The Psychological Record (2022) 72:247–261
https://doi.org/10.1007/s40732-021-00478-7

ORIGINAL ARTICLE

Chatbot-Delivered Cognitive Defusion versus Cognitive Restructuring for Negative Self-Referential Thoughts: A Pilot Study

Joseph Lavelle[1] · Neil Dunne[1] · Hugh E. Mulcahy[2] · Louise McHugh[1]

Accepted: 22 June 2021 / Published online: 24 August 2021

Abstract

Conversational agents or chatbots are a novel, highly accessible, and low-resource method of psychological intervention delivery. The present research aims to compare two brief chatbot interventions that delivered cognitive restructuring and defusion interventions, respectively. It was hypothesized that a defusion chatbot would lead to reduced cognitive fusion and decreased thought believability relative to cognitive restructuring and a nonactive control. Participants (N = 223; M age of 28.01 [SD = 10.29]; 47 identified as male, 174 as female, and 2 as nonbinary) were randomized into one of three conditions (defusion, restructuring, control), engaged for 5 days completing thought and mood measures pre- and postintervention. Sixty-two participants (M age of 25.98; SD = 8.647 years) completed measures again at time 2 (49 identified as female, 12 as male, and 1 as nonbinary). No statistically significant differences were observed among groups on believability of thoughts (F[2, 25] = .79, p = .47, ηp^2 = .06), negativity of thoughts (F[2,25] = 1.49, p = .25, η^2 = .11), discomfort associated with thoughts (F[2, 25] = .48, p = .62, ηp^2 = .04), and willingness (F[2, 25] = 3.00, p = .07, ηp^2 = .19) to have negative self-referential thoughts. Moreover, substantial attrition of 72% was observed. Acceptability and usability of the chatbots employed are discussed as contributing toward the limited effectiveness of interventions and elevated attrition. Various recommendations are presented to support researchers and clinicians in developing engaging and effective chatbots.

Keywords cognitive defusion · cognitive restructuring · chatbot-delivered interventions · negative self-referential thoughts

Negative self-referential thoughts are those self-relevant thoughts that are critical, derogatory, or self-defeating in nature (Clark & Rhyno, 2005). Cognitive models of psychopathology have long asserted the involvement of negative self-relevant cognitions, negative cognitions pertaining to the future, and negative appraisals of past events (Beck et al., 1976; Clark & Rhyno, 2005). Indeed, such thoughts have been posited to be involved in the onset of myriad psychological conditions including depression, anxiety, psychosis, and body image difficulties (Clark & Rhyno, 2005).

Cognitive behavioral therapy (CBT) is one therapeutic approach often employed to treat the above difficulties and that intervenes on patterns of negative thinking (including negative self-referential thoughts). CBT draws on cognitive restructuring (hereafter restructuring) to act on thoughts. Restructuring can be defined as "structured, goal-directed, and collaborative intervention strategies that focus on the exploration, evaluation, and substitution of the maladaptive thoughts, appraisals, and beliefs that maintain psychological disturbance" (Clark, 2014). Restructuring protocols typically involve challenging the veracity of negative cognitions, identifying cognitive distortions evident in negative cognitions, and subsequently reformulating the cognition in a more objective and less negative form (Beck & Beck, 2011; Ellis, 2008). The intention or premise of restructuring is that change to cognition will lead to changes in emotion and behavior (Clark & Beck, 2011).

An alternative approach to intervening on negative self-referential thoughts comes from acceptance and commitment therapy (ACT; Hayes et al., 1999). ACT is rooted in the philosophy of functional contextualism and has its theoretical underpinnings in relational frame theory (RFT; Hayes et al., 2001); a behavioral approach to language and cognition).

✉ Joseph Lavelle
joseph.lavelle@ucdconnect.ie

[1] School of Psychology, University College Dublin, Belfield, Dublin 4, Ireland

[2] Center for Colorectal Disease, St. Vincent's University Hospital, School of Medicine, University College Dublin, Dublin, Ireland

Thus, ACT adopts an approach to cognition that conceptualizes cognitions as behavior (behavior that only be observed by the individual engaging in that cognition). Although behavior cannot be said to cause other behavior, it may have a controlling effect on behavior in certain contexts (Hayes & Brownstein, 1986). ACT draws on cognitive defusion (hereafter defusion) to separate thoughts from their literal and undesirable functions and develop awareness of cognition as an active, ongoing, process (Cullen, 2008; Hayes et al., 1999; Hayes, 2004; Luoma & Hayes, 2008; Hayes et al., 2012). Thus, via defusion, ACT facilitates individuals in becoming aware of their own private verbal behavior and diminishes the controlling context created by thoughts (Healy et al., 2008). In summary, unlike CBT, which aims to reduce negativity and frequency of cognitions, ACT draws upon defusion to alter the relationship with a negative self-referential thought.

Studies comparing the brief defusion exercises entailing rapid repetition of a word (Hayes et al., 1999; Titchener, 1916) to distraction tasks and nonactive control groups consistently observe reduced discomfort and believability of thoughts; increased self-esteem; and reduced psychological distress (Hinton & Gaynor, 2010; Masuda et al., 2004; Masuda et al., 2010). Subsequent studies have investigated other defusion interventions such as "I am having the thought that . . ." wherein this prefix is placed before a thought (e.g., "I am having the thought that I am an idiot"). Healy et al. (2008) observed that said defusion exercise resulted in reduced discomfort with negative self-statements and undermined the believability of same while increasing willingness to experience such psychological content in future. Discomfort findings were replicated in a study that combined psychoeducation on cognitive fusion with a defusion technique that involved prefacing a negative self-referential thought with an acknowledgement that that cognition is a thought (e.g., "there I go with a thought that I am unattractive"; Pilecki & McKay, 2012, p. 27). Similar outcomes were observed by Foody et al. (2015), who observed that hierarchical relating facilitated via the ACT exercise commonly known as "leaves on a stream" led to reductions in discomfort.

A number of studies have compared brief interventions informed by defusion and restructuring, respectively. One such study by Barrera et al. (2015) infused defusion and restructuring respectively into in-vivo exposure for social anxiety (resulting in three intervention groups: defusion plus in-vivo exposure, restructuring plus in-vivo exposure, and standalone in- vivo exposure). They observed that all three conditions saw reductions in discomfort with thoughts but not believability or importance of thoughts. Other studies observe that both brief defusion and restructuring interventions have been observed to produce comparable reductions in body dissatisfaction (Deacon et al., 2011). Moreover, both defusion and restructuring produced reductions in believability of and discomfort with thoughts about weight.

Larsson et al. (2016) investigated the utility of defusion and restructuring as brief interventions for negative self-referential thoughts. Indeed, they proposed that these processes might be effective interventions for painful psychological content even is isolation from the other processes that ACT and CBT employ in conjunction with defusion and restructuring, respectively. They observed that defusion produced significantly greater decreases in believability and discomfort and an increase in willingness to experience such thoughts, again compared to restructuring and control. Defusion and restructuring saw comparable reductions in negativity of difficult thoughts. The defusion intervention employed also produced increases in positive affect relative to both the restructuring group and nonactive control group.

Much of the research to date has delivered defusion and restructuring in person. However, growing evidence suggests that both CBT and ACT-informed interventions are effective as online self-help orientated interventions for difficulties including depression, anxiety, and stress (Brown et al., 2016; Grist & Cavanagh, 2013; O'Connor et al., 2018; Rathbone, 2017). Although these interventions are economic in terms of finances, resources, and clinician time, dropout from such interventions tends to be high—due to poor acceptability and engagement—which may hamper effectiveness (Andersson, 2018). A novel method of delivering online psychological interventions is automated conversational agent (hereafter chatbot). Unlike traditional online self-help programs, which mirror workbooks in an online domain, chatbot-delivered psychological interventions aim to mirror human-to-human interaction via voice and text. Indeed, chatbots aim to recreate human interactional qualities and mirror some therapeutic processes such as accountability (Fitzpatrick et al., 2017). For example, a chatbot might assign homework to a user and in a subsequent conversation inquire if this assigned homework was completed and of the user's experience of and learning from same. Although chatbots largely depend upon text communication with users and mirror text or instant message conversations, they also draw upon rich text media, such as audio recordings, picture files, or video recordings. Early evidence suggests that chatbots delivering CBT are effective interventions for depression, anxiety, and stress with attrition observed to be lower than traditional online self-help (Fitzpatrick et al., 2017; Fulmer et al., 2018; Inkster et al., 2018; Ly et al., 2017). However, chatbot has yet to undergo a trial as a mode of delivery for ACT or its processes, such as defusion.

