AN EXPERIMENTAL TEST OF A COGNITIVE DEFUSION EXERCISE: COPING WITH NEGATIVE AND POSITIVE SELF-STATEMENTS

Hilary-Anne Healy, Yvonne Barnes-Holmes,
Dermot Barnes-Holmes, and Claire Keogh
National University of Ireland, Maynooth

Carmen Luciano
University of Almeria, Spain

Kelly Wilson
University of Mississippi

*This study investigated the impact of defusion on a nonclinical sample (*n = 60*) in the context of negative (e.g., "I am a bad person") and positive (e.g., "I am whole") self-statements. Participants were assigned to one of three experimental conditions (Pro-Defusion, Anti-Defusion, and Neutral) that manipulated instructions about the impact of a defusion strategy. Defusion was also manipulated through the visual presentation of the self-statements, with each presented in three formats (Normal, Defused, Abnormal). Participants rated each self-statement for comfort, believability, and willingness. Although the instructions did not affect ratings, negative statements presented in the defused format decreased discomfort and increased willingness and believability relative to the nondefused statements. The findings suggest using defusion strategies in coping with negative psychological content.*

In acceptance and commitment therapy (ACT: Hayes, Strosahl, & Wilson, 1999), an acceptance-based therapeutic regime, clients are encouraged to make willing contact with aversive psychological content. In order to achieve this therapeutic aim, ACT interventions are often bolstered with techniques that facilitate what is commonly referred to as *cognitive defusion* (Bach & Hayes, 2002; Hayes, Strosahl, et al., 1999). From an ACT perspective, clients are frequently "fused" with painful or negatively evaluated psychological content,

and defusion strategies create a "de-fused" perspective that permits greater behavioral flexibility.

Defusion

Defusion is a difficult concept to define in simple language. In clinical terms, clients are described as being *"fused"* with their thoughts when they believe that their thoughts say something important about who they are. For example, if they frequently had the thought "I am a horrible person," they might easily begin to have additional and more convincing thoughts like "I really am a horrible person and nobody could ever love me." ACT clinicians encourage clients to *"defuse"* from their negative content by viewing their thoughts as just thoughts, rather than considering them to have meaning. In this way, defusion involves a change in perspective that permits clients to see themselves as more than simply the sum of their thoughts and personal evaluations. In order to achieve this change in perspective, ACT therapists also employ techniques that focus on a process referred to as "self-as-context" (e.g., the Floating Leaf Exercise). This technique establishes a broader sense of self within which thoughts and internal events have less importance (Hayes, Barnes-Holmes, & Roche, 2001). In more technical terms, the process of defusion has also been defined as the disruption of existing verbal functions of language such that the "ongoing process of framing events relationally is evident in the moment" (Hayes, Strosahl, et al., 1999, p. 74). This process is thought to result in a breaking down of existing problematic verbal relations, or at least appears to permit the person to see the relations without having to act in accordance with them.

Some aspects of defusion and self-as-context resemble, or are derived from, practices within cognitive research such as those that establish a distanced self-perspective (e.g., Kross, Ayduk, & Mischel, 2005) and those that involve mindfulness (e.g., Teasdale, Segal, Williams, Ridgeway, Soulsby et al, 2000). But Titchener's (1916) rapid word-repetition technique (e.g., "milk, milk, milk") is perhaps the most well-known method for facilitating defusion. Within ACT, for example, clients might be encouraged to repeat negative content words rapidly (e.g., by saying "stupid, stupid, stupid"). The therapeutic aim of such a strategy is that during the repetition exercise the semantic functions, or meaning of the word, will be significantly reduced. Indeed, clients report that toward the end of the exercise they experienced the word simply as a strange sound (Hayes, Strosahl, et al., 1999, p. 154). The therapist then uses this experience to highlight that the client's negative content is also purely verbal and not a reflection of reality.

Empirical Evidence for Defusion

Although cognitive defusion strategies are frequently used within an ACT-based clinical context, the empirical evidence to support their efficacy is relatively limited. For example, indirect evidence for the utility of defusion may be derived from chronic pain analog studies involving the Cold Presser Task (Hayes, Bisset, et al., 1999; Takahashi, Muto, Tada, & Sugiyama, 2002). These studies have indicated that acceptance-based interventions that rely heavily on defusion strategies increase participants' pain tolerance. Some support for the utility of defusion can also be derived from ACT intervention research. For example, Bach & Hayes (2002) reported positive clinical outcomes for psychosis

with a brief ACT intervention in which defusion was a central component. Specifically, clients were encouraged to notice thoughts rather than to believe them. They engaged in a common defusion exercise in which they were instructed to "take their minds for a walk." Indeed, the researchers argued that the 50% reduction in rehospitalization rates for the ACT group could be attributed to a reduction in the believability (i.e., defusion) of their psychotic psychological content. However, since defusion in this case was only one aspect of a complete ACT package, the specific role of defusion is difficult to decipher.

More direct evidence for defusion was generated by the first empirical investigation of the technique when Titchener's (1916) word-repetition exercise was recently used in the context of negatively evaluated self-referential content (Masuda, Hayes, Sackett, & Twohig, 2004). Specifically, undergraduate students were asked to generate two self-relevant negative thoughts that they found particularly disturbing (e.g., "I am too fat") and then to restate each thought in a single word (e.g., "fat"). Participants then rated the levels of discomfort and believability associated with each word. In Experiment 1, the researchers compared the defusion strategy (i.e., word repetition) with a distraction technique (i.e., reading an unrelated article). In Experiment 2, defusion was compared with thought control (e.g., participants could use breathing, positive self-talk, or positive imagery). The results from both experiments indicated that the defusion rationale produced the largest reductions in the believability of the negative self-relevant words, as well as in the levels of discomfort the words produced.

