The impact of a cognitive defusion intervention on behavioral and psychological flexibility: An experimental evaluation in a spider fearful non-clinical sample ☆

Timothy R. Ritzert [a,*], John P. Forsyth [a], Christopher R. Berghoff [a], Dermot Barnes-Holmes [b], Emma Nicholson [b]

[a] University at Albany, State University of New York, 1400 Washington Ave, Albany, NY 12222, USA
[b] National University of Ireland Maynooth, Maynooth Co. Kildare, Ireland

A B S T R A C T

Acceptance and Commitment Therapy (ACT) utilizes cognitive defusion strategies to alter the function of unwanted, distressing thoughts so as to reduce their impact and foster greater psychological flexibility. Though defusion is linked to several positive outcomes (e.g., reduced thought believability, improved pain tolerance), it remains unclear whether such effects are associated with the behavioral process of defusion itself; namely, behavior that is relatively broad and flexible compared with behavior that is narrow and inflexible. To address this issue, the present study assessed the impact of a defusion intervention on behavior, as measured by the Implicit Relational Assessment Procedure (IRAP). Undergraduates high in spider fear ($N=65$) were randomly assigned to one of two interventions targeting phobic thoughts of spiders: a defusion intervention or a thought distraction intervention. A control condition (reading an article) was used to evaluate the relative impact of both interventions. Participants completed the IRAP and rating scales of believability and distress of target thoughts pre- and post-intervention. Consistent with expectation, defusion produced significantly greater pre- to post-intervention reductions in IRAP effects and thought believability relative to comparison conditions. The practical and scientific implications of the present findings, suggesting that defused behavior is broad and flexible relative to narrow and inflexible, are discussed.

© 2015 Association for Contextual Behavioral Science. Published by Elsevier Inc. All rights reserved.

1. Introduction

Unwanted distressing thoughts have a prominent role in cognitive behavior therapy (CBT) models describing the etiology, maintenance, and treatment of several forms of human suffering. Indeed, modification of negative thoughts is a central focus of many traditional CBTs (Beck, Rush, Shaw, & Emery, 1979; Clark, 1995; Hofmann & Asmundson, 2008; Hofmann, Asmundson, & Beck, 2013). Conversely, a group of newer, acceptance- and mindfulness-based CBTs emphasize the relationship people have with their thoughts and not necessarily the content of thoughts themselves. For example, acceptance and commitment therapy (ACT; Hayes, Strosahl, & Wilson, 1999, 2012) focuses on the context in which thoughts are experienced, and makes no effort to alter the form or frequency of

unwanted thoughts directly. Rather, a variety of behavioral interventions, such as cognitive defusion strategies, encourage mindful awareness and acceptance of unwanted private events. Such a context allows for engagement in value-guided behavior independent of what thoughts and emotions are experienced.

ACT is grounded in a modern, functional analytic theory of language and cognition called relational frame theory (RFT; Hayes, Barnes-Holmes, & Roche, 2001), which considers language and cognition to be composed of contextually situated patterns of behavioral responses. Specifically, language and thinking involve responding to stimuli in relation to one another, often in an arbitrary and/or derived manner based on contextual cues (i.e., arbitrarily applicable relational responding; Hayes et al., 2001). Importantly, such responding can result in the transformation of stimulus functions. That is, when stimuli are responded to in relation to each other, their functions may be altered based on the nature of the relation (Blackledge, 2003). Thus, suppose an individual is afraid of snakes, and thereafter learns that snakes can be found in the woods (i.e., woods are coordinated with snakes). Given the appropriate contextual cues, this individual may subsequently react with fear to the suggestion "Let's go for a walk in the woods," despite having no

---

☆ This article is based on the Master's thesis of the first author, under the supervision of the second author.
* Correspondence to: University at Albany, State University of New York, Department of Psychology, Social Sciences 399, 1400 Washington Ave, Albany, NY 12222, USA. Tel.: +1 518 442 4820; fax: +1 518 442 4867.
E-mail address: tritzert@albany.edu (T.R. Ritzert).

aversive experiences with snakes or walks in the woods. In this way, stimuli can acquire functions based on verbal processes (e.g., arbitrarily applicable relational responding). Patterns of past relational responding, and the verbal functions acquired by this history, build relational histories that may impact present behavior, given the right context.

ACT is seen as reticulating closely with RFT (see Hayes, Barnes-Holmes, & Wilson, 2012). According to ACT, suffering can occur when verbal processes organize behavior in unhelpful ways. For example, cognitive fusion describes a process by which the specific verbal functions of some stimuli may come to excessively control or inappropriately narrow behavior, leading to inflexible response repertoires (Hayes et al., 1999; Hayes & Wilson, 1994). When behavior is fused, people "buy into" their thoughts, respond to them in a literal manner, and are less sensitive to other potential sources of helpful behavioral regulation (Forsyth, Eifert, & Barrios, 2006; Gillanders et al. 2014). Fusion may be maintained by a context that supports responding to verbal stimuli as if they are the same as actual stimuli (Forsyth et al., 2006; Hayes et al., 2012). Extending the above example, fused behavior would occur when the snake-fearful individual reacts with fear to the word "woods," based on a coordinated relational response involving the words "woods" and "snakes" and actual snakes.

*De*fusion interventions, such as saying thoughts in a funny voice or labeling thoughts as thoughts, are designed to undermine the behavior regulatory functions of thoughts by altering the context in which thoughts occur (Masuda, Hayes, Sackett, & Twohig, 2004). From an RFT perspective, defusion weakens the influence of specific verbal functions of stimuli on behavior, increasing sensitivity to other stimulus functions that may shape behavior in more useful ways (Blackledge, 2007; Hayes et al., 1999; Hayes & Wilson, 1993).

