L., Gibbon, M., & Williams, J.B.W. (2002). *Structured Clinical Interview for DSM-IV Axis I Disorders, Research Version, Non-patient Edition. (SCID-I/NP).* New York: Biometrics Research, New York State Psychiatric Institute, November 2002.

Franklin, J.C., Glenn, C.R., & Nock, M.K. (2015). Meta-analysis of risk factors for suicidal behavior. Manuscript in preparation.

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. *Journal of Abnormal Psychology, 122,* 463-469.

Friese, M., Bluemke, M., & Wanke, M. (2007). Predicting voting behavior with implicit attitude measures: The 2002 German parliamentary election. *Experimental Psychology, 54,* 247-255.

Frost, R.O., & Green, M.L. (1982). Velten mood induction procedure effects: Duration and

postexperimental removal. *Personality and Social Psychology Bulletin, 8,* 341-347.

Gamer, J., Schmukle, S.C., Luka-Krausgrill, U., & Egloff, B. (2008). Examining the dynamics of the implicit and the explicit self-concept in social anxiety: Changes in the Implicit Association Test-Anxiety and the Social Phobia Anxiety Inventory following treatment. *Journal of Personality Assessment, 90*, 476-480.

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*, 282-289.

Gibson, S., Boden, S.V.R., Benson, O., & Brand, S.L. (2014). The impact of participating in suicide research online. *Suicide and Life-Threatening Behavior, 44,* 372-383.

Gilboa-Schechtman, E., Revelle, W., & Gotlib, I.H. (2000). Stroop interference following mood induction: Emotionality, mood congruence, and concern relevance. *Cognitive Therapy and Research, 24,* 491-502.

Giletta, M., Scholte, R.H., Engels, R.C., Ciairano, S., & Prinstein, M.J. (2012). Adolescent non-suicidal self-injury: A cross-national study of community samples from Italy, the Netherlands and the United States. *Psychiatry Research, 197,* 66–72.

Glenn, C.R., & Klonsky, E.D. (2011). Prospective prediction of nonsuicidal self-injury: A 1-year longitudinal study in young adults. *Behavior Therapy, 42,* 751-762.

Glenn, C.R., & Nock, M.K. (2014). Improving the short-term prediction of suicidal behavior. *American Journal of Preventive Medicine, 47,* S176-S180.

Gotlib, I.H., & Cane, D.B. (1987). Construct accessibility and clinical depression: A longitudinal investigation. *Journal of Abnormal Psychology, 96,* 199-204.

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: A randomized controlled trial. *Journal of the American Medical Association, 293,* 1635-1643.

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

Greenwald A.G., Nosek B.A., & Banaji M.R. (2003) Understanding and using the Implicit Association Test I: An improved scoring algorithm. *Journal of Personality and Social Psychology, 85*, 197–216.

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

Guan, K., Fox, K.R., & Prinstein, M.J. (2012). Nonsuicidal self-injury as a time-invariant predictor of adolescent suicide ideation and attempts in a diverse community sample. *Journal of Consulting and Clinical Psychology, 80,* 842-849.

Haeffel, G.J., Abramson, L.Y., Brazy, P.C., Shah, J.Y., Teachman, B.A., & Nosek, B.A. (2007). Explicit and implicit cognition: A preliminary test of a dual-process theory of cognitive vulnerability to depression. *Behavior Research and Therapy, 45,* 1155-1167.

Hofmann, W., Gschwendner, T., Nosek, B.A., & Schmitt, M. (2005). What moderates implicit-explicit consistency? *European Review of Social Psychology, 16,* 335-390.

Horesh, N., & Apter, A. (2006). Self-disclosure, depression, anxiety, and suicidal behavior in adolescent psychiatric inpatients. *Crisis, 27,* 66-71.

Horesh, N., Zalsman, G., & Apter, A. (2004). Suicidal behavior and self-disclosure in adolescent psychiatric inpatients. *The Journal of Nervous and Mental Disease, 192,* 837-842.

