Neural signatures of emotional biases predict clinical outcomes in difficult-to-treat depression

Diede Fennema[1], Gareth J. Barker[2], Owen O'Daly[2], Beata R. Godlewska[3,4], Ewan Carr[5], Kimberley Goldsmith[5], Allan H. Young[1,6], Jorge Moll[7] & Roland Zahn[1,6,7*]

[1] *Centre of Affective Disorders, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK*

[2] *Department of Neuroimaging, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK*

[3] *Psychopharmacology Research Unit, University Department of Psychiatry, University of Oxford, Oxford, UK*

[4] *Oxford Health NHS Foundation Trust, Warneford Hospital, Oxford, UK*

[5] *Department of Biostatistics and Health Informatics, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK*

[6] *National Service for Affective Disorders, South London and Maudsley NHS Foundation Trust, London, UK*

[7] *Cognitive and Behavioral Neuroscience Unit, D'Or Institute for Research and Education (IDOR), Pioneer Science Program, Rio de Janeiro, Brazil*

* Corresponding author

Professor Roland Zahn (see address above)
E-mail: roland.zahn@kcl.ac.uk
Phone: 0044-(0)20 7848 0348
Fax: 0044-(0)20 7848 0298

Word count (abstract): 200/200
Word count (article body): 3175
Number of figures: 1
Number of tables: 2
Supplemental information: 1

Keywords: *depression; prediction; fMRI; prognosis; difficult-to-treat*

Abstract

Background: Neural predictors underlying variability in depression outcomes are poorly understood. FMRI measures of subgenual cortex connectivity, self-blaming and negative perceptual biases have shown prognostic potential in treatment-naïve, medication-free, and fully remitting forms of major depressive disorder (MDD). However, their role in more chronic, difficult-to-treat forms of MDD is unknown.

Methods: Forty-five participants (n=38 meeting minimum data quality thresholds) fulfilled criteria for difficult-to-treat MDD. Clinical outcome was determined by computing percentage change at follow-up from baseline (four months) on the self-reported Quick Inventory of Depressive Symptomatology (16-item). Baseline measures included self-blame-selective connectivity of the right superior anterior temporal lobe with an *a priori* Brodmann Area 25 region-of-interest, blood-oxygen level-dependent *a priori* bilateral amygdala activation for subliminal sad vs happy faces, and resting-state connectivity of the subgenual cortex with an *a priori* defined ventrolateral prefrontal cortex/insula region-of-interest.

Findings: A linear regression model showed that baseline severity of depressive symptoms explained 3% of the variance in outcomes at follow-up ($F[3,34]=.33$, $p=.81$). In contrast, our three pre-registered neural measures combined, explained 32% of the variance in clinical outcomes ($F[4,33]=3.86$, $p=.01$).

Conclusion: These findings corroborate the pathophysiological relevance of neural signatures of emotional biases and their potential as predictors of outcomes in difficult-to-treat depression.

Introduction

Currently, treatment of major depressive disorder (MDD) is based on a trial-and-error approach, with only half of patients responding to their initial treatment (Rush et al. 2006). There is a clear need for improving treatment in patients with depression, which could be informed by standard clinical variables and biomarkers (Dunlop and Mayberg 2014; Fonseka, MacQueen and Kennedy 2018; Perlman et al. 2019). The field has started to identify various biomarkers showing promise, such as genetic markers (Breitenstein, Scheuer and Holsboer 2014; Laje et al. 2009), behavioural and cognitive markers (Groves, Douglas and Porter 2018; Park et al. 2018; Perna et al. 2020), metabolic and inflammatory markers (Lopresti et al. 2014; Schmidt, Shelton and Duman 2011) and neuroimaging markers (Breitenstein, Scheuer and Holsboer 2014; Dichter, Gibbs and Smoski 2015; Dunlop and Mayberg 2014; Fonseka, MacQueen and Kennedy 2018; Fu, Steiner and Costafreda 2013).

Such biomarkers are thought to represent underlying biological substrates of depression, which can be used to predict general prognosis regardless of treatment, better outcome with any treatment, or differential treatment response (Simon and Perlis 2010). For example, baseline metabolic profile was found to differentiate between responders and non-responders to sertraline or placebo (Kaddurah-Daouk et al. 2011), baseline C-reactive protein differentially predict response to escitalopram or nortriptyline (Uher et al. 2014), and baseline resting-state functional connectivity with the subgenual cortex differentially predicted response to antidepressant treatment or cognitive behavioural therapy (CBT) (Dunlop et al. 2017).

It is important to note, however, that MDD is a multifaceted disorder associated with a wide range of cognitive, behavioural, emotional, and physiological symptoms (Disner et al. 2011). As such, it is unlikely that a single clinical or biological marker can predict treatment outcome (Patel, Khalaf and Aizenstein 2016; Phillips et al. 2015). In fact, Lee et al. (2018) showed that models informed by multiple data types, such as a composite of clinical features,

neuroimaging, or genetic measures, were more accurate at predicting outcome than less complex models. Nonetheless, current clinical practice is mostly based on questionnaire- and interview-based assessments, which represent a wealth of clinical data which can be used to predict treatment outcome (Rost, Binder and Bruckl 2022).

In recent years, machine-learning methods have been increasingly employed to examine which clinical variables are most predictive of response or remission, allowing identification of patterns of information at an individual patient level (Chekroud et al. 2021; Jankowsky et al. 2024). Various studies have consistently identified baseline symptom severity, number of depressive episodes and co-morbid anxiety disorders as predictors of treatment outcome (Balestri et al. 2016; Bartova et al. 2019; Chekroud et al. 2016; Iniesta et al. 2016; Kautzky et al. 2018; Perlis 2013). However, standard clinical variables alone capture a limited amount of variance in clinical outcome, with estimates in the region of 5-10% (Iniesta et al. 2016; Perlis 2013), and they tend to perform worse than neuroimaging measures (Dunlop 2015; Jollans and Whelan 2016; Lee et al. 2018; Poirot et al. 2024; Schmaal et al. 2015).

Neuroimaging measures may be of particular interest, as dysfunctional neural processes are core to the development and maintenance of depressive symptoms (Godlewska 2020). They capture emotional biases associated with depression, such as the tendency to focus more on negative facial expressions than positive ones (Bourke, Douglas and Porter 2010; Krause et al. 2021), proneness to experience excessive self-blaming emotions, such as overgeneralised guilt and disgust/contempt towards oneself (Duan et al. 2022; Duan et al. 2023; Green et al. 2013; Weiner 1985; Zahn et al. 2015), as well as rumination, i.e. a tendency to engage in recursive, automatic thoughts often linked to self-critical thinking (Berman et al. 2014; Hamilton et al. 2015; Nolen-Hoeksema, Wisco and Lyubomirsky 2008).

Leading neuroanatomical models of MDD propose that impaired function within prefrontal-limbic neural circuits, particularly the subgenual cingulate cortex and amygdala,

explains disruptions of emotional processing and regulation associated with depression (Price and Drevets 2010; Ressler and Mayberg 2007). Neuroimaging biomarkers capturing the aforementioned – often implicit – emotional biases associated with depression have shown promise in predicting prognosis in MDD at an individual level, notably amygdala activation in response to emotional faces (Williams et al. 2015) and resting-state posterior subgenual cortex connectivity (Dunlop et al. 2017) in current MDD, and self-blame-selective anterior temporal-subgenual connectivity in remitted MDD (Lawrence et al. 2022). Despite these promising findings, studies tend to focus on treatment-naïve and treatment-free samples of MDD, and it is unclear whether these neural signatures generalise to pragmatic samples of patients encountered in clinical settings. Moreover, it is important to establish whether imaging measures provide added value in predicting clinical outcomes compared to standard baseline clinical variables.

Here, we probed the potential of these neural signatures of emotional biases in predicting clinical outcomes in a pragmatic sample of difficult-to-treat MDD after four months of primary care. These pre-registered (NCT04342299) neural signatures were selected based on their potential to predict response to treatment at an individual level and cover complementary neurocognitive aspects of MDD, i.e. self-blaming biases, negative perceptual biases, and dysfunction of task-independent subgenual networks.

Methods

The fMRI dataset reported here was collected as part of an observational sub-study within a feasibility trial, the Antidepressant Advisor Study (NCT03628027) (Harrison et al. 2020; Harrison et al. 2023). We have published tasked-based functional imaging (Fennema et al. 2023; Fennema, Barker, O'Daly, Duan, Godlewska, et al. 2024) and resting-state fMRI results

(Fennema, Barker, O'Daly, Duan, Carr, et al. 2024) from the same cohort previously, but here, we report on the prediction model for the first time.

Participants

Forty-five participants fulfilled criteria for current MDD according to the Diagnostic and Statistical Manual of Mental Health Disorders, Fifth Edition (First et al. 2015) and had not responded to at least two serotonergic antidepressants. Participants were encouraged to book an appointment with their GP to review their medication and followed up after receiving four months of standard care. For more information about inclusion/exclusion criteria, recruitment, and assessment, please see Supplementary Methods.

Prior to their medication review, participants attended an fMRI session, consisting of three paradigms: the moral sentiment task (assessing self-blame-related biases), the subliminal faces task (assessing bias in emotional processing), and a resting-state fMRI scan. As part of the moral sentiment task, participants viewed self- and other-blaming emotion-evoking statements. Participants were shown written statements describing actions counter to socio-moral values described by social concepts (e.g. impatient, dishonest) in which the agent was either the participant (self-agency) or their best friend (other-agency) (Fennema et al. 2023). As part of the subliminal faces task, participants were presented with a series of faces. The faces were shown in pairs, briefly displaying a “target” face (expressing sad, happy or neutral emotion) followed by another “mask” face of neutral expression (Fennema, Barker, O'Daly, Duan, Godlewska, et al. 2024). As part of the resting-state fMRI scan, participants were instructed to keep their eyes open and let their mind wander while focusing on a cross (Fennema, Barker, O'Daly, Duan, Carr, et al. 2024). For more details on the fMRI paradigms, please see Supplementary Materials.

Main outcome

As stated in our pre-registered protocol (NCT04342299), we used a continuous measure of clinical outcome rather than categorising participants into responders and non-responders using the standard definition of a 50% reduction (Nierenberg and DeCecco 2001) in self-reported Quick Inventory of Depressive Symptomatology (16-item; QIDS-SR16) (Rush et al. 2003) scores, due to an unbalanced split between the resulting groups (responders n=8; non-responders n=30). The outcome was defined as the percentage change at follow-up from baseline on our pre-registered primary outcome measure, QIDS-SR16, where negative scores corresponded to a reduction in depressive symptoms.

fMRI measures

Statistical Parametric Mapping 12 was used for blood-oxygen level-dependent (BOLD) effect analysis and psychophysiological interaction analysis, while Data Processing Assistant for Resting-State fMRI (DPARSF) was used for resting-state analysis (please see Supplementary Methods for more details). Regression coefficient averages (moral sentiment task and subliminal faces task) and cluster mean z-score (resting-state scan) over our pre-registered regions-of-interest (ROIs) were extracted for individual participants using the MarsBaR toolbox (Rorden and Brett 2000), i.e. self-blame-selective connectivity between the right superior anterior temporal lobe (RSATL) and posterior subgenual cortex (Brodmann Area [BA] 25), bilateral amygdala BOLD activation for subliminal sad vs happy faces, and resting-state functional connectivity between the bilateral posterior subgenual cortex and left ventrolateral prefrontal cortex (BA47; VLPFC)/insula. For more details, please see Supplementary Materials.