The Current Study

The current study was designed to test an alternative defusion technique to the one Masuda et al. (2004) used. The purpose was to ascertain if the reduction in believability was unique to the exercise used in the previous study, or if the same findings would be obtained across defusion strategies more generally. Rather than using word repetition, we presented a diffusion technique to undergraduate participants that ACT clinicians often use and is functionally similar to those described by Hayes, Strosahl, et al. (1999). Specifically, the phrase "I am having the thought that" was used to create a sense of cognitive distance between a participant and negative self-referential statements. For example, the effects of the statement "I am a bad person" (Normal presentation) were then compared with the effects produced by the same statement when prefixed with the defusion phrase ("*I am having the thought that* I am a bad person": Defused presentation). Similar to word repetition, the purpose of the prefix was to highlight the subsequent statement as simply a thought and not a fact.

In the current study, participants' reactions to negative and positive self-statements were measured with and without the defusion prefix, and with three self-report Likert-type rating scales that assessed levels of discomfort, believability, and willingness associated with the target statements. The three self-report scales were chosen on the basis of their inclusion in previous ACT research (e.g., Hayes, Bisset, et al., 1999; Levitt, Brown, Orsillo, & Barlow, 2004; Masuda et al., 2004). However, because the defusion prefix changed the length of the statement and might also be seen as creating an unusual verbalization, the self-statements were also prefixed with a phrase that matched the defusion statement in length and was equally unusual (i.e., "I have a wooden chair and

. . .": Abnormal presentation). This provided a comparison statement of similar length that was also unusual but did not have the same level of defusion as the cognitive distancing statements.

In the current study, we hypothesized that the defusion statements would likely impact directly upon the explicit ratings associated with the negative self-statements presented in the three formats (Normal, Abnormal, and Defused). Specifically, we hypothesized that the negative statements in the defused presentation format would likely decrease believability to see or say the statements, but would increase willingness, relative to the normal and abnormal statements. The believability prediction was consistent with ACT findings, in which the change of one's perspective about one's psychological content, by indicating that the person is greater than the sum of the content, reduces the believability of the content (Hayes, Luoma, Bond, Masuda, & Lillis, 2006). In other words, if I know that the thought "I am a horrible person" is only a thought, then I will be aware that the thought can say nothing about who I am as a real person. As a result, the thought will be less believable. Put simply, the believability of thoughts results from their potential to be viewed as having more meaning than they need to have (i.e., by saying something about who you are as a person). When this potential is reduced in the context of particular psychological content, that content, by definition, will be less believable. In the research by Masuda et al. (2004), the defusion exercise reduced the believability of the targeted psychological content.

The willingness prediction was also consistent with ACT; it simply asserted that if you are defused from psychological content such that it says nothing about who you are, you do not need to be unwilling to have it. In other words, if the content has no power, then you need not be afraid of what it can do to you. Although willingness had not been targeted directly by Masuda et al. (2004), other sources of empirical evidence have indicated that acceptance-based strategies increase willingness (Levitt et al., 2004). We therefore predicted that the defusion presentation format here would increase the willingness of participants to see or read the self-statements.

In the current research, we initially made no clear predictions about the impact of defusion on psychological discomfort. From an ACT perspective, discomfort is not targeted directly because one can still engage in valued action when levels of discomfort are high. Furthermore, because individuals generally have little or no control over their emotional states, ACT encourages clients not to focus upon these as a potential source of behavior change. According to this view then, discomfort may increase, decrease, or remain unchanged. Specifically, one might argue that increasing acceptance of content might *increase* discomfort because one is more willing to make experiential contact with it. Alternatively, one might predict that discomfort would *decrease* when one realizes that thoughts associated with emotion are only thoughts, and as such have limited control or impact on who you really are. Interestingly, Masuda et al. (2004) reported that their defusion technique resulted in decreased discomfort. Taken together then, we had no clear initial prediction about what would happen to the participants' levels of discomfort as a result of the defusion technique.

Traditionally, within a therapeutic context, cognitive defusion exercises have been employed as coping strategies for *negative* thoughts and feelings. Thus, we made no clear predictions about the impact of defusion on the positive self-statements. These statements were simply included here as

an experimental control to determine if the emotional impact of defusion depended on psychologically engaging with negative thoughts and feelings, or if the defusion effect simply reflected the process of distancing oneself from one's thoughts in general. Overall, the inclusion of the positive self-statements would allow a fuller understanding of the manner in which defusion works.

In order to control for demand compliance effects (Fernandez & Turk, 1994; Kanter, Kohlenberg, & Loftus, 2004), participants in the current study were also divided according to the types of explicit instructions they received about the utility of the defusion technique employed here. Each instruction stated that previous research had shown that the defusion prefix (1) increased (Pro-Defusion Condition); (2) decreased (Anti-Defusion Condition); or (3) had no effect upon (Neutral Condition) the emotional impact of self-statements. We hypothesized that although the defusion-relevant instructions might impact differentially upon participants' ratings of discomfort, believability, and willingness, the defusion prefix would still outperform the Normal and Abnormal statements.

An important caveat to the current research involves emphasizing that the primary aim of the work was to simply study the impact of a common ACT-based defusion technique, particularly on individuals' emotional reactions to negative self-statements. The research, therefore, was not designed to determine what defusion is or exactly how it works. With such limited existing evidence of the *utility* of defusion per se, an important starting point for research in this area seemed to be to simply identify defusion techniques, isolate them as much as possible, and determine whether or not they produce positive and predictable outcomes. Hence, the current work may be described as analog research in which some aspect of clinical practice is isolated and presented to a nonclinical sample in a controlled experimental environment. As a result, we anticipated that the current findings would contribute more to our faith in defusion rather than our understanding of it.

