Changes in positive and negative affect during pharmacological treatment and cognitive therapy for major depressive disorder - A secondary analysis of two randomized controlled trials

Barnaby D Dunn PhD[1], Ramaris E German PhD[2], Gabi Khazanov[3], Colin Xu[3], Steven D Hollon PhD[4], Robert J DeRubeis PhD[3]

[1]Mood Disorders Centre, University of Exeter, UK

[2]Neuroscience and Novel Therapeutics Unit, Emotion and Development Branch, National Institute of Mental Health, Bethesda, Maryland

[3]Department of Psychology, University of Pennsylvania, Philadelphia, USA

[4]Department of Psychology, Vanderbilt University, Nashville, USA

Target Journal: Clinical Psychological Science

Corresponding Author: Dr Barnaby Dunn, Mood Disorders Centre, University of Exeter, UK, telephone: +44 (0) 1392 724680, fax: +44 (0) 1392 724623, e-mail: b.d.dunn@exeter.ac.uk

Author contributions: Dunn had full access to all of the data in the studies and takes responsibility for the integrity of the data and the accuracy of the data analysis. Study concept and design: Dunn, DeRubeis, Hollon. Acquisition, analysis or interpretation of the data: Dunn, German, Kattan, Xu, DeRubeis, Hollon. Drafting of the manuscript: Dunn, German, Xu, Kattan, DeRubeis, Hollon. Statistical analysis: Dunn. Obtained funding: DeRubeis, Hollon.

Funding/Support: The grant was supported by grants MH55877 (DeRubeis) and MM555875 (Hollon) from the National Institute of Health (NIH)

Role of the Funder/Sponsor: The NIH had no role in the design and conduct of the study; collection, management, analysis or interpretation of the data; preparation, review or approval of the manuscript; and decision to submit the manuscript for publication.

Word count: 11866 words (excluding figures and tables)

Abstract

(150 words)

The cardinal symptoms of Major Depressive Disorder (MDD) are heightened depressed mood (negative affectivity; NA) and diminished interest or pleasure (positive affectivity; PA). It is unknown how well treatments for MDD repair each. Two secondary analyses of randomized controlled trials were therefore conducted. In Study One, 180 adult depressed outpatients received sixteen-weeks of antidepressant medication (ADM; n=120) or Cognitive Therapy (CT; n=60). In Study Two, adult depressed outpatients were treated until remission with ADM (n=225) or ADM and CT (n=227). Across trials and treatments, intake disturbances were more marked in PA than NA, there was smaller repair of PA than NA during treatment, and disturbances remained more pronounced for PA than NA post-treatment. Greater change in PA and NA were independently associated with depression symptom change. These findings suggest depression treatments more effectively repair NA than PA and that outcomes may be improved with more effective targeting of the latter.

Major Depressive Disorder (MDD) is a functionally debilitating and chronically recurrent condition that leads to substantial societal and economic costs (Kessler et al., 2003; Üstün, Ayuso-Mateos, Chatterji, Mathers, & Murray, 2004; Üstün, Ayuso-Mateos, Chatterji, Mathers, & Murray, 2004). Current pharmacological and psychological treatments are partly, but not fully, effective at treating MDD. At best only two thirds of patients respond (show at least a 50% drop in symptoms) and only about a third remit (show a complete normalization of symptoms) (Rush et al., 2006; Cuijpers et al, 2014). Functional impairment often lags behind symptomatic improvements (Rush, 2015; Sheehan et al., 2011, 2017). Of those who no longer meet diagnostic criteria for MDD at the end of treatment, over half will relapse within two years even if continued on maintenance antidepressant medication (Rush et al., 2006; Cuijpers, van Straten, Andersson, & van Oppen, 2008; Vittengl, Clark, Dunn, & Jarrett, 2007; Anderson et al., 2008). There is a pressing need to enhance treatment outcomes.

One way forward is to view MDD as a heterogeneous diagnostic construct and to consider it in terms of distinct underlying functional domains that may each require different intervention strategies (see Research Domains Criteria approach; Insel et al., 2010). Similarly, network analysis accounts argue it is beneficial to view the depressed state as an emergent property of patterns of inter-relationships among specific symptoms that become self-reinforcing (Fried et al., 2017; Borsboom, 2017; Hoffmann, Curtis & McNally, 2016). Treatment efficacy may be improved if core nodes in the depression network can be targeted, with different nodes potentially needing different intervention approaches.

An MDD diagnosis requires either a pervasive depressive mood (distress) or a loss of pleasure and interest in all or most activities (anhedonia). These symptoms result from disruptions to two underlying and partly dissociable neurobiological dimensions: up-regulation of a negative valence system that promotes withdrawal from punishing stimuli and drives negative affect (NA); and down-regulation of a positive valence system that guides

approach to rewarding stimuli and shapes positive affect (PA) (Watson, Wiese, Vaidya, & Tellegen, 1999; Gray, 1987; Paulus et al., 2017). This reflects the distinction drawn between the positive and negative valence systems in the Research Domains Criterion (RDoC) approach (Insel et al., 2010).

Client definitions of recovery from depression emphasise the importance of repairing PA as well as NA disturbances (Zimmerman et al., 2006; Demyttenaere et al., 2015), to allow them to function to the best of their ability in valued life domains (Slade, 2010). Network analyses consistently identify depressed mood (increased NA) and anhedonia (reduced PA) as central nodes in the networks maintaining a major depressive episode (Fried et al., 2016; van Borkulu et al., 2015). These depressed mood and anhedonic symptoms (along with low energy and fatigue) are the strongest concurrent predictors of functional impairment in depression (Fried & Nesse, 2014). Both PA and NA disturbances predict a sub-optimal treatment response and a poor future depression prognosis (Spijker, Bijl, De Graaf, & Nolen, 2001; Uher et al., 2012, McMakin et al., 2012).

The above analysis suggests that to effectively treat depression, to improve functioning, and to lead to sustained long-term recovery, treatments should simultaneously target both PA and NA disturbances. However, it has been proposed that existing depression psychological and pharmacological treatments place a greater emphasis on lowering NA than increasing PA (Treadway & Zald, 2011; Dunn, 2012; Dunn & Roberts, 2016). The failure to target PA deficits may contribute to suboptimal treatment outcomes.

This argument is based on a conceptual analysis of what the interventions target. Mainstream pharmacological treatments for depression predominantly target neurotransmitters linked to NA (e.g., selective serotonin or noradrenaline reuptake inhibitors; [SSRIs or SNRIs] and tricyclic anti-depressants [TCAs]) rather than neurotransmitters linked

to PA (e.g., dopamine and opioids; see Shelton, & Tomarken, 2001; Tomarken et al., 2007; Argyrpoulus & Nutt, 2013; Dunlop & Nemeroff, 2007). Similarly, mainstream psychological therapies focus on NA and neglect PA. For example, in cognitive therapy (CT; Beck, Rush Shaw & Emery, 1979) there is an initial emphasis on graded scheduling of positive activities to build a sense of mastery and pleasure. However, what is absent is a detailed theoretical model outlining the psychological mechanisms that drive reduced pleasure when engaging in positive activities and instructions about how to target these mechanisms in therapy (Dunn, in press; Dunn & Roberts, 2016). Subsequent sessions predominantly focus on identifying and challenging negative thoughts and beliefs that maintain a negative view of the self, world and future (the 'negative trial') and drive heightened NA, with little explicit focus on PA. CT represents one of a number of evidence based therapies for depression (including emerging 'third wave' cognitive treatments), all of which show equivalently sub-optimal treatment outcomes (Cuijpers et al., 2013; Hunot et al., 2013) and focus on NA to a greater extent than PA. However, conceptual analyses of this kind are subjective and empirical evaluation is required.

As far as we are aware, there are few if any data that have empirically examined how well current treatments repair PA relative to NA. In three observational studies of treatment seeking samples, greater changes in NA (relative to PA) over time, have been reported (Brown, 2007; Naragon-Gainey et al., 2013; Kring, Persons & Thomas, 2007). Interpretation of these findings is hindered by the heterogeneity of treatments offered, the absence of randomized comparison conditions, the lack of treatment fidelity assessment, and the use of indices of positive and negative temperament that combine data from both affect and personality measures.

A recent systematic review and meta-analysis examined the extent to which psychotherapeutic interventions repair PA versus NA (Boumparis, Karyotaki, Kleiboer,

Hofmann, & Cuijpers, 2016). The mean (Hedges g) effect size across the ten randomized controlled trials identified was 0.41 for PA (95% confidence interval 0.16-0.66) and 0.46 for NA (95% confidence interval 0.10-0.59) (Boumparis et al., 2016), both small to medium effect sizes according to rules of thumb (Cohen, 1988). Taken at face value, this suggests that existing treatments are equally (partially) effective at repairing PA and NA. However, inspection of the studies included in this meta-analysis indicates that this conclusion is premature due to issues of study quality and scope. Of the trials included, none delivered an adequate dose of a mainstream, evidence-based therapy to a diagnosed depressed population and evaluated outcomes using a well validated and clearly described measure of PA and NA (see supplementary materials Table S1). Moreover, this meta-analysis focused solely on psychotherapeutic interventions and did not consider pharmacological treatments.

A parallel literature has examined the extent to which interventions alter extraversion and neuroticism. Given there is some overlap of PA with extraversion and NA with neuroticism, this may indirectly cast light on how well existing treatments repair PA versus NA. Across presenting problems and treatments, there is consistently greater repair of neuroticism than extraversion (see meta-analysis by Roberts, Chow, Luo, Briley & Hill, 2017), perhaps suggesting treatments repair NA better than PA. However, whether this pattern of findings held in major depressive disorder specifically was not assessed in this meta-analysis. This is problematic as PA disturbances are relatively unique to depression (Watson & Naragon-Gainey, 2014) and a different pattern of PA change may be found for depression relative to other conditions as a result.