Statistical analysis

Multiple linear regression was used to assess potential predictors of QIDS-SR16 percentage change, as well as an exploratory logistic regression to determine likelihood of response vs.

non-response. The aim of the study was to estimate the effect size of using our pre-registered imaging measures as predictors of clinical outcomes, rather than tease out the importance of each predictor given the limitations of our sample size. As such, we ran our main "fMRI" multivariable model which assessed the contribution of our three pre-registered fMRI measures as outlined above, with baseline Maudsley Modified Patient Health Questionnaire, 9 items (MM-PHQ-9; measure of severity of depressive symptoms) (Harrison et al. 2021) as a covariate.

In addition, we ran a supplementary "clinical" multivariable model to compare the contribution of standard clinical measures (baseline MM-PHQ-9, baseline Generalised Anxiety Disorder, 7-items (GAD-7; measure of severity of anxiety symptoms) (Spitzer et al. 2006), and Maudsley Staging Method total score (proxy of treatment-resistance based on duration, severity and treatment failures) (Fekadu, Donocik and Cleare 2018); please see Supplementary Methods for more details on the clinical measures). Another supplementary "high-quality fMRI" multivariable model assessed the impact of suboptimal fMRI quality, i.e. signal drop-out and/or more motion, on the predictive value of the fMRI measures, including only participants with high-quality fMRI data for all three scans (n=30). Other supplementary models considered the individual contribution of the pre-registered fMRI measures (please see Supplementary Methods and Results).

Please note that our pre-registered imaging measures also included additional regions-of-interest: functional resting-state subgenual cortex connectivity with the left ventromedial prefrontal cortex (BA10) and with the dorsal midbrain (Fennema, Barker, O'Daly, Duan, Carr, et al. 2024), as well as pregenual anterior cingulate cortex BOLD activation for subliminal sad vs happy faces (Fennema, Barker, O'Daly, Duan, Godlewska, et al. 2024). However, as our sample size only allowed us to model a limited number of variables without risk of overfitting, for our primary prediction model, we solely included variables showing univariate prediction

effects in our previous analyses (Fennema, Barker, O'Daly, Duan, Carr, et al. 2024; Fennema, Barker, O'Daly, Duan, Godlewska, et al. 2024). For more details on the exploratory "pre-registration" model, please see Supplementary Methods.

All variables were Fisher *Z*-transformed to derive beta coefficients and corresponding standard error. Correlation analysis (Spearman's rho) was used to investigate the association between the pre-registered neural signatures. To test whether there is any link between treatment change and symptom change, a one-way analysis of variance was conducted (please see Supplementary Methods for a description of treatment change). All tests were carried out using IBM SPSS Statistics 27, using a significance threshold of p=.05, two-tailed.

Results

Subgroup characteristics

Table 1 presents participant characteristics at baseline, split by responders and non-responders. Of 45 included participants, 38 had usable fMRI data (31 [82%] female, mean [SD] age = 41.8 [14.8] years). Most participants fulfilled the DSM-5 anxious distress specifier criteria (82%) and met criteria for a life-time axis I co-morbidity (87%). Average baseline depression severity was severe according to MM-PHQ-9 (mean [SD] = 18.7 [4.7]) and QIDS-SR16 (mean [SD] = 17.3 [3.5]), and 82% of the participants were taking a selective serotonin-reuptake inhibitor. There were no significant differences between responders and non-responders at baseline in terms of demographic and clinical characteristics (t < 1.31 and p > .20), except for current major depressive episode duration (responders mean [SD] = 6.3 [5.3]; non-responders mean [SD] = 32.8 [50.2]; t[31.2] = -2.85, p = .01).

As part of the study, participants were encouraged to book an appointment with their GP to review their antidepressant medication. Even though UK care guidelines would recommend changing antidepressant medications in non-responders, unexpectedly, more than

half (55%) did not change their medication and some even stopped their medication (16%; Supplementary Table 1). Despite little change in treatment, on average, participants showed a significant reduction in depressive symptoms from baseline to follow-up in QIDS-SR16 scores (mean [95% CI] = -4.1 [-5.8, -2.4]). This was also the case for other self- and observer-rated scores (Supplementary Table 2).

There was a mean percentage change [SD] of -23.1 [30] in QIDS-SR16: those with a relevant change showed the most improvement in QIDS-SR16 (mean percentage change [SD] = -43.8 [20.3]), followed by participants with a minimal change (mean percentage change [SD] = -32.1 [32.4]) and participants with no change (mean percentage change [SD] = -17.6 [29.6]). However, there was no significant difference between the groups (F[2,37] = 1.78, p = .18).

Prediction models

The "fMRI" model using the pre-registered fMRI measures with baseline MM-PHQ-9 as a covariate explained 32% of the variance of QIDS-SR16 percentage change (F[4,33] = 3.86, p = .01, R^2 = .32, $R^2_{adjusted}$ = .24; Table 2). When including all previously pre-registered regions, the overall prediction effect for the "pre-registration" model was comparable (R^2 = 33%, please see Supplementary Results). When limiting to "high-quality fMRI", the model explained 43% of the variance of QIDS-SR16 percentage change (F[4,25] = 4.67, p = .01, R^2 = .43, $R^2_{adjusted}$ = .34; Supplementary Table 3). In contrast, the "clinical" model using standard clinical measures at baseline, i.e. MM-PHQ-9, GAD-7 and Maudsley Staging Method, explained only 3% of the variance of QIDS-SR16 percentage change (F[3,34] = .33, p = .81, R^2 = .03, $R^2_{adjusted}$ = -.06; Table 2).

Bilateral amygdala BOLD activation positively contributed to the variance in QIDS-SR16 percentage change (partial β = 11.11, t[33] = 2.21), while partial effects of resting-state functional connectivity between the posterior subgenual cortex and left VLPFC/insula as well as self-blame-selective RSATL-BA25 connectivity contributed negatively (resting-state:

partial β = -8.15, t[33] = -1.95; RSATL-BA25: partial β = -7.28, t[33] = -1.71; Figure 1). Please see Supplementary Results and Supplementary Table 3 for exploratory separate prediction models for each fMRI paradigm showing a maximum of 18% of variance in clinical outcomes explained, when using the bilateral amygdala BOLD signature.

Notably, there were no bivariate associations between the three pre-registered fMRI measures (self-blame-selective RSATL-BA25 connectivity and bilateral amygdala BOLD activation: r_s[38] = -.06, p = .71; self-blame-selective RSATL-BA25 connectivity and resting-state functional connectivity between posterior subgenual cortex and left VLPFC/insula: r_s[38] = .09, p = .61; bilateral amygdala BOLD activation and resting-state functional connectivity between posterior subgenual cortex and left VLPFC/insula: r_s[38] = -.09, p = .61).

Exploratory findings responders vs. non-responders

A logistic regression was performed to determine the effects of the pre-registered neural measures and baseline MM-PHQ-9 on the likelihood of response vs. non-response. The logistic regression model was statistically significant, $\chi^2(4) = 11.09$, p = .03. The model explained 39% (Nagelkerke R^2) of the variance in responders and correctly classified 81.6% of the cases. Increased functional connectivity between the bilateral subgenual cortex and left VLPFC/insula was associated with an increased likelihood of response. For more details, please see Supplementary Results.

Conclusions

Discussion

To our knowledge, this is the first study to combine complementary functional imaging measures of affective circuits in MDD and to probe their role in prospectively predicting clinical outcomes in a pragmatic setting. We show that neuroimaging markers hold promise: the model with the three pre-registered fMRI measures explained more variance in clinical

outcomes compared with the clinical model, i.e. 32% vs 3%. The model that only included participants with high-quality fMRI measures explained an even larger amount of variance (43%), highlighting the need to adequately account for signal drop-out and/or motion artifacts. However, it is important to acknowledge that no formal statistical tests were undertaken to compare the regression models as the study was not powered for such comparisons, which limits the interpretability of differences between the models.

Interestingly, the effects of the three pre-registered fMRI measures were uncorrelated, showing that these may capture distinct aspects of MDD pathophysiology, i.e. self-blaming biases (right superior anterior temporal-subgenual connectivity), negative perceptual biases (amygdala), and dysfunction of task-independent subgenual networks. If these neural signatures were to relate to specific subtypes rather than independently predicting the same underlying general pathophysiology, then this would offer the intriguing possibility of stratification for neuromodulation and neurofeedback studies based on distinct neural circuits of interest, by either modulating self-blaming or emotional perception biases in patients non-responsive to standard treatments. The feasibility of such interventions has recently been confirmed, with reports of a training-induced reduction in self-blame-selective connectivity (Jaeckle et al. 2023) and an enhancement of amygdala responsiveness to positive autobiographical memories (Young et al. 2019).

However, it is important to first determine whether these neural signatures represent a trait-like feature of a fully remitting subtype of MDD, or whether it is also modulated by depressive state. For example, both self-blame-related and emotional perception-related changes have been identified in remitted MDD (Joormann and Gotlib 2007; Lythe et al. 2020; Ruhe et al. 2019). It is unclear whether these changes are more pronounced when people develop a recurrent episode or are merely due to underlying vulnerabilities which are not modulated by symptomatic state. This question is key to a deeper pathophysiological

understanding of MDD in that little is known about how trait-related changes interact with precipitating biological and psychological trigger events to result in a depressive brain state, and how it affects subsequent response to treatment.

Limitations

Due to our relatively modest sample size, we were unable to use cross-validated and data-driven machine learning algorithms, which may have improved the prediction model performance. Moreover, our sample consisted of chronic MDD patients, often with anxious distress and other co-morbidities. In addition, treatment was not standardised and, unlike previous studies in randomised controlled trials, did not allow us to distinguish spontaneous remission and placebo effects from treatment-related effects. Given the selection biases in randomised controlled trials, however, it was important to investigate a pragmatic sample as we have undertaken in this study.

Clinical utility is complicated by the heterogenous nature of MDD, resulting in patients with a wide variety of symptoms, disease severity and treatment history (Strawbridge, Young and Cleare 2017), as well as patient response to treatment (Mayberg and Dunlop 2023). Further complementary predictive measures, such as novel cognitive markers (Lawrence et al. 2022), would be useful in addition to imaging markers to achieve clinically relevant levels of individual prediction of response to specific types of treatment.

Moreover, it is important to acknowledge that percentage-based reduction scores to define treatment response has been criticised, as it is biased towards more severe depressive symptoms at baseline (Rost, Binder and Bruckl 2022). As a result, it is plausible for a responder to still experience clinically significant distress or impairment when starting with a baseline score in the severe range, while a non-responder may show a clinically significant improvement – which was also observed in the current study.

Conclusions

Taken together, we reproduced clinically relevant neural signatures in an independent, pragmatic sample of difficult-to-treat MDD. The findings confirm the pathophysiological relevance and potential of the proposed candidate neural signatures to make relevant contributions to the prospective prediction of clinical outcomes in more chronic, difficult-to-treat forms of MDD and call for stratified neurofeedback and neuromodulation interventions.

Acknowledgements

We are grateful to Drs Mark Ashworth and Barbara Barrett who contributed to the trial study design, and to Drs Phillippa Harrison and Suqian Duan who collected trial data. We also thank the study participants for their support.

