Author/s:
Hitchcock, C;Funk, J;Cummins, R;Patel, SD;Catarino, A;Takano, K;Dalgleish, T;Ewbank, M

Title:
A deep learning quantification of patient specificity as a predictor of session attendance and treatment response to internet-enabled cognitive behavioural therapy for common mental health disorders

Date:
2024-04-01

Citation:
Hitchcock, C., Funk, J., Cummins, R., Patel, S. D., Catarino, A., Takano, K., Dalgleish, T. & Ewbank, M. (2024). A deep learning quantification of patient specificity as a predictor of session attendance and treatment response to internet-enabled cognitive behavioural therapy for common mental health disorders. Journal of Affective Disorders, 350, pp.485-491. https://doi.org/10.1016/j.jad.2024.01.134.

Persistent Link:
https://hdl.handle.net/11343/340263

Journal of Affective Disorders 350 (2024) 485–491

Contents lists available at ScienceDirect

Journal of Affective Disorders

journal homepage: www.elsevier.com/locate/jad

Research paper

A deep learning quantification of patient specificity as a predictor of session attendance and treatment response to internet-enabled cognitive behavioural therapy for common mental health disorders

Caitlin Hitchcock [a,b,*], Julia Funk [a,c], Ronan Cummins [d], Shivam D. Patel [a], Ana Catarino [d], Keisuke Takano [c], Tim Dalgleish [a], Michael Ewbank [d]

[a] Medical Research Council Cognition and Brain Sciences Unit, University of Cambridge, United Kingdom
[b] Melbourne School of Psychological Sciences, The University of Melbourne, Australia
[c] Department of Psychology, Ludwig-Maximilians-Universität München, Germany
[d] Ieso Digital Health, Jeffreys Building, Cowley Rd, Milton, Cambridge, United Kingdom

A B S T R A C T

Background: Increasing an individual's ability to focus on concrete, specific detail, thus reducing the tendency toward overly broad, decontextualised generalisations about the self and world, is a target within cognitive behavioural therapy (CBT). However, empirical investigation of the impact of within-treatment specificity on treatment outcomes is scarce. We evaluated whether the specificity of patient dialogue predicted a) end-of-treatment symptoms and b) session completion for CBT for common mental health issues.

Methods: This preregistered (https://osf.io/agr4t) study trained a deep learning model to score the specificity of patient dialogue in transcripts from 353,614 internet-enabled CBT sessions for common mental health disorders, delivered on behalf of UK NHS services. Data were from obtained from 65,030 participants ($n = 47,308$ female, $n = 241$ unstated) aged 18–94 years ($M = 34.69$, $SD = 12.35$). Depressive disorders were the most common (39.1 %) primary diagnosis. Primary outcome was end-of-treatment score on the Patient Health Questionnaire-9 (PHQ-9). Secondary outcome was number of sessions attended.

Results: Linear mixed-effects models demonstrated that increased patient specificity significantly predicted lower post-treatment symptoms on the PHQ-9, although the size and direction of the effect varied depending on the type of therapeutic activity being completed. Effect sizes were consistently small. Higher patient specificity was associated with completing a greater number of sessions.

Limitations: We are unable to infer causation from our data.

Conclusions: Although effect sizes were small, an effect of specificity was observed across common mental health disorders. Further studies are needed to explore whether encouraging patient specificity during CBT may provide an enhancement of treatment attendance and treatment effects.

1. Introduction

Cognitive behavioural therapy (CBT) is our leading treatment for common mental health problems. However, CBT fails to produce remission for 30–50 % of treated individuals (Clark et al., 2018), leaving them with significant symptoms and poor functional outcomes. Improving understanding of active therapy features implicated in good treatment outcomes is therefore of vital importance to further refine treatment.

In both traditional and third-wave CBTs, one potentially important mechanism of therapeutic change is movement away from an abstract, generalized cognitive style, toward concrete, specific processing (Beck and Haigh, 2014; van der Velden et al., 2015). A variety of treatment components seek to facilitate specific processing. Mindfulness exercises aim to improve ability to focus on concrete (e.g., sensory) detail and to decentre from abstracted thought patterns (e.g., worry) (Manjaly and Iglesias, 2020), cognitive restructuring aims to constrain generalized beliefs about the self and world to specific situations or events, and exposure techniques help to reinstate context (Craske et al., 2014). An emphasis on context-specific detail serves to isolate specific experiences in time and space (e.g., argument with partner after dinner last night), and illustrate that assumptions (e.g., my relationship is full of conflict) made on the basis of such events are only applicable within that context (e.g., when deciding who is going to do the dishes). A focus on the specific is thereby important to combat the overgeneralised beliefs of worthlessness, jeopardy, failure, and vulnerability which drive common mental health disorders (Beck and Haigh, 2014).

To this end, basic-science driven interventions have been offered as

* Corresponding author at: Melbourne School of Psychological Sciences, The University of Melbourne, Australia.
E-mail address: Caitlin.hitchcock@unimelb.edu.au (C. Hitchcock).

https://doi.org/10.1016/j.jad.2024.01.134
Received 12 April 2023; Received in revised form 9 January 2024; Accepted 14 January 2024
Available online 19 January 2024

adjuncts to CBT to enhance concrete, specific cognition, targeting rumination (Watkins et al., 2012), overgeneralised autobiographical memory (Hallford et al., 2021a), or mental imagery (Pile et al., 2021) as a training modality. Similarly, process-focussed adaptations of CBT have integrated a greater focus on being concrete (Jacobs et al., 2016). However, the specificity of narrative discourse during therapeutic exchanges has been largely unexplored as a treatment process, likely due to the quantity and unstructured format of such data.