While personality has some overlap with affect, there are important conceptual differences. Positive emotionality makes up only one component of extraversion (alongside experience seeking and sociability). These facets are only weakly correlated and show distinct (and sometimes diametrically opposed relationships) with psychopathology (Stasik,

Ellickson-Larew & Stanton, 2015; Watson et al., 2015). Similarly, neuroticism consists of multiple facets, not all of which directly overlap with NA and which can have distinct relationships with psychopathology (Schimmack, Oishi, Furr & Funder, 2004). Where different facets have different criterion validities, they can cancel each other out when combined into domain level scores (Paunonen, 2003). Therefore, it is potentially misleading to use global extraversion/neuroticism scores as a proxy for PA/NA respectively.

Overall, this means it is premature to conclude that mainstream depression treatments are better able to repair NA than PA and further examination of this topic is required. To gain traction on this issue, we conducted secondary analyses of existing trials that have collected but have yet not published PA and NA outcomes. In Study One, we analysed self-reported changes in PA and NA from a previously published RCT of treatment for outpatients with moderate to severe MDD in which ADM and CT were each superior to pill-placebo and not different from one another in reducing depression symptoms (CPT2 trial; DeRubeis et al, 2005). In Study Two, we analysed self-reported change in PA and NA in a previously published RCT for outpatients with chronic or recurrent depression in which combined (ADM + CT) treatment was superior to ADM alone in treating depression to remission (CPT3 trial; Hollon et al., 2014). Both of these were *post-hoc* secondary analyses planned after the data were collected.

Some thought is required about how to best measure repair of PA and NA in these analyses. When using symptom-focused measures like the Hamilton Depression Rating Scale (HDRS; Hamilton, 1960), the objective is to eliminate depression symptoms and to ensure individuals fall under some cut-off that indicates remission (ideally as close to zero as possible). Response is typically defined as showing a 50% reduction in depression symptom severity during treatment (Rush et al., 2006). However, for PA and NA it is less clear what counts as sufficient or optimal response and what cut-offs should be used to indicate

remission. A state devoid of any NA and with a total maximum possible of PA is unlikely to be adaptive to the individual.

One approach is to examine where an individual falls in the general population distribution of PA and NA, expressing these as Z-scores (0 indicating a general population average score and a score of +/-1 indicating a score one standard deviation above or below the general population average respectively). An additional advantage of this Z-score approach is that PA and NA are on a common scaling (with the same mean, standard deviation and theoretical maxima and minima), making it possible to directly compare PA and NA repair in analyses. Response can be defined as at least a 50% shift back towards the population mean (for example, moving from two to one SDs below the mean during treatment). Remission can be defined as being no more than half a standard deviation from the general population mean at treatment end (i.e., for PA$\geq$-0.5 and for NA$\leq$0.5), based on claims that half an SD is a useful proxy universal measure of minimum important difference for measures of health-related quality of life (Norman, Sloan & Wrywich, 2003). We will use this Z-score method to evaluate the extent to which PA and NA are repaired in the CPT2 and CPT3 trials. We will also examine if greater repair of PA and NA during treatment is associated with greater depression symptom reduction.

Study One: Secondary analysis of CPT2 trial

Method

Participants and trial design

Four hundred and thirty-seven adult participants were screened and 240 participants meeting criteria for the trial were recruited (59% female; mean age=40 [SD=12]; mean HDRS=23.4 [SD=2.9]) from sites at Vanderbilt University and the University of Pennsylvania. The primary inclusion criterion was currently meeting diagnostic criteria for

MDD with a HDRS score greater than 20 (indicating moderate to severe depression) at both the screening and baseline visits. The vast majority of patients in the recruited sample met criteria for recurrent depression and a sizable minority had chronic depression. Institutional review boards at both sites approved the study and all participants gave written informed consent. Participants were stratified by gender and number of prior depressive episodes and then randomised to 16 weeks of CT (n=60) or ADM (n=120), or 8 weeks of pill-placebo (n=60), with an equal number of participants in each condition at the Vanderbilt and Pennsylvania sites. ADM consisted of up to 50mg daily of paroxetine, augmented by lithium hydrochloride or desipramine hydrochloride if necessary. CT followed established procedures outlined in standard texts to treat depression (Beck, Rush, Shaw & Emery, 1979; J. Beck, 1995) and comorbid personality disorders (Beck & Freeman, 1990).

Patients and prescribing physicians were blind to pill-placebo versus ADM condition for the first 8 weeks of the trial and independent assessors were blind to condition throughout. There was 15% attrition in the CT arm and 16% attrition in the ADM arm across the 16 weeks of treatment. This RCT predated trial registration, so trial registration details cannot be provided. For a full summary of inclusion and exclusion criteria, the CONSORT diagram, sample characteristics, treatment conditions, and fidelity assessments, see DeRubeis et al. (2005). The present secondary analysis focused on changes in NA and PA in the two active arms at 16 weeks and how this related to concurrent change in depression symptoms during treatment (pill-placebo findings are not considered here).

Measures

The Positive and Negative Affect Schedule (PANAS; Watson, Clark, & Tellegen, 1988) was used to assess PA (10 items; e.g., excited; Cronbach's α=.84) and NA (10 items; e.g., distressed; α=.85) over the past week. The PANAS was administered at intake, mid-

treatment (8 weeks), and post-treatment (16 weeks). We concentrated on the intake to post-treatment data. To benchmark individuals' PA and NA scores against general population scores, PA and NA were Z transformed relative to data collected from a US general population sample (328 adults from the Dallas area; Watson & Clark, 1999). This comparison sample had a PA mean of 31.1 (SD=7.5) and a NA mean of 18.0 (SD=7.1). All subsequent analyses were conducted on these Z-scores.

Depression severity was measured using the 17-item HDRS (Hamilton, 1960), a clinician administered interview that is frequently seen as the "gold-standard" outcome measure in depression clinical trials.

Results

Alpha was set at .05 and all tests were two-tailed. Analyses were conducted in the Statistical Package for Social Science version 25 (SPSS; IBM Corp, 2017). Intake data were available for 117/120 of the ADM participants (98%) and 59/60 of the CT participants (98%), with no significant difference in the availability of data between the conditions, $\chi^2<1$. Data were available at sixteen weeks for 102/120 (85%) of those in the ADM arm and 52/60 (87%) of those in the CT arm, again with no significant difference in proportion of missing data between arms, $\chi^2<1$. There were no significant differences in intake PA, intake NA, and HDRS severity between those who had and did not have 16 week PANAS data, independent sample t-test, ps> .326.

Figure 1 plots PA (panel a) and NA (panel b) Z scores for each condition at intake and after 16 weeks of treatment. Clinical improvement is represented by an increase in PA and a decrease in NA. To aid visual comparison of the magnitude of PA and NA deficits, the Y-axis of the PA graph has been reversed.

[INSERT FIGURE 1 ABOUT HERE]

We used multiple imputation to simulate missing values prior to statistical analysis. Guidance recommends that the number of imputations should exceed the percentage of data missing (White, Royston & Wood, 2011), so we used 20 imputation runs given that we had a maximum of 15% of missing data. We included all variables used in subsequent analysis models (intake and sixteen-week PA, NA, and HDRS; group) and also variables that might predict variables with missing data (age, gender, site, condition, number of previous episodes, and first age of onset). Imputation was conducted using a Markov Chain Monte Carlo algorithm (MCMC). All subsequent analyses (run on an intent-to-treat basis) use pooled data across these 20 imputations.

Intake analyses

PA and NA levels were not significantly associated with one another at intake, simple Pearson's correlation r =-.084, p= .272, attenuated correlation =-.099, indicating they are dissociable constructs. HDRS depression severity at intake was significantly positively associated with NA, r=.235, p=.002, and negatively associated with PA at the level of a non-significant trend, r=-.126, p=.099. In all subsequent analyses, we reverse scored PA to make it possible to compare the magnitude of the deviation from general population averages for NA and PA. The magnitude of the NA and PA (reverse scored) associations with HDRS did not significantly differ, Z=1.110, p=.272.

A repeated measures Analysis of Variance (ANOVA) was run, with emotion (PA reverse scored, NA) as the within-subjects factor and condition (CT, ADM) as the between-subjects factor. The ANOVA revealed a main effect of emotion, $F(1,178)=33.643$, $p<.001$, $\eta^2_p=.159$, with PA deficits (Z-mean=-1.981, SD=0.736) being more marked than NA elevations (Z-mean=1.360, SD=1.110) at intake. There was no significant main effect of

condition, $F(1,178)=2.140$, $p=.151$, $\eta^2_p=.001$, and no significant condition by emotion interaction, $F(1,178)=1.027$, $p=.322$, $\eta^2_p=.006$.

At intake, on average 132.4 participants met clinical criteria for both PA (Z-score<-0.5) and NA (Z-score>0.5). Five participants did not meet clinical criteria for either NA or PA, 37.6 met the clinical criterion just for PA, and 5.1 met the clinical criterion just for NA. In total, 170 met the clinical criterion for PA and 137.4 met the clinical criterion for NA, with a significantly greater proportion for PA relative to NA, McNemar $p<.001$.

Sixteen-week analyses

To compare the magnitude of PA relative to NA change brought about by treatment, we calculated a simple difference score between Z-scores at intake and week 16 for NA and PA. Change scores are seen as a valid way to achieve this goal (see Jamieson, 2004). Similar to analyses of intake data, a repeated measure ANOVA was conducted, specifying emotion (reverse scored ΔPA, ΔNA) as the within-subjects factor and condition (CT, ADM) as the between-subjects factor. A significant main effect of emotion emerged, $F(1,178)=5.362$, $p=.032$, $\eta^2_p=.029$. There was no significant main effect of condition, $F(1,178)=3.096$, $p=.105$, $\eta^2_p=.017$, and no interaction between emotion and condition, $F<1$. There was a greater reduction in NA (ΔZ-mean=-1.442, SD=1.310) than there was an increase in PA (ΔZ-mean =1.209, SD=1.237).

We analysed absolute levels of NA and PA at week 16, again running a repeated measures ANOVA specifying emotion (reverse scored PA, NA) as the within-subjects factor and condition (CT, ADM) as the between-subjects factor. Analysis found a significant main effect of emotion, $F(1,178)=70.931$, $p<.001$, $\eta^2_p=.284$. PA deficits still remained, with mean PA levels continuing to fall below general population averages (Z-mean=-0.772, SD=1.97). NA elevations had now normalized and mean NA now fell below general population average

(Z-mean=-0.082, SD=0.987) at 16 weeks. There was no significant main effect of condition, $F(1,178)=1.046$, $p=.375$, $\eta^2_p = .006$, and no significant condition by emotion interaction, $F<1$.