The present study replicates some of the methodological aspects of Larsson et al. (2016) but delivers the brief defusion and restructuring interventions via chatbot. Past studies employing defusion as a brief intervention have tended to use a single defusion technique only (see Hinton & Gaynor, 2010; Masuda et al., 2004; Masuda et al., 2010; Healy et al., 2008). However, in practice, clients may be introduced to

multiple techniques. The present study will introduce participants to multiple defusion techniques with the intention of elucidating if this is facilitative or disruptive of developing awareness of private verbal behavior and undermining the controlling context of thoughts. Focus will be given to how these interventions affect outcomes consistent with ACT and CBT, namely the believability of, discomfort with, negativity of, and willingness to have negative self-referential thoughts. It is predicted that, per Healy et al. (2008, p. 638), defusion will alter the relationship with negative self-referential thoughts (i.e., willingness to experience thoughts, discomfort with thoughts, and believability of thoughts). Informed by Larsson et al. (2016), we predict that the defusion intervention will result in reduced believability of and discomfort with thoughts and greater increases in willingness to have such thoughts relative to restructuring and nonactive control, significant decreases in psychological inflexibility and cognitive fusion relative to both restructuring and nonactive control. It is expected that defusion and restructuring will result in reduced negative affect and increased positive affect relative to nonactive control, and those who engage with chatbots will show greater completion than nonactive control. It is hypothesized that those in the cognitive restructuring group will show reduced thought negativity relative to defusion and control. The present study also aims to introduce practitioners and researchers to chatbots as modalities of intervention delivery.

Materials and Method

Participants and Recruitment

A sample size of approximately 150 participants was targeted based on an a-priori sample size calculation to detect a small effect (F = .10; see Brown et al., 2016) and 80% power to observe a significant mixed (time*group) interaction effect on an ANOVA. This a-priori calculation was conducted with focus on the coprimary outcomes of thought believability, negativity, discomfort and willingness. Recruitment was via university study pools and social media. Inclusion criteria included being aged 18 or over and proficiency in English (implied). Participants were encouraged to self-exclude if they had a current diagnosis of depression or anxiety and/or if they were currently receiving any psychological or psychiatric treatment. Participants were reached via online participant recruitment platforms within the school of psychology at the host university, poster advertisement at same university, and via social media advertisement through Facebook and Twitter. Prospective participants consulted an online information sheet and consent form, which included details of support services.

This recruitment strategy saw 223 participants complete measures at the preintervention stage. Participants ranged in age from 18 to 68 with a mean age of 28.01 (SD = 10.29). Forty-seven participants identified as male, 174 as female, and 2 as nonbinary. Sixty-two participants (M age: 25.98; SD = 8.647 years) completed measures again at time two (49 identified as female, 12 as male, and 1 as nonbinary).

Design

The study employed a 3 x 2 experimental design with participants randomized to condition. The between-subjects variable was group (with three levels: defusion, restructuring, and inactive control) and the within-subjects variable was time (with two levels). Dependent variables were negativity of self-relevant thoughts, believability of thoughts, discomfort with thoughts, and willingness to have thoughts; psychological inflexibility; cognitive fusion; and positive and negative affect. Measurement of dependent variables occurred immediately before engagement with interventions (i.e., on day one of the study) and immediately after engagement with interventions (i.e., day 5 of the study).

Measures

Target Thought Measure

The Target Thought Measure (see Appendix A; Larsson et al., 2016) is a four-item instrument that asks participants to pick a negative self-referential thought (i.e., the target thought) that they are experiencing as extremely negative, extremely believable, extremely uncomfortable, and which they are extremely unwilling to have. Participants then rate—via four 10-point visual analogue scales—the negativity, believability, discomfort associated with and willingness to have said thought. A score of 1 indicates that a thought is extremely negative; extremely believable; extremely uncomfortable; and extremely unwilling [to have the thought]. A score of 10 indicates a thought is extremely positive; extremely unbelievable; extremely comfortable; and that the participant is extremely willing to have said thought. Of note, higher scores on the believability item indicate that a thought is less believable (i.e., scores closer to "10" denote a thought that is not particularly believable). Meanwhile, lower scores (i.e., scores closer to 1) on the negativity item indicate that a thought is extremely negative, whereas scores closer to 10 denote a thought as being extremely positive (see Appendix A). Participants were not required to state the same thought at the postintervention stage.

Acceptance and Action Questionnaire-II (AAQ-II)

The AAQ-II (Bond et al., 2011) is a seven-item measure of psychological inflexibility. Participants rate statements pertaining to psychological flexibility and experiential avoidance on a seven-point scale where one indicates that a

statement is "never true" of the individual and seven is "always true" of the individual. Although initial validation of the measure suggested it to be psychometrically sound with high validity and reliability (Bond et al., 2011), more recent findings have called into question the discriminant validity of the AAQ-II and its appropriateness as a measure of psychological inflexibility (Tyndall et al., 2019).

Cognitive Fusion Questionnaire-7 (CFQ-7)

CFQ-7 (Gillanders et al., 2014) is a seven-item measure of fusion. Participants rate statements on a seven-point scale ranging from "never true" to "always true." Statements are reflective of cognitive fusion (e.g., "I get upset with myself for having certain thoughts"). The measure has been shown to be psychometrically sound with high validity (Gillanders et al., 2014).

Positive and Negative Affect Schedule (PANAS)

The PANAS (Watson et al., 1988) is a 20-item measure measuring positive and negative affect, respectively. Participants are asked to rate the extent they feel the emotion in question (e.g., interested, strong, upset) at present on a scale of 1 (very slightly or not at all) to 5 (extremely). It has been demonstrated to have high construct validity (Crawford & Henry, 2004).

Adherence Quiz

At time 2, participants completed a brief adherence quiz. The multiple-choice quiz (see Appendix B) asked participants how their intervention encouraged them to respond to negative thoughts. Potential responses included those that were consistent with the defusion and restructuring interventions, respectively, and responses that were fundamentally inconsistent with the teaching of both interventions. A participant was considered to have adhered to the intervention and its material if they selected the response consistent with their assigned group. For example, a participant assigned to the defusion group would be considered to have adhered if they selected the response indicating that the chatbot encouraged them to treat their thoughts as thoughts rather than truth or reality. Any other response was treated as though the participant had not adhered to and/or understood the intervention. On the other hand, a participant assigned to the restructuring group would need to choose the response "consider evidence for and against my thought and come up with a more balanced thought."

Interventions

Chatbot Development

Chatbot scripts were written by the first author and aimed to model the conversational and informal style of existing chatbots that deliver psychological content (e.g., Woebot and Wysa; Fitzpatrick et al., 2017; Inkster et al., 2018). The defusion and restructuring content of both chatbots was chosen by the last author, a peer-reviewed ACT trainer, in view of what may be of use to chatbot-users, what had previously been suggested to be effective, and what might be engaging as a mode of delivery (i.e., picture delivery, audio recording, YouTube video). The content for the restructuring chatbot aimed to model a typical restructuring protocol (e.g., Ellis, 2008; Larsson et al., 2016). The content for the defusion chatbot aimed to introduce participants to a variety of defusion strategies including classic vocalization strategies (e.g., stating a thought in the voice of a cartoon character; Hayes et al., 1999; Larsson et al., 2016) and visualization strategies (e.g., "leaves on a stream"). Defusion strategies were selected that had been evaluated previously (or a conceptually similar strategy in the case of "the sushi train"; Harris, 2008).

Two chatbots were subsequently developed by a freelance computer scientist and tested by the first and last authors, respectively. Based on this first stage of testing, changes were made to correct typographical errors in the chatbot script, to script phrasing (to make language more conversational), and in the chatbot's coding to correct issues with the functioning of the chatbot. At the next phase of testing, the chatbots were completed by members of the Contextual Behavioral Science lab, coordinated by the last author. Feedback on user experience and suitability of psychological content was invited and implemented before trials began with participants of the present study.