Machine learning techniques can overcome these challenges, providing an unprecedented opportunity to examine factors at play *during* therapeutic exchanges between therapist and patient, at scale (Chekroud et al., 2021). Recent research has trained deep learning models to identify different features of treatment (e.g., topics of discussion, type of therapeutic tasks completed), and regressed these features on treatment outcomes (Ewbank et al., 2020), expanding our ability to identify features of successful CBT across large numbers of therapists, treatment sites, and treatment-seekers (for discussion see Chekroud et al., 2021). Here, our primary aim was to determine whether the degree of patient specificity within therapeutic interactions predicted post-treatment symptoms. We expected that increased patient specificity during therapy would predict better treatment outcomes for at least two reasons; patient specificity may reflect engagement in the more concrete processing style promoted within CBT, and/or more specific detail during therapeutic activities (e.g., when designing behavioural experiments) may lead to increased problem solving (Jing et al., 2016) and completion of planned therapeutic tasks. To explore whether specificity may be more important in some therapeutic tasks than others, we separately considered specificity within different therapy feature categories (as per Ewbank et al., 2020), including goal setting, CBT tasks, and management of homework.

Our secondary aim was to determine whether patient specificity predicted treatment engagement or dropout; notably, the number of CBT sessions completed. As many CBT tasks require the individual to provide specific information from their past (e.g., evidence gathering tasks), we hypothesised that individuals who tended to be more specific may find CBT easier to complete, and thus complete a greater number of sessions.

We trained a deep learning model to quantify the specificity of patient dialogue within internet-enabled CBT sessions delivered for UK NHS Improving Access to Psychological Therapy (IAPT) services. Our primary hypothesis was that the specificity of patients' dialogue would predict treatment outcome, such that a greater number of specific features in patients' responses would predict lower post-treatment scores on the Patient Health Questionnaire (PHQ-9; Kroenke et al., 2001). We focussed on the PHQ-9, a measure of depression, as it is administered to all IAPT service-users, is linked to post-treatment quality of life (Furukawa et al., 2021), and is a key outcome used when evaluating the efficacy of IAPT services (Gyani et al., 2013). Further, there is a well-established prognostic relationship between specificity and depressive symptoms (Hallford et al., 2021b). Our second hypothesis was that greater specificity of patient dialogue would predict a higher number of CBT sessions completed.

2. Method

2.1. Approvals

Cambridge Psychology Research Ethics committee granted ethics approval (PRE.2017.100). Hypotheses and analyses were pre-registered, https://osf.io/agr4t

2.2. Participants

Participants (n = 65,030) were receiving treatment from NHS Improving Access to Psychological Therapies (IAPT) services in the UK – a public health service which treats common mental health issues. Individuals can self-refer, online. Inclusion criteria were age $\geq$ 18 years and registration with a general practitioner in the services' NHS Trust. Exclusion criteria, determined by a clinician during the intake assessment, were being deemed unsuitable for online treatment (e.g., experiencing cognitive impairment, brain injury or dementia); already receiving psychological intervention; no access to an internet-enabled device; extremely low literacy levels indicating difficulty reading or writing; visual impairment preventing use of an electronic screen; lack of English; unsuitable for IAPT as in need of more intensive support (e.g., actively suicidal).

2.3. Internet-enabled CBT

CBT was administered via instant synchronous messaging in a secure online therapy room by a British Association for Behavioural and Cognitive Psychotherapies–accredited CBT therapist or NHS Psychological Wellbeing Practitioner, per National Institute of Health and Care Excellence (NICE) recommendations. That is, CBT was delivered on an individual basis, in real-time, by a trained therapist in a secure chat room, on behalf of NHS IAPT services by Ieso (www.iesogroup.com), in accordance with international standards for information security (ISO 27001; https://www.iesogroup.com/iso-certificates). Disorder-specific CBT protocols were used to guide treatment, which consisted of a range of CBT tasks, including cognitive restructuring, planning behavioural experiments, or behavioural activation. Sessions were approximately 45–60 min in length, and delivered on a weekly-basis, unless less frequent sessions were deemed appropriate by both therapist and patient.

2.4. Specificity scoring

We analyzed transcripts for 353,614 internet-enabled CBT sessions completed June 2012 to May 2021. Each therapy transcript recorded which text entries were submitted by the user, and which text entries were submitted by the therapist. Therapy transcripts were parsed for linguistic features indicating that a therapist was eliciting information (e.g., 'can you think of an example/the last time/a recent time'); full details are presented in the Supplementary Materials. Therapist elicitations included both explicit requests for specific information (e.g., 'can you tell me about last time this happened?'), and more open-ended requests for information (e.g., 'how do you know that is true?'). We considered specificity in patient responses during different types of therapeutic exchanges, as per Ewbank et al. (2020), to ensure that results were clinically informative (see Fig. 1). It is plausible that, for example, being specific when completing core therapy tasks may be more important for clinical outcome than specificity in setting the agenda. We therefore separately examined patient specificity to therapist elicitations during agenda setting, reviewing homework, setting goals, planning for the future, discussing perceptions of change, completing core therapeutic tasks (included any cognitive or behavioural reattribution, skill teaching, conceptualization, or psychoeducation), assessing risk of suicide, summarizing the session, or setting homework. Full details on the different task types are in Ewbank et al. (2020).