Author contribution statement

Diede Fennema: Conceptualisation, Methodology, Formal analysis, Investigation, Data curation, Writing – original draft, Visualisation, Funding acquisition. **Gareth Barker:** Conceptualisation, Methodology, Writing – review & editing, Supervision. **Owen O'Daly:** Conceptualisation, Methodology, Writing – review & editing. **Beata Godlewska:** Methodology, Resources, Writing – review & editing. **Ewan Carr:** Conceptualisation, Methodology, Writing – review & editing. **Kimberley Goldsmith:** Conceptualisation, Methodology, Writing – review & editing. **Allan Young:** Conceptualisation, Methodology, Writing – review & editing, Supervision, Project administration, Funding acquisition. **Jorge Moll:** Conceptualisation, Writing – review & editing. **Roland Zahn:** Conceptualisation, Methodology, Formal analysis, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Funding statement

This study represents independent research funded by the National Institute for Health and Care Research (NIHR) under its Research for Patient Benefit (RfPB) Programme (Grant Reference Number PB-PG-0416-20039) and independent research part funded by the National Institute for Health and Care (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London (Profs Zahn, Young, Goldsmith; Dr Carr). Prof Zahn was also partly funded by a Medical Research Council grant (ref. MR/T017538/1), while Prof Goldsmith was also supported by the National Institute for Health and Care Research (NIHR) Applied Research Collaboration South London (NIHR ARC South London) at King's College Hospital NHS Foundation Trust. This study was also supported by the Rosetrees Trust (M816) awarded to Prof Zahn. Dr Fennema was funded by a Medical Research Council Doctoral Training Partnership Studentship (ref. 2064430) and partly supported by a KCL/IDOR Pioneer Science Fellowship, funded by Scients Institute and the IDOR Pioneer Science Initiative. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care. Additional support was provided to the study by the South London Clinical Research Network and sponsorship by Lambeth CCG.

Conflicts of interest

Prof Zahn is a private psychiatrist service provider at The London Depression Institute and co-investigator on a Livanova-funded observational study of Vagus Nerve Stimulation for Depression. Prof Zahn has received honoraria for talks at medical symposia sponsored by Lundbeck as well as Janssen. Prof Zahn has collaborated with EMOTRA, EMIS PLC and Depsee Ltd. Prof Zahn is affiliated with the D'Or Institute of Research and Education, Rio de Janeiro and advises the Scients Institute, USA. Prof Barker receives honoraria for teaching

from GE Healthcare. Prof Young is employed by King's College London as an honorary consultant in the South London and Maudsley Trust (NHS UK) and is a consultant to Johnson & Johnson and Livanova. Prof Young has given paid lectures and sat on advisory open access boards for the following companies with drugs used in affective and related disorders: Astrazenaca, Eli Lilly, Lundbeck, Sunovion, Servier, Livanova, Janssen, Allegan, Bionomics, Sumitomo Dainippon Pharma. Prof Young has received honoraria for attending advisory boards and presenting talks at meetings organised by LivaNova. Prof Young is the Principal Investigator of the following studies: Restore-Life VNS registry study funded by LivaNova, ESKETINTRD3004: 'An Open-label, Long-term, Safety and Efficacy Study of Intranasal Esketamine in Treatment-resistant Depression', 'The Effects of Psilocybin on Cognitive Function in Healthy Participants' and 'The Safety and Efficacy of Psilocybin in Participants with Treatment-Resistant Depression (P-TRD)'. Prof Young has received grant funding (past and present) from the following: NIMH (USA); CIHR (Canada); NARSAD (USA); Stanley Medical Research Institute (USA); MRC (UK); Wellcome Trust (UK); Royal College of Physicians (Edin); BMA (UK); UBC-VGH Foundation (Canada); WEDC (Canada); CCS Depression Research Fund (Canada); MSFHR (Canada); NIHR (UK); Janssen (UK). Prof Young has no shareholdings in pharmaceutical companies. Prof Goldsmith reports grants from NIHR, Stroke association, National Institutes of Health (US) and Juvenile Diabetes Research Foundation (US) during the conduct of the study. None of the other authors report biomedical financial interests or potential conflicts of interested related to the subject of this paper.

Ethics statement

Ethical approval was obtained from the NHS Health Research Authority and National Research Ethics Service London – Camberwell St Giles Committee (REC reference: 17/LO/2074). Written informed consent was obtained from all participants.

Data availability statement

The data that support the findings of this study are available on request from the corresponding author, RZ. We will only be able to share fully anonymised, no pseudonymised data and requests will have to go through a King's College London repository.

Prior publication

Part of the study has been published in a PhD thesis available on the King's College London institutional repository, Pure, see Fennema (2022): https://kclpure.kcl.ac.uk/portal/en/studentTheses/neural-signatures-of-emotional-biases-and-prognosisin-treatment-.

Rights Retention

Connections references

Hickie I and Williams L (2023) Will new brain circuit focused methods (EEG, fMRI etc) lead to more personalized care options? *Research Directions: Depression* **1**(E12). https://doi.org/10.1017/dep.2023.22.

References

Balestri M, Calati R, Souery D, Kautzky A, Kasper S, Montgomery S, Zohar J, Mendlewicz J and Serretti A (2016) Socio-demographic and clinical predictors of treatment resistant depression: a prospective European multicenter study. *J Affect Disord* **189,** 224-232. https://doi.org/10.1016/j.jad.2015.09.033.

Bartova L, Dold M, Kautzky A, Fabbri C, Spies M, Serretti A, Souery D, Mendlewicz J, Zohar J, Montgomery S, Schosser A and Kasper S (2019) Results of the European Group for the Study of Resistant Depression (GSRD) - basis for further research and clinical practice. *World J Biol Psychiatry* **20**(6)**,** 427-448. https://doi.org/10.1080/15622975.2019.1635270.

Berman MG, Misic B, Buschkuehl M, Kross E, Deldin PJ, Peltier S, Churchill NW, Jaeggi SM, Vakorin V, McIntosh AR and Jonides J (2014) Does resting-state connectivity reflect depressive

rumination? A tale of two analyses. *Neuroimage* **103,** 267-279. https://doi.org/10.1016/j.neuroimage.2014.09.027.

Bourke C, Douglas K and Porter R (2010) Processing of facial emotion expression in major depression: a review. *Australian & New Zealand Journal of Psychiatry* **44**(8)**,** 681-696. https://doi.org/10.3109/00048674.2010.496359.

Breitenstein B, Scheuer S and Holsboer F (2014) Are there meaningful biomarkers of treatment response for depression? *Drug Discov Today* **19**(5)**,** 539-561. https://doi.org/10.1016/j.drudis.2014.02.002.

Chekroud AM, Bondar J, Delgadillo J, Doherty G, Wasil A, Fokkema M, Cohen Z, Belgrave D, DeRubeis R, Iniesta R, Dwyer D and 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.

Chekroud AM, Zotti RJ, Shehzad Z, Gueorguieva R, Johnson MK, Trivedi MH, Cannon TD, Krystal JH and Corlett PR (2016) Cross-trial prediction of treatment outcome in depression: a machine learning approach. *The Lancet Psychiatry* **3**(3)**,** 243-250. https://doi.org/10.1016/s2215-0366(15)00471-x.

Dichter GS, Gibbs D and Smoski MJ (2015) A systematic review of relations between resting-state functional-MRI and treatment response in major depressive disorder. *Journal of Affective Disorders* **172,** 8-17. https://doi.org/10.1016/j.jad.2014.09.028.

Disner SG, Beevers CG, Haigh EA and Beck AT (2011) Neural mechanisms of the cognitive model of depression. *Nature Reviews Neuroscience* **12**(8)**,** 467-477. https://doi.org/10.1038/nrn3027.

Duan S, Lawrence AJ, Valmaggia L, Moll J and Zahn R (2022) Maladaptive blame-related action tendencies are associated with vulnerability to major depressive disorder. *J Psychiatr Res* **145,** 70-76. https://doi.org/10.1016/j.jpsychires.2021.11.043.

Duan S, Valmaggia L, Fennema D, Moll J and Zahn R (2023) Remote virtual reality assessment elucidates self-blame-related action tendencies in depression. *J Psychiatr Res* **161,** 77-83. https://doi.org/10.1016/j.jpsychires.2023.02.031.

Dunlop BW (2015) Prediction of treatment outcomes in major depressive disorder. *Expert Rev Clin Pharmacol* **8**(6)**,** 669-672. https://doi.org/10.1586/17512433.2015.1075390.

Dunlop BW and Mayberg HS (2014) Neuroimaging-based biomarkers for treatment selection in major depressive disorder. *Dialogues in Clinical Neuroscience* **16**(4)**,** 479-490. https://doi.org/10.31887/DCNS.2014.16.4/bdunlop.

Dunlop BW, Rajendra JK, Craighead WE, Kelley ME, McGrath CL, Choi KS, Kinkead B, Nemeroff CB and Mayberg HS (2017) Functional connectivity of the subcallosal cingulate cortex and differential outcomes to treatment with cognitive-behavioral therapy or antidepressant medication for major depressive disorder. *Am J Psychiatry* **174**(6)**,** 533-545. https://doi.org/10.1176/appi.ajp.2016.16050518.

Fekadu A, Donocik JG and Cleare AJ (2018) Standardisation framework for the Maudsley staging method for treatment resistance in depression. *BMC Psychiatry* **18**(1)**,** 100. https://doi.org/10.1186/s12888-018-1679-x.

Fennema D (2022) Neural signatures of emotional biases and prognosis in treatment-resistant depression. PhD, King's College London.

Fennema D, Barker GJ, O'Daly O, Duan S, Carr E, Goldsmith K, Young AH, Moll J and Zahn R (2023) Self-blame-selective hyper-connectivity between anterior temporal and subgenual cortices predicts prognosis in major depressive disorder. *Neuroimage Clin* **39,** 103453. https://doi.org/10.1016/j.nicl.2023.103453.

Fennema D, Barker GJ, O'Daly O, Duan S, Carr E, Goldsmith K, Young AH, Moll J and Zahn R (2024) The role of subgenual resting-state connectivity networks in predicting prognosis in major depressive disorder. *Biol Psychiatry Glob Open Sci* **4**(3)**,** 100308. https://doi.org/10.1016/j.bpsgos.2024.100308.

Fennema D, Barker GJ, O'Daly O, Duan S, Godlewska BR, Goldsmith K, Young AH, Moll J and Zahn R (2024) Neural responses to facial emotions and subsequent clinical outcomes in difficult-to-treat depression. *Psychol Med*. https://doi.org/10.1017/S0033291724001144.

First MB, Williams JBW, Karg RS and Spitzer RL (2015) *Structured Clinical Interview for DSM-5 - Research Version (SCID-5 for DSM-5, Research Version; SCID-5-RV, Version 1.0.0)*. Arlington, VA: American Psychiatric Association.

Fonseka TM, MacQueen GM and Kennedy SH (2018) Neuroimaging biomarkers as predictors of treatment outcome in major depressive disorder. *Journal of Affective Disorders* **233,** 21-35. https://doi.org/10.1016/j.jad.2017.10.049.

Fu CH, Steiner H and Costafreda SG (2013) Predictive neural biomarkers of clinical response in depression: a meta-analysis of functional and structural neuroimaging studies of pharmacological and psychological therapies. *Neurobiology of Disease* **52,** 75-83. https://doi.org/10.1016/j.nbd.2012.05.008.

Godlewska BR (2020) Neuroimaging as a tool for individualized treatment choice in depression: the past, the present and the future. *Current Behavioral Neuroscience Reports* **7**(1)**,** 32-39. https://doi.org/10.1007/s40473-020-00198-2.