Defusion and Restructuring Chatbots

Participants assigned to active intervention groups engaged with a chatbot (delivering a defusion or restructuring intervention) in the form of brief daily conversations and mood tracking. Engagements took place over 5 days and lasted approximately 10 min each. The intervention was platform agnostic and delivered via the instant messaging function in Facebook. Each engagement was commenced by the chatbot with an inquiry into current mood and with the intention of establishing rapport with the participant. Participant responses were in text or emoji format (and were not recorded). Participants were then introduced to a defusion or restructuring technique via text and rich media such as image, video, or audio recording. The first day of the intervention orientated participants to the chatbot by explaining that the buttons provided should be used to engage (unless asked to do otherwise, e.g., to provide a typed response)

and that the chatbot would be providing daily tips. To support engagement with techniques, participants were

asked to complete exercises such as defusing or restructuring the thought of a fictional character, completing a quiz, and practicing the technique using a negative self-referential thought that they experience by typing said thought into the chat. The chatbot script was unchanged throughout the study. Participants who had disengaged were manually prompted to reengage once via the message: "Do you still want to chat, if so press the last button above." or "Uh oh, sometimes I don't always work as I should. Just press the last button above to kick me back into action" (the latter was used when disengagement was due to a chatbot malfunction). The techniques and exercises for each day of both chatbots are presented in Table 1. Table 2 shows the script for day 1 of the defusion chatbot with Figure 1 displaying how this might appear on the participant's smartphone.

Nonactive Control

Participants in this group did not receive an intervention and completed measures at time 1 and 2.

Procedure

Participants who provided informed consent self-generated an ID code and were assigned to a group via a randomizer function (1:1:1 block randomization) in the online survey platform (this also ensured blinding of researchers and participants to group allocation). Upon completing all measures detailed above, participants who had been randomized into active intervention groups were required to click a hyperlink bringing them to the chatbot corresponding to their group. Participants remained blind to the intervention (defusion or restructuring) being delivered. Following five engagements (spread over at least 5 consecutive days) with the intervention or 5 chronological days in the case of the nonactive control group, participants were sent a link to posttreatment outcome measures (automatically via chatbot or SMS) and completed all of the measures detailed above for a second time.

Data Analytic Plan

Preliminary Analyses

Post data collection, participant data was collated and cleaned in IBM SPSS version 21. Descriptives for each measure detailed above were calculated for both the pre- and postintervention phases of the study. A series of independent *t*-tests were conducted examining for differences, at the preintervention phase, in completers and noncompleters on all outcomes measured. Pearson's chi square test was employed to screen for differences between completers and noncompleters based on gender and experimental group assignment.

Effectiveness of Interventions

A series of ANOVAs were conducted to examine for differences between groups at time 1 on each dependent variable with no statistically significant differences observed (see Tables 4 and 5 for descriptives for each measure). A series of 3 x 2 mixed ANOVAs were conducted to examine the experimental research question wherein the between-subjects variable was group (with three levels: defusion, restructuring, and inactive control) and the within-subjects variable was time (with two levels).

In view of the large number of outcomes being examined in the present analyses, we acknowledge the inflated risk to Type I error in the present study. Due to the relatively small sample size in the present study and informed by Perneger (1998) and similar studies (see Deacon et al., 2011), results are reported without Bonferroni correction to guard against inflated Type II error. In light of this, we encourage tentative interpretation of the statistics presented and advise readers to consider the effect sizes presented (in this case partial eta squared). However, we advise caution here also and remind the reader that effect sizes are not immune to inflation or impact due to sample size.

Exclusion from Analyses

During data collection it was noted that, on occasion, participants did not choose thoughts that (per the instructions of the target thought measure) they considered extremely negative,

Table. 1 Defusion and Restructuring Techniques Presented via Chatbots over 5-Day Intervention Period

Defusion	Restructuring	
Day 1	"I'm having the thought that . . ." (see Figure 1)	Cognitive Distortions Psychoeducation (See Figure 1)
Day 2	Say it in a cartoon character's voice	Identifying evidence against thought
Day 3	"I'm noticing that I'm having the thought that . . (video)	." Generating a balanced thought
Day 4	Sushi Train Metaphor (video)	Combining skills from days 1–3
Day 5	Leaves on a stream (audio)	Combining skills from days 1–3

Note. See appendix C and D for further detail on stimuli.

Table. 2 Day one script of the defusion-based chatbot

Chatbot Message	User Response option 1	User Response option 2
Hey Steve Look at my word buttons. Can you see them?	*Yes*	*No*
They are the main way that you will communicate with me	*Cool*	
I'm going to give you a tip each day! [to help around negative self-referential thoughts]	*Okay*	
The very first tip I have for you is this:		
Some thoughts are really horrible to have such as "I am a failure".		
One thing that can help is to put "I am having the thought that..." before the thought to help you notice that itis just a thought		
So for example I would now say "I am having the thought that I am a failure".		
Let's try out today's tip! Susan sometimes has the thought: "I am annoying".		
Try helping Susan to defuse her thought by putting "I am having the thought…" first.	*User enters typed text response (i.e. "I am having the thought I am annoying")*	
That's wonderful helping Susan to defuse her negative thought. Now it's your turn to try today's tip! Think of a negative thought you've had or are having about yourself that feels okay to be with for the next few moments. Got one?		
Write the thought here:	*User enters typed text response (e.g. "I am stupid").*	
Now let's try writing the thought and putting "I am having the thought that…" on front! Type it below.	*User enter typed text response (e.g. "I am having the thought I am stupid").*	
Well done doing something different with your thought today. Until we talk tomorrow why not keep trying out today's tip when you notice yourself having negative thoughts. Let's chat again tomorrow!	*Bye*	

extremely believable, extremely uncomfortable, and that they were extremely unwilling to have. At this point in the experiment, it was not possible to insert a mechanism that would direct participants to choose an alternative negative self-referential thought (which met the above criteria) or to prevent their advancement in the study. Informed by Larsson et al. (2016), a noncomprehension exclusion procedure was instated. Under this procedure, participants who rated their thoughts as neutral, positive, with low believability, as comfortable to have, and as willing to have said thoughts, were excluded from final analyses. To minimize undue burden on participants and loss of data, a score of 5 or above on each of the above respective items was employed to inform exclusion. A score of 5 or above denoted that a thought was: positive, not believable, comfortable to experience, and/or that the participant was willing to have that thought.

Missing Data

Final analyses included only completers who had met the noncomprehension exclusion criteria detailed above. Where possible, missing data was imputed via multiple imputation of the mean (Sterne et al., 2009; Jakobsen et al., 2017). This procedure was enacted for the AAQ-II, CFQ-7, and PANAS providing 70% of items had been completed by the participant (Sterne et al., 2009; Jakobsen et al., 2017). As such, differing numbers of participants included in analyses reflect that a participant's data could not be imputed.

Results

Descriptive Statistics and Psychometric Properties

Descriptive statistics and internal consistency at time 1 and 2 for the measures used are presented in Table 3. All measures showed high internal consistency at both time points (Cronbach's $\alpha = .90$ or above).

Attrition

The study observed a rate of attrition of 72% among those who entered randomization (see Figure 2). No significant difference on completion was observed between groups ($\chi^2 = .92$, $p = .63$). It was observed that, of those allocated to receive an active intervention ($n = 152$), 48 participants did not subsequently engage with their allocated chatbot intervention (i.e., said participants completed the above measures but did not subsequently click the hyperlink to begin engagement with their allocated intervention). One hundred four participants assigned to the active chatbot intervention groups engaged with their assigned intervention.