Method

Participants

A total of 60 participants (31 female, 29 male) with an age range of 18–57 years volunteered for the study (mode = 21). All participants were undergraduate students and were recruited through Faculty announcements in the Department of Psychology. None had prior exposure to similar experimental procedures. Participants were divided into three groups, each composed of 20 students that differed only in the nature of the defusion-relevant instructions they received (Pro-Defusion, Anti-Defusion, and Neutral).

Materials

The 10 negative statements used in the study were rated the most negative from a larger pool, rated from 0 (extremely negative) to 20 (extremely positive); none of the 32 independent raters was thereafter recruited as an experimental participant. The mean ratings for the 10 negative statements ranged from 1.45 for "I am a failure" to 3.23 for "No one will ever love me." Table 1 presents all 10 negative self-statements and their mean ratings.

Table 1
Mean Ratings of the 10 Negative Self-Statements Employed in the Current Study

I am a failure.	1.45
My life is pointless.	2.39
I am stupid.	2.54
I am a bad person.	2.66
I make a mess of everything.	2.69
I am broken.	2.71
I am ugly.	2.85
I am helpless.	3.07
Sometimes I wish I wasn't me.	3.21
No-one will ever love me	3.23

The 10 positive statements used here were selected in a similar manner by the same 32 independent raters. The mean ratings for the 10 positive statements ranged from 18.66 ("I love life") to 16.37 ("I am whole"). Table 2 presents all 10 self-statements and their mean ratings.

Table 2
Mean ratings of the 10 Positive Self-Statements Employed in the Current Study

I love life.	18.65
I know that I am loved.	17.79
I am happy with who I am.	17.77
There is so much that I can do with my life.	17.48
There is so much for me to be happy about.	17.41
I am part of a beautiful world.	17.35
When things go wrong I know that I will always have friends.	16.92
I have no problems that can't be solved.	16.73
I am proud of myself.	16.68
I am whole.	16.37

Measures

Each participant first completed three self-report questionnaires. These were composed of the Acceptance and Action Questionnaire (AAQ-37, see Hayes et al., 2004); the Beck Depression Inventory (BDI-II, see Beck, Steer, & Brown, 1996); and the State Trait Anxiety Inventory (STAI Form Y-2, see Spielberger, Gorsuch, Luschene, Vagg, & Jacobs, 1983). The assignment of participants to experimental conditions was randomized and did not depend upon their scores on the questionnaires. The three measures were simply included as a means of determining any differences between the three experimental groups (Pro-Defusion, Anti-Defusion, and Neutral) that might influence their performance during the experiment. For example, participants who score high in acceptance

(i.e., low in avoidance) on the AAQ might respond more readily to defusion than those who score low in acceptance (high in avoidance). Furthermore, low AAQ responders may be less willing to engage with the negative self-statements, and this might considerably undermine their full participation in the study.

Experiential Avoidance

The AAQ 37-item version is a self-report measure of an individual's general level of emotional avoidance. Individuals are asked to rate the truth of each statement as it applies to themselves, on a scale of 1 (never true) to 7 (always true). Low scores on the AAQ indicate high avoidance and low acceptance, whereas high scores indicate low avoidance and high acceptance. Although the AAQ 37-item is a relatively new measure, initial research on similar versions of the AAQ indicate good psychometric properties, with good evidence of convergent, criterion-related, and construct validity (see Bond & Bunce, 2003; Hayes et al., 2004).

Depression

The BDI-II consists of 21 groups of four statements (all scored from 0 to 3). Individuals are asked to select one statement from each group that best describes how they have been feeling during the previous two weeks. The BDI-II is scored according to four categories: 0–13: Minimal Depression; 14–19: Mild; 20–28: Moderate; and 29–63: Severe. The BDI-II is a widely used measure of depression and has good, well-established psychometric properties. The scales indicate good internal consistency (Cronbach coefficient alpha = 0.93 for college students) and correlate well with several clinical assessment ratings of depression (Beck et al., 1996).

Trait Anxiety

The **STAI Form Y-2** is used to assess levels of trait anxiety in both clinical and nonclinical populations. It is composed of 20 self-statements that ask individuals to rate how they feel in general, on a scale of 1 (almost never) to 4 (almost always). The STAI Form Y is a valid measure of anxiety indicating good test-retest reliability for a sample of college students (0.73–0.86) and high internal consistency (Cronbach coefficient alpha is 0.90). The STAI Form Y is also reliable and correlates with other established anxiety measures (see Spielberger, et al., 1983).

Procedure

This study used a design of $3 \times 3 \times 2$ mixed-between-within participants with instructions (about the impact of defusion as a coping strategy) as the between-participant variable (Pro-Defusion, Anti-Defusion, and Neutral), and the visual presentation format of the self-statements (Normal, Defused, and Abnormal), and statement-type (negative and positive) as within-participant variables. Participants completed each aspect of the experiment independently, while the experimenter remained seated outside the experimental room. All participants were seated on a standard plastic office swivel chair (rather than a wooden chair). Each individual took approximately 90 minutes to complete the study.

Phase 1: Psychological Measures

Participants completed the AAQ, the BDI-II, and the STAI. The psychological measures were presented in this order and in written format in a booklet.

Phase 2: Defusion-related Instructions

Prior to Phase 1, each participant was randomly assigned to one of three groups (Pro-Defusion, Anti-Defusion, and Neutral). The defusion-related instructions were presented in written format, in a booklet the experimenter provided. Each instruction booklet differed as to the specific instructions pertaining to the emotional impact of defusion. Participants in the *Pro-Defusion Group* received instructions supporting a defusion rationale. That is, they were instructed that adding the prefix "I am having the thought that" to a negative self-statement *decreases* its emotional impact, thereby making the statement *easier* to cope with. In contrast, participants in the *Anti-Defusion Group* were given an instruction undermining a defusion strategy. The instruction stated that the prefix "I am having the thought that" *increases* the emotional impact of negative self-statements, thereby making them *harder* to cope with. The *Neutral Group* simply functioned as an experimental control. Participants in this group were instructed that the prefix would have *no effect* on one's emotional reaction to negative self-statements. Having read the defusion-related instructions, participants were required to indicate they had understood the instructions by selecting the correct answer from a multiple choice question.