Green S, Moll J, Deakin JF, Hulleman J and Zahn R (2013) Proneness to decreased negative emotions in major depressive disorder when blaming others rather than oneself. *Psychopathology* **46**(1)**,** 34-44. https://doi.org/10.1159/000338632.

Groves SJ, Douglas KM and Porter RJ (2018) A systematic review of cognitive predictors of treatment outcome in major depression. *Front Psychiatry* **9,** 382. https://doi.org/10.3389/fpsyt.2018.00382.

Hamilton JP, Farmer M, Fogelman P and Gotlib IH (2015) Depressive rumination, the default-mode network, and the dark matter of clinical neuroscience. *Biol Psychiatry* **78**(4)**,** 224-230. https://doi.org/10.1016/j.biopsych.2015.02.020.

Harrison P, Carr E, Goldsmith K, Young AH, Ashworth M, Fennema D, Barrett B and Zahn R (2020) Study protocol for the antidepressant advisor (ADeSS): a decision support system for antidepressant treatment for depression in UK primary care: a feasibility study. *BMJ Open* **10**(5)**,** e035905. https://doi.org/10.1136/bmjopen-2019-035905.

Harrison P, Carr E, Goldsmith K, Young AH, Ashworth M, Fennema D, Duan S, Barrett B and Zahn R (2023) The Antidepressant Advisor (ADeSS): a decision support system for antidepressant treatment for depression in UK primary care - a feasibility study. *BMJ Open* **13**(3)**,** e060516. https://doi.org/10.1136/bmjopen-2021-060516.

Harrison P, Walton S, Fennema D, Duan S, Jaeckle T, Goldsmith K, Carr E, Ashworth M, Young AH and Zahn R (2021) Development and validation of the Maudsley Modified Patient Health Questionnaire (MM-PHQ-9). *BJPsych Open* **7**(4)**,** e123. https://doi.org/10.1192/bjo.2021.953.

Iniesta R, Malki K, Maier W, Rietschel M, Mors O, Hauser J, Henigsberg N, Dernovsek MZ, Souery D, Stahl D, Dobson R, Aitchison KJ, Farmer A, Lewis CM, McGuffin P and Uher R (2016) Combining clinical variables to optimize prediction of antidepressant treatment outcomes. *J Psychiatr Res* **78,** 94-102. https://doi.org/10.1016/j.jpsychires.2016.03.016.

Jaeckle T, Williams SCR, Barker GJ, Basilio R, Carr E, Goldsmith K, Colasanti A, Giampietro V, Cleare A, Young AH, Moll J and Zahn R (2023) Self-blame in major depression: a randomised pilot trial comparing fMRI neurofeedback with self-guided psychological strategies. *Psychological Medicine* **53**(7)**,** 2831-2841. https://doi.org/10.1017/S0033291721004797.

Jankowsky K, Krakay L, Schroeders U, Zwerenz R and Beutel ME (2024) Predicting treatment response using machine learning: A registered report. *Br J Clin Psychol* **63,** 137-155. https://doi.org/10.1111/bjc.12452.

Jollans L and Whelan R (2016) The clinical added value of imaging: a perspective from outcome prediction. *Biol Psychiatry Cogn Neurosci Neuroimaging* **1**(5)**,** 423-432. https://doi.org/10.1016/j.bpsc.2016.04.005.

Joormann J and Gotlib IH (2007) Selective attention to emotional faces following recovery from depression. *J Abnorm Psychol* **116**(1)**,** 80-85. https://doi.org/10.1037/0021-843X.116.1.80.

Kaddurah-Daouk R, Boyle SH, Matson W, Sharma S, Matson S, Zhu H, Bogdanov MB, Churchill E, Krishnan RR, Rush AJ, Pickering E and Delnomdedieu M (2011) Pretreatment metabotype as a predictor of response to sertraline or placebo in depressed outpatients: a proof of concept. *Transl Psychiatry* **1**(7)**,** e26. https://doi.org/10.1038/tp.2011.22.

Kautzky A, Dold M, Bartova L, Spies M, Vanicek T, Souery D, Montgomery S, Mendlewicz J, Zohar J, Fabbri C, Serretti A, Lanzenberger R and Kasper S (2018) Refining prediction in treatment-resistant depression: results of machine learning analyses in the TRD III sample. *J Clin Psychiatry* **79**(1)**,** 16m11385. https://doi.org/10.4088/JCP.16m11385.

Krause FC, Linardatos E, Fresco DM and Moore MT (2021) Facial emotion recognition in major depressive disorder: A meta-analytic review. *Journal of Affective Disorders* **293,** 320-328. https://doi.org/10.1016/j.jad.2021.06.053.

Laje G, Perlis RH, Rush AJ and McMahon FJ (2009) Pharmacogenetics studies in STAR*D: strengths, limitations, and results. *Psychiatr Serv* **60**(11)**,** 1446-1457. https://doi.org/10.1176/appi.ps.60.11.1446.

Lawrence AJ, Stahl D, Duan S, Fennema D, Jaeckle T, Young AH, Dazzan P, Moll J and Zahn R (2022) Neurocognitive measures of self-blame and risk prediction models of recurrence in major depressive disorder. *Biol Psychiatry Cogn Neurosci Neuroimaging* **7**(3)**,** 256-264. https://doi.org/10.1016/j.bpsc.2021.06.010.

Lee Y, Ragguett RM, Mansur RB, Boutilier JJ, Rosenblat JD, Trevizol A, Brietzke E, Lin K, Pan Z, Subramaniapillai M, Chan TCY, Fus D, Park C, Musial N, Zuckerman H, Chen VC, Ho R, Rong C and McIntyre RS (2018) Applications of machine learning algorithms to predict therapeutic outcomes in depression: a meta-analysis and systematic review. *J Affect Disord* **241,** 519-532. https://doi.org/10.1016/j.jad.2018.08.073.

Lopresti AL, Maker GL, Hood SD and Drummond PD (2014) A review of peripheral biomarkers in major depression: the potential of inflammatory and oxidative stress biomarkers. *Prog Neuropsychopharmacol Biol Psychiatry* **48,** 102-111. https://doi.org/10.1016/j.pnpbp.2013.09.017.

Lythe KE, Gethin JA, Workman CI, Lambon Ralph MA, Deakin JFW, Moll J and Zahn R (2020) Subgenual activation and the finger of blame: individual differences and depression vulnerability. *Psychol Med***,** 1-9. https://doi.org/10.1017/S0033291720003372.

Mayberg H and Dunlop BW (2023) Balancing the beautiful and the good in pursuit of biomarkers for depression. *World Psychiatry* **22**(2)**,** 265-267. https://doi.org/10.1002/wps.21081.

Nierenberg AA and DeCecco L (2001) Definitions of antidepressant treatment response, remission, nonresponse, partial response, and other relevant treatment outcomes: a focus on treatment-resistant depression. *Journal of Clinical Psychiatry* **62,** 5-9.

Nolen-Hoeksema S, Wisco BE and Lyubomirsky S (2008) Rethinking rumination. *Perspect Psychol Sci* **3**(5)**,** 400-424. https://doi.org/10.1111/j.1745-6924.2008.00088.x.

Park C, Pan Z, Brietzke E, Subramaniapillai M, Rosenblat JD, Zuckerman H, Lee Y, Fus D and McIntyre RS (2018) Predicting antidepressant response using early changes in cognition: a systematic review. *Behav Brain Res* **353,** 154-160. https://doi.org/10.1016/j.bbr.2018.07.011.

Patel MJ, Khalaf A and Aizenstein HJ (2016) Studying depression using imaging and machine learning methods. *Neuroimage Clin* **10,** 115-123. https://doi.org/10.1016/j.nicl.2015.11.003.

Perlis RH (2013) A clinical risk stratification tool for predicting treatment resistance in major depressive disorder. *Biol Psychiatry* **74**(1)**,** 7-14. https://doi.org/10.1016/j.biopsych.2012.12.007.

Perlman K, Benrimoh D, Israel S, Rollins C, Brown E, Tunteng JF, You R, You E, Tanguay-Sela M, Snook E, Miresco M and Berlim MT (2019) A systematic meta-review of predictors of antidepressant treatment outcome in major depressive disorder. *J Affect Disord* **243,** 503-515. https://doi.org/10.1016/j.jad.2018.09.067.

Perna G, Alciati A, Dacco S, Grassi M and Caldirola D (2020) Personalized psychiatry and depression: the role of sociodemographic and clinical variables. *Psychiatry Investig* **17**(3)**,** 193-206. https://doi.org/10.30773/pi.2019.0289.

Phillips ML, Chase HW, Sheline YI, Etkin A, Almeida JR, Deckersbach T and Trivedi MH (2015) Identifying predictors, moderators, and mediators of antidepressant response in major depressive disorder: neuroimaging approaches. *Am J Psychiatry* **172**(2)**,** 124-138. https://doi.org/10.1176/appi.ajp.2014.14010076.

Poirot MG, Ruhe HG, Mutsaerts HMM, Maximov II, Groote IR, Bjørnerud A, Marquering HA, Reneman L and Caan MWA (2024) Treatment response prediction in major depressive disorder using multimodal MRI and clinical data: Secondary analysis of a randomized clinical trial. *Am J Psychiatry* **181**(3)**,** 223-233. https://doi.org/10.1176/appi.ajp.20230206.

Price JL and Drevets WC (2010) Neurocircuitry of mood disorders. *Neuropsychopharmacology* **35**(1)**,** 192-216. https://doi.org/10.1038/npp.2009.104.

Ressler KJ and Mayberg HS (2007) Targeting abnormal neural circuits in mood and anxiety disorders: from laboratory to the clinic. *Nat Neurosci* **10**(9)**,** 1116-1124.

Rorden C and Brett M (2000) Stereotaxic display of brain lesions. *Behavioural Neurology* **12**(4)**,** 191-200. https://doi.org/10.1155/2000/421719.

Rost N, Binder EB and Bruckl TM (2022) Predicting treatment outcome in depression: an introduction into current concepts and challenges. *Eur Arch Psychiatry Clin Neurosci*. https://doi.org/10.1007/s00406-022-01418-4.

Ruhe HG, Mocking RJT, Figueroa CA, Seeverens PWJ, Ikani N, Tyborowska A, Browning M, Vrijsen JN, Harmer CJ and Schene AH (2019) Emotional biases and recurrence in major depressive disorder. Results of 2.5 years follow-up of drug-free cohort vulnerable for recurrence. *Front Psychiatry* **10,** 145. https://doi.org/10.3389/fpsyt.2019.00145.

Rush AJ, Trivedi MH, Ibrahim HM, Carmody TJ, Arnow B, Klein DN, Markowitz JC, Ninan PT, Kornstein S, Manber R, Thase ME, Kocsis JH and Keller MB (2003) The 16-item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. *Biological Psychiatry* **54**(5)**,** 573-583. https://doi.org/10.1016/s0006-3223(02)01866-8.

Rush AJ, Trivedi MH, Wisniewski SR, Nierenberg AA, Stewart JW, Warden D, Niederehe G, Thase ME, Lavori PW, Lebowitz BD, McGrath PJ, Rosenbaum JF, Sackeim HA, Kupfer DJ, Luther J and Fava M (2006) Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. *American Journal of Psychiatry* **163**(11)**,** 1905-1917. https://doi.org/10.1176/ajp.2006.163.11.1905.