Next, response (>50% Z-score repair) and remission (Z-score$\geq$-0.5 for PA; Z-score$\leq$0.5 for NA) rates were examined, collapsing across treatments given that there were no significant differences between the CT and ADM arm at 16 weeks. For response, on average 82.8 individuals responded for both NA and PA, 56.4 individuals responded for NA only, 14.1 individuals responded for PA only, and 26.7 individuals responded for neither NA nor PA. In total, 139.2 individuals met the NA response criterion and only 96.9 individuals met the PA response criterion, with these proportions significantly differing, McNemar $p<.001$.

For remission, on average 68.9 participants met remission criteria for both NA and PA, 67.0 participants met the remission criterion for NA only, 7.2 participants met the remission criterion for PA only, and 37.0 participants met remission criteria for neither PA nor NA. In total, 135.9 participants met the NA criterion for remission, while only 76.0 participants met the PA criterion for remission, with these proportions significantly differing, McNemar, $p<.001$.

Further, the number of participants showing reliable- and clinically-significant change (Jacobson & Truax, 1991; using criterion c) was computed. On average, 54.0 individuals failed to improve for either NA or PA, 65.5 individuals improved for both PA and NA, 22.5 individuals improved just for PA, and 38.0 individuals improved just for NA. This resulted in

87.9 individuals in total improving on PA and 103.5 individuals improving in total on NA, with this proportion differing at the level of a non-significant trend, McNemar, p=.074[1].

Are changes in NA and PA related to depression outcomes?

To assess if changes in NA and PA related to acute depression outcomes, we computed standardized residual change scores from intake to sixteen-weeks for the HDRS, NA and PA scales. We examined whether NA change and PA change correlated with HDRS change. Greater HDRS reduction was associated with greater PA increase, r=-.440, p<.001, and greater NA decrease, r=.559, p<.001.There was a difference in the magnitude of these associations at the level of a non-significant trend, Z=1.832, p=.067 (first reverse coding PA residual change score). We also simultaneously entered PA change and NA change into a regression model. Greater NA decrease, r_p=.452, p<.001, and greater PA increase, r_p=-.255, p<.010, were independently associated with greater reduction in HDRS.

Additional Analyses

To examine if the response and remission findings would vary using a different general population comparison sample, we reran key analyses compared to a sample of 2527 Scottish adults (1441 female) from Aberdeen with a mean age of 42.15 (*SD*=16.52) (Crawford et al., 2009). The same pattern of findings emerged, with a smaller proportion of participants meeting response and remission criteria for PA relative to NA. It is also possible that the remission findings may be different if using a percentile cut-off to define remission (for example, <75% for NA and >25% for PA), as these do not make any assumptions about an underlying normal distribution. Individual participant data were available for the Crawford

et al (2009) normative sample, allowing us to compute the interquartile range for these norms. Using this revised definition of remission, an identical pattern of findings emerged.

Discussion

A secondary analysis of the CPT2 trial established that PA deficits were more marked than NA elevations at intake; that NA elevations were repaired to a greater extent than PA reductions during treatment; and that PA deficits remain more marked than NA elevations at the end of treatment. A greater proportion of the sample met response and remission criteria at post-treatment for NA than PA. There was also a non-significant trend for a greater number of participants to show reliable and clinically significant change for NA than PA. These findings support the claim that ADM and CT do a better job of repairing NA than PA in depressed individuals. Increase in PA and reductions in NA during acute treatment were both uniquely associated with concurrent reduction in depression symptoms during acute treatment (although the association tended to be more marked for NA than PA).

Study One had a number of limitations that mean these findings should be considered preliminary. The sample size was limited, which means that estimates of differences between conditions and changes in positive versus negative affect may have wide confidence intervals. The dose and duration given of both ADM and CT may not have been sufficient to fully repair PA and NA. Combination treatment (giving individuals both CT and ADM together) may be more effective than either treatment alone but this possibility was not examined. Finally, it is potentially circular logic to examine whether affect change relates to depression change, given that affective symptoms are core components of depression. An alternative approach could be to examine whether affect change relates to measures of functional improvement, as functional measures have no direct content overlap with affect measures.

In addition to these limitations, it is important to replicate findings to have confidence in the conclusions reached, particularly when analyses are post-hoc. An independent replication is required on a trial with a larger sample size, where there is a sufficient dose of treatment given (ideally including a combined treatment arm), and where functional as well as symptom severity outcomes are measured.

Therefore, we next examined if the same findings emerged in the CPT3 trial, where 452 individuals with chronic or recurrent depression were randomised to either ADM alone versus combined ADM and CT (Hollon et al., 2014). The CPT3 trial additionally included the Global Assessment of Functioning (GAF; American Psychiatric Association, 2004) as a measure of functional impairment. Based on the findings of Study One, we hypothesized that intake levels of PA would be more impaired than intake levels of NA; that both treatments would lead to a greater change in NA relative to PA; and that levels of PA would remain more impaired than levels of NA at the end of treatment. We predicted that PA and NA change would each be independently associated with improvement in depression symptoms and functional outcomes. We had no *a priori* predictions about differential effects of ADM alone versus combined treatment on PA versus NA.

Study Two: Secondary analysis of CPT3 trial

Method

Participants and trial design

452 treatment seeking adult outpatients with recurrent or chronic major depressive disorder (MDD) were recruited (59% female; mean age=43.16 [SD=13.10]; mean HDRS=22.08 [SD=4.21]) from outpatient clinics run at the University of Vanderbilt, Nashville, Tennessee; the University of Pennsylvania, Philadelphia; and Rush Medical

Centre, Chicago, Illinois. The primary inclusion criteria were meeting diagnostic criterion for recurrent or chronic (episode duration $\geq$ 2 years) depression and a 17-item HDRS score $\geq$ 14. Institutional review boards at both sites approved the study and all participants gave written informed consent.

Participants were randomly assigned in a 1:1 ratio to antidepressant medication treatment alone (ADM group; n=225) or combined ADM and CT (COM group; n=227), with allocation stratified by sex, marital status, symptom severity, history of recurrence, chronicity, and comorbid axis II disorders. In the acute phase of treatment, participants were treated until they met criterion for remission (four consecutive weeks of minimal symptoms; assessed at least monthly during the trial by interviewers blind to condition). Median time to remission was 39 weeks in the ADM arm and 31 weeks in the combined arm. The maximum treatment offered was 42 months in total. Pharmacotherapy followed a principle-based algorithm, aiming to deliver personalized antidepressant therapy using best clinical practices. The algorithm allowed for up to four different classes of ADM (SSRIs, SNIRs, TCAs, and MAOIs) and the use of any of the augmenting agents commonly used in clinical practice. The first line treatment was typically an SSRI or SNRI. Cognitive therapy followed the treatment manual for CT for depression (Beck et al., 1979), augmented as necessary for patients with comorbid personality disorders (Beck & Freeman, 1990). For a full summary of inclusion and exclusion criteria, sample characteristics, treatment conditions, fidelity assessments, trial registration and the trial CONSORT diagram, see Hollon et al (2014).

Measurements

The Mood and Anxiety Symptom Questionnaire (MASQ; Watson & Clark, 1991) measured affect change during treatment. Participants were asked to judge for each of 90 items how much they have felt the way described over the past week, ranging from 1 (not at

all) to 5 (extremely). The general distress (GD) subscale serves as a measure of NA and the anhedonic depression subscale serves as a measure of PA. The MASQ was administered at each assessment point during the trial (at least monthly). We focused on the MASQ taken at the end of acute treatment (remission for those who remitted and termination for those who did not remit within 18 months)[2].

The original factor structure of the MASQ proposed by Clark and Watson (1991) has not been replicated in recent studies, with many of the negatively keyed 'loss of interest' items originally included in the anhedonia subscale loading more clearly on general distress (Bedford, 1997; Keogh & Reidy, 2000; Kendall et al., 2016). Therefore, we used the revised factor structure proposed by Keogh and Reidy (2000), in which the anhedonia scale consists solely of positively keyed 'high positive affect' items. Reliability in the present sample was high (intake: GD α=.937; AD α=.938). As in Study One, AD and GD scores were Z-transformed relative to a general population sample (534 UK undergraduate students; Keogh & Reidy, 2000). This sample had a GD mean of 40.92 (SD=16.26) and an AD mean of 66.02 (SD=17.99). All subsequent analyses were conducted on these Z-scores.

As in Study One, the 17-item HDRS (Hamilton, 1960) was administered to assess depression severity. Additionally, the Global Assessment of Functioning (GAF; American Psychiatric Association, 2004) was used to assess day-to-day functioning (in psychological, social and occupational functioning domains) over the past week.

Results

Alpha was set at .05, all tests were two-tailed, and analyses were conducted in SPSS version 25 except where otherwise stated. Intake MASQ data were available for 215/225 participants [96%] in the ADM arm and 221/225 participants [98%] in the combined arm, with the proportion of complete data not differing between arms, $\chi^2<1$. There was MASQ

data for 210/225 participants [93%] in the ADM arm and 216/227 participants [95%] in the combined arm at the end of acute treatment assessment, again with the proportion of complete data not differing between arms, $\chi^2<1$. There were no significant differences in intake depression severity, intake MASQ AD, or intake MASQ GD between those included in the intake analyses with and without complete MASQ data at the acute end follow-up, ps>.094. Figure 2 plots GD and AD scores for participants in each arm at intake and the end of acute treatment.

[INSERT FIGURE 2 ABOUT HERE]

As in Study One, multiple imputation (implemented via a MCMC algorithm, entering all variables used in the analyses and also age, gender, site, condition, number of previous episodes and intake depression severity) was used to simulate missing data and the data was analysed on an intent-to-treat basis. This time 10 imputation runs were used, as the maximum level of missingness was 7%. All subsequent analyses average across these ten imputation runs.