Phase 3: Automated Presentation and Ratings of Self-Statements

In each of the three groups, participants were exposed to an identical automated procedure that was used to deliver a number of independent computer trials. Each computer trial presented 1 of the 10 negative self-statements or 1 of the 10 positive statements on the screen. In addition, each negative and positive self-statement was presented in three different presentation formats (Normal, Abnormal, and Defused). Thus, each participant was presented with a total of 60 single trials (i.e., 10 negative statements presented in three formats and 10 positive statements presented in three formats). In the *Normal* presentation format, each self-statement appeared without a prefix (e.g., "I am a bad person"). In the *Defused* presentation, each statement was prefixed with "I am having the thought that" (e.g., "I am having the thought that I am a bad person"). In *Abnormal* presentation, each statement was prefixed with "I have a wooden chair and" (e.g., "I have a wooden chair and I am a bad person").

Each experimental trial commenced with the instruction to read the following single self-statement carefully and to think about it. On each of the 60 independent trials, one of the negative or positive statements in one of the three formats appeared on the screen for 6 s. Immediately after the presentation of each self-statement, a new screen automatically appeared with three rating scales. Participants were asked to provide three ratings, as follows: (1) "Rate the extent to which you felt comfortable or uncomfortable reading and thinking about the previous statement" (Comfort); (2) "Rate the extent to which you found the previous statement believable" (Believability); and (3) "Rate the extent to which you were willing to read and think about the previous statement" (Willingness). All three rating scales ranged from 0 (extremely comfortable; extremely believable; and extremely willing) to 100 (extremely uncomfortable;

extremely unbelievable; and extremely unwilling). Participants rated their reactions to the statements by using the computer's mouse to drag a pointer along each of the three scales. The trial ended when the participant clicked on an end-of-trial button; this cleared the screen and after a 1 s. inter-trial interval, the next trial commenced immediately (i.e., the next self-statement appeared on the screen). Each of the 10 negative and 10 positive self-statements, in their three formats, was presented once in random order for each participant. After the last trial (60), a message appeared on screen asking the participant to report to the experimenter.

Phase 4: Postexperimental Measures

After the 60 trials, participants completed two rating scales that asked (1) "To what extent did you read the statements and ratings and answer them honestly?" (honesty) and (2) "To what extent, do you think your ratings have been affected by the instructions you received at the beginning of the study?" (control of defusion-related instructions). Participants provided each rating on a 7-point Likert scale, ranging from 1 (Not at All) to 7 (Entirely). Participants were then thanked and debriefed.

Results

Psychological Measures

Table 3 presents the mean (and standard deviation) ratings for each of the psychological measures (AAQ, BDI-II, and STAI) for each group of participants. In each case, the measures were broadly similar across groups. Indeed, three separate one-way analyses of variance (ANOVAs) for each measure revealed no significant differences (AAQ 37-item [$p = 0.67$]; BDI [$p = 0.25$]; STAI (Y2) [$p = 0.84$]).

Table 3
Means (and Standard Deviations) for Psychological Measures for Each Instruction Group

	AAQ-37	BDI-II	STAI-Y2
Pro-Defusion	160.95 (17.37)	7.25 (4.87)	36.45 (8.42)
Anti-Defusion	158.25 (12.54)	6.10 (5.77)	37.25 (11.15)
Neutral	163.25 (21.81)	8.90 (5.26)	38.40 (11.31)

Adherence Measures

High ratings were obtained on the honesty measure across all three groups (Pro-Defusion: M = 6.20, SD = 0.61; Anti-Defusion: M = 5.95, SD = 0.99; and Neutral: M = 5.90, SD = 0.85). Furthermore, a one-way ANOVA revealed no significant differences between groups, $F(2, 57) = 0.74$, $p = 0.48$. These findings indicated that participants on the whole had responded honestly to the self-statements.

In contrast, the mean ratings on adherence to the defusion-related instructions appeared to be relatively low for each of the three groups (Pro-Defusion:

M = 3.40, SD = 1.90; Anti-Defusion: M = 2.95, SD = 1.98; and Neutral: M = 3.90, SD = 2.07). The results from a one-way ANOVA revealed no significant difference between groups in this regard, $F(2, 57) = 1.14$, $p = 0.33$. Although these findings suggested that the defusion-related instructions did not greatly influence the participants, perhaps a more important result was of no between-group differences.

Self-Report Measures

For the purposes of analysis, the 30 negative and 30 positive self-statements were divided according to the three presentation formats: Normal, Abnormal, and Defused. The three ratings (comfort, believability, and willingness) for each of the three statement-types, summed across the 10 trials for each participant, ranged from 0 to 1000. The data obtained for the negative and positive self-statements were analyzed separately and are presented in separate sections below.

Negative Self-Statements

The mean overall scores were calculated for the negative self-statements for each of the three ratings, across participants, for each of the three instruction groups and divided by the three statement-types. In the interests of clarity, the three sets of means are referred to as negative comfort, negative willingness, and negative believability. Table 4 presents the means and standard deviations for each of these ratings, calculated across participants, for each of the three instruction groups divided by the three statement-types.