Schmaal L, Marquand AF, Rhebergen D, van Tol MJ, Ruhe HG, van der Wee NJ, Veltman DJ and Penninx BW (2015) Predicting the naturalistic course of major depressive disorder using clinical and multimodal neuroimaging information: a multivariate pattern recognition study. *Biol Psychiatry* **78**(4)**,** 278-286. https://doi.org/10.1016/j.biopsych.2014.11.018.

Schmidt HD, Shelton RC and Duman RS (2011) Functional biomarkers of depression: diagnosis, treatment, and pathophysiology. *Neuropsychopharmacology* **36**(12)**,** 2375-2394. https://doi.org/10.1038/npp.2011.151.

Simon GE and Perlis RH (2010) Personalized medicine for depression: can we match patients with treatments? *Am J Psychiatry* **167**(12)**,** 1445-1455. https://doi.org/10.1176/appi.ajp.2010.09111680.

Spitzer RL, Kroenke K, Williams JBW and Lowe B (2006) A brief measure for assessing generalised anxiety disorder: the GAD-7. *Archives of Internal Medicine* **166**(10)**,** 1092-1097. https://doi.org/10.1001/archinte.166.10.1092.

Strawbridge R, Young AH and Cleare AJ (2017) Biomarkers for depression: recent insights, current challenges and future prospects. *Neuropsychiatr Dis Treat* **13,** 1245-1262. https://doi.org/10.2147/NDT.S114542.

Uher R, Tansey KE, Dew T, Maier W, Mors O, Hauser J, Zvezdana Dernovsek M, Henigsberg N, Souery D, Farmer A and McGuffin P (2014) An inflammatory biomarker as a differential predictor of outcome of depression treatment with escitalopram and nortriptyline. *Am J Psychiatry* **171**(12)**,** 1278-1286. https://doi.org/10.1176/appi.ajp.2014.14010094.

Weiner B (1985) An attributional theory of achievement motivation and emotion. *Psychol Rev* **92**(4)**,** 548-573.

Williams LM, Korgaonkar MS, Song YC, Paton R, Eagles S, Goldstein-Piekarski A, Grieve SM, Harris AW, Usherwood T and Etkin A (2015) Amygdala reactivity to emotional faces in the prediction of general and medication-specific responses to antidepressant treatment in the randomized iSPOT-D trial. *Neuropsychopharmacology* **40**(10)**,** 2398-2408. https://doi.org/10.1038/npp.2015.89.

Young KD, Siegle GJ, Zotev V, Phillips R, Misaki M, Yuan H, Drevets WC and Bodurka J (2019) Randomized clinical trial of real-time fMRI amygdala neurofeedback for major depressive disorder: Effects on symptoms and autobiographical memory recall. *American Journal of Psychiatry* **174,** 748-755. https://doi.org/10.1176/appi.ajp.2017.16060637.

Zahn R, Lythe KE, Gethin JA, Green S, Deakin JF, Young AH and Moll J (2015) The role of self-blame and worthlessness in the psychopathology of major depressive disorder. *J Affect Disord* **186,** 337-341. https://doi.org/10.1016/j.jad.2015.08.001.

Tables

Table 1. Baseline demographic and clinical characteristics by responders and non-responders (n=38).

Characteristic	**Responders (n=8)**	**Non-responders (n=30)**
	n (%) or mean ± SD; range	
Age, in years	42.9 ± 16.1; 19-66	41.6 ± 14.8; 20-62
Sex		
Female	7 (88)	24 (80)
Male	1 (13)	5 (17)
Other	0 (0)	1 (3)
Ethnicity[a]		
Asian	1 (13)	24 (80)
Black	0 (0)	2 (7)
White	5 (63)	3 (10)
Other ethnicity	1 (13)	1 (3)
Years of education, in years	17.4 ± 3.3; 12-22	16.9 ± 3.6; 10-24
Depression severity		
Current MDE duration, in months	6.3 ± 5.3; 1-15	32.8 ± 50.2; 1-176
Number of MDEs	7.3 ± 5.6; 3-20	6.2 ± 4.8; 1-20
MM-PHQ-9 total score	20.0 ± 3.8; 13-25	18.4 ± 4.9; 8-27
QIDS-SR16 total score	17.9 ± 3.9; 11-22	17.2 ± 3.5; 10-23
MADRS total score	29.5 ± 5.0; 23-38	32.0 ± 4.9; 22-42
SOFAS total score	55.9 ± 3.5; 52-61	53.1 ± 5.7; 33-61
Maudsley Staging Method		
Mild	3 (38)	12 (40)
Moderate	5 (63)	18 (60)
Severe	0 (0)	0 (0)
MDD DSM-5 subtype		
Anxious distress only	0 (0)	7 (23)
Melancholic features + anxious distress	1 (13)	4 (13)
Atypical features only	0 (0)	1 (3)
Atypical features + anxious distress	3 (38)	15 (50)
No specific subtype	4 (50)	3 (10)
Treatment at baseline		
SSRI	6 (75)	25 (83)
SNRI	1 (13)	3 (10)
Other class	1 (13)	2 (7)
Non-pharmacological treatment	4 (50)	6 (20)
GAD-7 total score	10.4 ± 6.6; 1-21	11.6 ± 3.6; 5-20

Table 1. Continued.

Life-time axis-I co-morbidity		
Posttraumatic stress disorder	2 (25)	15 (50)
Other anxiety disorder	4 (50)	12 (40)
Obsessive-compulsive disorder	0 (0)	3 (10)
Eating disorder	3 (38)	10 (33)
None	2 (25)	3 (10)

[a] Missing data for one participant. Ethnicity categories have been combined: "White" includes White: British, Other, and Polish; "Asian" includes Asian or Asian British: Indian, Chinese, and Other Asian; "Black" includes Black or Black British: Caribbean.

Percentages may not add up to 100 due to rounding. MDD = major depressive disorder; DSM-5 = Diagnostic and Statistical Manual for Mental Disorders 5th edition; MDE = major depressive episode; SD = standard deviation; MM-PHQ-9 = Maudsley Modified Patient Health Questionnaire, 9 items; QIDS-SR16 = Quick Inventory Depressive Symptomatology, self-rated, 16 items; MADRS = Montgomery-Åsberg Depression Rating Scale; SOFAS = Social and Occupational Functioning Scale; SSRI = selective serotonin reuptake inhibitor; SNRI = selective norepinephrine reuptake inhibitor; GAD-7 = Generalised Anxiety Disorder, 7 items.

Table 2. Prediction models of clinical outcomes in depression (n=38).

	Model parameters				Overall model	
	β	**SE**	***t***	***p***	R^2	***p***
Standard clinical variables model					.03	.81
Baseline MM-PHQ-9	.15	10.86	.01	.99		
Baseline GAD-7	3.76	8.53	.44	.66		
Maudsley Staging Method	3.37	5.35	.63	.53		
fMRI measures model					.32	.01*
Baseline MM-PHQ-9	4.19	6.97	.60	.55		
Self-blame-selective RSATL-BA25 connectivity	-7.28	4.25	-1.71	.10		
Bilateral amygdala BOLD activation for sad vs happy subliminal faces	11.11	5.04	2.21	.04*		
Resting-state posterior subgenual cortex-VLPFC/insula functional connectivity	-8.15	4.18	-1.95	.06		

* significant at $p < .05$ threshold, two-tailed. SE = standard error; MM-PHQ-9 = Maudsley Modified Patient Health Questionnaire, 9 items; GAD-7 = Generalised Anxiety Disorder, 7 items; RSATL = right superior anterior temporal lobe; BA = Brodmann Area; BOLD = blood-oxygen level-dependent; VLPFC = ventrolateral prefrontal cortex.

Figure Legends

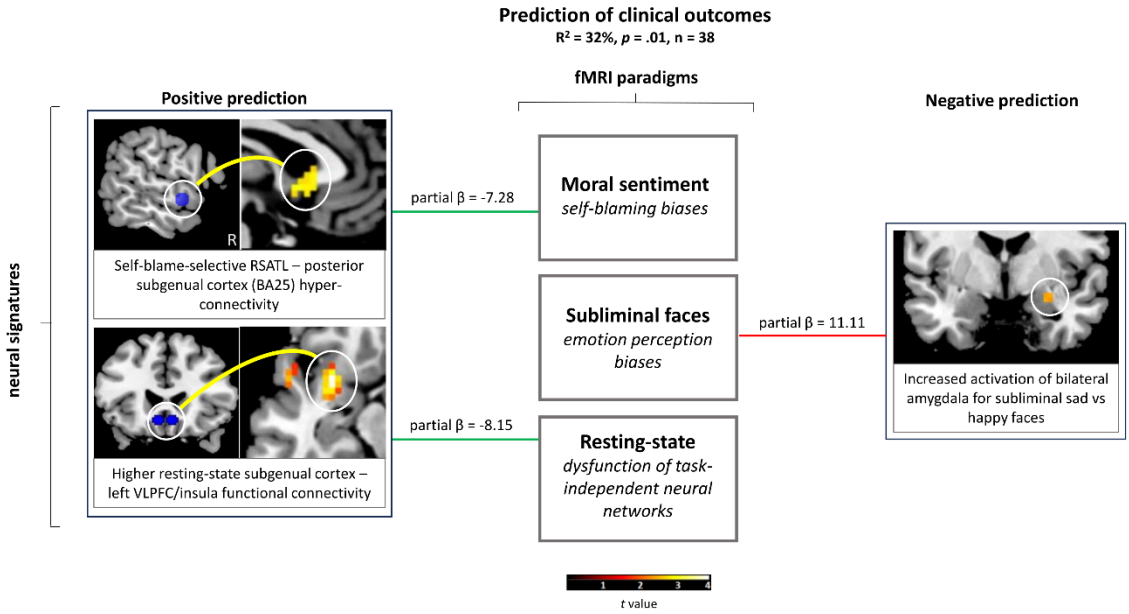

Figure 1 | Neural signatures of emotional biases associated with clinical outcomes in difficult-to-treat MDD.

Three neural signatures of emotional biases were associated with clinical outcomes in UK primary care. More specifically, it shows cropped sections of voxel-based analyses illustrating the respective pre-registered *a priori* regions-of-interest, i.e. self-blame-selective right superior anterior temporal lobe-posterior subgenual cortex (BA25) connectivity, resting-state functional connectivity between the subgenual cortex and ventrolateral prefrontal cortex/insula, and bilateral amygdala blood-oxygen level-dependent activation in response to subliminal sad vs happy faces. These cropped sections are displayed using MRIcron at an uncorrected voxel-level threshold of p=.005, with no cluster-size threshold (the colour bar represents t values) and adapted from figures previously published (Fennema et al. 2023; Fennema, Barker, O'Daly, Duan, Carr, et al. 2024; Fennema, Barker, O'Daly, Duan, Godlewska, et al. 2024). A linear model using the pre-registered fMRI measures with baseline Maudsley Modified Patient Health Questionnaire (9 items) as a covariate explained 32% of the variance of QIDS-SR16 percentage change. The red and green lines display the partial effects of the fMRI measures on

the variance of QIDS-SR16 percentage change after four months of standard primary care. MDD = major depressive disorder; BA = Brodmann Area; RSATL = right superior anterior temporal lobe; VLPFC = ventrolateral prefrontal cortex; BOLD = blood-oxygen level-dependent; QIDS-SR16 = Quick Inventory of Depressive Symptomatology, self-rated (16 items).