Intake Analyses

MASQ-GD and MASQ-AD were significantly positively correlated, r=.423, p<.001, correction for attenuation r =.451 (a medium effect; Cohen, 1988). This indicates that AD and GD are less clearly orthogonal than PANAS PA and NA used in Study One, but nevertheless are still dissociable. Greater AD, r=.219, p<.001, and GD, r=.382, p<.001, were significantly associated with greater HDRS at intake, with the magnitude of this association being significantly greater for GD than AD, Z=3.395, p<.001. When both were entered into the same regression, greater levels of GD, r_p=.327, p<.001, but not AD, r_p=.069, p=.151, were uniquely associated with greater levels of depression.

Mean functioning score at intake was 56.003 (SD=7.370) (moderate difficulty). Lower functioning was significantly related to greater intake AD, r=-.215, p<.001, and GD, r=-.243, p<.001, with no difference in the magnitude of the associations, Z<1. When both were entered into the same regression, greater levels of AD, r_p=-.128, p=.009, and GD, r_p=-.172, p<.001, were each uniquely associated with lower levels of functioning.

A repeated measure ANOVA was run on the Z-transformed intake scores, with MASQ-factor (AD, GD) as the within-subjects factor and condition (ADM, combined) as the between-subjects factor. This found a main effect of MASQ-factor, F(1,450)=72.292, p<.001, η^2_p = .138. Replicating the pattern of findings from Study One, AD symptoms (Z-mean=1.816, SD=0.731) were more marked than GD symptoms (Z-mean=1.406, SD=1.092) at intake. There was no significant main, F(1,450)=2.357, p=.134, η^2_p = .005, or interactive, F<1, effect of condition.

We also determined the proportion of the sample showing clinical levels of AD and GD at intake (Z-scores>0.5). On average, 346.9 participants met clinical criteria for AD and GD, 80.3 participants met the clinical criterion just for AD, 6.5 participants met the clinical criterion just for GD, and 18.3 participants met clinical criteria for neither AD nor GD. In total, 427.2 participants met clinical criteria for AD and 353.4 participants met clinical criteria for GD at intake, with a significantly greater proportion meeting criteria for AD than GD, McNemar, p<.001.

End of acute treatment analysis

A repeated measure ANOVA examined simple change in AD and GD during treatment, specifying MASQ-factor (AD, GD change) as the within-subjects factor and condition (ADM, COM) as the between-subjects factor. A significant main effect of MASQ-factor emerged, F(1,450)=26.556, p<.001, η^2_p = .056. Mirroring findings from Study One,

there was a greater repair of GD (ΔZ-mean=-1.350, SD=1.299) than AD (ΔZ-mean=-1.057, SD=1.228). There was a greater repair of overall symptoms in the combined arm (relative to the ADM only arm) at the level of a weak, non-significant trend, $F(1,450)=2.894$, $p=.097$, $\eta^2_p = .006$. There was no significant interaction between condition and MASQ-factor, $F<1$.

We also analysed absolute levels of AD and GD at acute treatment end. Repeated measures ANOVA found a significant main effect of MASQ-factor, $F(1,450)=201.400$, $p<.001$, $\eta^2_p=.309$. There were no significant main or interactive effects of condition, $Fs<1$. Again replicating Study One, AD symptoms (Z-mean=0.759, SD=1.213) were more marked than GD symptoms (Z-mean=0.056, SD=1.146) at the end of treatment.

Next we collapsed across conditions and looked at the proportion of individuals meeting response (>50% Z-score change), remission (Z-score$\leq$0.5), and reliable and clinically-significant change criteria. For response, on average 189.5 participants met criteria for both GD and AD, 109.0 participants met criterion for just GD, 41.7 participants met criterion just for AD, and 111.8 participants failed to meet either criteria. In total, 298.5 participants met the response criterion for GD and 231.2 participants met the response criterion for AD, with the proportion being greater for GD than AD, McNemar $p<.001$.

For remission, 183.9 participants met criteria for both GD and AD, 7.5 participants met criterion just for AD, 151.9 participants met criterion just for GD, and 108.7 participants met neither remission criteria. In total, 335.8 participants met the remission criterion for GD and 191.4 participants met the remission criterion for AD, with the proportion being greater for GD than AD, McNemar $p<.001$.

For reliable and clinically significant change (Jacobson & Truax, 1991; criterion c), on average 200 participants met criteria for both AD and GD, 121.9 participants failed to meet criteria for either AD or GD, 26.9 participants met criteria just for AD, and 103.2

participants met criteria just for GD. In total, 226.9 individuals met criteria for AD and 303.2 individuals met criteria for GD, with the proportion being greater for GD than AD, McNemar $p<.001$[3].

Are changes in AD and GD related to depression and functional outcomes?

As in Study One, we computed standardized residual change scores for the HDRS, AD and GD scales (in this case from intake to the end of acute treatment) and examined the associations between these change scores. Greater repair in HDRS depression severity was significantly associated with greater AD reduction, Pearson's $r=.454$, $p<.001$, and greater GD reduction, $r=.544$, $p<.001$. The correlation with depression severity was significantly stronger for GD than AD, $Z=2.625$, $p=.004$. Both AD residual change, $r_p=.180$, $p=.001$, and GD residual change, $r_p=.379$, $p<.001$, continued to predict depression change when entered simultaneously into a regression model.

We also examined the associations between residual change in affective and functional outcomes. Greater reduction in AD, $r=-.506$, $p<.001$, and GD, $r=-.489$, $p<.001$, were both associated with a greater increase in functioning. There was no difference in the strength of these associations, $Z<1$. When both were entered in the same regression, greater reductions in AD, $r_p=-.299$, $p<.001$, and GD, $r_p=-.262$, $p<.001$, each uniquely predicted greater increases in functioning.

Additional Analyses

Due to the fact that participants were treated until remission, acute treatment end varied between participants. We repeated key analyses when looking at the MASQ assessment point closest to six months after acute treatment started (excluding cases where that assessment point was not within plus or minus 30 days of six months). We chose six

months as this is often a standard treatment dose in depression psychotherapy trials (e.g. Wiles et al.; 2013; Richards et al., 2016). An identical pattern of findings emerged. Unlike Study One, we were not able to repeat the response and remission analysis using a different normative data set, as we are not aware of other published normative data on the MASQ factor structure put forward by Keogh and Reidy (2000). Similarly, we could not replicate remissions findings using percentile cut-offs as we did not have individual item data from Keogh and Reidy (2000) to allow us to calculate the interquartile range in this sample.

The MASQ was administered over repeated occasions in the CPT3 trial, making it possible to examine if the slope of change over time differed for GD versus AD. A hierarchical linear model was run using the mixed command in Stata (StataCorp, 2015), with affect (Z-transformed AD, GD), nested within time (each point MASQ was administered), nested within individual participant. We modelled random slopes for time and participant, only including MASQ data collected during the intake and acute treatment phase of the trial. Affect was binary coded (0 for AD; 1 for GD). We person-mean centered the time variable as is generally recommended in longitudinal models of this kind (Wang & Maxwell, 2015), particularly where there is significant heterogeneity in the number and timings of assessments between participants (Blozis & Cho, 2008). Preliminary analyses found that model fit (based on the Akaike and Bayesian information criteria) was best when time was log transformed to reflect the fact that change in MASQ symptoms was more marked earlier than later in treatment. Therefore, we report results from this log transformed model. Data were available for 442 participants, with 2,696 observations for each of MASQ AD and GD (5,392 in total). There was a significant main effect of time, β=-.193(SE=.012), Z=-16.21, p<.001, and affect, β=.083 (SE=.026), Z=-3.15, p=.002, which was qualified by a significant time by affect interaction, β=-.037(SE=.011), Z=-3.49, p<.001. There was a reduction in symptoms over time, which was more marked for GD than AD.

Discussion

Study Two fully replicated the findings of Study One in a different sample. AD deficits were more marked than GD deficits at intake; AD deficits changed to a lesser degree than GD deficits during treatment; and as a result post-treatment AD deficits were more marked than GD deficits. Improvement in AD and GD each were uniquely associated with concurrent improvement in depression symptoms and functioning outcomes.

General Discussion

We examined the extent to which current mainstream MDD treatments repair elevations in NA and deficits in PA across two different randomised controlled trials. The CPT2 trial compared sixteen weeks of CT and ADM for moderate to severe depression (DeRubeis et al., 2005), using the PANAS as a measure of affect. The CPT3 trial compared ADM to combined ADM and CT (treating to remission) for chronic or recurrent depression (Hollon et al., 2014), using the MASQ to measure affect.

In both trials, PA deficits were more marked than NA deficits at intake, relative to comparison sample averages. This is consistent with the view that disturbances to the PA system are particularly prominent in MDD and therefore should be an explicit intervention target (Dunn, 2012; Argyropoulos & Nutt, 2013; Treadway & Zald, 2011). PA and NA improved during treatment in both trials, with no difference between treatment arms. However, the magnitude of PA repair was significantly smaller than the magnitude of NA repair. In Study Two, hierarchical liner modelling analyses showed a slower repair of AD relative to GD over time. This is despite the fact that in both studies PA was more disturbed than NA at intake, meaning regression to the mean should have favoured greater change in PA than NA. At the end of acute treatment PA disturbances remained significantly more pronounced relative to NA disturbances in both studies. As a result, PA levels remained

beneath general population average levels at these time points. That is, PA improved but never fully normalised. In contrast, NA levels largely normalized in both trials (with average NA scores now falling close to general population averages). In both trials, a greater proportion of participants met response (50% reduction in symptoms) and remission (falling within half a standard deviation of the general population mean) criterion for NA than PA. A greater proportion of participants also showed reliable and clinically significant change for NA than PA (albeit the Study One findings were only a non-significant trend in that direction). Overall, this suggests that neither ADM, CT nor combined treatment were satisfactorily effective in repairing PA deficits in depression.

The present results are the first to delineate the absolute levels of PA and NA disturbance in depression (relative to general population averages), showing that PA deficits are more marked than NA deficits at both intake and post-treatment assessments. The treatment outcome findings parallel results reported in Roberts et al (2017) that treatments are more effective at repairing neuroticism than extraversion, extending them into the affective domain and focusing specifically on mainstream treatments of major depression. The results deviate from the meta-analytic results of Boumparis et al (2016), who found that depression interventions produced comparably small to medium effects on both PA and NA. However, none of the studies included in Boumparis et al (2016) were of current mainstream treatments delivered with an optimal dose and format to a diagnosed depressed population. Therefore, the conclusions in Boumparis et al (2016) that depression therapies are similarly ineffective at repairing PA and NA should now be revised on the basis of the current results.