Patient responses were quantified by a supervised deep learning model, which yielded a probability that the response was specific. The model was trained to identify specificity on a dataset of text describing personal experiences (10,000 text entries), written by individuals of varying mental health status. These memories had been previously scored as specific or non-specific by a human rater using the scoring criteria for the Autobiographical Memory Test (Williams and Broadbent, 1986). We used 8000 examples to train our model and 1000 to validate model performance using early stopping (set to 10 epochs). The remaining 1000 examples were used as the held-out test set. We then further validated the model on another separate dataset of held-out text describing personal experiences (1632 entries) written by individuals

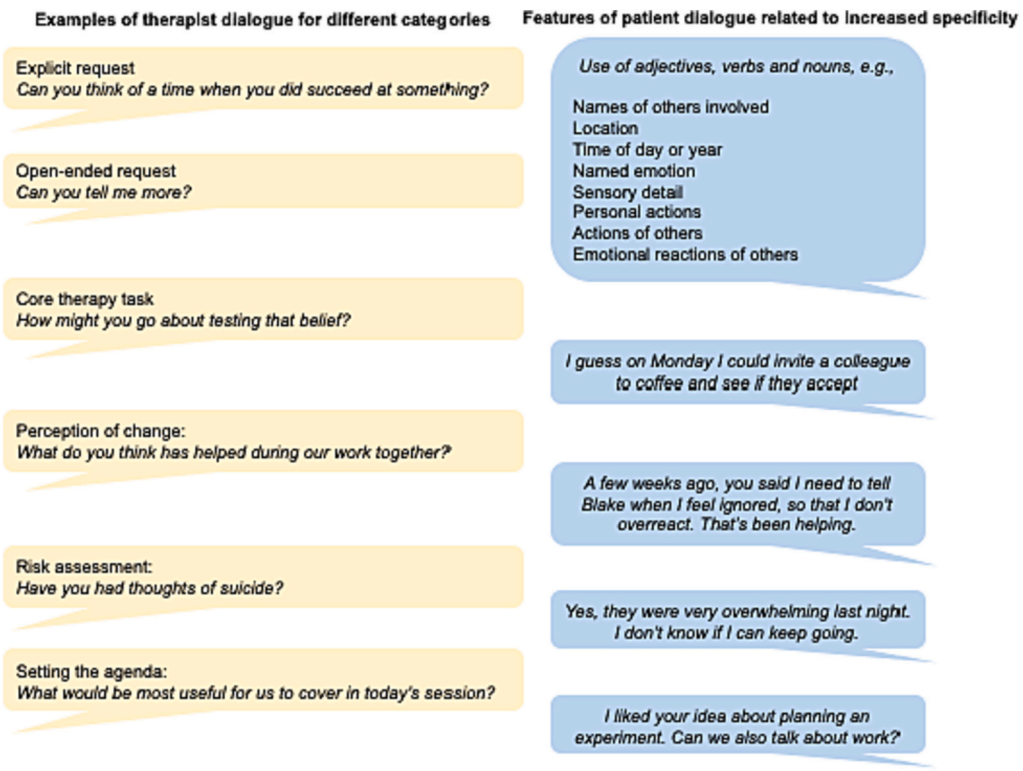

Fig. 1. Examples of therapist and patient dialogue used to score specificity.

with diagnosed depressive disorders and again scored for specificity by humans. Full details on model training are presented in Supplementary Materials. Mean specificity (sum of feature specificity/number of sessions completed) was calculated for each participant, for each therapy task.

2.5. Symptom measures

The Patient Health Questionnaire-9 (PHQ-9; Kroenke et al., 2001) is a reliable and valid index of depressive symptoms, and was used as our primary clinical outcome. Within IAPT, the PHQ-9 is routinely administered at each CBT session, as is the Generalized Anxiety Disorder 7-item scale (GAD-7; Spitzer et al., 2006)), which we pre-registered as a secondary clinical outcome. We had also intended to examine Work and Social Adjustment Scale (Mundt et al., 2002), but were unable to obtain this data. As we used real-world service data, our conceptualisation of clinical outcomes was limited by the available data.

2.6. Data analysis

As pre-registered, hypotheses were analyzed with linear mixed-effects models using lmer in the R computing environment, employing Satterthwaite approximation and restricted maximum likelihood estimation. Last observed PHQ-9 scores were used as end-of-treatment outcome. Random effects of therapist were considered within treatment site. For all models, variance inflation factors were small, $\leq$1.43, as were correlations between the different specificity variables, $rs \leq 0.32$. Each model included total number of sessions completed, mean number of participant words across all sessions (to account for length of discourse), PHQ-9 and GAD-7 scores at the beginning of therapy (to control for baseline individual differences), age-at-referral, gender identity (male, female, undisclosed), use of psychotropic medicine and primary diagnosis (see Table 1). In addition to specificity over the course of treatment, we also, as preregistered, considered specificity in the first session only, to provide an indication of whether adjunctive training programmes to boost patient specificity may be useful if completed prior to beginning CBT. Results did not differ thus we report specificity over the course of treatment. We also considered therapist factors by including the mean numbers of direct requests for specific information

Table 1
Sample characteristics.

Primary diagnosis	Prevalence	Psychotropic medication use	Prevalence
Depressive disorder	39.1 %	Not recorded	1.6 %
Generalized anxiety disorder	23.7 %	Not prescribed	50.8 %
Social anxiety disorder	7.2 %	Prescribed but not taking it	8.3 %
Unspecified anxiety disorder	4.9 %	Taking medication	39.2 %
Obsessive compulsive disorder	4.2 %		
Panic disorder	3.8 %		
Posttraumatic stress disorder	3.7 %		
Hypochondriacal disorder	3.0 %		
Mixed depression and anxiety	2.7 %		
Adjustment disorder	1.7 %		
Specific phobia	1.5 %		
Agoraphobia	0.9 %		
Eating disorder	0.6 %		
Relationship difficulties	0.5 %		
Anger and irritability	0.4 %		
Somatoform disorder	0.2 %		
Chronic pain	0.2 %		
Bipolar disorder	0.1 %		
Alcohol use disorder	0.1 %		
Sexual dysfunction	0.1 %		
Complicated grief	0.1 %		
Mental disorder not otherwise specified	0.9 %		
Not recorded	0.3 %		

and of open-ended requests for information.