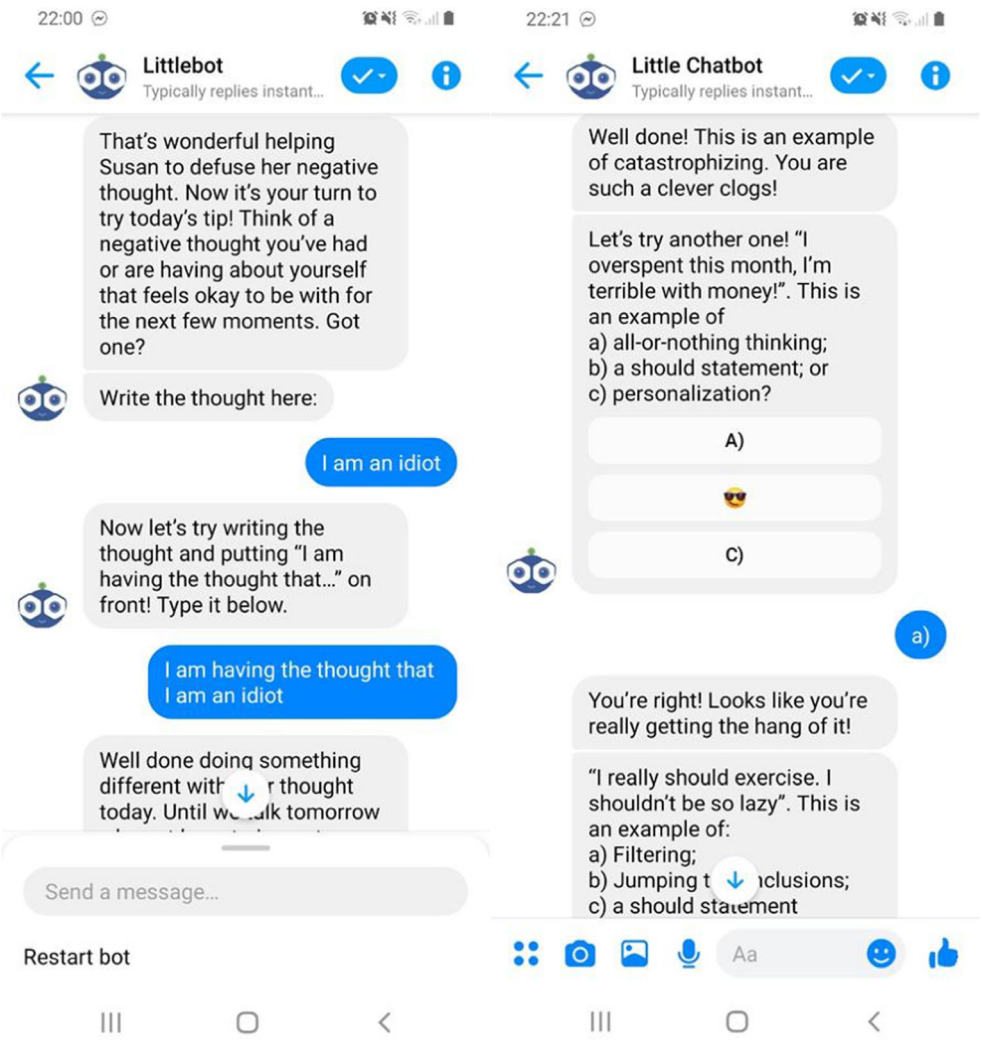

Fig. 1 Defusion (left; Littlebot) and restructuring (right; Little Chatbot) interventions as presented to participants via Facebook Messenger

Fifty-one participants subsequently completed their assigned chatbot intervention in full. Of those who did not complete (n = 53), 36 disengaged from their assigned chatbot intervention without apparent explanation (i.e., they ignored the chatbot's prompts to engage and commence conversation). Seventeen disengaged immediately following a chatbot malfunction (e.g., the chatbot became stuck in a loop of messages and did not advance; the chatbot message scheduler did not trigger and did not continue to engage with the participant; the participant responded in a manner outside of the prescribed manner [entered a text response when a button response was expected]). Of those who completed their assigned chatbot, six did not subsequently complete outcome measures postintervention. It was not possible to follow up with noncompleters in the active and nonactive groups given that this was an online study and all data was anonymous. Sixty-two participants completed outcome measures at postintervention, 34 of whom were excluded from the final analyses under the noncomprehension exclusion criteria detailed under the data analytic plan. Further breakdown of these figures is presented in Figure 2.

A series of independent sample t-tests (Table 4) were conducted to examine if there were significant differences

Table. 3 Descriptive Statistics and Psychometric Properties for Time 1 and Time 2

Variable	Time 1				Time 2			
	n	*M*	*SD*	α	n	*M*	*SD*	α
AAQ-II	202	26.91	9.99	0.91	60	27.30	9.93	0.93
CFQ-7	207	30.59	9.93	0.92	53	29.60	9.61	0.92
PANAS Pos.	206	26.60	8.79	0.92	60	25.38	9.19	0.94
PANAS Neg.	206	21.11	8.87	0.92	60	21.65	9.67	0.93
TTQ Negativity	210	2.78	1.97		62	3.00	1.75	
TTQ Believability	210	4.05	2.43		62	4.34	2.42	
TTQ Comfort	210	2.70	2.19		62	2.92	1.95	
TTQ Willingness	210	4.24	2.51		62	3.74	2.04	

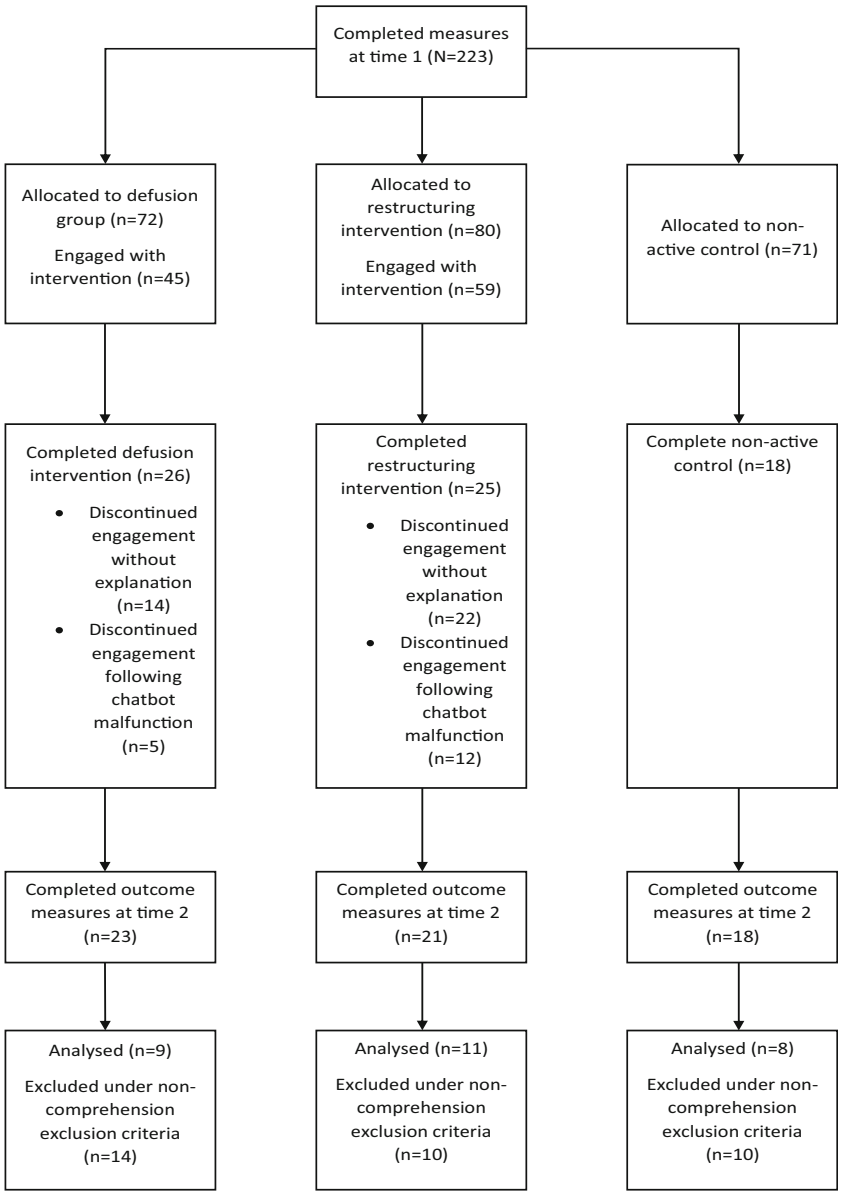

Fig. 2 Flowchart of participant progression through study

between completers and noncompleters on dependent variables. Noncompleters were observed to rate their thoughts as significantly less believable and with significantly greater positive affect than completers at time one.

Pearson's Chi Square test was conducted to examine for differences in completion by gender and experimental group, respectively. It was observed that there was no difference in completion by gender ($\chi^2 = 0.17$, $p = 0.68$) or experimental groups ($\chi^2 = .92$, $p = .63$).

Effectiveness of Interventions

A series of ANOVAs were conducted to examine for differences between groups at time one on each dependent variable with no statistically significant differences observed (see Tables 5 and 6 for descriptives for each measure for those included in final analyses). Per the data analytic plan a series of 3 x 2 mixed ANOVAs were conducted to examine the experimental research questions.