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., Sanislow, C., & Wang,

P. (2010). Research domain criteria (RDoC): Toward a new classification framework for research on mental disorders. *American Journal of Psychiatry, 167,* 748–751.

Jamieson, J.P., Koslov, K., Nock, M.K., & Mendes, W.B. (2013). Experiencing discrimination increases risk-taking. *Psychological Science, 24,* 131-139.

Janis, I.B., & Nock, M.K. (2008). Behavioral forecasting does not improve the prediction of future behavior: A prospective study of self-injury. *Journal of Clinical Psychology, 64*, 1164-1174.

Jollant, F., Lawrence, N.S., Giampietro, V., Brammer, M.J., Fullana, M.A., Drapier, D.,…Phillips, M.L. (2008). Orbitofrontal cortex response to angry faces in men with histories of suicide attempts. *American Journal of Psychiatry, 165,* 740-748.

Kensinger, E.A., & Schacter, D.L. (2006). Processing emotional pictures and words: Effects of valence and arousal. *Cognitive, Affective, & Behavioral Neuroscience, 6,* 110-126.

Kessler, R.C., Berglund, P., Borges, G., Nock, M.K., & Wang, P.S. (2005). Trends in suicide ideation, plans, gestures, and attempts in the United States 1990-92 to 2001-2003. *Journal of the American Medical Association, 293,* 2487-2495.

Klonsky, E.D. (2011). Non-suicidal self-injury in United States adults: Prevalence, sociodemographics, topography and functions. *Psychological Medicine, 41,* 1981-1986.

Klonsky, E.D., May, A.M., & Glenn, C.R. (2013). The relationship between nonsuicidal self-injury and attempted suicide: Converging evidence from four samples. *Journal of Abnormal Psychology, 122,* 231-237.

Lakeman, R., & Fitzgerald, M. (2009). The ethics of suicide research: The views of ethics committee members. *Crisis, 30,* 13-19.

Lang, P.J., Bradley, M.M., & Cuthbert, B.N. (2008). *International Affective Picture System (IAPS): Affective Ratings of Pictures and Instruction Manual, Technical Report A-8.* University of Florida, Gainesville, FL.

Lang, P.J., Greenwald, M.K., Bradley, M.M., & Hamm, A.O. (1993). Looking at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology, 30,* 261-273.

Lau, M.A., Segal, Z.V., & Williams, J.M.G. (2004). Teasdale's Differential Activation Hypothesis: Implications for mechanisms of depressive relapse and suicidal behavior. *Behavior Research and Therapy, 42,* 1001-1017.

Lindgren, K.P., Westgate, E., Kilmer, J., Kaysen, D., & Teachman, B.A. (2012). Pick your poison: Stimuli selection in implicit measures. *Addictive Behaviors, 37,* 990-993.

Linehan, M.M., Comtois, K.A., Brown, M.Z., Heard, H.L., & Wagner, A. (2006). Suicide Attempt Self-Injury Interview (SASII): development, reliability, and validity of a scale to assess suicide attempts and intentional self-injury. *Psychological Assessment, 18,* 303-312.

Lloyd, E.E., Kelley, M.L., & Hope, T. (1997, April). *Self-mutilation in a community sample of adolescents: Descriptive characteristics and provisional prevalence rates.* Poster session presented at the annual meeting of the Society for Behavioral Medicine, New Orleans, LA.

Martin, M. (1990). On the induction of mood. *Clinical Psychology Review, 10,* 669-697.

Mathias, C.W., Furr, R.M., Sheftall, A.H., Hill-Kapturczak, N., Crum, P., & Dougherty, D.M. (2012). What's the harm in asking about suicidal ideation? *Suicide and Life-Threatening Behavior, 42,* 341-351.

Millner, A., D'Angelo, E.J., & Nock, M.K. (2015). Using implicit measures to assess and

predict suicidal thoughts and behaviors among adolescents. Manuscript in preparation.

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,* 26-37.

Nisbett, R.E., & Wilson, T.D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84,* 231-259.