Defusion strategies create a non-literal context for thoughts which brings the process of thinking itself into awareness (e.g., how words sound to the ear; Hayes et al., 1999; Hayes & Wilson, 1994). This new context does not erase the verbal functions of stimuli *per se*, but rather allows more flexible behavior because a broader range of stimulus functions begins to influence responding (Eifert & Forsyth, 2005; Hayes & Wilson, 1994). In short, individuals can learn to experience their thoughts as thoughts (which are nothing more than products of their unique history) and engage in actions in the service of what matters.

Several lines of research suggest defusion is a core process in ACT-based interventions. For instance, mediation studies have demonstrated that defusion is a mechanism of change in ACT (e.g., Arch, Wolitzky-Taylor, Eifert, & Craske, 2012; Forman et al. 2012; Zettle, Rains, & Hayes, 2011). Moreover, several studies have examined the effects of defusion in isolation from other ACT treatment components. The most commonly studied defusion exercise, known as the *Milk, Milk, Milk* task (Hayes et al., 1999; originally appearing in Titchener, 1916), involves quickly repeating a negatively evaluated thought out loud for 30 s. A series of studies by Masuda and colleagues indicates the exercise reduces self-reported believability and distress regarding negatively evaluated self-relevant thoughts, relative to comparison conditions such as thought suppression and distraction (e.g., Masuda et al., 2004. 2010). Such effects, in turn, have been replicated by others using this and other defusion tasks (for a meta-analysis, see Levin, Hildebrandt, Lillis, & Hayes, 2012).

However, several key questions remain. First, most studies have relied on self-report measures of believability as a proxy variable for defusion. Of interest would be the use of more objective, behavioral measures (e.g., Hooper & McHugh, 2013; Hooper, Sandoz, Ashton, Clarke, & McHugh, 2012). Second, from an RFT point of view, defusion is thought to weaken the influence of a narrow or specific range of verbal stimuli on behavior, thus increasing sensitivity to a

broader range of verbal functions, including direct-acting contingencies. If this specific behavioral process were assessed, it could provide evidence that defusion operates in a manner corresponding with underlying theory. This would be an important undertaking given the need not only to assess the overall efficacy of therapies, but also to critically evaluate their theoretical models by examining links between theory and specific treatment components (David & Montgomery, 2011; Hayes, Levin, Plumb-Vilardaga, Villatte, & Pistorello, 2013; Lohr, 2011).

To address these issues, the present study was undertaken to provide an experimental evaluation of defusion, targeting spider phobic thoughts in a sample of spider fearful individuals. The investigation assessed the relative effects of defusion using both self-report measures of thought believability and distress, and the behavioral process of defusion as indexed by the Implicit Relational Assessment Procedure (IRAP; Barnes-Holmes et al., 2006). The IRAP, a computerized measure assessing properties of relational responding, has gained interest as a behavioral assessment of defusion (e.g., Kishita, Muto, Ohtsuki, & Barnes-Holmes, 2014).[1]

The IRAP presents trials of stimulus pairings to which participants respond by selecting one of two responses. On half of the trials, correct responses, as determined by feedback contingencies, are consistent with pre-existing relational histories (often referred to as consistent trials). On the other half of trials, correct responses are inconsistent with pre-existing relational histories (often referred to as inconsistent trials). Individuals tend to evidence shorter response latencies when responding in a manner that is consistent, rather than inconsistent, with their relational history (Barnes-Holmes et al., 2006). This phenomenon is termed the *IRAP effect* (for a more detailed description of the IRAP and an RFT account of the IRAP effect, see Barnes-Holmes, Barnes-Holmes, Stewart, & Boles, 2010).

The present study used an IRAP which was designed to assess spider fear, originally developed by Nicholson and Barnes-Holmes (2012b). On this IRAP, photographs of either spiders or pleasant scenes are paired with either evaluative (e.g., "disgusts me") or approach-based (e.g., "I would approach") phrases. Participants respond to the pairings by selecting either *true* or *false*. On consistent trials, correct responses are consistent with a history of responding to spiders with fear/disgust (e.g., it is correct to respond *true* when a spider photograph is paired with "disgusts me"). On inconsistent trials, correct responses are inconsistent with such a history (e.g., it is correct to respond *false* to the previous pairing). Results of Nicholson and Barnes-Holmes (2012b) suggested that spider fearful individuals displayed large IRAP effects relative to those with low spider fear. Moreover, this pattern of responding only occurred on trials with spider photographs, and not those with pictures of pleasant scenes.

From an RFT perspective, the IRAP effect (i.e., differential response latencies on consistent and inconsistent trials) can be seen as a measure of response flexibility versus inflexibility. For example, on the IRAP used in the present study, relatively large difference scores on the IRAP could be interpreted as indicative of a narrow and inflexible repertoire with regard to spider-related

---

[1] Although we refer to the "behavioral process of defusion," it seems important to note here that this so-called process remains ill-defined, at least in terms of functional-analytic abstractive behavioral principles (for a detailed treatment of this issue, see Barnes-Holmes, Hussey, McEnteggart, Barnes-Holmes, & Foody, in press). A relatively precise description of how the IRAP was used in the current study, therefore, might be as follows: as an index of the impact of a defusion exercise on a particular property (response latency) of the entailment functions that obtain between pictures of spiders and a class of negatively valenced words. Or, to put it another way, demonstrating the positive impact of a defusion *exercise* on an IRAP performance does not provide evidence for defusion as a well-defined behavioral process. Having noted this caveat, however, we will continue when convenient, and in the interests of linguistic brevity, to refer to the (ill-defined) process of defusion.