Thought Negativity

No significant interaction effect between experimental group and time was observed, $\Lambda = .82$, $F(2, 25) = 2.79$, $p = .08$, $\eta p^2 = .18$. The main effect for time was significant, $\Lambda = .69$, $F(1, 25) = 11.30$, $p = .002$, $\eta p^2 = .31$. The main effect for group was not significant, $F(2,25) = 1.49$, $p = .25$, $\eta p^2 = .11$. Moreover, post-hoc analyses did not suggest between-group differences at time 2.

Table. 4 T-Tests Examining Potential Differences between Completers and Noncompleters in Study Dependent Variables at Preintervention

Variable	t	Sig.	Complete		Non- Complete	
			M	*SD*	*M*	*SD*
Negativity	0.53	0.60	2.89	2.21	2.73	1.87
Believability	-2.02	0.04*	3.5	2.27	4.27	2.47
Comfort	-0.83	0.41	2.53	1.55	2.77	2.01
Willingness	-1.80	0.07	3.76	2.12	4.44	2.65
AAQ	1.65	0.10	28.66	9.31	26.15	10.20
CFQ-7	0.76	0.45	31.39	8.98	30.25	10.33
PANAS Pos.	-2.65	0.01*	24.16	8.55	27.65	8.71
PANAS Neg.	0.42	0.67	21.51	8.57	20.94	9.02

Note. "*" indicates significance at the .05 level; Scores closer to 10 on the negativity item indicate that a thought is more positive while scores closer to 10 on the believability item indicate that a thought is less believable.

Thought Believability

No significant interaction effect between experimental group and time was observed, $\Lambda = .84$, $F(2, 25) = .2.44$, $p = .11$, $\eta p^2 = .16$. The main effect for time was significant, $\Lambda = .77$, $F(1, 25) = 7.60$, $p = .011$, $\eta p^2 = .23$. The main effect for group was not significant, $F(2, 25) = .79$, $p = .47$, $\eta p^2 = .06$. Post-hoc analyses did not suggest between-group differences at time 2.

Thought Discomfort

The interaction effect between time and experimental group was not significant, $\Lambda = .83$, $F(2, 25) = 2.51$, $p = .10$, $\eta p^2 = .17$. The main effect for time was not significant, $\Lambda = .94$, $F(1, 25) = 1.54$, $p = .23$, $\eta p^2 = .06$. The main effect of group was not significant, $F(2, 25) = .48$, $p = .62$, $\eta p^2 = .04$. Post-hoc analyses did not suggest between-group differences at time 2.

Table. 6 Descriptive Statistics for ACT Process Measures and Measures of Mood by Group at Times 1 and 2

Group	AAQII		CFQ-7		PA		NA	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Time 1								
Defusion	27.75	9.08	29.17	10.72	24.75	7.25	18.75	5.97
Restructuring	30.09	8.37	31.30	5.81	20.27	7.73	21.55	8.43
Control	29.35	8.18	30.61	7.88	28.29	6.9	23.00	10.91
Time 2								
Defusion	29.25	11.03	31.33	7.42	25.50	8.19	20.50	12.49
Restructuring	28.09	8.58	31.00	8.64	19.18	7.83	22.00	7.90
Control	30.00	9.27	31.71	8.44	28.00	7.98	22.71	10.75

Note. PA = positive affect as measured by PANAS; NA = negative affect as measured by PANAS

Thought Willingness

The interaction effect between time and experimental group was significant, $\Lambda = .74$, $F(2, 25) = 4.38$, $p = .023$, $\eta p^2 = .26$. The main effect for time was not significant, $\Lambda = .91$, $F(1, 25) = .2.46$, $p = .13$, $\eta p^2 = .09$. The main effect for group was not significant, $F(2, 25) = 3.00$, $p = .07$, $\eta p^2 = .19$. Post-hoc analyses did not suggest between-group differences at time 2.

AAQ-II

The interaction effect between time and group was not significant, $\Lambda = .95$, $F(2, 23) = .57$, $p = .57$, $\eta p^2 = .05$. The main effect for time was not significant, $\Lambda = .999$, $F(1, 23) = .01$, $p = .91$, $\eta p^2 = .001$. The main effect for group was not significant, $F(2, 23) = .06$, $p = .94$, $\eta p^2 = .005$. Post-hoc analyses did not suggest between-group differences at time 2.

Table. 5 Descriptive Statistics for Target Thought Questionnaire by Group at Times 1 and 2

Scale	Defusion				Restructuring				Control			
	T1		T2		T1		T2		T1		T2	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Negativity	2.11	0.93	3.11	0.93	2.00	1.00	2.73	1.01	1.88	1.13	1.88	0.99
Believability	2.56	1.13	2.67	1.73	2.18	1.17	4.55	2.70	2.25	1.17	3.38	2.13
Discomfort	1.67	0.87	2.67	1.23	2.00	1.00	2.27	1.19	2.00	1.20	1.63	0.52
Willingness	1.78	0.83	2.44	1.94	3.09	1.14	2.36	1.50	2.75	1.49	4.38	1.92

Note: T1 = time 1; T2 = time 2; Scores closer to 10 on the negativity item indicate that a thought is more positive while scores closer to 10 on the believability item indicate that a thought is less believable.

CFQ-7 Measured Cognitive Fusion

The interaction effect between time and group was not significant, $\Lambda = .98$, $F(2, 20) = .17$, $p = .84$, $\eta p^2 = .017$. The main effect for time was not significant, $\Lambda = .99$, $F(1, 20) = .27$, $p = .61$, $\eta p^2 = .013$. The main effect for group was not significant, $F(2, 20) = .40$, $p = .96$, $\eta p^2 = .004$. Post-hoc analyses did not suggest between-group differences at time 2.

Positive Affect

The interaction effect between time and group was not significant, $\Lambda = .99$, $F(2, 23) = .11$, $p = .899$, $\eta p^2 = .009$. The main effect for time was not significant, $\Lambda = .999$, $F(1, 23) = .015$, $p = .90$, $\eta p^2 = .001$. The main effect for group was significant, $F(2, 23) = 4.01$, $p = .032$, $\eta p^2 = .258$.

Post-hoc analysis (Bonferroni) indicated that significant differences lay between the restructuring group ($M = 19.19$, $SD = 7.83$) and the control group ($M = 28.00$, $SD = 7.98$), $p = .037$. However, these differences existed at time 1 and thus maintained at time 2.

Negative Affect

The interaction effect between time and group was not significant, $\Lambda = .99$, $F(2, 23) = .08$, $p = .92$, $\eta p^2 = .007$. The main effect for time was not significant, $\Lambda = .996$, $F(1, 23) = .10$, $p = .75$, $\eta p^2 = .004$. The main effect for group was not significant, $F(2, 23) = .33$ $p = .73$, $\eta p^2 = .028$. Post-hoc analyses did not suggest between-group differences at time 2.

Intervention Adherence

Adherence to intervention was tested via Chi-square analysis among completers from the defusion and restructuring groups. No significant differences were observed between the defusion and restructuring groups on adherence to intervention, $\chi^2 = .2.92$, $p = .6$.

Discussion

The present study sought to investigate a novel, cost-effective, and scalable modality of intervention delivery namely conversational agent or chatbot. In doing so, we compared the effectiveness of two chatbot-delivered interventions (delivering cognitive defusion and cognitive restructuring, respectively) to a nonactive control group. It was predicted that the defusion intervention would lead to significantly reduced believability of thoughts and increased willingness to experience and comfort with thoughts relative to both cognitive restructuring and nonactive control while restructuring would significantly reduce negativity of thoughts relative to defusion and control. This prediction was not supported with no significant differences observed between-groups on any facet measured by the Target Thought Measure. The second prediction stated that significant differences in psychological flexibility and fusion would exist between the defusion group and both other groups at time 2, was not supported. The third prediction stated that significant differences would be observed positive affect and negative effect within both active-intervention groups and relative to nonactive control, was not supported. Among the most salient of findings was the high levels of attrition observed. In view of this and that the targeted sample size was not achieved, much of the discussion that follows focuses on reducing attrition and developing more engaging and effective chatbot interventions.