Table 4
Means (and Standard Deviations) for Discomfort, Willingness, and Believability Ratings of the Three Statement Types for the Three Conditions for Negative Self-Statements

Statement type	Pro-Defusion group	Anti-Defusion group	Neutral group
		Comfort ratings	
Normal	326.80 (293.18)	465.65 (322.41)	390.95 (268.59)
Abnormal	284.70 (261.68)	421.50 (275.46)	412.75 (293.69)
Defused	219.90 (220.62)	393.70 (297.62)	323.05 (232.14)
		Willingness ratings	
Normal	347.85 (336.91)	453.40 (314.58)	481.80 (266.26)
Abnormal	346.95 (325.22)	433.65 (255.67)	505.75 (286.82)
Defused	245.30 (259.04)	369.65 (281.98)	400.25 (262.42)
		Believability ratings	
Normal	849.250 (99.66)	709.50 (264.92)	837.10 (132.21)
Abnormal	835.05 (145.51)	706.35 (259.01)	817.25 (149.89)
Defused	630.80 (199.70)	625.90 (270.13)	726.00 (127.79)

Note. Lower scores indicate greater comfort, greater willingness, and greater believability.

The descriptive statistics indicated that within each group, the Defused presentation format, relative to Normal and Abnormal, produced lower levels of discomfort, higher levels of willingness, but—contrary to our prediction—higher

levels of believability. Across-group comparisons indicated lowest levels of discomfort for Pro-Defusion and highest for Anti-Defusion. For willingness, Pro-Defusion produced greatest willingness and Neutral lowest. For believability, the Anti-Defusion group produced the highest levels; the differences between Pro-Defusion and Neutral were unsystematic. The statistical analyses conducted for each of the three types of ratings are discussed separately below.

Comfort ratings. A 3×3 mixed repeated measures ANOVA was conducted on the comfort ratings, with group (Pro-Defusion, Anti-Defusion, and Neutral) as the between-participant variable, and presentation format (Normal, Abnormal, and Defused) as the within-participant variable. A significant main effect was identified for presentation format, $F(2, 57) = 19.05$, $p < 0.0001$, $\eta_p^2 = 0.25$, but not for group ($p = 0.21$); and the interaction was also nonsignificant ($p = 0.17$). Post hoc (Scheffe) tests revealed a significant difference when comparing Defusion with Normal ($p < 0.0001$) and with Abnormal ($p = 0.0001$) presentation formats but not between Normal and Abnormal. In short, defusion significantly decreased discomfort relative to the other two presentation formats, and this effect was not significantly modulated by the defusion-related instructions (see Table 4).

Willingness ratings. Another 3×3 ANOVA was conducted on the willingness ratings. Similar to the comfort measure, a significant main effect for the presentation format was obtained, $F(2, 57) = 20.91$, $p < 0.0001$, $\eta_p^2 = 0.26$; but once again, no other significant effects were found for group ($p = 0.23$) or interaction ($p = 0.74$). Scheffe post hoc tests revealed the same pattern as for comfort (Defusion vs. Normal, $p < 0.0001$; Defusion vs. Abnormal, $p < 0.0001$; Normal vs. Abnormal, not significant). In effect, defusion significantly increased willingness relative to the other two presentation formats, with no modulation by the defusion-related instructions (see Table 4).

Believability ratings. The results of a third 3×3 ANOVA on the believability data also revealed a significant main effect for presentation format, $F(2, 57) = 44.30$, $p < 0.0001$, $\eta_p^2 = 0.43$, and no effect for group ($p = 0.12$). However, a significant interaction was obtained between format and group, $F(4, 114) = 4.18$, $p = 0.003$, $\eta_p^2 = 0.12$. In order to examine this effect, three separate one-way repeated measures ANOVAs were conducted, one for each group. The ANOVA for the Pro-Defusion group was significant, $F(2, 19) = 24.19$, $p < 0.0001$, $\eta_p^2 = 0.56$, and Sheffe post hoc tests indicated that believability was greater for Defusion relative to both Normal and Abnormal presentation formats ($p < 0.0001$, $p < 0.0001$, respectively), with no significant difference between the latter ($p = 0.92$). The Anti-Defusion ANOVA was also significant, $F(2, 19) = 6.33$, $p = 0.004$, $\eta_p^2 = 0.25$; and post hoc tests revealed greater believability for Defusion relative to Normal and Abnormal ($p = 0.0125$, $p = 0.0168$, respectively), with no significant difference between the latter ($p = 0.99$). The final ANOVA for the Neutral group was also significant, $F(2, 19) = 16.98$, $p < 0.0001$, $\eta_p^2 = 0.47$, with the post hoc tests again revealing greater believability for Defusion relative to Normal and Abnormal ($p < 0.0001$, $p = 0.0003$, respectively), but no significant difference between the latter ($p = 0.62$). In summary, and contrary to predictions, Defusion significantly increased believability relative to the Normal and Abnormal presentation formats for each of the three groups (see Table 4).

Positive Self-Statements

The mean overall scores were calculated for the positive self-statements for each of the three ratings, across participants, for each of the three instruction

groups and divided by the three statement-types. In the interests of clarity, the three sets of means are referred to as positive comfort, positive willingness, and positive believability. Table 5 presents the means and standard deviations for each of the ratings, calculated across participants, for each of the three instruction groups divided by the three statement-types.

Table 5
Means (and Standard Deviations) for Discomfort, Willingness, and Believability Ratings of the Three Statement Types for the Three Conditions for Positive Self-Statements

Statement Type	Pro-Defusion group	Anti-Defusion group	Neutral group
		Comfort ratings	
Normal	112.55 (101.09)	141.40 (126.61)	158.25 (153.05)
Abnormal	122.15 (117.04)	189.90 (155.74)	220.55 (173.78)
Defused	119.40 (137.72)	195.50 (147.08)	166.95 (142.43)
		Willingness ratings	
Normal	125.05 (118.75)	172.25 (144.99)	216.95 (175.51)
Abnormal	162.15 (178.62)	227.95 (184.08)	285.10 (198.43)
Defused	135.15 (157.85)	190.85 (154.00)	219.25 (158.75)
		Believability ratings	
Normal	196.80 (140.65)	196.30 (143.53)	227.75 (156.34)
Abnormal	249.35 (191.89)	301.75 (190.77)	278.60 (125.09)
Defused	218.95 (204.13)	246.50 (149.09)	258.40 (156.77)

Note. Lower scores indicate greater comfort, greater willingness, and greater believability.