3. Results

3.1. Sample and treatment characteristics

Data were from obtained from 65,030 participants (n = 47,308 female, n = 241 unstated) treated within 78 IAPT sites across the UK. Participants completed an average of 5.64 (SD = 3.57, range = 1–39) internet-enabled CBT sessions. As is common for IAPT, primary presenting difficulties were varied, with depressive disorders being most common (see Table 1). Mean age of participants was 34.69 years (SD = 12.35, range 18–94 years). Mean pre-treatment PHQ-9 scores (M = 13.15, SD = 5.98) were in the 'moderately severe' range (Kroenke et al., 2001).

3.2. Treatment outcome

Our primary hypothesis was that the specificity of patients' responses would predict end-of-treatment PHQ-9 scores. The intraclass correlation for the random effect (0.02) was small. Results for all predictors are presented in Table 2. As model predictors were on very different scales, the continuous predictors were rescaled using the standardize package in R, such that the mean for all variables was 0 and the standard deviation was 1. End-of-treatment PHQ-9 scores were significantly predicted by the specificity of patient utterances when discussing perceptions of change, completing core therapeutic tasks, setting the agenda, reviewing homework, and assessing risk of suicide, with very small effect sizes (see Fig. 2).[1] Contrary to predictions, patient specificity during risk checking was associated with higher PHQ-9 scores. Sensitivity analysis with only those with a primary depressive presentation (n = 25,448) yielded similar results (see Supplementary Materials).

3.3. Session attendance

Our secondary hypothesis was that patient specificity would predict the number of CBT sessions completed. We included the same predictors as the primary analysis (ICC = 0.10). As hypothesised, number of completed sessions was significantly predicted by specificity on all tasks other than summarizing the session (see Table 3). Completion of a greater number of CBT sessions was associated with higher specificity when discussing perceptions of change, planning for the future, completing core therapy tasks, and agenda setting. Higher specificity when assessing risk and setting goals was associated with a lower number of completed sessions. An additional analysis predicting attendance at ≥2 sessions (a binary metric which the IAPT services used to index whether a service-user is 'engaged' in treatment) was also completed, as number of completed sessions may be associated other factors (e.g., symptom severity), which yielded similar results (see Supplementary Materials).

4. Discussion

This study utilized machine learning to identify the specificity of patient dialogue during CBT, and explored how specificity related to session attendance and symptom response. Results demonstrated that greater specificity during some, but not all, types of therapeutic interaction were associated with better treatment response. Results suggested lower post-treatment depressive symptoms were associated with higher patient specificity when completing core therapeutic tasks (e.g., cognitive restructuring, evidence gathering), setting the agenda, reviewing homework, and discussing changes observed so far. As hypothesised, greater specificity across a variety of treatment tasks was associated with completion of a greater number of CBT sessions. Together, these findings suggest that greater patient specificity across therapy may be associated with increased session attendance – a well-established treatment barrier (Batterham et al., 2020) and potentially, better treatment outcomes.

There are two key factors which must be considered. First, this is observational data, and causation cannot be inferred. Our primary goal was to identify potential features of successful CBT, to guide future research. Specificity on some therapy tasks (controlling for effect of therapist) was associated with greater symptom improvement, but this could be influenced by other factors (e.g., quality of delivered CBT). Similarly, the use of data from active treatment services ensures that our results are applicable to real-world treatment delivery, but there is greater between-person variability in treatment features (e.g., therapeutic components, such as whether behavioural activation was completed; planned number of sessions) relative to randomised controlled trials, which will have impacted our results. Second, the observed effects are very small. However, small effects should not be automatically dismissed (Funder and Ozer, 2019; Lakens, 2013). The typical effect size for weeks of CBT (versus placebo pill) on depressive symptoms is only 0.30 (Furukawa et al., 2017, 2021). Indeed, a decrease of just one additional point on the PHQ-9 at post-treatment corresponds to an improvement in quality of life (Furukawa et al., 2021). Critically, small effects can have important accumulative effects, when applied by an individual consistently and over time (e.g., across many CBT sessions), or when applied on a single occasion by a high number (i.e., the thousands treated with CBT) of individuals (Funder and Ozer, 2019). For example, small improvements in symptoms observed over a large number of people can reduce costs to service provision. Finally, patient specificity was not purposely targeted in the delivered CBT – the effect size may be larger in therapies with a particular focus on specificity (e.g., rumination-focused CBT). It is quite simple to encourage a patient to be more specific or to edit CBT worksheets to directly instruct the individual to include specific detail, and these actions do not require extensive resources as some other proposed CBT enhancements (e.g., brain stimulation). Thus, although effects were small, it would not be overly difficult or expensive to explore the utility of promoting further specificity within CBT, potentially for some gain.

In accordance with the cognitive model guiding CBT (Beck and Haigh, 2014), results seem to support encouragement of specific detail during therapeutic exchanges. Lower post-treatment depressive symptoms were observed when i) patient specificity was high when reviewing homework and reflecting on what they have learnt so far (e.g., what strategy/new knowledge in particular has been useful), ii) patients demonstrated a greater focus on concrete, specific detail when completing structured CBT activities which challenge maladaptive beliefs, iii) there were specific suggestions from patients when setting session agendas, and iv) more frequent therapist-use of explicit requests for specific information (e.g., 'can you think of one time', 'can you give me an example'). These findings are also consistent with prior assertions that enhancing the context-specific and episodic features of information learnt during therapy will improve patient outcomes (Harvey et al., 2014). Experimental manipulation of these factors in future research could produce useful insight for improving therapist (and potentially patient) practices that may enhance treatment outcomes.