The present findings suggest minimal effectiveness of chatbots as a delivery method of brief defusion and restructuring interventions to change form, frequency, and relationship with negative self-referential thoughts. However, this absence of the predicted effects may be explained by the omission of a rationale and the broad range of defusion techniques employed. Unlike previous studies examining brief restructuring and defusion interventions (e.g., Larsson et al., 2016; Masuda et al., 2004; Masuda et al., 2010) the present study did not present a rationale for the defusion and restructuring interventions that were presented. Presentation of a rationale has been observed to increase the effectiveness of brief defusion interventions (Masuda et al., 2010). Assaz et al. (2018) suggest an effective rationale supports the creation of a context in which defusion can be understood and integrated by participants and thus effective. A further difference to previous studies centers on the use of five defusion techniques. Previous studies examining the effect of brief defusion interventions have focused on vocalizing-based techniques (e.g., "I'm having the thought that"; rapid repetition; singing the thought) and visualization techniques (e.g., "leaves on a stream"; Foody et al., 2015), and have tended to use of a maximum of two defusion techniques in the context of brief intervention (Hinton & Gaynor, 2010; Larsson et al., 2016; Masuda et al., 2004, 2010). The present study introduced five defusion techniques including both vocalization and visualization strategies. This provision of five different defusion techniques that did not coherently integrate with each other may have hindered effective implementation of defusion for this group.

During the present study, a high rate of attrition (72%) was observed. This was much larger than observed attrition of 17% observed in previous studies on chatbots (see Fitzpatrick et al., 2017) and meta-analyses of online and face-to-face ACT (O'Connor et al., 2018; Ong et al., 2018). One explanation for the high rate of noncompletion is that those who dropped out of the study may have been experiencing little need to or benefit from participation. Indeed, it was observed that those who did not complete the study rated their

negative self-referential thoughts as less believable and had higher positive affect those who went on to the complete the study. A further factor that contributed to attrition within the chatbot groups was the anonymous nature of the study. For those in the chatbot groups, contact details were not collected with follow-up measures being delivered via the assigned chatbot. As such, it was not possible to deliver postintervention measures to those who completed some but not all of their assigned chatbot (the link to postintervention measures were included in the last message sent by the chatbot to the participant).

The substantial rate of attrition and ineffectiveness of the chatbots may also be explained by factors pertinent to user experience including usability and acceptability. Indeed, a systematic review by Gaffney et al. (2019) identified a number of factors which impact on the effectiveness of and continued engagement with chatbot interventions including repetitive content, a shallow or superficial relationship, and inflexibility in the agent's ability to understand and respond appropriately. Many of these challenges were evident in the present study and may have contributed to dropout and the ineffectiveness of chatbots. Indeed, the chatbots employed in the present study were rudimentary with no integrated artificial intelligence (AI), and no natural language processing (NLP) capabilities. As such, chatbots were unable to respond flexibly to participants when participants responded in unexpected ways (e.g., providing a text-based response when a prescribed button response was expected). In such instances the chatbots issued the error response: "Uh oh, I'm still learning how to communicate with my human friends. It is best to use my word buttons." If the participant persisted with text responses, the chatbot continued to reissue the above error responses resulting in conversational loops. As noted above, 17 participants disengaged from their assigned chatbot immediately following such a malfunction. Those who continued to engage following said malfunction may have experienced an affected relationship with the chatbot. Indeed, Langevin et al. (2021) identify cognitive flexibility as an important usability issue encountered by chatbots that hampers engagement and effectiveness. This incorporates the ability of chatbots to respond flexibly, minimize errors, and respond appropriately when errors do occur. Thus, engagement with and the effectiveness of the present chatbots may have been impact by their reduced capability to respond flexibly to unexpected responses and errors.

Given the absence of AI and NLP, broad and generic empathetic responses were scripted as responses to participant text responses. For example, the chatbot might thank the participant for sharing a negative self-referential thought or provide generic validation after practicing the assigned exercise (e.g., "That was brilliant work doing something different with your thoughts"). This limited ability to respond with specific and empathetic responses is likely to have inhibited participant rapport and working alliance with chatbots (Bendig et al., 2019; Morris et al., 2018; Abd-alrazaq et al., 2019). Indeed, a recent study by Prochaska et al. (2021) observed that a stronger working alliance with the chatbot employed was associated with key therapeutic outcomes and greater engagement with the intervention. Given the potential impact to working alliance of generic and nonspecific responding, and chatbot malfunction, an affected working alliance may, in part, explain the ineffectiveness of chatbots and attrition observed in the present study.

Strengths of the present study include the randomized design, blinding of participants and researchers to group allocation, and participants being naïve to the intervention being delivered. A further strength of the present study was the inclusion of outcomes reflective of the treatment and philosophical approaches of ACT (believability, discomfort, willingness) and CBT (negativity; Larsson et al., 2016). Indeed, some previous studies investigating brief defusion and restructuring interventions have focused on outcomes consisted with ACT or defusion only that may hide or obfuscate the effect of a restructuring intervention. A limitation of the present study is the omission of a measure of acceptability. Although some inferences relating to acceptability can be drawn from the high rate of attrition, it is unclear how acceptable the intervention was to completers and by extension how acceptability may have affected intervention effectiveness. Future research might address this limitation through the use of a Likert scale measuring acceptability supplemented by qualitative data on acceptability such as that employed by Fitzpatrick et al. (2017).

The present study is of note as it is among the first to deliver an ACT process via the modality of chatbot. Growing evidence suggests that chatbots are feasible, acceptable, and effective modalities to deliver psychological intervention (Gaffney et al., 2019). Moreover, these interventions are economic in terms of time and finances. In light of the present findings, we propose a number of recommendations to researchers and practitioners who intend to develop and investigate chatbots to deliver psychological content. First, we recommend extensive testing to gain feedback not only on psychological content but also to acquire usability and acceptability data. This might include testing by those with appropriate psychological expertise, those with knowledge in computer science and/or information and communication studies, and potential end-users. Such testing should guard against challenges encountered by the present users, namely confusion in integrating psychological content, chatbot malfunctions, and inflexible chatbot responding. Second, and in line with this, future researchers should employ varied approaches to

measurement of engagement, usability, and acceptability. As noted by Gaffney et al. (2019), measurement of these constructs is not standard but might include recording numbers of the engagements with the chatbot intervention, Likert scales on acceptability and usability (see Fitzpatrick et al., 2017), qualitative interviews (see Ly et al., 2017), and questionnaires (see Langevin et al., 2021). Third, that future researchers employ more sophisticated chatbots that avail of AI and NLP. Such functionality should ensure that chatbot interventions are better equipped to respond flexibly to participants, which should improve usability, acceptability, and working alliance. Fourth, that researchers script a wider variety of empathetic responses. This recommendation is proposed to facilitate deepened relationship and working alliance with chatbots with in view of evidence suggesting this is an important factor influencing effectiveness and engagement (Prochaska et al., 2021).

Although the present study provides minimal evidence for the effectiveness of chatbots as a modality to deliver brief defusion and restructuring interventions, the lack of provision of a rationale, presentation of five defusion techniques, and poor acceptability of chatbot interventions are presented as alternative explanations for the present findings. Indeed, poor user experience rather than the psychological content presented may explain the absence of the predicted effects in the present study with future research necessary to provide greater insight. Despite this, the present study makes a novel contribution because it is among the first to deliver an ACT process via chatbot. Moreover, the present research introduces this mode of intervention to a wider audience and makes several recommendations intended to facilitate researchers and clinicians in developing and evaluating chatbots delivering psychological content.

Appendices

Appendix A. Target Thought Questionnaire

Now pick a negative thought about yourself that would rate as **EXTREMELY NEGATIVE, EXTREMELY BELIEVABLE, EXTREMELY UNCOMFORTABLE** and that you are **EXTREMELY UNWILLING** to be thinking.