The descriptive statistics indicated that within each group, the Defused presentation format produced medium levels of willingness and believability relative to the Normal and Abnormal formats, with the Normal format producing the highest level for each measure. The Normal format also produced the highest level of comfort; however, the differences between Abnormal and Defused failed to reveal the same pattern as the other two measures (i.e., willingness and believability). Across group comparisons indicated highest levels of comfort for Pro-Defusion, whereas the differences between Neutral and Anti-Defusion groups were unsystematic. The Pro-Defusion group also produced the highest levels of willingness, with the Neutral group producing the lowest levels of willingness. For believability, across group differences were overall unsystematic, although there is some indication that the Pro-Defusion instruction overall produced the highest levels of believability. The differences between believability ratings for the remaining instruction groups appeared random.

Comfort ratings. A 3×3 mixed repeated measures ANOVA was conducted on the comfort ratings, with defusion-related instruction group as the between-participant variable and statement presentation format as the within-participant variable. A significant main effect was identified for statement presentation format, $F(2, 57) = 6.64$, $p = 0.002$, $\eta_p^2 = 0.10$, but not for instruction group ($p = 0.25$). Furthermore, there was no significant interaction effect ($p = 0.07$). Post hoc (Scheffe) tests revealed a significant difference when comparing Normal and Abnormal statements ($p = 0.002$), with Normal statements

rated as more comfortable than Abnormal statements. However, no significant difference was obtained when comparing Defusion with Normal ($p = 0.12$) and with Abnormal ($p = 0.31$) presentation formats. Taken together, these results suggest that the Defused presentation format had little impact on the comfort ratings relative to the other two presentation formats. These findings were not significantly modulated by the defusion-related instructions (see Table 5).

Willingness ratings. Another 3×3 mixed repeated measures ANOVA was conducted on the willingness ratings. This analysis revealed a significant main effect for statement presentation format, $F(2, 57) = 9.54$, $p = 0.0001$, $\eta_p^2 = 0.14$. However, the main effect for defusion-related instruction group was nonsignificant ($p = 0.13$), as was the interaction effect ($p = 0.76$). Post hoc (Scheffe) tests revealed a significant difference between Normal and Abnormal ($p = 0.0004$) and between Defusion and Abnormal ($p = 0.005$) presentation format, with participants being more willing to experience statements in the Normal and Defused presentation format and least willing to experience statements presented in the Abnormal format. The difference between Normal and Defusion statements was nonsignificant ($p = 0.73$). Thus, the Defused format and the Normal format appeared to impact on participants' willingness to experience the positive self-statements in a similar manner. Once again, these findings were not modulated by the defusion-related instructions (see Table 5).

Believability ratings. A 3×3 mixed repeated measures ANOVA was also conducted on the positive believability ratings. Similar to both the positive comfort and positive willingness data, the analysis of the believability ratings revealed a significant main effect for statement presentation format, $F(2, 57) = 9.79$, $p = 0.0001$, $\eta_p^2 = 0.15$, but the effects for instruction group and the interaction effects were both nonsignificant ($p = 0.76$, $p = 0.62$, respectively). Similar to the comfort ratings, post hoc (Scheffe) tests revealed a significant difference between Normal and Abnormal statements only ($p = 0.0001$), with Normal statements rated as more believable than Abnormal statements. No significant differences were obtained when comparing Defusion with Normal ($p = 0.10$) and with Abnormal ($p = 0.09$) presentation formats.

In short, the Defused presentation format had little impact on the believability ratings relative to the other two presentation formats, and these findings could not be attributed to the defusion-related instructions (see Table 5).

In summary, a different pattern of responding was observed for the negative relative to the positive self-statements. With respect to the negative self-statements, the Defused presentation format produced significantly higher levels of comfort (i.e., reduced discomfort), willingness, and believability relative to the Normal and Abnormal presentation formats. In contrast, the Defused presentation format had relatively little impact on the emotional ratings of the positive self-statements and overall did not significantly differ from the Normal or Abnormal presentation formats.

Discussion

The current study primarily investigated the impact of cognitive defusion on participants' self-reported levels of discomfort, willingness, and believability in the context of negative self-statements that appeared randomly on a computer screen. For all participants, the impact of the negative statements presented in the Defused form (relative to Normal and Abnormal presentations) was consistent with experimental hypotheses in that they increased

participants' willingness to read and think about them. Although we had not made firm predictions about the potential impact of the defused presentations on participant discomfort associated with the negative statements, the data indicated that discomfort decreased in that context but not when the same statements were presented as Normal or Abnormal. In contrast, we had made firm predictions that believability of the negative statements would also be reduced by defusion, but the findings were inconsistent with these predictions and indicated that believability of the statements actually increased as a result of defusion. The results for the discomfort ratings here are consistent with those reported by Masuda et al. (2004), but the believability data are not (Masuda et al. reported *decreases* in believability). The current research was the first empirical analysis of the impact of defusion on experiential willingness.

In attempting to explain the discrepancy between the current and previous findings on believability of the negative self-statements, it might be argued that the target process of defusion did not actually occur here. However, this seems unlikely, because increases in willingness are consistent with a defusion effect. In other words, defusion should increase willingness to experience negative thoughts and feelings: if the content is perceived to be meaningless (i.e., if you are fully defused), then there should be no reason to avoid it.