However, not all observed effects were in the expected direction. Interestingly, the largest (although still small) effect size on treatment

[1] In light of the different direction of effects across the specificity variables, as a sensibility check we examined correlations between end of treatment PHQ-9 score and the different specificity variables, which may indicate whether the direction of the individual effects was only reversed due to the multiple analysis. Pearson correlations were in the same direction as the estimates from the mixed effects model.It is also worth noting that the positive association between number of completed sessions and end PHQ-9 scores is likely influenced by the common triage practice of allocating a greater number of sessions to those with more severe pre-treatment symptoms.

Table 2
Fixed effects predicting PHQ-9 scores at end-of-treatment.

Fixed effect	*t*	*df*	*b*	*SEb*
Total number of sessions completed	20.04	60,698.01	0.28	0.01***
Number of patient words	−1.31	56,729.13	−0.02	0.01
Number of therapist specific requests	−3.82	55,256.60	−0.05	0.01***
Number of therapist open-ended requests	0.05	50,481.21	0.01	0.01
Start GAD-7 score	26.90	63,112.01	0.48	0.02***
Start PHQ-9 score	246.31	63,132.14	4.65	0.002***
Primary diagnoses (range; individual diagnoses in Supplement)	−4.11–4.83	50,680.41–63,128.48	−0.51–0.82	0.02–0.82**
Gender: male	−1.63	63,102.72	−0.05	0.03
Gender: non binary or undisclosed	2.30	63,037.83	0.50	0.21*
Age	−14.86	63,121.81	−0.21	0.01***
Medication use: Prescribed but not taking	5.03	63,113.44	0.25	0.05***
Medication use: currently taking	10.05	63,162.68	0.29	0.03***
Medication use: undisclosed	4.47	46,718.17	0.49	0.11***
Specificity of perceptions of change	−6.06	62,372.13	−0.08	0.01***
Specificity of planning for future	−1.08	63,098.56	−0.01	0.01
Specificity of core therapeutic tasks	−4.14	55,416.27	−0.07	0.02***
Specificity of setting goals	0.74	51,207.10	0.01	0.01
Specificity of summarizing session	0.94	63,067.10	0.01	0.01
Specificity of reviewing homework	−3.66	51,253.06	−0.05	0.04***
Specificity of setting homework	1.89	58,180.01	0.03	0.01
Specificity of setting agenda	−4.05	48,155.55	−0.06	0.01***
Specificity of risk checking	21.71	48,155.55	0.29	0.01***

Note. *** $p < .001$. Intercept $b = 12.28$, $SE_b = 0.13$, $t(509.60) = 95.55$, $p < .001$. Continuous predictors (i.e., all other than medication, gender, and primary diagnosis) are on the same scale. For gender, the reference is female. For medication, the reference is no medication prescribed. For Diagnosis Name, reference is no diagnosis recorded. As there were 24 different primary diagnoses, the range of estimates is presented and the full list is in the Supplementary materials.

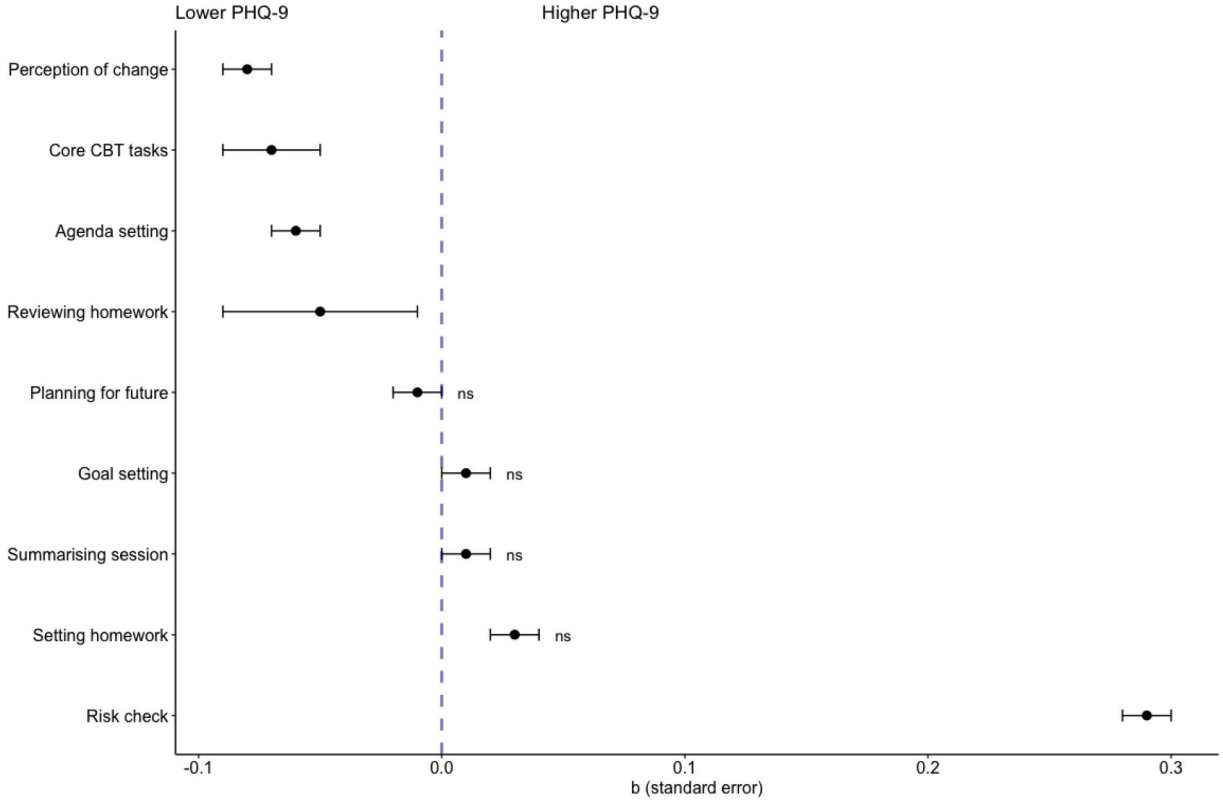

Fig. 2. Association of specificity variables with end of treatment PHQ-9 scores.
Note. ns = non significant. Error bars represent standard error.