In both trials greater repair of PA and NA were each significantly associated with greater concurrent repair of depression. While depression change was more strongly related to NA than PA in both trails (a trend significant difference in Study One and a fully significant difference in Study Two), when both PA and NA change were entered into the

same analyses each were independently associated with depression repair. Causal conclusions cannot be drawn from association data of this kind (i.e. change in PA and NA is concurrent with change in depression symptoms, so temporal precedence is not established). Therefore, future studies should conduct mediation or cross-lagged analyses, in particular examining if early change in PA or NA predicts subsequent change in depression to more robustly test this hypothesis.

Another potential criticism of these association analyses is that they are based on circular logic, given that anhedonia (PA) and depressed mood (NA) form central components of the depression construct. One way to evaluate this critique is to consider the overlap of individual items of the depression scale (17-item HDRS) with PA and NA. Two items in the HDRS directly measure NA (item one indexing depressed mood and item ten measuring psychic anxiety), while one item indirectly measures PA (item seven on work and activities mentions loss of interest in the scoring key). Therefore there is moderate but not high item content overlap. Another way to evaluate this issue is to examine the strength of the association of individual depression symptoms with PA and NA. In the present samples the associations between affect scores and individual depression items were generally non-significant and of small magnitude (see supplementary materials Table S2). While this differs from the at least moderate strength relationships reported in some previous studies (e.g., Watson, Clark & Carey, 1988), this does not suggest a high degree of overlap in the CPT2 or CPT3 datasets. Moreover, in Study Two it is encouraging that change in PA and NA both independently predicted functional improvement, as this outcome measure has no direct overlap with affect. Therefore, in our view the present association results are not substantially undermined by problems of circular logic.

The key implication of these findings is that better outcomes may result if treatments can target PA as effectively as they do NA, given that anhedonia symptoms predict future

prognosis, functional impairments and suicide completion rates (Spijker, Bijl, De Graaf, & Nolen, 2001; Uher et al., 2012, McMakin et al., 2012; Geschwind et al., 2011; Fried & Nesse, 2014; Fawcett, Scheftner, Fogg, Clark & Young, 1990). Moreover, studies suggest that in the eyes of patients repair of PA is at least as important as reductions in NA in recovery from depression (Zimmerman et al., 2006; Demyttenaere et al., 2015). This perspective resonates with a broader recovery literature arguing mental health treatments should place a greater emphasis on patient defined recovery goals relating to positive functioning (Slade, 2010) and that a complete state of positive mental health involves both an alleviation of symptoms of mental illness and the cultivation of wellbeing (Provencher & Keyes, 2011). The fact that existing mainstream treatments fail to normalise PA to general population levels therefore indicates there is significant room for improvement.

It is conceivable that PA, relative to NA, is inherently less amenable to change (Brown, 2007; Naragon-Gainey, Gallagher, & Brown, 2013) and therefore that treatment efficacy is already at ceiling. However, there are promising treatment advances indicating that improving PA outcomes may be achievable. There is preliminary evidence that drugs that act primarily on the dopamine system (e.g., buproprion and ketamine) can be effective in alleviating anhedonia in mood disorders (Jamerson, Krishnan, Roberts, Krishen, & Modell, 2003; Tomarken, Dichter, Freid, Addington, & Shelton, 2004; Lally et al., 2015). Adapted forms of psychotherapy targeting PA and broader wellbeing are emerging, including Positive CBT (Geschwind, Arntz, Bannink & Peeters, 2019), Positive Affect Treatment (Craske et al., in press), Wellbeing Therapy (Ruini & Fava, 2012), Augmented Depression Therapy (Dunn et al., 2019), and adaptations of Positive Psychology Interventions (e.g., Chaves, Lopez-Gomez, Hervas & Vazquez, 2017). These novel treatments are increasingly informed by a better understanding of the underlying psychological mechanisms driving PA deficits, including elevated use of dampening appraisals (e.g., thinking ‘this is too good to last’) and

reduced experiential processing (e.g., Burr et al., 2017; Dunn et al., 2018; Gadeikis et al., 2017), opening up new avenues for intervention.

Given that a majority of depressed clients presents with impairments in both affective systems, optimal depression outcomes are likely to emerge from universal treatment protocols that are able to simultaneously target both PA and NA (rather than a proliferation of separate treatments for NA and PA).These universal treatment protocols should be flexible enough to tailor the relative focus on PA and NA based on the presentation of each individual client.

The present findings highlight the explanatory benefits of fractionating depression into underlying dimensions or symptom clusters, as recommended both by the Research Domains Criteria approach (Insel et al., 2010) and network models of psychopathology (Fried et al., 2017; Borsboom, 2017; Hoffmann, Curtis & McNally, 2016). This perspective also fits with recent recommendations that the field should move to a 'process-based therapy' perspective, whereby treatments should aim to target theoretically derived and empirically validated core processes that maintain key symptoms/dimensions using empirically tested treatment procedures (see Hofmann & Hayes, 2018).

A concern voiced by patients regarding antidepressants that target serotonin is that such drugs numb their positive emotion experience, thereby exacerbating anhedonia (Price, Cole & Goodwin, 2009). The present findings are not consistent with this viewpoint. In both trials, ADM treatment did improve levels of PA from pre- to post- treatment assessment at the group level, but failed to normalize them to general population average levels. Very few participants showed a clinically significant deterioration in PA when using antidepressants in either trial. It is plausible that patients misattribute blunted levels of PA as a side effect of

ADM treatment rather than as a residual feature of their condition that persists after only partially successful treatment.

The use of a benchmarking approach (expressing measures in Z-score units relative to comparison sample distributions) is novel in that it makes it possible to test degree of normalization of the outcome variable. This benchmarking approach could be utilized when analysing other RCT outcome data where adequate normative data are available.

That an identical pattern of findings emerged across two different trials and using different measures of positive and negative affectivity (and in Study One across different comparison samples) suggests this is a robust, replicable result that is unlikely to be an artifact of the outcome measures or comparison sample chosen.

There are various limitations of the present analyses. First, CT reflects only one example of an evidence-based psychological treatment for depression and we cannot rule out that this is a class effect. It is conceivable that other psychological therapies (e.g. Behavioural Activation; Martell, Dimidjian & Herman-Dunn, 2010) may be more successful at repairing PA. However, given that the initial positive activity scheduling of CT has substantial overlap with Behavioural Activation, this seems unlikely. Second, the criteria used for remission (falling within half a standard deviation of general population averages), despite having a precedent in the broader literature (Norman et al., 2003), is equally arbitrary as any other choice of cut-off point. It is reassuring in this regard that an identical pattern of findings emerged if using percentile rather than standard deviation definitions of remission in Study One. Third, the 50% Z-score response criterion could be seen as more stringent for PA than NA, given that PA disturbances were more marked at intake. However, this mirrors the 50% response criterion routinely used to determine depression response (Rush et al., 2006). Further supporting the use of percentage change criterion, there is evidence to suggest that

where baseline impairments are more marked depressed participants report needing to change a greater amount to feel they have reliably improved (e.g., Button et al, 2015). Fourth, the validity of the present findings depends on the underlying tools used to measure PA and NA being robust and replicable. While the PANAS factor structure has been extensively validated, the optimal MASQ factor structure remains open to debate. However, that results were identical for Study One using the PANAS and Study Two using the MASQ is encouraging in this regard. Fifth, both the PANAS and MASQ are measures of dispositional positive and negative mood rather than an index of positive and negative reactivity to stimuli. A different pattern of results may emerge if looking at reactivity, for example using the Snaith Hamilton Pleasure Scale (Snaith et al., 1995) as a measure of positive reactivity. Finally, PA can be fractionated into motivational ('wanting'), consummatory ('liking'), and cognitive ('learning') elements (Berridge & Kringlebach, 2008; Treadway & Zald, 2011) and here we have focused on the consummatory aspect only. Future studies should measure how treatments repair these various components of PA.

In summary, individuals with major depressive disorder show more marked abnormalities in PA than NA, and existing depression treatment like antidepressants and cognitive therapy repair NA more effectively than PA. As a result, depressed individuals are left with residual deficits in PA post-treatment. There is potential to improve depression treatment outcomes by targeting PA more systematically in pharmacological and psychological treatment approaches.

Acknowledgements

The CPT2 trial was supported by grants MH55877 (DeRubeis) and MM555875 (Hollon) from the National Institute of Mental Health. GlaxoSmithKline provided medications and pill-placebos for the trial. The CPT3 trial was supported by grants MH60713 and MH01697 (K02) (Hollon), MH60998 (DeRubeis), and MH060768 (Fawcett and Zajecka) from the National Institute of Mental Health. Wyeth Pharmaceuticals provided venlafaxine, and Pfizer Inc provided sertraline for the trial. We declare no conflicts of interest. The authors would like to thank Andrew J. Tomarken for his comments on an earlier version of the manuscript, and Alice Price, Sophie Dunn and Merle Kock for help with referencing and proof reading the manuscript.

Footnotes

1- We also examined reliable and clinically significant deterioration for PA and NA. Very few people deteriorated (on average 4.6 participants for PA only, 3.5 participants for NA only, and 1.2 participants for both PA and NA), with no significant difference between NA and PA, McNemar, ns.
2- The CPT3 trial did not include the PANAS as an additional outcome measure, precluding a direct replication of the Study One results.
3- As in Study One, we examined reliable and clinically significant deterioration for AD and GD. Very few people deteriorated (on average 4.7 participants for AD only, 3.5 participants for GD only, and 4.1 participants for both AD and GD), with no significant difference between AD and GD, McNemar, ns.

References

American Psychiatric Association. (2000). *Diagnostic and Statistical Manual of Mental Disorders* (4th ed., text rev.). Washington, DC: Author.

Anderson, I. M., Ferrier, I. N., Baldwin, R. C., Cowen, P. J., Howard, L., Lewis, G., Matthews, K., McAllister-Williams, R.H., Peveler, R.C., Scott, J.& Tylee, A. (2008). Evidence-based guidelines for treating depressive disorders with antidepressants: a revision of the 2000 British Association for Psychopharmacology guidelines. *Journal of Psychopharmacology*, *22*, 343-396.