Consequently, the believability measure probably did not operate here in the way we had anticipated. In retrospect, the believability findings obtained in the current study seem highly likely. Consider first that participants were required to respond to the discomfort and willingness scales by responding to the *whole* defusion statement (e.g., how uncomfortable does the statement "I am having the thought that I am a bad person" make you feel?). It seems likely, therefore, that participants responded to the believability of the defusion statements in the same way (e.g., how believable is it that you are having the thought that you are a bad person?). In effect, the increased believability ratings for the defused statements indicated that the participants believed they were indeed having that thought, rather than indicating that the thought was true. In this sense, therefore, the increased believability ratings could be seen as evidence *for* a defusion effect as specified by ACT, especially when taken in the context of decreased discomfort and increased willingness for the same set of statements. Although we can only speculate about whether participants here were responding to the self-statement per se or to the fully defused statement, the latter is a possibility and in any case raises an interesting empirical issue for future research.

The differences in believability across the two studies may also be accounted for by the fact that different defusion exercises may operate in different ways. Specifically, Masuda et al. (2004) employed a semantic satiation exercise (i.e., word repetition) as opposed to the more cognitive distancing exercise employed here. Although both types of exercise altered believability in different ways, believability did change, and thus one can reasonably argue that changes in believability are an important mechanism for the impact of defusion. However, whether on any instance believability should increase or decrease perhaps depends on how believability is explicitly assessed. The current research, therefore, highlights the need for future research on how best to assess the truth or meaning of an individual's psychological content.

The current study manipulated instructions on the putative effects of

defusion in order to compare the relative utility of defusion instructions against more active experiential defusion through the visual presentation of the statements. Across all three defusion-related instruction groups, the Defusion presentation format continued to produce positive effects (relative to the other two formats) in terms of comfort, believability, and willingness, whereas the instructions overall on defusion appeared to have little or no impact. The primary aim of this minor manipulation was to begin to compare the impact of defusion-based instructions versus exercises, because in ACT, defusion is rarely instructed and more often is delivered via experiential work. Indeed, the findings recorded here provided some very preliminary evidence that the least visual experiential contact with the defused statements generated greater defusion than instructions.

One issue that may be raised about the current manipulation of defusion instructions is that all three types of instructions actually contained the defusion prefix ("I'm having the thought that"). And it might be argued that this contaminated the potential impact of the prefix when presented with the statements in the Defused presentation format. Although this possibility remains, the data indicated that the instructions were of little impact, and so the observed defusion effects were not likely to depend upon them. Indeed, what is remarkable is that the observed defusion effects on the ratings occurred for participants who had been instructed that defusion has either no impact, or that it is counter-productive in the context of negative self-statements. Furthermore, our experimental target here was to compare defusion instructions versus experiential contact. Nonetheless, one avenue for future research might involve an alternative instruction manipulation that targeted all three types of presentation format and would provide greater clarity on the potential interactions between instructions and presentation.

In effect, the results suggest the superiority of experientially engaging with the statements over simple Pro-Defusion instructions. Indeed, the data here do not indicate that the provision of Pro-Defusion instructions actually enhanced the effect of the Defused presentation format, relative to the other two types of instruction. Although clinical concerns over the relative utility of instructions versus experiential exercises are commonplace (e.g., Eifert & Forsyth, 2005, p. 124), the current research contributes to this debate with empirical support for experiential contact over instructions, at least in the context of defusion. However, it is important to note that the defusion instructions provided here were not presented as a full therapeutic intervention; nor were they designed to be, and, as such, an entirely different preparation may be necessary to fully explore this issue.

The low adherence ratings for the defusion-related instructions are entirely consistent with their lack of impact on the ratings. If participants in all three groups had not attended well to the information on defusion, they were probably not influenced by this subsequently. However, it was not the case that participants did not understand these instructions, because on the instruction booklet, they had been asked to tick a box to indicate that they had fully understood and all participants had done so. It also remains possible that participants interpreted the adherence question (presented at the end of the experiment) as a reference to their adherence to the main instructions on the presentations of the statements and the ratings. However, this seems unlikely because the strong changes in ratings suggested no lack of

adherence during this part of the experiment, whereas the limited impact of the defusion instructions does imply lack of adherence at that earlier point. Although instructions encouraging participants to attend fully to them are difficult to provide, while at the same time avoiding them doing any experiential exercises, the current study, at the very least, highlights the importance of including adherence measures.

The current research suggests that defusion is less psychologically active when used in conjunction with the positive self-statements relative to the negative self-statements. Specifically, although statements presented in the Defused format were significantly different from the Normal and Abnormal forms for the *negative* self-statements, the same effect was not recorded for the *positive* statements. Indeed, positive statements presented in the Defused format did not significantly differ from the other presentation formats, suggesting that the defusion effect did not occur here. The absence of any change in levels of believability about the Defused presentation format further suggests the absence of a defusion effect with the positive self-statements if we assume that changes in levels of believability are a useful measure of defusion-related change (e.g., Masuda et al., 2004). These differential outcomes are consistent with our original hypotheses, and indeed with ACT, in suggesting the greater impact of defusion on psychological content that is perhaps more susceptible to emotional avoidance.

However, it could be argued that participants simply failed to experientially engage with the positive self-statements, but did so with the more challenging negative statements. This may be particularly so because all the defusion-based instructions contained a negative, rather than a positive, self-statement. Put simply, the instructions manipulation may have oriented participants toward the negative statements (similar to the previous possibility that the instructions also oriented participants toward the defusion prefix). Yet again this seems unlikely because of the significant differences between the positive Normal and Abnormal presentation formats and because of the general lack of impact of the instructions. Indeed, the inclusion of a positive statement within the instructions would not likely have generated different results.