Nixon, M.K., Cloutier, P.F., & Aggarwal, S. (2002). Affect regulation and addictive aspects of repetitive self-injury in hospitalized adolescents. *Journal of American Academy of Child and Adolescent Psychiatry, 41,* 1333-1341.

Nock, M.K. (2009a). Why do people hurt themselves? New insights into the nature and functions of self-injury. *Current Directions in Psychological Science, 18,* 78-83.

Nock, M.K. (2009b). Self-Injury. *Annual Review of Clinical Psychology, 6,* 339-363.

Nock, M.K., & Banaji, M.R. (2007). Assessment of self-injurious thoughts using a behavioral test. *American Journal of Psychiatry 164,* 820-823.

Nock, M.K., & Cha, C.B. (2009). Psychological models of nonsuicidal self-injury. *In M.K. Nock (Ed.), Understanding nonsuicidal self-injury: Origins, assessment, and treatment (pp. 65-78).* Washington, DC: American Psychological Association

Nock, M.K., Holmberg, E.B., Photos, V.I., & Michel, B.D. (2007). The Self-Injurious Thoughts and Behaviors Interview: Development, reliability, and validity in an adolescent sample. *Psychological Assessment, 19,* 309-317.

Nock, M.K., Joiner, T.E., Gordon, K.H., Lloyd-Richardson, E., & Prinstein, M.J. (2006). Non-suicidal self-injury among adolescents: Diagnostic correlates and relation to suicide attempts. *Psychiatry Research, 144,* 65-72.

Nock, M.K., & Mendes, W.B. (2008). Physiological arousal, distress tolerance, and social problem solving deficits among adolescent self-injurers. *Journal of Consulting and Clinical Psychology, 76,* 28-38.

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,* 511-517.

Nock, M.K., & Prinstein, M.J. (2004). A functional approach to the assessment of self-mutilative behavior. *Journal of Consulting and Clinical Psychology, 72,* 885-890.

Peterson, J., Skeem, J., & Manchak, S. (2011). If you want to know, consider asking: How likely is it that patients will hurt themselves in the future? *Psychological Assessment, 23,* 626-634.

Phillips, W.J., Hine, D.W., & Thorsteinsson, E.B. (2010). Implicit cognition and depression: A meta-analysis. *Clinical Psychology Review, 30,* 691-709.

Plenar, P.L., Bubalo, N., Fladung, A.K., Ludolph, A.G., Lulé, D. (2012). Prone to excitement: Adolescent females with non-suicidal self-injury (NSSI) show altered cortical pattern o emotional and NSS-related material. *Psychiatry Research: Neuroimaging, 203,* 146-152.

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,* 335-343.

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*, 522-526.

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,* 714-721.

Reisch, T., Seifritz, E., Esposito, F., Wiest, R., Valach, L., & Michel, K. (2010). An fMRI study on mental pain and suicidal behavior. *Journal of Affective Disorders, 126,* 321-325.

Robinson, J., Yuen, H.P., Martin, C., Hughs, A., Baksheev, G.N., Dodd, S.,…Yung, A.R. (2011). Does screening high school students for psychological distress, deliberate self-harm, or suicidal ideation cause distress—and is it acceptable? *Crisis, 32,* 254-263.

Rudd, M.D., Madrusiak, M., Joiner, T.E., Berman, A.L., Van Orden, K.A., & Hollar, D. (2006). The emotional impact and east of recall of warning signs for suicide: A controlled study. *Suicide and Life-Threatening Behavior, 36,* 288-295.

Russell, J.A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39,* 1161-1178.

Russell, J.A., Weiss, A., Mendelsohn, G.A. (1989). Affect Grid: A single-item scale of pleasure and arousal. *Journal of Personality and Social Psychology, 57,* 493-502.

Sagaspe, P., Sanchez-Ortuno, M., Charles, A., Taillard, J., Valtat, C., Bioulac, B., & Philip, P. (2006). Effects of sleep deprivation on color-word, emotional, and specific Stroop interference on self-reported anxiety. *Brain and Cognition, 60,* 76-87.