Argyropoulos, S. V., & Nutt, D. J. (2013). Anhedonia revisited: is there a role for dopamine-targeting drugs for depression? *Journal of Psychopharmacology, 27*, 869-877.

Bannink, F. (2012). *Practicing Positive CBT: From Reducing Distress to Building Success*. John Wiley & Sons

Beck, A. T., Rush, A.J., Shaw, B.F. & Emery, G. (Ed.). (1979). *Cognitive Therapy of Depression*. Guilford press.

Beck, A.T., Freeman, A. (1990). *Cognitive Therapy of Personality Disorders*. New York, NY: Guilford Press

Beck, J.S. (1995) *Cognitive Therapy: Basics and Beyond*. New York, NY: Guilford Press.

Bedford, A. (1997). On Clark-Watson's tripartite model of anxiety and depression. *Psychological Reports*, *80*, 125-126.

Berridge, K. C., & Kringlebach, M. L. (2008). Affective neuroscience of pleasure: Reward in humans and animals. *Psychopharmacology*, 199, 457-480.

Borsboom, D. (2017) A network theory of mental disorders. *World Psychiatry, 16*, 5-13.

Boumparis, N., Karyotaki, E., Kleiboer, A., Hofmann, S. G., & Cuijpers, P. (2016). The effect of psychotherapeutic interventions on positive and negative affect in depression: A systematic review and meta-analysis. *Journal of Affective Disorders, 202*, 153-162.

Blozis, S. A., & Cho, Y. (2008). Coding and centering of time in latent curve models in the presence of interindividual time heterogeneity. *Structural Equation Modelling: A Multidisciplinary Journal, 15*, 413-433.

Brown, T. A. (2007). Temporal course and structural relationships among dimensions of temperament and DSM-IV anxiety and mood disorder constructs. *Journal of Abnormal Psychology, 116*, 313-328.

Burr, L.A., Javaid., M., Jell, G., Werner-Seidler, A. (2017). Turning lemonade into lemons: Dampening appraisals reduce positive affect and increase negative affect during positive activity scheduling. *Behaviour Research and Therapy*, *91*, 91-101.

Chaves, C., Lopez-Gomez, I., Hervas, G., & Vazquez, C. (2017). A comparative study on the efficacy of a positive psychology intervention and a cognitive behavioural therapy for clinical depression. *Cognitive Therapy and Research, 41*, 417-433.

Cipriani, A., Furukawa, T. A., Salanti, G., Chaimani, A., Atkinson, L. Z., Ogawa, Y., Leucht, S., Ruhe, H.G., Turner, E.H., Higgins, J.P. & Egger, M. (2018). Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. *The Lancet*, *391*, 1357-1366.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioural Sciences.* New York: Academic Press.

Cole, D. A., Martin, N. C., & Steiger, J. H. (2005). Empirical and conceptual problems with longitudinal trait–state models: Introducing a trait–state–occasion model. *Psychological Methods, 10*, 3–20.

Craske, M. et al (in press). Positive affect treatment for depression and anxiety. A randomized controlled trial for a core feature of anhedonia. *Journal of Consulting and Clinical Psychology.*

Crawford, J. R., Garthwaite, P. H., Lawrie, C. J., Henry, J. D., MacDonald, M. A., Sutherland, J., & Sinha, P. (2009). A convenient method of obtaining percentile norms and accompanying interval estimates for self-report mood scales (DASS, DASS-21, HADS, PANAS, and sAD). *British Journal of Clinical Psychology*, *48*, 163-180.

Cuijpers, P., van Straten, A., Andersson, G., & van Oppen, P. (2008). Psychotherapy for depression in adults: a meta-analysis of comparative outcome studies. *Journal of Consulting and Clinical Psychology*, *76*, 909-922.

Cuijpers, P., Berking, M., Andersson, G., Quigley, L., Kleiboer, A., & Dobson, K. S. (2013). A meta-analysis of cognitive-behavioural therapy for adult depression, alone and in comparison with other treatments. *The Canadian Journal of Psychiatry, 58*, 376-385.

Cuijpers, P., Karyotaki, E., Weitz, E., Andersson, G., Hollon, S. D., & van Straten, A. (2014). The effects of psychotherapies for major depression in adults on remission, recovery and improvement: a meta-analysis. *Journal of Affective Disorders*, *159*, 118-126.

Demyttenaere, K., Donneau, A. F., Albert, A., Ansseau, M., Constant, E., & Van Heeringen, K. (2015). What is important in being cured from depression? Discordance between physicians and patients (1). *Journal of Affective Disorders*, *174*, 390-396.

DeRubeis, R. J., Hollon, S. D., Amsterdam, J. D., Shelton, R. C., Young, P. R., Salomon, R. M., O'Reardon, J.P., Lovett, M.L., Gladis, M.M., Brown, L.L. & Gallop, R. (2005). Cognitive therapy vs medications in the treatment of moderate to severe depression. *Archives of General Psychiatry, 62*, 409-416.

Dunlop, B. W., & Nemeroff, C. B. (2007). The role of dopamine in the pathophysiology of depression. *Archives of General Psychiatry, 64*, 327-337.

Dunn, B.D., & Roberts, H. (2012) Improving the Capacity to Treat Depression using Talking Therapies. *The Wiley Handbook of Positive Clinical Psychology,* 181-204.

Dunn, B. D. (2012). Helping depressed clients reconnect to positive emotion experience: current insights and future directions. *Clinical Psychology & Psychotherapy, 19*, 326-340.

Dunn, B. D. (2017) Opportunities and challenges for the emerging field of positive emotion regulation: A commentary on the special edition on positive emotions and cognitions in clinical psychology. *Cognitive Therapy & Research, 41*, 469-478.

Dunn, B. D. (in press) Augmenting Cognitive Behavioural Therapy to Target the Anhedonic Symptoms of Depression. In J. Gruber (Eds.) *Oxford Handbook of Positive Emotion and Psychopathology.*

Dunn, B. D., et al. (2018) Turning gold into lead: Dampening appraisals reduce happiness and pleasantness and increase sadness during anticipation and recall of pleasant activities in the laboratory. *Behaviour Research and Therapy, 107*, 19-33.

Dunn, B. D., Widnall, E. W., Reed, N., Taylor, R. S., Owens, C., Spencer, A…..Kuyken, W. (2019). Evaluating Augmented Depression Therapy (ADepT): Study protocol for a pilot randomised controlled trial. *Pilot & Feasibility Trials, 5:* 63.

Fawcett, J., Scheftner, W. A., Fogg, L., Clark, D. C., & Young, M. A. (1990). Time-related predictors of suicide in major affective disorder. *The American Journal of Psychiatry, 147*, 1189-1194.

Fried, E. I., & Nesse, R. M. (2014). The impact of individual depressive symptoms on impairment of psychosocial functioning. *PLOS One, 9*, e90311

Fried, E. I., van Borkulo, C. D., Epskamp, S., Schoevers, R. A., Tuerlinckx, F., & Borsboom, D. (2016). Measuring depression over time... Or not? Lack of unidimensionality and

longitudinal measurement invariance in four common rating scales of depression. *Psychological Assessment*, *28*, 1354-1367.

Fried, E. I., Epskamp, S., Nesse, R. M., Tuerlinckx, F., Borsboom, D. (2016). What are 'good' depression symptoms? Comparing the centrality of DSM and non-DSM symptoms of depression in a network analysis. *Journal of Affective Disorders, 189*, 314-320.

Fried, E. I., van Borkulo, C. D., Cramer, A. O. J., Boschloo, L., Schoevers, R. A., & Borsboom, D. (2017). Mental disorders as networks of problems: a review of recent insights. *Social Psychiatry and Psychiatric Epidemiology, 51*, 1-10.

Gadeikis, D., Bos, N., Schweizer, S., Murphy, F., Dunn, B. D. (2017). Engaging in an experiential processing mode increases positive emotional response during recall of pleasant autobiographical memories. *Behaviour Research and Therapy*, 92, 68-76.

Geschwind, N., Nicolson, N. A., Peeters, F., van Os, J., Barge-Schaapveld, D., & Wichers, M. (2011). Early improvement in positive rather than negative emotion predicts remission from depression after pharmacotherapy. *European Neuropsychopharmacology, 21*, 241-247.

Geschwind, N., Arntz, A., Bannink, F., & Peeters, F. (2019). Positive cognitive behaviour therapy in the treatment of depression: A randomized order within-subject comparison with traditional cognitive behaviour therapy. *Behaviour Research and Therapy, 116*, 119-130.

Gray, J. A. (1987). *The Psychology of Fear and Stress*. McGraw-Hill.

Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery, and Psychiatry*, *23*, 56-62.

Hofmann, S. G., Curtiss, J., & McNally, R. J. (2016). A complex network perspective on clinical science. *Perspectives in Psychological Science, 11*, 597-605.

Hofmann, S. G., & Hayes, S. C. (2019). The future of intervention science: Process-Based Therapy. *Clinical Psychological Science*, 7, 37-50.

Hollon, S. D., DeRubeis, R. J., Fawcett, J., Amsterdam, J. D., Shelton, R. C., Zajecka, J., Young, P.R & Gallop, R. (2014). Effect of cognitive therapy with antidepressant medications vs antidepressants alone on the rate of recovery in major depressive disorder: a randomized clinical trial. *JAMA Psychiatry, 71*, 1157-1164.

Hunot, V., Moore, T.H.M., Caldwell, D. M., Davies, P., Jones, H., Honyashiki, M., Chen, P., Lewis, G., & Churchill, R. (2013). 'Third wave' cognitive and behavioural therapies versus other psychological therapies for depression. *Cochrane Database of Systematic*

Reviews 2013, Issue 10. Art. No.: CD008704. DOI: 10.1002/14651858.CD008704.pub2.

IBM Corp (2017).IBM SPSS Statistics for Windows, Version 25.0. Armonk, NY: IBM Corp.