outcome was seen for specificity during risk checking, where greater specificity was associated with higher post-treatment symptoms and completion of fewer CBT sessions. Increased specificity during risk assessment may reflect that service-users had more detailed plans of suicide, and/or that suicide safety plans needed to be more detailed (e.g., include location of nearest hospital, name of friend to turn to) due to strong suicidal ideation (Green et al., 2018). In the current sample, high suicidality also triggered on-referral for higher-intensity therapy, potentially contributing to completion of fewer online-CBT sessions. Fewer completed sessions were also associated with greater specificity during goal setting. Specific goal setting is thought to be important in motivating and encouraging patient adherence to treatment, however it

Table 3
Fixed effects predicting the number of CBT sessions completed.

Fixed effect	*t*	*df*	*b*	*SEb*
Number of patient words	26.61	64,933.07	0.40	0.01***
Number of therapist specific requests	−8.91	64,759.76	−0.12	0.02***
Number of therapist open-ended requests	9.54	63,095.10	0.13	0.01***
Start GAD-7 score	9.52	63,758.99	0.17	0.02***
Start PHQ-9 score	5.47	63,686.20	0.10	0.02***
Primary diagnoses (range; individual diagnoses in supplement)	−4.23–11.55	58,994.46–64,200.36	−1.13–1.24	0.10–0.46***
Gender: male	−3.96	63,454.40	−0.12	0.03***
Gender: non binary or undisclosed	0.92	63,069.07	0.19	0.21
Age	7.55	63,734.14	0.10	0.01***
Medication use: prescribed but not taking	−6.05	63,469.57	−0.30	0.05***
Medication use: currently taking	−1.65	63,677.35	−0.05	0.03
Medication use: undisclosed	−3.73	52,570.94	−0.40	0.11***
Specificity of perceptions of change	13.34	64,354.18	0.18	0.01***
Specificity of planning for future	7.06	63,438.72	0.09	0.01***
Specificity of core therapeutic tasks	10.26	64,631.4	0.17	0.02***
Specificity of setting goals	−26.46	64,714.84	−0.36	0.01***
Specificity of summarizing session	0.85	61,398.14	0.02	0.01
Specificity of reviewing homework	17.96	64,444.39	0.25	0.01***
Specificity of setting homework	4.43	64,944.53	0.06	0.01***
Specificity of setting agenda	8.28	63,787.33	0.12	0.01***
Specificity of risk checking	−29.38	64,943.23	−0.40	0.01***

Note. *** $p < .001$; ** $p = .02$. Intercept $t(375.32) = 40.69$, $p < .001$. Continuous predictors (i.e., all other than medication, gender, and primary diagnosis) are on the same scale. For gender, the reference is female. For medication, the reference is no medication prescribed. For Diagnosis Name, reference is no diagnosis recorded. As there were 24 different primary diagnoses, the range of estimates is presented and the full list is in the Supplementary materials.

is possible that patients who identified very specific goals were more likely to reach those goals earlier and therefore attend fewer sessions. Thus, there are a number of potential confounds that need to be evaluated in future research.

Results did suggest that specificity on a range of other treatment tasks had a positive, significant association with completion of CBT sessions. Our exploration of treatment engagement was limited by the available data – notably, individuals could attend sessions but not be highly engaged in the treatment content. Further application of machine learning to treatment transcripts could help to explore this further. Our findings do indicate that exploration of the role of specificity in enhancing treatment engagement might be useful, as at a minimum, our results suggest that patients using a greater degree of concrete, specific detail (e.g., when, where, names, times, locations, sensory detail, and actions) in their dialogue attended a greater number of treatment sessions, above the effects of other established predictors of treatment attendance, such as diagnosis, gender, age, concurrent medication use, and baseline depression and anxiety symptom severity.

The mix of positive and negative effect sizes between therapeutic tasks suggests the role of specificity during treatment is complex. We have previously proposed flexible movement between specific, context-dependent information and generalized summaries may be more important for mental health rather than a sole focus on concrete detail (Hitchcock et al., 2018; Dalgleish and Hitchcock, 2023). Future research may benefit from exploration of flexible application of specificity during therapy (e.g., when evaluating evidence for adaptive versus maladaptive beliefs). Limitations of this study should also be addressed, including that dialogue specificity was not matched to other measures of cognitive concreteness commonly explored in the literature (e.g., rumination), and we did not examine changes in specificity over the course of therapy. Our deep learning model was also validated on text provided by those experiencing depression, and thus our measure of specificity may be biased toward language characteristic of depressive presentations. Clearly, our use of observational data rules out establishing causality, and results should be interpreted in line with this. Further research is needed to determine whether increased specificity casually improves continued treatment attendance or outcomes.

There are a number of novel interventions translated from cognitive science which target either memory specificity (Barry et al., 2019; Raes et al., 2009), flexible movement between general and specific levels of detail (Hitchcock et al., 2018; Moradi et al., 2021), concreteness of information processing (Watkins et al., 2011, 2012), or level of sensory detail in mental imagery (Pile et al., 2021), each of which have demonstrated promising treatment effects as stand-alone interventions on symptoms of common mental health disorders. Our results warrant further studies to explore whether experimental induction of specificity, instructing therapist to encourage specificity, or patient completion of specificity interventions prior to beginning (or perhaps while completing) CBT might help to improve the number of CBT sessions completed, and potentially, depressive outcomes. As these interventions are primarily self-guided, they could be easily and cost-effectively used as adjuncts to regular service delivery. Our results suggest that further evaluation of patient specificity during CBT, and therapist behaviours promoting specificity, may help to emphasise pathways for improving treatment outcomes.