In summary, the research presented here provides empirical evidence of the utility of cognitive defusion in reducing the negative emotional impact of psychologically painful content and encouraging willingness to experience it. These findings are consistent with the clinical assumptions of ACT (Hayes, Strosahl, et al., 1999), and they provide indirect evidence that the positive outcomes reported for ACT result from its strong emphasis on defusion (e.g., Gifford et al., 2004; Gutierrez, Luciano, & Fink, 2004; Twohig, Hayes, & Masuda, 2006). At the very least, the evidence here supports the view that defusion functions by changing an individual's levels of discomfort, believability, and willingness associated with painful content, rather than by altering the content directly. The clinical implications of the work suggest that defusion should be delivered through experiential exercises rather than solely by instructions, and that the success of defusion techniques may be measured by assessing concurrent changes in discomfort, believability, and willingness. Although the process of defusion itself remains obscure, the current research offers further insight into how it can be measured and how it appears to operate and, as such, makes a useful contribution to further research on the topic and more insightful clinical practice.

References

BACH, P., & HAYES, S. C. (2002). The use of acceptance and commitment therapy to prevent the rehospitalization of psychotic patients: a randomized control trial. *Journal of Consulting and Clinical Psychology, 70,* 1129–1139.

BECK, A. T., STEER, R., & BROWN, G. (1996). *Beck Depression Inventory: Second Edition.* Psychological Corporation.

BOND, F. W., & BUNCE, D. (2003). The role of acceptance and job control in mental health, job satisfaction and work performance. *Journal of Applied Psychology, 88,* 1057–1067.

EIFER, G. H., & FORSYTH, J. P. (2005). Acceptance and commitment therapy for anxiety disorders: a practitioner's guide to using mindfulness, acceptance, and values-based behavior change strategies. U.S.A.: New Harbinger Publications.

FERNANDEZ, E., & TURK, D. C. (1994). Demand characteristics underlying ratings of sensory versus affective components of pain. *Journal of Behavioural Medicine, 17,* 375–390.

GIFFORD, E. V., KOHLENBERG, B. S., HAYES, S. C., ANTONUCCIO, D. O., PIASECKI, M. M., RASMUSSEN-HALL, M. L. ET AL. (2004). Acceptance-based treatment for smoking cessation. *Behavior Therapy, 35,* 689–705.

GUTIERREZ, O., LUCIANO, C., RODRIGUEZ, M., & FINK, B. C. (2004). Comparison between an acceptance-based and a cognitive-control-based protocol for coping with pain. *Behavior Therapy, 35,* 767–784.

HAYES, S. C., BARNES-HOLMES, D., & ROCHE, B. (2001). *Relational frame theory: A post-Skinnerian account of human language and cognition.* New York: Kluwer Academic/Plenum Publishers.

HAYES, S. C., BISSET, R. T., KORN, Z., ZETTLE, R. D., ROSENFARD, I. S., COOPER, L. D. ET AL. (1999). The impact of acceptance versus control rationales on pain tolerance. *The Psychological Record, 49,* 33– 47.

HAYES, S. C., LUOMA, J. B., BOND, F. W., MASUDA, A., & LILLIS, J. (2006). Acceptance and commitment therapy: model, processes and outcomes. *Behaviour Research and Therapy, 44,* 1–25.

HAYES, S. C., STROSAHL, K. D., & WILSON, K. G. (1999). *Acceptance and commitment therapy: an experiential approach to behavior change.* New York: Guilford Press.

HAYES, S. C., STROSAHL, K., WILSON, K. G., BISSET, R. T., PISTORELLO, J., TOARMINO, D. ET AL. (2004). Experiential avoidance: a preliminary test of a working model. *The Psychological Record, 54,* 553–578.

KANTER, J. W., KOHLENBERG, R. J., & LOFTUS, E. F. (2004). Experimental and psychotherapeutic demand characteristics and the cognitive therapy rationale: an analogue study. *Cognitive Therapy and Research, 28,* 229–239.

KROSS, E., AYDUK, O., & MISCHEL, W. (2005). When asking "why" does not hurt: distinguishing rumination from reflective processing of negative emotions. *Psychological Science, 16,* 709–715.

LEVITT, J. T., BROWN, T. A., ORSILLO, S. M., & BARLOW, D. H. (2004). The effects of acceptance versus suppression of emotion on subjective and psychophysiological response to carbon dioxide challenge in patients with panic disorder. *Behavior Therapy, 35,* 747–766.

MASUDA, A., HAYES, S. C., SACKETT, C. F., & TWOHIG, M. P. (2004). Cognitive defusion and self-relevant negative thoughts: examining the impact of a ninety year old technique. *Behavior Research and Therapy 42,* 477– 485.

SPIELBERGER, C. D., GORSUCH, R. L., LUSCHENE, R., VAGG, P. R., & JACOBS, G. A. (1983). *State-Trait Anxiety Inventory.* Mind Garden Inc.

TAKAHASHI, M., MUTO, T., TADA, M., & SUGIYAMA, M. (2002). Acceptance rationale and increasing pain-tolerance: acceptance-based and FEAR-based practice. *Japanese Journal of Behavior Therapy, 28,* 35– 46.

TEASDALE, J. D., SEGAL, Z. V., WILLIAMS, J. M. G., RIDGEWAY, V. A., SOULSBY, J. M., LAU, M. A. (2000). Prevention of relapse/recurrence in major depression by mindfulness-based cognitive therapy. *Journal of Consulting and Clinical Psychology, 68,* 615– 623.

TITCHENER, E. B. (1916). *A text-book of psychology.* New York: MacMillan.

TWOHIG, M.P., HAYES, S. C., & MASUDA, A. (2006). Increasing willingness to experience obsessions: acceptance and commitment therapy as a treatment for obsessive-compulsive disorder. *Behavior Therapy, 37,* 3–13.