Negative Thought:_____

1. How **negative** is the thought?

Extremely negative									Extremely positive
1	2	3	4	5	6	7	8	9	10

2. How **believable** is the thought?

Extremely believable									Extremely unbelievable
1	2	3	4	5	6	7	8	9	10

3. How **comfortable** is the thought?

Extremely uncomfortable									Extremely comfortable
1	2	3	4	5	6	7	8	9	10

4. How **willing** are you to have the thought?

Extremely unwilling									Extremely willing
1	2	3	4	5	6	7	8	9	10

Appendix B. Adherence Quiz

The chatbot suggested, when dealing with thoughts I don't want, I should:
a) Notice my thoughts as just thoughts rather that truth or reality
b) Treat my thoughts as reality
c) Consider evidence for and against my thought and come up with a more balanced thought
d) Focus solely on evidence that suggests my thought is true

Appendix C. Defusion Chatbot Intervention Stimuli

Day 1: Some thoughts are really horrible to have such as "I am a failure." One thing that can help is to put "I am having the thought that" before the thought to help you notice that it is just a thought. For example, "I am having the thought that I am a failure."
Day 2: Find a thought that bothers or upsets you. Focus on that thought for 10 s believing it as much as possible. Notice how it affects you. Then pick an animated cartoon character with a humorous voice such as Mickey Mouse, Bugs Bunny, Shrek, or Homer Simpson. Now bring the troubling thought to mind but hear it in the cartoon character's voice as though the cartoon character was speaking your thoughts out loud. Notice what happens.
Day 3: A video clip elaborating on day one by adding "I am noticing . . . ," i.e., "I am noticing that I am having the thought that I am a failure." https://www.youtube.com/watch?v=kwlYXupjoaI
Day 4: A video clip of an ACT metaphor comparing thoughts and the mind to dishes on a sushi train. https://www.youtube.com/watch?v=tzUoXJVI0wo
Day 5: An audio clip of "leaves on a stream," a mindfulness meditation that asks that one visualize their thoughts as leaves on a stream.

Appendix D. Cognitive Restructuring Chatbot Intervention Stimuli

Day 1: Cognitive distortions/thinking errors, a psychoeducational information sheet detailing common cognitive distortions including "all-or-nothing thinking," "catastrophizing," and "should statements."
Day 2: Examining the thought—Often when we have negative thoughts we find it very easy to think of reasons and evidence that support that negative thought. Though we may believe something to be true, this does not necessarily mean that it is. It is often valuable to see if the facts of the situation back up what you are thinking, or whether they contradict what you are thinking. A good question to ask is: "would other people accept my thoughts as true?"
Day 3: Balancing a thought—Jenny sometimes makes mistakes or fumbles her words when she answers questions aloud in class. She often has the thought "I am an idiot." When she thought about the evidence against this thought she found that: (1) she usually gets good grades; (2) she won student of the year in school; and (3) her classmates often ask for her help in understanding coursework. Jenny decided it would be fairer to think "I sometimes make mistakes but I usually get good grades and am quite clever." Like Jenny, we too can generate a more balanced or fair thought when we think about evidence that contradicts our negative thought.
Days 4 & 5: Combining the above skills

Authors' contributions All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Joseph Lavelle, Neil Dunne and Louise McHugh. The first draft of the manuscript was written by Joseph Lavelle and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding Open Access funding provided by the IReL Consortium. The first author (Joseph Lavelle) is funded by a UCD Foundation Scholarship donated by the Centre for Colorectal Disease (Grant numbers 28335 & 31189; Centre for Colorectal Disease, Saint Vincent's University Hospital, Elm Park, Dublin 4, Ireland).

Data availability The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflicts of Interest The lead author (Mr. Joseph Lavelle) is supported by funding from the UCD Foundation. Said funding was donated by the Centre for Colorectal Disease, Saint Vincent's University Hospital, Elm Park, Dublin 4, Ireland. However, neither the UCD foundation nor the Centre for Colorectal Disease had any involvement in study design; in the collection, analysis, and interpretation of data; in the writing of the report; and in the decision to submit the article for publication. None of other authors have any declarations in respect to funding or potential conflicts of interest.

Ethics Approval The study was approved by the University College Dublin Humanities and Social Sciences Ethics Committee and was in line with the 1964 Declaration of Helsinki and its later amendments.

Consent to Participate All participants provided informed consent per the procedures approved in the above ethics approval.

Consent for publication All authors are aware of and consent to the submission of this manuscript to the Psychological Record. Further, all participants were made aware that their analysed data would contribute toward a manuscript for peer-revied publication and consented to same.

Code availability N/A

References

Abd-alrazaq, A. A., Alajlani, M., Alalwan, A. A., Bewick, B. M., Gardner, P., & Househ, M. (2019). An overview of the features of chatbots in mental health: A scoping review. *International Journal of Medical Informatics, 132*, 103978. https://doi.org/10.1016/j.ijmedinf.2019.103978.

Andersson, G. (2018). Information technology and the changing role of practice. In S. C. Hayes & S. G. Hoffman (Eds.), *Process-based CBT: The science and core clinical competencies of cognitive behavioral therapy (pp. 67–82)*. Context Press.

Assaz, D. A., Roche, B., Kanter, J. W., & Oshiro, C. K. B. (2018). Cognitive defusion in acceptance and commitment therapy: What are the basic processes of change? *The Psychological Record, 68*(4), 405–418. https://doi.org/10.1007/s40732-017-0254-z.

Barrera, T. L., Szafranski, D. D., Ratcliff, C. G., Garnaat, S. L., & Norton, P. J. (2015). An experimental comparison of techniques: Cognitive defusion, cognitive restructuring, and in-vivo exposure for social anxiety. *Behavioral & Cognitive Psychotherapy, 44*(2), 249–254. https://doi.org/10.1017/S1352465814000630.

Beck, A. T., Steer, R. A., Beck, J. S., & Newman, C. F. (1976). Hopelessness, depression, suicidal ideation, and clinical diagnosis of depression. *Suicide & Life-Threatening Behavior, 23*(2), 139–145.

Beck, J., & Beck, A. T. (2011). *Cognitive behavior therapy: Basics and beyond*. Guilford Press.

Bendig, E., Erb, B., Schulze-Thuesing, L., & Baumeister, H. (2019). The next generation: Chatbots in clinical psychology and psychotherapy to foster mental health: A scoping review. *Verhaltenstherapie*, 1–13. https://doi.org/10.1159/000501812.

Bond, F. W., Hayes, S. C., Baer, R. A., Carpenter, K. M., Guenole, N., Orcutt, H. K., Waltz, T., & Zettle, R. D. (2011). Preliminary psychometric properties of the Acceptance and Action Questionnaire-II: A revised measure of psychological inflexibility and experiential avoidance. *Behavior Therapy, 42*(4), 676–688. https://doi.org/10.1016/j.beth.2011.03.007.

Brown, M., Glendening, A., Hoon, A., & John, A. (2016). Effectiveness of web-delivered acceptance and commitment therapy in relation to mental health and well-being: A systematic review and meta-analysis. *Journal of Medical Internet Research, 18*(8), 1–40. https://doi.org/10.2196/jmir.6200.

Clark, D. A. (2014). Cognitive restructuring. In S. G. Hoffman (Ed.), *The Wiley handbook of cognitive behavioral therapy (pp. 2–22)*. John Wiley & Sons. https://doi.org/10.1002/9781118528563.wbcbt02.

Clark, D. A., & Beck, A. T. (2011). *Cognitive therapy of anxiety disorders*. Guilford Press.

Clark, D. A., & Rhyno, S. (2005). Unwanted intrusive thoughts in nonclinical individuals: Implications for clinical disorders. In D. A. Clark (Ed.), *Intrusive thoughts in clinical disorders: Theory, research, and treatment (pp. 1–29)*. Guilford Press.

Crawford, J., & Henry, J. (2004). The positive and negative affect schedule (PANAS): Construct validity, measurement properties and normative data in a large non-clinical sample. *British Journal of Clinical Psychology, 43*, 245–265. https://doi.org/10.1348/0144665031752934.

Cullen, C. (2008). Acceptance and commitment therapy (ACT): A third wave behaviour therapy. *Behavioural and Cognitive Psychotherapy, 36*(6), 667–673. https://doi.org/10.1017/S1352465808004797.

Deacon, B. J., Fawzy, T., & Lickel, J. J. (2011). Cognitive defusion versus cognitive restructuring in the treatment of negative self-referential thoughts: An investigation of process and outcome. *Journal of Cognitive Psychotherapy, 25*(8), 218–232. https://doi.org/10.1891/0889-8391.25.3.218.

Ellis, A. (2008). Cognitive restructuring of the disputing of irrational beliefs. In W. T. O'Donohue & J. E. Fisher (Eds.), *Cognitive behavior therapy: Applying empirically supported techniques in your practice (pp. 91–95)*. John Wiley & Sons.

Fitzpatrick, K., Darcy, A., & Vierhile, M. (2017). Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): A randomized controlled trial. *Journal of Medical Internet Research Mental Health, 4*(2), e19. https://doi.org/10.2196/mental.7785.

Foody, M., Barnes-Holmes, Y., Barnes-Holmes, D., Rai, L., & Luciano, C. (2015). An empirical investigation of the role of self, hierarchy, and distinction in a common ACT exercise. *The Psychological Record, 65*(2), 231–243. https://doi.org/10.1007/s40732-014-0103-2.