Jamerson, B. D., Krishnan, K. R., Roberts, J., Krishen, A., & Modell, J. G. (2003). Effect of bupropion SR on specific symptom clusters of depression: analysis of the 31-item Hamilton Rating Scale for depression. *Psychopharmacology Bulletin*, *37*, 67-78.

Jamieson, J. (2004). Analysis of covariance (ANCOVA) with difference scores. *International Journal of Psychophysiology, 52*, 277-283.

Johnson, J. & Wood, A. M. (2015). Integrating positive and clinical psychology: Viewing human functioning as continua from positive to negative can benefit clinical assessment, interventions and understandings of resilience. *Cognitive Therapy & Research, 41*, 335-349.

Keller, M. B., Lavori, P. W., Friedman, B., Nielsen, E., Endicott, J., McDonald-Scott, P., & Andreasen, N. C. (1987). The Longitudinal Interval Follow-up Evaluation: a comprehensive method for assessing outcome in prospective longitudinal studies. *Archives of General Psychiatry, 44*, 540-548.

Kendall, A. D., Zinbarg, R. E., Bobova, L., Mineka, S., Revelle, W., Jason M. Proveau., Craske, M. G. (2016). Measuring positive emotion with the mood and anxiety symptom questionnaire: Psychometric properties of the anhedonic depression scale. *Assessment, 23*, 86-95.

Keogh, E., & Reidy, J. (2000). Exploring the factor structure of the Mood and Anxiety Symptom Questionnaire (MASQ). *Journal of Personality Assessment, 74*, 106-125.

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Koretz, D., Merikangas, K. R., Rush, A.J., Walters, E.E, & Wang, P. S. (2003). The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). *Journal of the American Medical Association, 289*, 3095-3105.

Kring, A. M., Persons, J. B., & Thomas, C. (2007). Changes in affect during treatment for depression and anxiety. *Behaviour Research and Therapy, 45*, 1753-1764.

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., Sanislow, C. & Wang, P. (2010). Research domain criteria (RDoC): Toward a new classification framework for research on mental disorders. *American Journal of Psychiatry, 167*, 748-751.

Lally, N., Nugent, A. C., Luckenbaugh, D. A., Niciu, M. J., Roiser, J. P., & Zarate Jr, C. A. (2015). Neural correlates of change in major depressive disorder anhedonia following open-label ketamine. *Journal of Psychopharmacology*, *29*, 596-607.

Martell, C. R., Dimidjian, S., & Herman-Dunn, R. (2010). *Behavioral Activation for Depression: A Clinician's Guide*. Guilford Press.

McMakin, D. L., Olino, T. M., Porta, G., Dietz, L. J., Emslie, G., Clarke, G., Wagner, K.D., Asarnow, J.R., Ryan, N.D., Birmaher, B. & Shamseddeen, W. (2012). Anhedonia predicts poorer recovery among youth with selective serotonin reuptake inhibitor treatment–resistant depression. *Journal of the American Academy of Child & Adolescent Psychiatry*, *51*, 404-411.

Muchinsky, P. M. (1996). The correction for attenuation. *Educational and Psychological Measurement*, *56*, 63-75.

Naragon-Gainey, K., Gallagher, M. W., & Brown, T. A. (2013). Stable "trait" variance of temperament as a predictor of the temporal course of depression and social phobia. *Journal of Abnormal Psychology*, *122*, 611-623

Norman, G. R., Sloan, J. A., & Wyrwich, K. W. (2003). Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Medical Care*, *41*, 582-592.

Padesky, C. A., & Mooney, K. A. (2012). Strengths-based cognitive–behavioural therapy: A four-step model to build resilience. *Clinical Psychology & Psychotherapy*, *19*, 283-290.

Paulus, M. P., Stein, M. B., Craske, M. G., Bookheimer, S., Taylor, C. T., Simmons, A. N., Sidhu, N., Young, K. S., Fan, B. (2017). Latent variable analysis of positive and negative valence processing focused on symptom and behavioural units of analysis in mood and anxiety disorders. *Journal of Affective Disorders*, *216*, 17-29.

Paunonen, S. V. (2003). Big Five factors of personality and replicated predictions of behavior. *Journal of Personality and Social Psychology*, *84*, 411–424.

Price, J., Cole, V., & Goodwin, G. M. (2009). Emotional side-effects of selective serotonin reuptake inhibitors: qualitative study. *The British Journal of Psychiatry*, *195*, 211-217.

Provencher H. L., Keyes C. L. M. (2011). Complete mental health recovery: Bridging mental illness with positive mental health. *Journal of Public Mental Health*, *10*, 57–69.

Richards, D. A., Ekers, D., McMillan, D., Taylor, R. S., Byford, S., Warren, F. C., Barrett, B., Farrand, P.A., Gilbody, S., Kuyken, W. & O'Mahen, H. (2016). Cost and

Outcome of Behavioural Activation versus Cognitive Behavioural Therapy for Depression (COBRA): a randomised, controlled, non-inferiority trial. *The Lancet*, *388*, 871-880.

Roberts, B. W., Luo, J., Briley, D. A., Chow, P. I., Su, R., Hill, P. L. (2017). A systematic review of personality trait change through intervention. *Psychological Bulletin, 143*, 117-141.

Ruini, C., & Fava, G. A. (2012). Role of well-being therapy in achieving a balanced and individualized path to optimal functioning. *Clinical Psychology & Psychotherapy*, *19*, 291-304.

Rush, A. J., Trivedi, M. H., Wisniewski, S. R., Nierenberg, A. A., Stewart, J. W., Warden, D., Niederehe, G., Thase, M.E., Lavori, P.W., Lebowitz, B.D. & McGrath, P. J. (2006). Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR* D report. *American Journal of Psychiatry*, *163*, 1905-1917.

Rush, A. J., Kraemer, H. C., Sackheim, H. A., Fava, M., Trivedi, M.H, Frank E., et al (2006). Report by the ACNP Task Force on response and remission in major depressive disorder. *Neuropsychopharmacology*, *31*, 1841-1853.

Schimmack, U., Oishi, S., Furr, R. M., & Funder, D. C. (2004). Personality and life satisfaction: A facet-level analysis. *Personality and Social Psychology Bulletin, 30*, 1062-1075.

Shelton, R. C., & Tomarken, A. J. (2001). Can recovery from depression be achieved?. *Psychiatric Services*, *52*, 1469-1478.

Slade, M. (2010). Mental illness and well-being: The central importance of positive psychology and recovery approaches. *BMC Health Services Research, 10*: 26.

Snaith, R. P., Hamilton, M., Morley, S., Humayan, A., Hargreaves, D., & Trigwell, P. (1995). A scale for the assessment of hedonic tone the Snaith–Hamilton Pleasure Scale. *The British Journal of Psychiatry*, *167*, 99-103.

Spijker, J., Bijl, R. V., De Graaf, R., & Nolen, W. A. (2001). Determinants of poor 1-year outcome of DSM-III-R major depression in the general population: results of the Netherlands Mental Health Survey and Incidence Study (NEMESIS). *Acta Psychiatrica Scandinavica*, *103*, 122-130.

StataCorp. 2015. *Stata Statistical Software: Release 14*. College Station, TX: StataCorp LP

Tomarken, A. J., Dichter, G. S., Freid, C., Addington, S., & Shelton, R. C. (2004). Assessing the effects of bupropion SR on mood dimensions of depression. *Journal of Affective Disorders*, *78*, 235-241.

Tomarken, A. J., Shelton, R. C., & Hollon, S. D. (2007). Affective Science as a Framework for Understanding the Mechanisms and Effects of Antidepressant Medications. In J. Rottenberg & S. L. Johnson (Eds.), *Emotion and Psychopathology: Bridging Affective and Clinical Science* (pp. 263-283). Washington, DC, US: American Psychological Association.

Treadway, M. T., & Zald, D. H. (2011). Reconsidering anhedonia in depression: lessons from translational neuroscience. *Neuroscience & Biobehavioral Reviews*, *35*, 537-555.

Uher, R., Perlis, R. H., Henigsberg, N., Zobel, A., Rietschel, M., Mors, O., Hauser, J., Dernovsek, M.Z., Souery, D., Bajs, M. & Maier, W. (2012). Depression symptom dimensions as predictors of antidepressant treatment outcome: replicable evidence for interest-activity symptoms. *Psychological Medicine*, *42*, 967-980.

Üstün, T. B., Ayuso-Mateos, J. L., Chatterji, S., Mathers, C., & Murray, C. J. (2004). Global burden of depressive disorders in the year 2000. *The British Journal of Psychiatry*, *184*, 386-392.

Vittengl, J. R., Clark, L. A., Dunn, T. W., & Jarrett, R. B. (2007). Reducing relapse and recurrence in unipolar depression: A comparative meta-analysis of cognitive-behavioral therapy's effects. *Journal of Consulting and Clinical Psychology*, *75*, 475-488.

Von Borkulo, C., Boschioo, L., Borsboom, D., Penninx, B. W., Waldorp, L. J., & Schoevers, R. A. (2015). Association of symptom network structure with the course of depression. *JAMA Psychiatry*, *72*, 1219-1226.

Wang, L., & Maxwell, S. E. (2015). On disaggregating between-person and within-person effects with longitudinal data using multilevel models. *Psychological Methods, 20*, 63-83.

Watson, D. & Naragon-Gainey, K. (2014). Personality, emotions and the emotional disorders. *Clinical Psychological Science, 2*, 422-442.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of Personality and Social Psychology*, *54*, 1063-1070

Watson, D., & Clark, L. A. (1991). The Mood and Anxiety Symptom Questionnaire. *Unpublished manuscript, University of Iowa, Department of Psychology, Iowa City*.

Watson, D., Wiese, D., Vaidya, J., & Tellegen, A. (1999). The two general activation systems of affect: Structural findings, evolutionary considerations, and psychobiological evidence. *Journal of Personality and Social Psychology*, *76*, 820-838.

Watson, D., & Clark, L.A (1999). The PANAS-X: Manual for the Positive and Negative Affect Schedule – Expanded form. University of Iowa.

Watson, D., Stasik, S. M., Ellickson-Larew, S., & Stanton, K. (2015). Extraversion and psychopathology: A facet level analysis. *Journal of Abnormal Psychology, 124*, 432-446.

White, I .R., Royston, P., & Wood, A. M. (2010). Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine, 30*, 377-399.