Role of the funding source

This work was supported by a UK Economic and Social Research Council award (ES/R010781/1) to CH, and a UK Medical Research Council Award to TD (SUAG/043). CH was also supported by the Australian Research Council (DE200100043).

CRediT authorship contribution statement

Caitlin Hitchcock: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Julia Funk:** Data curation, Formal analysis, Investigation, Project administration, Resources, Software, Validation, Writing – original draft, Writing – review & editing. **Ronan Cummins:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Resources, Software, Validation, Writing – original draft, Writing – review & editing. **Shivam D. Patel:** Data curation, Investigation, Methodology, Project administration, Writing – review & editing. **Ana Catarino:** Conceptualization, Methodology, Writing – review & editing. **Keisuke Takano:** Conceptualization, Data curation, Investigation, Methodology, Resources, Supervision, Validation, Writing – review & editing. **Tim Dalgleish:** Conceptualization, Methodology, Resources, Supervision,

Writing – review & editing. **Michael Ewbank:** Conceptualization, Investigation, Methodology, Project administration, Resources, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare no conflict of interest.

Acknowledgements

We thank the service-users and therapists who provided the data analyzed in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jad.2024.01.134.

References

Barry, T.J., Sze, W.Y., Raes, F., 2019. A meta-analysis and systematic review of Memory Specificity Training (MeST) in the treatment of emotional disorders. Behav. Res. Ther. 116, 36–51. https://doi.org/10.1016/j.brat.2019.02.001.

Batterham, P.J., Han, J., Mackinnon, A.J., Werner-Seidler, A., Calear, A.L., Wong, Q., Torok, M., Christensen, H., 2020. Factors associated with engagement in online self-help programs among people with suicidal thoughts. J. Affect. Disord. 265, 402–409. https://doi.org/10.1016/j.jad.2020.01.044.

Beck, A.T., Haigh, E.A.P., 2014. Advances in cognitive theory and therapy: the generic cognitive model. Annu. Rev. Clin. Psychol. 10 (1), 1–24. https://doi.org/10.1146/annurev-clinpsy-032813-153734.

Chekroud, A.M., Bondar, J., Delgadillo, J., Doherty, G., Wasil, A., Fokkema, M., Cohen, Z., Belgrave, D., DeRubeis, R., Iniesta, R., Dwyer, D., Choi, K., 2021. The promise of machine learning in predicting treatment outcomes in psychiatry. World Psychiatry 20 (2), 154–170. https://doi.org/10.1002/wps.20882.

Clark, D.M., Canvin, L., Green, J., Layard, R., Pilling, S., Janecka, M., 2018. Transparency about the outcomes of mental health services (IAPT approach): an analysis of public data. Lancet 391 (10121), 679–686. https://doi.org/10.1016/S0140-6736(17)32133-5.

Craske, M.G., Treanor, M., Conway, C.C., Zbozinek, T., Vervliet, B., 2014. Maximizing exposure therapy: an inhibitory learning approach. Behav. Res. Ther. 58, 10–23. https://doi.org/10.1016/j.brat.2014.04.006.

Dalgleish, T., Hitchcock, C., 2023. Transdiagnostic distortions in autobiographical memory recollection. Nat. Rev. Psychol. 2 (3), 166–182. https://doi.org/10.1038/s44159-023-00148-1.

Ewbank, M.P., Cummins, R., Tablan, V., Bateup, S., Catarino, A., Martin, A.J., Blackwell, A.D., 2020. Quantifying the association between psychotherapy content and clinical outcomes using deep learning. JAMA Psychiatry 77 (1), 35–43. https://doi.org/10.1001/jamapsychiatry.2019.2664.

Funder, D.C., Ozer, D.J., 2019. Evaluating effect size in psychological research: sense and nonsense. Adv. Methods Pract. Psychol. Sci. 2 (2), 156–168. https://doi.org/10.1177/2515245919847202.

Furukawa, T.A., Weitz, E.S., Tanaka, S., Hollon, S.D., Hofmann, S.G., Andersson, G., Twisk, J., DeRubeis, R.J., Dimidjian, S., Hegerl, U., Mergl, R., Jarrett, R.B., Vittengl, J.R., Watanabe, N., Cuijpers, P., 2017. Initial severity of depression and efficacy of cognitive–behavioural therapy: individual-participant data meta-analysis of pill-placebo-controlled trials. Br. J. Psychiatry 210 (3), 190–196. https://doi.org/10.1192/bjp.bp.116.187773.

Furukawa, T.A., Levine, S.Z., Buntrock, C., Ebert, D.D., Gilbody, S., Brabyn, S., Kessler, D., Björkelund, C., Eriksson, M., Kleiboer, A., van Straten, A., Riper, H., Montero-Marin, J., Garcia-Campayo, J., Phillips, R., Schneider, J., Cuijpers, P., Karyotaki, E., 2021. How can we estimate QALYs based on PHQ-9 scores? Equipercentile linking analysis of PHQ-9 and EQ-5D. Evid. Based Mental Health 24 (3), 97. https://doi.org/10.1136/ebmental-2020-300240.

Green, J.D., Kearns, J.C., Rosen, R.C., Keane, T.M., Marx, B.P., 2018. Evaluating the effectiveness of safety plans for military veterans: do safety plans tailored to veteran characteristics decrease suicide risk? Integration of Technological Advances in Cognitive-Behavior Therapy 49 (6), 931–938. https://doi.org/10.1016/j.beth.2017.11.005.

Gyani, A., Shafran, R., Layard, R., Clark, D.M., 2013. Enhancing recovery rates: lessons from year one of IAPT. Behav. Res. Ther. 51 (9), 597–606. https://doi.org/10.1016/j.brat.2013.06.004.