Sanislow, C. A., Pine, D. S., Quinn, K. J., Kozak, M. J., Garvey, M. A., Heinssen, R. K., … Cuthbert, B. N. (2010). Developing constructs for psychopathology research: Research domain criteria. *Journal of Abnormal Psychology, 119,* 631–639.

Schnabel, K., Asendorpf, J.B., & Greenwald, A.G. (2008). Assessment of individual differences in implicit cognition: A review of IAT measures. *European Journal of Psychological Assessment, 24,* 210-217.

Segal, Z.V., & Ingram, R.E. (1994). Mood priming and construct activation in tests of cognitive

vulnerability to unipolar depression. *Clinical Psychology Review, 14,* 663-695.

Shaffer, D., Vieland, V., Garland, A., Rojas, M., Underwood, M., & Busner, C. (1990). Adolescent suicide attempters: Response to suicide-prevention programs. *Journal of the American Medical Association, 24,* 3151-3155.

Sheehan, D.V., Lecrubier, Y., Harnett-Sheehan, K., Amorim, P., Janavs, J., Weiller, E.,…Dunbar, G.C. (1998). The MINI-International Neuropsychiatric Interview (M.I.N.I): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *Journal of Clinical Psychiatry, 59,* 22-33.

Skeem, J.L., Manchak, S.M., Lidz, C.W., & Mulvey, E.P. (2013). The utility of patients' self-perceptions of violence risk: Consider asking the person who may know best. *Psychiatric Services, 64,* 401-415.

Stoppe, G., Sandholzer, H., Huppertz, C., Duwe, H., & Staedt, J. (1999). Family physicians and the risk of suicide in the depressed elderly. *Journal of Affective Disorders, 54,* 193-198.

Teachman, B.A., Marker, C.D., & Smith-Janik, S.B. (2008). Automatic associations and panic disorder: Trajectories of change over the course of treatment. *Journal of Consulting and Clinical Psychology, 76,* 988-1002.

Teachman, B.A., & Woody, S.R. (2003). Automatic processing in spider phobia: Implicit fear associations over the course of treatment. *Journal of Abnormal Psychology, 112,* 100-109.

Teasdale, J.D. (1988). Cognitive vulnerability to persistent depression. *Cognition and Emotion, 2,* 247-274.

Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin, 133,* 859-883.

Tuisku, V., Kiviruusu, O., Pelkonen, M., Karlsson, L., Strandholm, T., & Marttunen, M. (2014). Depressed adolescents as young adults – Predictors of suicide attempt and non-suicidal self-injury during an 8-year follow-up. *Journal of Affective Disorders, 152-154,* 313-319.

Wallenstein, M.B., & Nock, M.K. (2007). Physical exercise for the treatment of non-suicidal self-injury. Evidence from a single-case study. *American Journal of Psychiatry, 164,* 350-351.

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,* 1063-1070.

Werntz, A.J., Steinman, S.A., Glenn, J.J., Nock, M.K., & Teachman, B.A. (2015). Characterizing Implicit Mental Health Associations across Clinical Domains. Manuscript under review.

Whitlock, J., Eckenrode, J., & Silverman, D. (2006). Self-injurious behaviors in a college population. *Pediatrics, 117,* 1939-1948.

Whitlock, J.L., Purington, A., & Gershkovich, M. (2009). Influence of the media on self-injurious behavior. In M.K. Nock (Ed.), *Understanding non-suicidal self-injury: Origins, assessment, and treatment* (pp. 139-156). Washington, DC: American Psychological Association Press.

Whitlock, J., Pietrusza, C., & Purington, A. (2013). Young adult respondent experiences of disclosing self-injury, suicide-related behavior, and psychological distress in a web-based survey. *Archives of Suicide Research, 17,* 20-32.

Wilkinson, P., Kelvin, R., Roberts, C., Dubicka, B., & Goodyer, I. (2011). Clinical and

psychosocial predictors of suicide attempts and nonsuicidal self-injury in the Adolescent Depression and Antidepressants and Psychotherapy Trial (ADAPT). *American Journal of Psychiatry, 168,* 495–501.