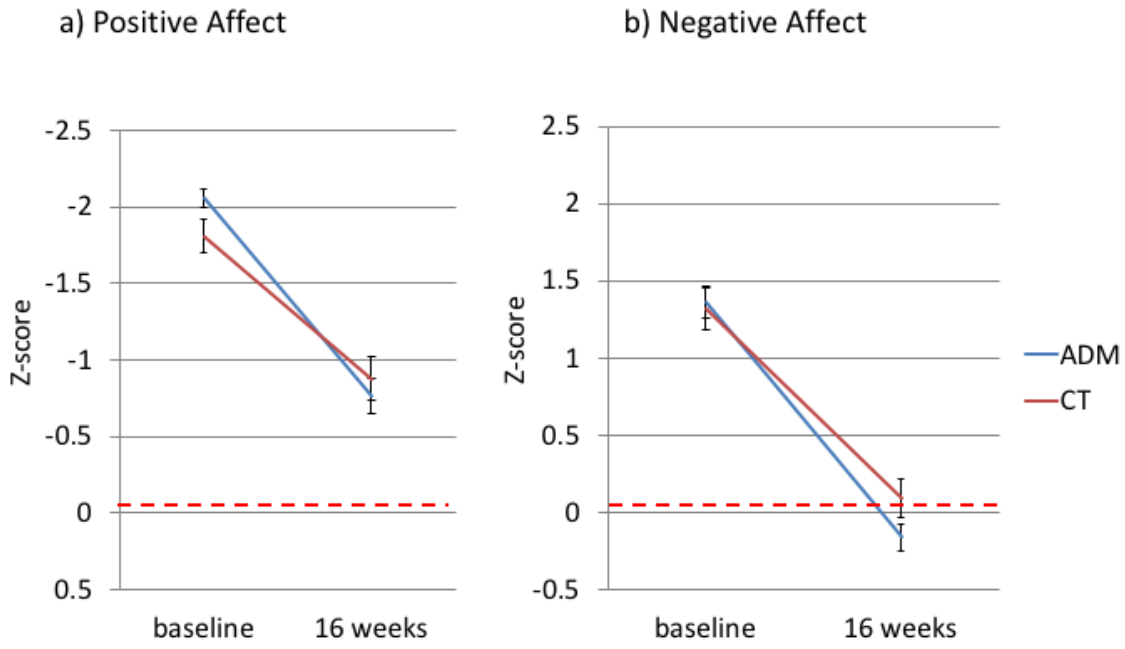
Wiles, N., Thomas, L., Abel, A., Ridgway, N., Turner, N., Campbell, J., Garland, A., Hollinghurst, S., Jerrom, B., Kessler, D. & Kuyken, W. (2013). Cognitive behavioural therapy as an adjunct to pharmacotherapy for primary care based patients with treatment resistant depression: results of the CoBalT randomised controlled trial. *The Lancet*, *381*, 375-384.

Zimmerman, M., McGlinchey, J. B., Posternak, M. A., Friedman, M., Attiullah, N., & Boerescu, D. (2006). How should remission from depression be defined? The depressed patient's perspective. *American Journal of Psychiatry*, *163*, 148-150.

Figures

Figure 1.

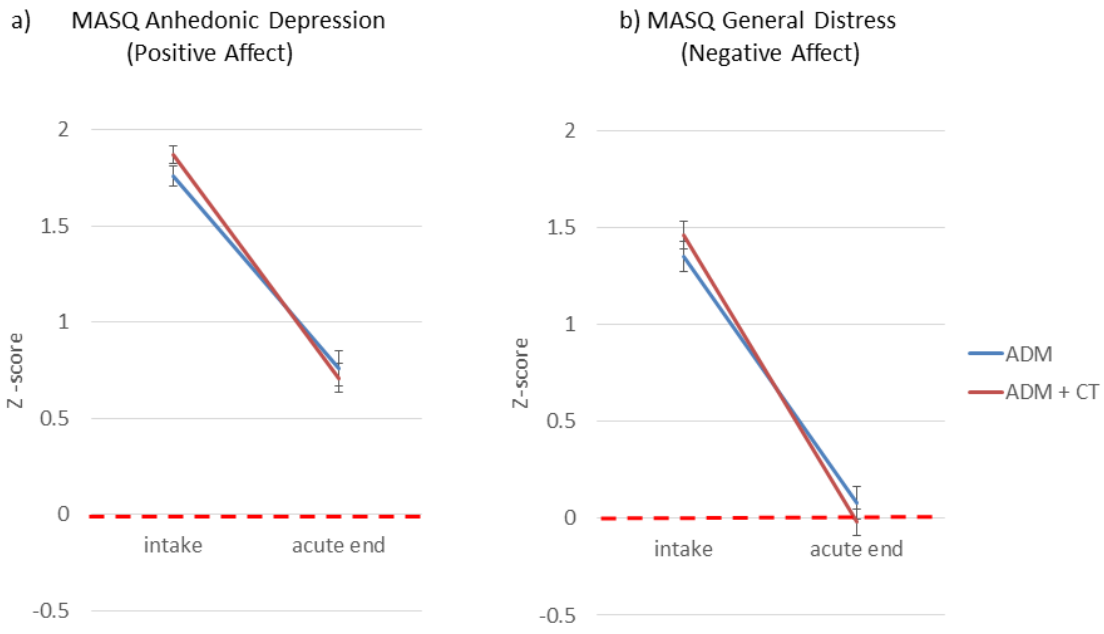
Positive Affect (panel a) and Negative Affect (panel b) at intake and end of treatment (16 weeks) in the antidepressant (ADM) and Cognitive Therapy (CT) arms of the CPT2 trial.



Note – data are mean (one standard error of the mean) Z-score values. To allow visual comparison with Negative Affect, the Positive Affect axis is reverse scored. Therefore, moving downwards represents clinical improvement for both Negative Affect and Positive Affect. Zero on each vertical axis (highlighted with bold dotted line) represents US adult general population mean levels; -1/+1 represent one standard deviation below/above this mean respectively.

Figure 2

MASQ Anhedonic Depression (panel a) and General Distress (panel b) at intake- and acute treatment end in the antidepressant only (ADM) and antidepressant and cognitive therapy combined (ADM+CT) arms of the CPT3 trial.



Note – data are mean (one standard error of the mean) Z-score values. Moving downwards represents clinical improvement for both Anhedonic Depression and General Distress. Zero on each vertical axis (highlighted with bold dotted line) represents UK adult general population mean levels; +1 represents one standard deviation above this mean.

Supplementary Materials Table S1: Summary of studies included in Boumparis et al meta-analysis.

Study	Sample	Intervention	Comparator	Measure of Affect
Delgado-Pastor et al (2015)	Chronic worry in female university students (n=45), based on > 80th % on Penn State Worry Questionnaire	Bespoke mindfulness cognitive training course (length of treatment unclear)	No intervention control	PANAS (time scale unclear)
Ehde et al (2015)	163 adults with multiple sclerosis, with fatigue, chronic pain and/or depressive symptoms	Eight week telephone delivered self-management Intervention	Eight week multiple sclerosis education intervention	PANAS (time scale unclear)
Graziano et al (2015)	82 patients with relapsing remitting multiple sclerosis (MS)	Four 2 hour cognitive behavioural group based sessions over two months, with an additional booster session at six months	Three information sessions about stem cells, complementary therapies and nourishment	PANAS (time scale unclear)
Lee and Bang (2010)	75 women reporting depression symptoms, based on Beck Depression Inventory – revised scores (unclear what cut off used)	Adapted Mindfulness Based Cognitive Therapy, with additional self-compassion exercises	Wait list control	PANAS (time scale unclear)
Perini et al (2009)	45 individuals meeting diagnostic criteria for depression	Six week sadness programme (six online CBT lessons, homework assignments, participation in a discussion forum, and email contact with mental health clinician)	Wait list Control	PANAS (time scale unclear)
Penton-Voak et al (2012)	193 young adults scoring > 14 on Beck Depression Inventory-revised	Four days practicing online cognitive bias modification procedure (judging if faces happy or sad)	Matched control condition	PANAS (time scale unclear)
Newby et al (2014)	60 dysphoric individuals, based on scores on Beck Depression Inventory-revised.	Single session cognitive bias modification to train positive appraisals of intrusive memories	Waitlist control condition	PANAS (present moment version)
Walker and Lampropoulos (2014)	94 dysphoric undergraduates, based on scoring > 10 on the Centre for Epidemiological Studies Depression Scale	Practice of CBT homework exercises over two weeks (at least four hours in total)	No instruction control and positive psychology homework control	PANAS (past two weeks version)
Yiend et al (2014)	40 adults meeting diagnostic criteria for major depressive disorder	Single session cognitive bias modification training	Matched control condition	PANAS trait version
Zhou et al (2012)	125 older adults scoring between 11 and 25 on the Geriatric Depression Scale	Three Health education sessions over six weeks and six weekly sessions of group reminiscence therapy	Three health education sessions over six weeks	Affect Balance Scale

Table S2:

Correlations of individual item depression symptoms with measures of PA and NA

	Study 1 (n=180)		Study 2 (n=440)	
	PA	NA	MASQ AD	MASQ GD
1. Depressed mood	-.06	.08	.09	.13**
2. Feelings of guilt	-.04	.30*	.12*	.28**
3. Suicide	-.12	.13	.13*	.20**
4. Insomnia – initial	.07	-.12	.08	.14**
5. Insomnia – middle	.12	.03	.02	.00
6. Insomnia – delayed	.06	-.02	.05	.03
7. Work and interests	-.25*	.10	.13*	.13*
8. Retardation	-.18*	-.07	.09	.16*
9. Agitation	.21*	.19*	-.03	.05
10. anxiety psychic	.01	.26*	.09	.26**
11. anxiety - somatic	.09	.04	.13*	.18**
12. somatic symptoms – GI	.02	-.09	.06	.10*
13. somatic symptoms - general	-.21*	.11	.11*	.07
14. genital symptoms	-.07	-.01	.15**	.07
15. hypochondriasis	.02	.17*	-.01	.09
16. weight loss	.03	-.01	.04	.07
17. insight	-.05	-.07	-.03	.08