Supplementary Online Content

Neural signatures of emotional biases predict clinical outcomes in difficult-to-treat depression

Diede Fennema[1], Gareth J. Barker[2], Owen O'Daly[2], Beata R. Godlewska[3,4], Ewan Carr[5], Kimberley Goldsmith[5], Allan H. Young[1,6], Jorge Moll[7] & Roland Zahn[1,6,7*]

[1] *Centre of Affective Disorders, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK*

[2] *Department of Neuroimaging, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK*

[3] *Psychopharmacology Research Unit, University Department of Psychiatry, University of Oxford, Oxford, UK*

[4] *Oxford Health NHS Foundation Trust, Warneford Hospital, Oxford, UK*

[5] *Department of Biostatistics and Health Informatics, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK*

[6] *National Service for Affective Disorders, South London and Maudsley NHS Foundation Trust, London, UK*

[7] *Cognitive and Behavioral Neuroscience Unit, D'Or Institute for Research and Education (IDOR), Pioneer Science Program, Rio de Janeiro, Brazil*

* Corresponding author

Professor Roland Zahn (see address above)
E-mail: roland.zahn@kcl.ac.uk
Phone: 0044-(0)20 7848 0348
Fax: 0044-(0)20 7848 0298

Keywords: *depression; prediction; fMRI; prognosis; difficult-to-treat*

Parts of this Supplementary Online Content have been adapted from a previously published one in Neuroimage:Clinical (doi: 10.1016/j.nicl.2023.103453).

Supplementary Methods

Additional inclusion/exclusion criteria

In addition to the criteria mentioned in the main manuscript, participants were included if they met any of the following: aged 18 years and over, currently experiencing a major depressive episode (MDE) and at least moderately severe depressive syndrome on the Patient Health Questionnaire (PHQ-9; score ≥ 15) (Spitzer et al. 1999), and non-responders to at least two serotonergic antidepressants from the following list in current or previous episodes: citalopram, fluoxetine, sertraline, escitalopram, paroxetine, venlafaxine or duloxetine.

Participants were excluded if they met any of the following: previous prescription of mirtazapine or vortioxetine at therapeutic dose, MRI contraindications, currently receiving specialist psychiatric treatment, high suicide risk on the Mini International Neuropsychiatric Interview (MINI) suicidality screen (Sheehan et al. 1998), past diagnosis of schizophrenia or schizo-affective disorder, psychotic symptoms using clinical screening questions, bipolar disorder, at risk of being violent, drug or alcohol abuse over the last six months, suspected neurological condition, pregnancy or insufficient contraception in women of childbearing age and breastfeeding or within six months of giving birth.

Recruitment and clinical assessment

We recruited participants from September 2018 to March 2020 partly through a cluster-randomised feasibility clinical trial, the Antidepressant Advisor Study (ADeSS; NCT03628027), which evaluated the feasibility of a novel computerised decision support algorithm for antidepressant medications in patients with major depressive disorder (MDD) in primary care (Harrison et al. 2020; Harrison et al. 2022). Participants enrolled in the ADeSS trial were assigned to either i) use of a computerised decision-support tool by their general practitioner (GP) to assist with antidepressant choices, or ii) treatment-as-usual, and were asked to attend an optional MRI session. The computerised decision-support tool implemented National Institute for Health and Care Excellence guidelines, prompting GPs to increase the dose or switch to another antidepressant, and resembled standard care. Recruitment was halted due to the COVID-19 pandemic and recommenced in October 2020, using online advertising only, and was completed in August 2021.

As described in the trial protocol (Harrison et al. 2020), GP practices screened for patients with a history of treatment-resistance to antidepressant medications within their practice, i.e. non-responders to at least two serotonergic antidepressants in the current or previous episodes. Potential participants were approached for consent and, if given, asked to fill in a pre-screening questionnaire. Potentially eligible participants were invited for an in-depth assessment by the study team, which included a clinical assessment using the Structured Clinical Interview for DSM-5 (SCID) to establish a current MDD (First et al. 2015), a history of participants' depressive episodes, their current and past antidepressant medications, and completing various clinical, behavioural and experimental measures.

A follow-up assessment was conducted to establish whether any changes in baseline measures had occurred. This visit took place around 14-18 weeks after enrolling in the study, which should allow observation of any treatment effect if there is one. The assessment included questions related to medications taken in the study period as well as various clinical and behavioural measures. The main clinical measures collected at baseline and follow-up were the Quick Inventory of Depressive Symptomology (16 items, self-rated; QIDS-SR16) (Rush et al. 2003), Maudsley Modified Patient Health Questionnaire (9 items; MM-PHQ-9) (Harrison et al. 2021), Generalised Anxiety Disorder (7 items; GAD-7) (Spitzer et al. 2006), Montgomery-Åsberg Depression Rating Scale (MADRS) (Montgomery and Asberg 1979), and Social and Occupational Functioning Assessment Scale (SOFAS, part of SCID) (First et al. 2015). Please refer to the ADeSS trial protocol for more details regarding these procedures (Fennema 2022; Harrison et al. 2020).

As the ADeSS trial was stopped due to the COVID-19 pandemic, an alternative recruitment route was employed to continue recruitment for the observational fMRI study. Trial adverts were posted online, with further dissemination of study adverts via university and institutional recruitment circulars. Interested participants were asked to complete a similar pre-screening questionnaire as those approached for the ADeSS trial. If potentially eligible, participants were invited for an in-depth assessment to confirm their eligibility. This group resembled the treatment-as-usual arm in the ADeSS trial, receiving standard care. For more details, please see Fennema (2022).

A total of 1,755 participants with a history of MDD showed interest in participating and completed a pre-screening questionnaire. Potentially eligible MDD participants (n = 89) for the ADeSS trial and the fMRI study were invited to attend an in-depth assessment. Of those, 45 participants enrolled in the fMRI study, attended their MRI session and completed the study.

Of those 45 participants, ten participants were also part of the ADeSS trial (support tool arm: n = 4; treatment-as-usual arm: n = 6).

Sample size

As there was no previous comparable study from which effect sizes could be drawn, we carried out a sensitivity analysis showing that for determining an at least 20% above chance level prediction model performance, a minimum of n = 44 MDD patients was required to achieve 85% power at $p = .05$ using a binomial test.

fMRI paradigms

Participants were asked to complete three fMRI paradigms: the moral sentiment task, the subliminal faces task and a resting-state scan:

1) *Moral sentiment task, as described in Fennema et al. (2023)*

Participants were shown an optimised and shortened version of the fMRI paradigm outlined by Green et al. (2012) and Lythe et al. (2015). For details on the optimisation, please see Duan et al. (2023) and Fennema (2022). In brief, participants were shown 54 short written statements describing actions counter to social and moral values described by social concepts (e.g. impatient, dishonest) in which the agent was either the participant (self-agency condition [number of stimuli = 27]) or their best friend (other-agency condition [number of stimuli = 27]). Participants were asked to name their best friend prior to the scanning session to allow personalisation of the statements. Self- and other-agency conditions used the same social concepts (e.g. self-agency "Tom is dishonest towards Pete" and e.g. other-agency "Pete is dishonest towards Tom"). In addition, there were 27 low-level null events as a baseline condition, i.e. fixation of a visual pattern with no button press or other response required.

Stimuli were presented in an event-related design for a maximum of 5 seconds, within which time participants had to decide whether they would feel that the imagined behaviours were "quite unpleasant" or "mildly unpleasant" from their own perspective. The stimuli were presented in a pseudo-random order, presented at jittered intervals with a mean of 4000ms (with steps of 500ms). The total task time was 12 minutes and 9 seconds.

2) *Subliminal faces task, as described in Fennema, Barker, O'Daly, Duan, Godlewska, et al. (2024)*

Participants completed a backward masking task based on the fMRI paradigm outlined by Godlewska et al. (2018). Participants were shown pairs of faces, with a first "target" face (expressing a sad, happy, or neutral emotion), displayed for 34 milliseconds, and then immediately "masked" by a face of neutral expression, displayed for 66 milliseconds. The task followed a block design, with each participant being shown four blocks with sad faces, four blocks with happy faces and nine blocks with neutral faces. Each block cycled through ten target-mask pairs of faces, with the order varying for each block. The neutral (N) blocks were interleaved with sad (S) and happy (H) blocks, in one of two orders: N-S-N-H-N-S-H-N or N-H-N-S-N-H-N-S-N. The order of blocks was determined by pseudo-randomisation, with an even split within the MDD and control groups and across the total sample. After each block, there was a 10-second block of baseline fixation. The total task time was 8 minutes and 47 seconds.

3) Resting-state scan, as described in Fennema, Barker, O'Daly, Duan, Carr, et al. (2024)

The resting-state scan was based on the methodology as outlined by Dunlop et al. (2017). Participants were shown a fixation cross on the screen and were instructed to keep their eyes open and let their mind wander while focusing on the cross. The total scan time was 7 minutes and 24 seconds.

Image acquisition

Image acquisition was carried out on an MR750 3T MR system (GE Healthcare, Chicago, USA), using a Nova Medical 32-channel head coil. The scanning session started with approximately 20 minutes of structural imaging, acquiring T1-weighted, T2*-weighted and Fluid-Attenuated Inversion Recovery (FLAIR) images, followed by approximately 30 minutes of fMRI paradigms (moral sentiment, subliminal faces and resting-state). While in the MRI scanner, the participant's head motion was restricted using padding, and heart rate and respiration rate measurements were recorded via a manufacturer-supplied finger pulse sensor (peripheral plethysmograph) and respiratory belt, respectively. A mirror fitted to the head coil allowed participants to view visual stimuli presented during image acquisition, as stimuli were projected onto a screen located behind the participant's head. Verbal instructions were communicated via the MRI intercom, using a pre-defined script to ensure consistency between participants.

High-resolution anatomical images were acquired with a 3D Inversion Recovery prepared Spoiled Gradient Echo sequence (IR-SPGR; repetition time (TR) = 7.3 ms; echo time (TE) = 3.02 ms; inversion time (TI) = 400 ms; matrix = 256 x 256; excitation flip angle = 11 degrees; field-of-view (FOV) = 270 mm; slice thickness = 1.2 mm, 196 slices). Images for incidental findings review were acquired using a 2D Fast-Recovery Fast Spin-Echo (TR = 4380 ms; TE = 64.85 ms; matrix = 320 x 256; refocusing flip angle = 111 degrees; FOV = 240; 2 mm contiguous slices, 72 slices) and 2D FLAIR sequence (TR = 8000 ms; TE = 128.41 ms; matrix = 256 x 128; refocusing flip angle = 111 degrees; FOV = 220; 4 mm continuous slices, 36 slices) and checked by a neuroradiologist at King's College London Hospital for any significant brain abnormalities that might warrant follow up, independent of additional, internal checks by the study team.

For all three paradigms, shimming was automatically applied as part of the scanner's "pre-scan" procedures, and four additional volumes were acquired and automatically discarded at the start of each fMRI run, allowing for T1 equilibration effects:

1) *Moral sentiment task, as described in Fennema et al. (2023)*

Functional image acquisition was obtained in the anterior commissure – posterior commissure plane, with slices running top to bottom, using a T2*-weighted echo-planar imaging blood-level oxygen-dependent (BOLD) sequence (TR = 2000ms; TE = 20ms; matrix = 64x64; FOV = 211mm; flip angle = 75 degrees; slice thickness = 2.9mm, slice gap = 0.1mm, inter-slice distance = 3mm, 41 slices, 368 volumes).