Williams, J.M.G., Barnhofer, T., Crane, C., & Beck, A.T. (2005). Problem solving deteriorates following mood challenge in formerly depressed patients with a history of suicidal ideation. *Journal of Abnormal Psychology, 114,* 421-431.

Williams, J.M.G., & Broadbent, K. (1986). Distraction by emotional stimuli: Use of a Stroop task with suicide attempters. *British Journal of Clinical Psychology, 25,* 101-110.

Williams, J.M.G., Mathews, A., & MacLeod, C. (1996). The emotional Stroop task and psychopathology. *Psychological Bulletin, 120,* 3-24.

Williams, J.M.G., Van der Does, A.J.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,* 83-104.

Yates, T.M. (2004). The developmental psychopathology of self-injurious behavior: Compensatory regulation in posttraumatic adaptation. *Clinical Psychology Review, 24,* 35-74.

You, J., & Lin, M. (in press). Predicting suicide attempts by time-varying frequency of nonsuicidal self-injury among Chinese community adolescents. *Journal of Consulting and Clinical Psychology*.

Zlotnick, C., Donaldson, D., Spirito, A., & Pearlstein, T. (1997). Affect regulation and suicide attempts in adolescent inpatients. *Journal of the American Academy of Child and Adolescent Psychiatry*, *36*, 793–798.

APPENDIX:

Studies 1a-d

Studies 1a-d were each designed to test other hypotheses about SITBs; however, questions relevant to iatrogenic risk were administered in each study immediately before and after participants viewed SITB-related stimuli.

Study 1a Method. Importantly, data for Study 1a were taken from a larger sample of 5,261 consenting participants. Part way through data collection, a question asking for participant age was added, and individuals who reported being under the age of 18 were excluded from all analyses. Adult participants who did not complete the questions about iatrogenic effects also were excluded. Compared to those excluded (n=1,957), included participants were older, p=.004, d=.12, more likely to be Caucasian/White, and more likely to have SITB history, ps<.001, Φs=.09-.17. There were no differences across gender and ethnicity, ps=.08-.27, Φs=.02-.03. Visitors to the PIMH website elected to complete a task relevant to suicide or self-injury from a list of mental health-relevant topics (e.g., anxiety, depression), but were then randomly assigned to complete only one of three different SITB-related IATs. Those completing the different IATs did not vary across demographic factors or SITB history, ps=.06-.95, nor did they differ on baseline ratings of desire to self-injure, desire to die, or mood, ps=.40-.97. Responses across the three groups were collapsed to increase statistical power to detect potential iatrogenic effects.

To recruit online participants, the PIMH website was publicized throughout public and academic talks, news stories, and academic websites. Visitors to the PIMH website underwent an informed consent process, during which they learned about the IAT, the feedback they would receive, and resources for mental health care referral and help. Participants then completed measures in the aforementioned order.

At the beginning of the IAT, the respondent was first presented with a screen on which *target-attribute* category pairs were shown in the upper left and right corners. Category pairs presented in the first block (e.g., *Cutting-Me, Not Cutting-Not Me*) would then be switched in the second block (e.g., *Cutting-Not Me, Not Cutting-Me*), in random order. The specific category labels depended on the type of IAT administered. For this study, respondents completed one of three possible SITB-related IATs: (1) *Cutting/No Cutting* target images with *Me/Not Me* attribute words, (2) *Death/Life* target words with *Me/Not Me* attribute words, and (3) *Suicide/Life* target words with *Me/Not Me* attribute words. Once the trials began, words or images appeared in the middle of the screen, one at a time. The respondent was instructed to sort these words or images into the appropriate category, either by pressing the left or right key. Participants were asked to go as fast as they could while making as few mistakes as possible. If the respondent categorized correctly, the next trial began. If the respondent categorized incorrectly, a red "X" appeared and s/he was asked to try again. The IAT is designed such that respondents' speed of classifying stimuli into correct categories helps determine the associative strength between category pairs. IAT data were not directly used in analyses.