Fulmer, R., Joerin, A., Gentile, B., Lakerink, L., & Rauws, M. (2018). Using psychological artificial intelligence (Tess) to relieve symptoms of depression and anxiety: Randomized controlled trial. *Journal of Medical Internet Research Mental Health, 5*(4), e64. https://doi.org/10.2196/mental.9782.

Gaffney, H., Mansell, W., & Tai, S. (2019). Conversational agents in the treatment of mental health problems: Mixed-method systematic review. *Journal of Medical Internet Research Mental Health, 6*(10), e14166. https://doi.org/10.2196/14166.

Gillanders, D. T., Bolderston, H., Bond, F. W., Dempster, M., Flaxman, P. E., Lindsey Campbell, S. K., Tansey, L., Noel, P., Ferenbach, C., Masley, S., Roach, L., Joda Lloyd, L. M., Clarke, S., & Remington, B. (2014). The development and initial validation of the cognitive fusion questionnaire. *Behavior Therapy, 45*(1), 83–101. https://doi.org/10.1016/j.beth.2013.09.001.

Grist, R., & Cavanagh, K. (2013). Computerised cognitive behavioural therapy for common mental health disorders, what works, for whom, and under what circumstances? A systematic review and meta-analysis. *Journal of Contemporary Psychotherapy, 43*(4), 243–251. https://doi.org/10.1007/s10879-013-9243-y.

Harris, R. (2008). *The happiness trap: How to stop struggling and start living*. Trumpeter Books.

Hayes, S. (2004). Acceptance and commitment therapy, relational frame theory, and the third wave of behavioral and cognitive therapies. *Behavior Therapy, 35*, 639–665. https://doi.org/10.1016/S0005-7894(04)80013-3.

Hayes, S. C., & Brownstein, A. J. (1986). Mentalism, behavior-behavior relations, and a behavior-analytic view of the purposes of science. *The Behavior Analyst, 9*(2), 175–190. https://doi.org/10.1007/BF03391944.

Hayes, S. C., Barnes-Holmes, D., & Roche, B. (Eds.). (2001). *Relational frame theory: A post-Skinnerian account of human language and cognition.* Kluwer Academic/Plenum Publishers.

Hayes, S. C., Pistorello, J., & Levin, M. E. (2012). Acceptance and commitment therapy as a unified model of behaviour change. *The*

Counseling Psychologist, 40(7), 976–1002. https://doi.org/10.1177/0011000012460836.

Hayes, S. C., Strosahl, K. D., & Wilson, K. G. (1999). *Acceptance and commitment therapy: An experiential approach to behavior change*. Guilford Press.

Healy, H., Barnes-Holmes, Y., Barnes-Holmes, D., Keogh, C., Luciano, C., & Wilson, K. (2008). An experimental test of cognitive defusion exercise: Coping with negative and positive self-statements. *The Psychological Record, 58*(4), 623–640.

Hinton, M. J., & Gaynor, S. T. (2010). Cognitive defusion for psychological distress, dysphoria, and low self-esteem: A randomized technique evaluation trial of vocalizing strategies. *International Journal of Behavioural Consultation & Therapy, 6*(3), 164–185. https://doi.org/10.1037/h0100906.

Inkster, B., Sarda, S., & Subramanian, V. (2018). An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: Real-world data evaluation mixed-methods study. *Journal of Medical Internet Research mHealth uHealth, 6*(11), e12106. https://doi.org/10.2196/12106.

Jakobsen, J. C., Gluud, C., Wetterslev, J., & Winkel, P. (2017). When and how should multiple imputation be used for handling missing data in randomised clinical trials: A practical guide with flowcharts. *BMC Medical Research Methodology, 17*(162). https://doi.org/10.1186/s12874-017-0442-1.

Langevin, R., Lordon, R., Avrahami, T., Cowan, B., Hirsch, T., & Hsieh, G. (2021). Heuristic evaluation of conversational agents. Paper presented at CHI '21: ACM CHI Conference on Human Factors in Computing Systems, 2021, Yokohama, Japan. https://faculty.washington.edu/garyhs/docs/langevin-CHI2021-caheuristics.pdf

Larsson, A., Hooper, N., Osborne, L. A., Bennett, P., & McHugh, L. (2016). Using brief cognitive restructuring and cognitive defusion techniques to cope with negative thoughts. *Behavior Modification*, *40*(3), 1–31. https://doi.org/10.1177/0145445515621488.

Luoma, J. B., & Hayes, S. C. (2008). Cognitive defusion. In W. J. Donohue & J. E. Fischer (Eds.), *Cognitive behavior therapy: Empirically supported techniques in your practice (2nd ed.; pp. 83–90)*. John Wiley and Sons.

Ly, K. H., Ly, A. M., & Andersson, G. (2017). A fully automated conversational agent for promoting mental well-being: A pilot RCT using mixed methods. *Internet Interventions, 10*(10), 39–46. https://doi.org/10.1016/j.invent.2017.10.002.

Masuda, A., Feinstein, A. B., Wendell, J. W., & Sheehan, S. T. (2010). Cognitive defusion versus thought distraction: A clinical rationale, training, and experiential exercise in altering psychological impacts of negative self-referential thoughts. *Behavior Modification, 34*(6), 520–538. https://doi.org/10.1177/0145445510379632.

Masuda, A., Hayes, S. C., Sackett, C. F., & Twohig, M. P. (2004). Cognitive defusion and self-relevant negative thoughts: Examining the impact of a 90-year-old technique. *Behavior Research & Therapy, 42*(4), 477–485. https://doi.org/10.1016/j.brat.2003.10.008.

Morris, R. R., Kouddous, K., Kshirsagar, R., & Schueller, S. M. (2018). Towards an artificially empathic conversational agent for mental health applications: System design and user perceptions. *Journal of Medical Internet Research, 20*(6), e10148. https://doi.org/10.2196/10148.

O'Connor, M., Munnelly, A., Whelan, R., & McHugh, L. (2018). The efficacy and acceptability of third-wave behavioral and cognitive eHealth treatments: A systematic review and meta-analysis of randomized controlled trials. *Behavior Therapy, 49*(3), 459–475. https://doi.org/10.1016/j.beth.2017.07.007.

Ong, C. W., Lee, E. B., & Twohig, M. P. (2018). A meta-analysis of dropout rates in acceptance and commitment therapy. *Behavior Research & Therapy, 104*, 14–33. https://doi.org/10.1016/j.brat.2018.02.004.

Perneger, T. V. (1998). What's wrong with Bonferroni adjustments. *British Medical Journal, 316*(7139), 1236–1238. https://doi.org/10.1136/bmj.316.7139.1236.

Pilecki, B., & McKay, D. (2012). An experimental investigation of cognitive defusion. *The Psychological Record, 62,* 19–40.

Prochaska, J. J., Vogel, E. A., Chieng, A., Kendra, M., Baiocchi, M., Pajarito, S., & Robinson, A. A. (2021). Therapeutic relational agent for reducing problematic substance use (woebot): Development and usability study. *Journal of Medical Internet Research, 23*(3), e24850. https://doi.org/10.2196/24850.

Rathbone, A. L. (2017). The use of mobile apps and SMS messaging as physical and mental health interventions: Systematic review. *Journal of Medical Internet Research, 19*(8), e295. https://doi.org/10.2196/jmir.7740.

Sterne, J. A. C., White, I. R., Carlin, J. B., Spratt, M., Royston, P., Kenward, M. G., Wood, A. M., & Carpenter, J. R. (2009). Multiple imputation for missing data in epidemiological and clinical research: Potential and pitfalls. *British Medical Journal, 338*, 157–160. https://doi.org/10.1136/bmj.b2393.

Titchener, E. B. (1916). *A textbook of psychology.* Macmillan.

Tyndall, I., Waldeck, D., Pancani, L., Whelan, R., Roche, B., & Dawson, D. L. (2019). The Acceptance and Action Questionnaire-II (AAQ-II) as a measure of experiential avoidance: Concerns over discriminant validity. *Journal of Contextual Behavioral Science, 12,* 278–284. https://doi.org/10.1016/j.jcbs.2018.09.005.

Watson, D., Clark, L., & Carey, G. (1988). Positive and negative affectivity and their relation to anxiety and depressive disorders. *Journal of Abnormal Psychology, 97*, 346–353. https://doi.org/10.1037/0021-843X.97.3.34.