2) *Subliminal faces task, as described in Fennema, Barker, O'Daly, Duan, Godlewska, et al. (2024)*

Functional image acquisition was obtained parallel to the anterior commissure – posterior commissure plane, with slices running top to bottom, using a standard T2*-weighted echo-planar imaging BOLD sequence (TR = 2000 ms; TE = 30 ms; matrix = 64 x 64; FOV = 240 mm; flip angle = 75 degrees; slice thickness = 3 mm, slice gap = 0.3 mm, inter-slice distance = 3.3 mm, 41 slices, 267 volumes).

3) *Resting-state scan, as described in Fennema, Barker, O'Daly, Duan, Carr, et al. (2024)*

Resting-state echo-planar images were acquired using a sequence which was optimised for the detection of ventral frontal signal (222 volumes; 41 slices; descending sequential acquisition;

TR = 2000ms; TE = 20ms; matrix = 64 x 64; FOV = 211mm; flip angle = 75 degrees; slice thickness = 2.9mm, slice gap = 0.1mm, inter-slice distance = 3mm).

Image analysis

Statistical Parametric Mapping (SPM) 12 (http://www.fil.ion.ucl.ac.uk/spm12) was used for BOLD effect analysis and psychophysiological interaction (PPI) analysis (moral sentiment paradigm, subliminal faces paradigm), while Data Processing Assistant for Resting-State fMRI (DPARSF) (Chao-Gan and Yu-Feng 2010) was used for resting-state analysis.

1) Moral sentiment task, as described in Fennema et al. (2023)

Standard pre-processing steps were followed: functional images were realigned, unwarped and co-registered to the participant's T1 images. These images were normalised to the co-registered T1 image and resliced at a voxel size of 3 x 3 x 3 mm. A smoothing kernel of full-width half-maximum equal to 6 mm was used. No slice timing correction was applied. Motion correction was applied in the form of censoring, i.e. identifying outliers based on framewise displacement and regressing them from the fMRI timeseries. Framewise displacement was calculated using Brain and Mind Lab (BRAMILA) tools (https://github.com/spunt/bspm/blob/master/thirdparty/bramila/bramila_framewiseDisplacement.m) to identify outliers regarding motion. Any framewise displacement of ≥ 0.5 mm was marked as a spike in movement and scan nulling regressors were added to the standard six motion parameters describing movement by rotation and translation to account for the spike(s). Participants with spikes in more than 25% of the functional images overall were deemed to have moved too much and were excluded from the analysis. We chose a threshold of 25% of motion-contaminated volumes in combination with the threshold of any framewise displacement of ≥ 0.5 mm as a trade-off between retaining patient data with reasonable quality and avoiding overfitting with too many scanning nulling regressors.

At the individual level, BOLD effects were modelled for the self-agency condition, other-agency condition and null event, with an event duration of 0 seconds. Movement parameters (i.e. six parameters describing movement by rotation and translation in three dimensions each, plus any scan nulling regressors) were included as covariates. No time and dispersion derivatives were modelled.

Connectivity was determined using PPI analysis. We extracted the signal from our pre-registered seed region, i.e. the right superior anterior temporal lobe (RSATL; Montreal Neurological Institute [MNI] coordinates: x = 58, y = 0, z = -12; 6 mm sphere), and created

interaction terms for the psychological variable (main effect of condition, i.e. self-agency vs. fixation and other-agency vs. fixation) with the physiological variable (the right superior anterior temporal lobe signal time course irrespective of condition).

2) *Subliminal faces task, as described in Fennema, Barker, O'Daly, Duan, Godlewska, et al. (2024)*

Standard pre-processing steps were followed: functional images were realigned, unwarped and co-registered to the participant's T1 images. These images were normalised to the co-registered T1 image and resliced at a voxel size of 3 x 3 x 3 mm. A smoothing kernel of full-width half-maximum equal to 6 mm was used. No slice timing correction was applied. Motion correction was applied in the form of censoring. Framewise displacement was calculated using BRAMILA tools to identify outliers regarding motion. Any framewise displacement of ≥ 1 mm was marked as a spike in movement and scan nulling regressors were added to the standard six motion parameters describing movement by rotation and translation to account for the spike(s). Participants with spikes in more than 10% of the functional images overall were deemed to have moved too much and were excluded from the analysis. We chose a threshold of 10% of motion-contaminated volumes in combination with the threshold of any framewise displacement of $\geq$ 1 mm as a trade-off between retaining patient data with reasonable quality and avoiding overfitting with too many scanning nulling regressors.

Additional noise correction was applied: the MATLAB PhysIO toolbox was used to partially mitigate the impact of physiological noise (Kasper et al. 2017) (version R2021a-v8.0.0, open-source code available as part of the Translational Algorithms for Psychiatry-Advancing Science software collection (Frassle et al. 2021): https://www.translationalneuromodeling .org/tapas). Heart rate and respiration rate measurements were used in a retrospective image correction (RETROICOR) model, using Fourier expansions of different orders for the estimated phases of cardiac pulsation (third order), respiration (fourth order) and cardio-respiratory interactions (first order) (Harvey et al. 2008).

BOLD effects were modelled for each of the emotion blocks, i.e. sad, happy and neutral. Baseline fixation was not modelled to avoid overspecification of the model. Nuisance regressors created by the PhysIO toolbox, i.e. physiological noise regressors and motion-related regressors, were included as covariates. Contrasts were created to examine the subtraction-based difference between sad and happy faces (sad vs. happy).

3) Resting-state scan, as described in Fennema, Barker, O'Daly, Duan, Carr, et al. (2024)

The resting-state fMRI pre-processing followed a similar approach to that outlined in Workman et al. (2016), using DPARSF and Artifact Detection Tools (ART). SPM8 was used for pre-processing steps to ensure compatibility with DPARSF.

Functional resting-state echo-planar images (EPIs) and IR-SPGR anatomical images underwent standard pre-processing steps in DPARSF. ART was used to flag spikes in motion, i.e. framewise signal intensity > 3 standard deviation from the global mean and framewise head displacement > 1mm, and to create nulling regressors. Participants with spikes in more than 10% of the functional images were deemed to have moved too much and were excluded from the analysis.

In addition, the MATLAB PhysIO toolbox was used to partially mitigate the impact of physiological noise (Kasper et al. 2017). Heart rate and respiration rate measurements were used in a RETROICOR model, using Fourier expansions of different orders for the estimated phases of cardiac pulsation (second order), respiration (second order) and cardio-respiratory interactions (first order) (Glover et al. 2000).

Following this initial pre-processing, EPIs underwent linear detrending and nuisance covariates regression (6 motion parameters (Bright and Murphy 2015), white matter signal, cerebrospinal fluid signal, ART regressors and PhysIO regressors) and normalisation using non-linear transformation parameters derived during segmentation. Band-pass filtering was applied to retain frequencies between 0.01 and 0.08 Hz.

Functional connectivity maps were computed using the fully pre-processed functional images for each participant by correlating the average time course within the seed region (i.e. subgenual frontal cortex [Brodmann Area [BA] 25]) with the time course of each voxel within the brain, which were Fisher Z-transformed.

Description of treatment change

As part of the study, participants were encouraged to book an appointment with their GP to review their treatment. Exploratory analysis was undertaken to attribute a reduction in depressive symptoms to a change in pharmacological treatment, i.e. an increase in dose or a change to another medication. More specifically, participants were classified as follows: i) no change, i.e. participants who did not make any changes to their treatment, stopped taking their antidepressant, or lowered the dose of their current antidepressant; ii) minimal change, i.e. participants who increased their current antidepressant from an effective dose to a higher dose, or who changed to another antidepressant at an ineffective dose; and iii) relevant change, i.e.

participants who increased their current antidepressant from an ineffective dose to an effective dose, or who changed to another antidepressant at an effective dose.

Some participants had more than one change in their treatment during the follow-up period. In those cases, the change most relevant for the clinical outcome measure was used, which was usually the change occurring closest to the follow-up assessment. Moreover, the change to another antidepressant at an effective dose or an increase in dose had to have lasted at least two weeks prior to the follow-up assessment to be counted a treatment trial, which is in line with the Maudsley Staging Method (Fekadu et al. 2018).

Exploratory prediction models

In addition to our main multivariate prediction model, we ran exploratory models to assess the contribution of the pre-registered fMRI measures individually, with baseline MM-PHQ-9 as a covariate: "self-blaming biases" with a focus on self-blame-selective connectivity between the right superior anterior temporal lobe and the posterior subgenual cortex (BA25), "negative perceptual biases" with a focus on bilateral amygdala BOLD activation for subliminal sad vs happy faces, and "subgenual resting-state networks" with a focus on subgenual resting-state connectivity with the ventrolateral prefrontal cortex/insula (BA47).

Moreover, we ran an exploratory "pre-registration" model which included three additional neural measures as originally outlined in our pre-registered protocol (NCT04342299), i.e. functional resting-state subgenual cortex connectivity with the left ventromedial prefrontal cortex (BA10) and with the dorsal midbrain, as well as pregenual anterior cingulate cortex BOLD activation for subliminal sad vs happy faces.

Supplementary Results

Exploratory prediction models

There was no association between baseline MM-PHQ-9 and the pre-registered fMRI measures, i.e. self-blame-selective connectivity between the right superior anterior lobe and the posterior subgenual cortex (BA25; r_s [30] = -.19, p = .32), bilateral amygdala BOLD activation to sad vs happy (r_s [30] = .03, p = .86) and resting-state connectivity between the subgenual cortex and ventrolateral prefrontal cortex/insula (r_s [30] = -.18, p = .33).

The "self-blaming biases" model explained 12% of the variance in QIDS-SR16 percentage (F[2,35] = 2.45, p = .10, R^2 = .12, R^2_{adjusted} = .07; Supplementary Table 3). Self-blame-selective connectivity between the right superior anterior temporal lobe and the posterior

subgenual cortex (BA25) negatively contributed to the variance in QIDS-SR16 percentage change (partial β = -9.59, t[35] = -2.10), while there was a positive contribution by baseline MM-PHQ-9 (partial β = 3.17, t[35] = .42).

The "negative perceptual biases" model explained 18% of the variance in QIDS-SR16 percentage (F[2,35] = 3.79, p = .03, R^2 = .18, $R^2_{adjusted}$ = .13; Supplementary Table 3). Bilateral amygdala BOLD activation to sad vs happy subliminal faces positively contributed to the variance in QIDS-SR16 percentage change (partial β = 13.87, t[35] = 2.66), as did baseline MM-PHQ-9 (partial β = 7.19, t[35] = .98).

The "subgenual resting-state networks" model explained 12% of the variance in QIDS-SR16 percentage (F[2,35] = 2.28, p = .12, R^2 = .12, $R^2_{adjusted}$ = .07; Supplementary Table 3). Resting-state connectivity between the subgenual cortex and ventrolateral prefrontal cortex/insula negatively contributed to the variance in QIDS-SR16 percentage change (partial β = -9.27, t[35] = -2.01), while there was a positive contribution by baseline MM-PHQ-9 (partial β = 4.21, t[35] = .56).