Hallford, D.J., Austin, D., Takano, K., Yeow, J., Rusanov, D., Fuller-Tyszkiewicz, M., Raes, F., 2021a. Improving Usual Care Outcomes in Major Depression in Youth by Targeting Memory Specificity: A Randomized Controlled Trial of Adjunct Computerised Memory Specificity Training (c-MeST). https://doi.org/10.31234/osf.io/vmurs.

Hallford, D.J., Rusanov, D., Yeow, J., Barry, T., 2021b. Overgeneral and specific autobiographical memory predict the course of depression: an updated meta-analysis. Psychol. Med. 51 (6), 909–926. https://doi.org/10.1017/s0033291721001343.

Harvey, A.G., Lee, J., Williams, J., Hollon, S.D., Walker, M.P., Thompson, M.A., Smith, R., 2014. Improving outcome of psychosocial treatments by enhancing memory and learning. Perspect. Psychol. Sci. 9 (2), 161–179. https://doi.org/10.1177/1745691614521781.

Hitchcock, C., Gormley, S., Rees, C., Rodrigues, E., Gillard, J., Panesar, I., Wright, I.M., Hammond, E., Watson, P., Werner-Seidler, A., Dalgleish, T., 2018. A randomised controlled trial of memory flexibility training (MemFlex) to enhance memory flexibility and reduce depressive symptomatology in individuals with major depressive disorder. Behav. Res. Ther. 110, 22–30. https://doi.org/10.1016/j.brat.2018.08.008.

Jacobs, R.H., Watkins, E.R., Peters, A.T., Feldhaus, C.G., Barba, A., Carbray, J., Langenecker, S.A., 2016. Targeting ruminative thinking in adolescents at risk for depressive relapse: rumination-focused cognitive behavior therapy in a pilot randomized controlled trial with resting state fMRI. PLoS One 11 (11), e0163952. https://doi.org/10.1371/journal.pone.0163952.

Jing, H.G., Madore, K.P., Schacter, D.L., 2016. Worrying about the future: an episodic specificity induction impacts problem solving, reappraisal, and well-being. J. Exp. Psychol. Gen. 145 (4), 402–418. https://doi.org/10.1037/xge0000142.

Kroenke, K., Spitzer, R.L., Williams, J.B.W., 2001. The PHQ-9. J. Gen. Intern. Med. 16 (9), 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x.

Lakens, D., 2013. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. Front. Psychol. 4 https://doi.org/10.3389/fpsyg.2013.00863.

Manjaly, Z.-M., Iglesias, S., 2020. A computational theory of mindfulness based cognitive therapy from the "Bayesian brain" perspective. *Frontiers*. Psychiatry 11. https://doi.org/10.3389/fpsyt.2020.00404.

Moradi, A.R., Piltan, M., Choobin, M.H., Azadfallah, P., Watson, P., Dalgleish, T., Hitchcock, C., 2021. Proof of concept for the autobiographical memory flexibility (MemFlex) intervention for posttraumatic stress disorder. Clin. Psychol. Sci. 9 (4), 686–698. https://doi.org/10.1177/2167702620982576.

Mundt, J.C., Marks, I.M., Shear, M.K., Greist, J.M., 2002. The work and social adjustment scale: a simple measure of impairment in functioning. Br. J. Psychiatry 180 (5), 461–464. https://doi.org/10.1192/bjp.180.5.461.

Pile, V., Smith, P., Leamy, M., Oliver, A., Bennett, E., Blackwell, S.E., Meiser-Stedman, R., Stringer, D., Dunn, B.D., Holmes, E.A., Lau, J.Y.F., 2021. A feasibility randomised controlled trial of a brief early intervention for adolescent depression that targets emotional mental images and memory specificity (IMAGINE). Behav. Res. Ther. 143, 103876 https://doi.org/10.1016/j.brat.2021.103876.

Raes, F., Williams, J.M.G., Hermans, D., 2009. Reducing cognitive vulnerability to depression: a preliminary investigation of MEmory specificity training (MEST) in inpatients with depressive symptomatology. J. Behav. Ther. Exp. Psychiatry 40 (1), 24–38. https://doi.org/10.1016/j.jbtep.2008.03.001.

Spitzer, R.L., Kroenke, K., Williams, J.B.W., Löwe, B., 2006. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch. Intern. Med. 166 (10), 1092–1097. https://doi.org/10.1001/archinte.166.10.1092.

van der Velden, A.M., Kuyken, W., Wattar, U., Crane, C., Pallesen, K.J., Dahlgaard, J., Fjorback, L.O., Piet, J., 2015. A systematic review of mechanisms of change in mindfulness-based cognitive therapy in the treatment of recurrent major depressive disorder. Clin. Psychol. Rev. 37, 26–39. https://doi.org/10.1016/j.cpr.2015.02.001.

Watkins, E.R., Mullan, E., Wingrove, J., Rimes, K., Steiner, H., Bathurst, N., Eastman, R., Scott, J., 2011. Rumination-focused cognitive–behavioural therapy for residual depression: phase II randomised controlled trial. Br. J. Psychiatry 199 (4), 317–322. https://doi.org/10.1192/bjp.bp.110.090282.

Watkins, E.R., Taylor, R.S., Byng, R., Baeyens, C., Read, R., Pearson, K., Watson, L., 2012. Guided self-help concreteness training as an intervention for major depression in primary care: a phase II randomized controlled trial. Psychol. Med. 42 (7), 1359–1371. https://doi.org/10.1017/S0033291711002480.

Williams, J.M., Broadbent, K., 1986. Autobiographical memory in suicide attempters. J. Abnorm. Psychol. 95 (2), 5. https://doi.org/10.1037/0021-843X.95.2.144.