Study 1b Method. Participants were recruited through an online study pool system in which they read a brief description of the study and voluntarily signed up to participate. Exclusion criteria for the Study 1b sample included: any factor that would impair one's ability to consent, comprehend, or effectively participate in all parts of the research study (e.g., inability to speak English fluently, severe cognitive impairment, extremely agitated/violent behavior). Participants were compensated with course credit.

Each participant completed ten consecutive IATs: (1) a practice *Flower/Insect* target images with *Good/Bad* attribute words, (2) *Death/Life* target words with *Me/Not Me* attribute

words (3) *Despair/Hope* target words with *Me/Not Me* attribute words, (4) *Escape/Stay* target words with *Me/Not Me* attribute words, (5) *Suicide/Life* target words with *Me/Not Me* attribute words, (6) single-category *Death* target words with *Me/Not Me* attribute words, (7) single-category *Despair* target words with *Me/Not Me* attribute words, (8) single-category *Escape* target words with *Me/Not Me* attribute words, (9) single-category *Suicide* target words with *Me/Not Me* attribute words, and (10) single-category *Suicide* target images with *Me/Not Me* attribute words.

Study 1c Method. All patients admitted to the psychiatric inpatient unit were provided with a recruitment brochure. Those who indicated interest and received approval from their treatment team were approached for consent. Patients and their guardians provided assent and consent, respectively, prior to participation. Inclusion criteria were admission to the adolescent psychiatric inpatient unit, and hospital staff members' approval of the patients' participation. Exclusion criteria were the same as Study 1b, extending to both patients and their parents.

The current sample was drawn from a larger sample of 131 recruited patients. Patients were excluded from analyses either because they did not receive all necessary measures due to a mid-study change in procedures (n=44), or because they withdrew from participation (n=4). Compared to those excluded, included participants did not differ across most demographic or most psychiatric factors (ps=.17-.96). They also did not differ across most types of SITB (ps=.28-.60). There were two exceptions: a greater proportion of included patients were diagnosed with eating disorders, $\chi^2_{(1)}$=4.36, p=.04, Φ=-0.19, and had attempted suicide, $\chi^2_{(1)}$=5.03, p=.03, Φ=0.20.

Study 1d Method. Participants were recruited from the community via printed and electronic advertisements. Consenting individuals who appeared to meet eligibility criteria based on a brief phone screen, completed one laboratory visit and a subsequent fMRI scanning session.

All participants were required to be right handed, and had no history of head injury or neurological problems.

The PANAS is a 20-item measure consisting of 10 positive and 10 negative mood items. All items are rated using a five-point scale where 1=very slightly or not at all and 5=extremely. PANAS scales are internally consistent and demonstrate excellent convergent and discriminant validity (Watson et al., 1988).

During the scanning portion of the study, each block contained 12 pictures of the same type (neutral, NSSI-related, negative, positive). Across valence type (positive, negative and neutral), pictures were matched for arousal intensity. Some of the NSSI pictures showed a female about to engage in self injury (e.g., pressing a razor onto her skin). Others, obtained through an online image search, showed completed NSSI and depicted cut and bloodied skin. Each picture's owner granted permission for her picture to be used in the study.

Supplemental Data

Study 1a Post-Hoc Analyses. The observed increase in desire to self-injure in Study 1a, albeit slight, warrants further investigation. Post-hoc analyses focused on the subsample of respondents reporting an increase in desire to self-injure (n=239, 7.2% of overall sample). Compared to respondents who did not report increased desire to self-injure, this subsample was more likely to have had history of suicide plan or attempt, χ^2s(1)=4.01-4.16, ps<.05, Φs=.13. They also reported stronger desire to self-injure at baseline, t(237)=2.78, p=.01, d=.36. Most of these 239 respondents indicated no higher than moderate desire to hurt themselves after the IAT (M=1.71, SD=0.86). The minority of them (n=48, 20.1%) reported *increased* desire to die, and fewer also reported mood decline (n=17, 7.5%). In sum, only 0.5% of all web-based respondents displayed consistently concerning trends across the three ratings. These post-hoc analyses reveal

that increased desire to self-injure appears to be infrequent, slight, and typically unaccompanied by other concerning patterns.