The "pre-registration" model explained 33% of the variance in QIDS-SR16 percentage (F[7,30] = 2.13, p = .07, R^2 = .33, $R^2_{adjusted}$ = .18; Supplementary Table 3). Baseline MM-PHQ-9 positively contributed to the variance in QIDS-SR16 (partial β = 3.20, t[30] = .43), while there was a negative contribution by self-blame-selective connectivity between the right superior anterior temporal lobe and the posterior subgenual cortex (BA25) (partial β = -7.53, t[30] = -1.65). Bilateral amygdala and pregenual anterior cingulate cortex BOLD to sad vs happy subliminal faces both positively contributed to the variance in QIDS-SR16 percentage (partial β = 11.06, t[30] = 1.99 and partial β = 2.62, t[30] = .49, respectively), as well as resting-state subgenual functional connectivity with the ventromedial prefrontal cortex (partial β = .64, t[30] = .17) and with the dorsal midbrain (partial β = 2.69, t[30] = .63). Resting-state subgenual functional connectivity with the ventrolateral prefrontal cortex/insula negatively contributed to the variance in QIDS-SR16 (partial β = -7.58, t[30] = -1.70).

Exploratory findings responders vs. non-responders

Of the predictor variables to explore response vs. non-response (Supplementary Table 4), increased functional connectivity between the bilateral subgenual cortex and left VLPFC/insula was significantly associated with an increased likelihood of response (OR [95% CI] = 2.87 [1.05, 7.84], p = .04). Increased self-blame-selective connectivity between the right superior anterior temporal lobe and the posterior subgenual cortex (BA25) (OR [95% CI] = 1.54 [.53, 4.45], p = .08) and higher baseline MM-PHQ-9 scores (OR [95% CI] = 6.36 [.65, 61.82]) were

associated with an increased likelihood of response, whereas lower bilateral amygdala BOLD activation for sad vs happy subliminal faces was associated with a lower likelihood of response (OR [95% CI] = .226 [.04, 1.21]).

Supplementary Tables

Supplementary Table 1. Treatment during follow-up period (n=38).

Characteristic	**n (%)**
Main change	
No change in antidepressant	22 (58%)
Stopped antidepressant	5 (13%)
Lowered dose of antidepressant	0 (0%)
Increase from effective dose to higher effective dose	5 (13%)
Increase from ineffective dose to effective dose	0 (0%)
Change to another antidepressant at effective dose	4 (11%)
Change to another antidepressant at ineffective dose	2 (5%)
Main antidepressant	
SSRI	31 (82%)
Sertraline	8 (21%)
Citalopram	5 (13%)
Escitalopram	4 (11%)
Fluoxetine	3 (8%)
Venlafaxine (≤ 150mg)	6 (16%)
SNRI	4 (11%)
Duloxetine	2 (5%)
Venlafaxine (> 150mg)	2 (5%)
Mirtazapine	3 (8%)
Tricyclic antidepressant	1 (3%)
Other antidepressant	0 (0%)
Add-on treatment	5 (13%)
Change in mental health service use	
Started accessing mental health service	8 (21%)
Continued care in mental health service	8 (21%)
Stopped mental health treatment	2 (5%)
Type of mental health service use	
CBT	3 (8%)
Psychotherapy	5 (13%)
Psychoanalysis	1 (3%)
Counselling	2 (5%)
Other	5 (13%)

Percentages may not add up to 100 due to rounding. MDD = major depressive disorder; SSRI = selective serotonin reuptake inhibitor; SNRI = selective noradrenaline reuptake inhibitor; CBT = cognitive behavioural therapy.

Supplementary Table 2. Descriptive statistics for clinical symptom measures at baseline and follow-up (n=38).

	Baseline (mean ± SD; min – max)	**Follow-up (mean ± SD; min – max)**	**Difference [95% CI]**	***t* value**	***p* value**
QIDS-SR16	17.3 ± 3.5; 10 – 23	13.2 ± 5.6; 4 – 24	-4.1 [-5.8, -2.4]	-4.80	<.001
MM-PHQ-9	18.7 ± 4.7; 8 – 27	14.0 ± 7.8; 0 – 27	-4.7 [-6.9, -2.5]	-4.28	<.001
GAD-7 [a]	11.3 ± 4.3; 1 – 21	10.4 ± 5.7; 0 – 21	-1.1 [-3.0, 0.8]	-1.21	.12
MADRS	31.5 ± 4.9; 22 – 42	23.8 ± 10.9; 3 – 44	-7.7 [-10.6, -4.7]	-5.30	<.001
SOFAS	53.7 ± 5.4; 33 – 61	58.3 ± 11.2; 33 – 85	4.6 [1.7, 7.5]	3.26	.001

[a] Missing follow-up data for one participant.
MDD = major depressive disorder; CI = confidence interval; QIDS-SR16 = Quick Inventory of Depressive Symptomatology – self-rated, 16 items; MM-PHQ-9 = Maudsley Modified Personal Health Questionnaire, 9 items; GAD-7 = Generalised Anxiety Disorder, 7 items; MADRS = Montgomery-Åsberg Depression Rating Scale; SOFAS = Social and Occupational Functioning Assessment Scale. M = mean; SD = standard deviation; min = minimum; max = maximum.

Supplementary Table 3. Exploratory prediction models of clinical outcomes in depression.

	Model parameters				Overall model	
	β	SE	*t*	*p*	R^2	*p*
Pre-registration (n=38)					.33	.07
Baseline MM-PHQ-9	3.20	7.40	.43	.67		
Self-blame-selective RSATL-BA25 connectivity	-7.53	4.56	-1.65	.11		
Bilateral amygdala BOLD activation for sad vs happy subliminal faces	11.06	5.56	1.99	.06		
Pregenual anterior cingulate cortex BOLD activation for sad vs happy subliminal faces	2.62	5.31	.49	.63		
Resting-state subgenual cortex-VLPFC/insula functional connectivity	-7.58	4.45	-1.70	.10		
Resting-state subgenual cortex-VMPFC functional connectivity	.64	3.69	.17	.86		
Resting-state subgenual cortex-dorsal midbrain functional connectivity	2.69	4.28	.63	.53		
Self-blaming biases (n=38)					.12	.10
Baseline MM-PHQ-9	3.17	7.60	.42	.68		
Self-blame-selective RSATL-BA25 connectivity	-9.59	4.57	-2.10	.04*		
Negative perceptual biases (n=38)					.18	.03*
Baseline MM-PHQ-9	7.19	7.32	.98	.33		
Bilateral amygdala BOLD activation for sad vs happy subliminal faces	13.87	5.22	2.66	.01*		
Subgenual resting-state networks (n=38)					.12	.12
Baseline MM-PHQ-9	4.21	7.58	.56	.58		
Resting-state subgenual cortex-VLPFC/insula functional connectivity	-9.27	4.60	-2.01	.05		
High quality fMRI (n=30)					.43	.01*
Baseline MM-PHQ-9	4.11	7.59	.54	.59		
Self-blame-selective RSATL-BA25 connectivity	-9.88	4.90	-2.02	.06		
Bilateral amygdala BOLD activation for sad vs happy subliminal faces	9.56	6.25	1.53	.14		
Resting-state subgenual cortex-VLPFC/insula functional connectivity	-11.86	4.68	-2.54	.02*		

* significant at $p < .05$ threshold, two-tailed. SE = standard error; MM-PHQ-9 = Maudsley Modified Patient Health Questionnaire, 9 items; GAD-7 = Generalised Anxiety Disorder, 7 items; RSATL = right superior anterior temporal lobe; BA = Brodmann Area; BOLD = blood-oxygen level-dependent; VLPFC = ventrolateral prefrontal cortex; VMPFC = ventromedial prefrontal cortex.

Supplementary Table 4. Logistic regression of response vs. non-response in depression (n=38).

Predictor variable	***β***	**SE**	**Wald**	***p***	**Odds Ratio [95% CI]**
Baseline MM-PHQ-9	1.85	1.16	2.54	.11	6.36 [.65, 61.82]
Self-blame-selective RSATL-BA25 connectivity	.43	.54	.64	.43	1.54 [.53, 4.45]
Bilateral amygdala BOLD activation for sad vs happy subliminal faces	-1.49	.86	3.03	.08	.23 [.04, 1.21]
Resting-state subgenual cortex-VLPFC/insula functional connectivity	1.05	.51	4.21	.04*	2.87 [1.05, 7.84]

* significant at $p < .05$ threshold, two-tailed. SE = standard error; CI = confidence interval; MM-PHQ-9 = Maudsley Modified Patient Health Questionnaire, 9 items; GAD-7 = Generalised Anxiety Disorder, 7 items; RSATL = right superior anterior temporal lobe; BA = Brodmann Area; BOLD = blood-oxygen level-dependent; VLPFC = ventrolateral prefrontal cortex.

References

Bright MG and Murphy K (2015) Is fMRI "noise" really noise? Resting state nuisance regressors remove variance with network structure. *Neuroimage* **114,** 158-169. https://doi.org/10.1016/j.neuroimage.2015.03.070.

Chao-Gan Y and Yu-Feng Z (2010) DPARSF: A MATLAB Toolbox for "pipeline" data analysis of resting-state fMRI. *Front Syst Neurosci* **4,** 13. https://doi.org/10.3389/fnsys.2010.00013.

Duan S, Valmaggia L, Fennema D, Moll J and Zahn R (2023) Remote virtual reality assessment elucidates self-blame-related action tendencies in depression. *J Psychiatr Res* **161,** 77-83. https://doi.org/10.1016/j.jpsychires.2023.02.031.

Dunlop BW, Rajendra JK, Craighead WE, Kelley ME, McGrath CL, Choi KS, Kinkead B, Nemeroff CB and Mayberg HS (2017) Functional connectivity of the subcallosal cingulate cortex and differential outcomes to treatment with cognitive-behavioral therapy or antidepressant medication for major depressive disorder. *Am J Psychiatry* **174**(6)**,** 533-545. https://doi.org/10.1176/appi.ajp.2016.16050518.

Fekadu A, Donocik JG and Cleare AJ (2018) Standardisation framework for the Maudsley staging method for treatment resistance in depression. *BMC Psychiatry* **18**(1)**,** 100. https://doi.org/10.1186/s12888-018-1679-x.

Fennema D (2022) Neural signatures of emotional biases and prognosis in treatment-resistant depression. PhD, King's College London.

Fennema D, Barker GJ, O'Daly O, Duan S, Carr E, Goldsmith K, Young AH, Moll J and Zahn R (2023) Self-blame-selective hyper-connectivity between anterior temporal and subgenual cortices predicts prognosis in major depressive disorder. *NeuroImage: Clinical* **39,** 103453. https://doi.org/10.1016/j.nicl.2023.103453.

Fennema D, Barker GJ, O'Daly O, Duan S, Carr E, Goldsmith K, Young AH, Moll J and Zahn R (2024) The role of subgenual resting-state connectivity networks in predicting prognosis in major depressive disorder. *Biological Psychiatry: Global Open Science* **4**(3)**,** 100308. https://doi.org/10.1016/j.bpsgos.2024.100308.

Fennema D, Barker GJ, O'Daly O, Duan S, Godlewska BR, Goldsmith K, Young AH, Moll J and Zahn R (2024) Neural responses to facial emotions and subsequent clinical outcomes in difficult-to-treat depression. *Psychol Med*. https://doi.org/10.1017/S0033291724001144.

First MB, Williams JBW, Karg RS and Spitzer RL (2015) *Structured Clinical Interview for DSM-5 - Research Version (SCID-5 for DSM-5, Research Version; SCID-5-RV, Version 1.0.0)*. Arlington, VA: American Psychiatric Association.

Frassle S, Aponte EA, Bollmann S, Brodersen KH, Do CT, Harrison OK, Harrison SJ, Heinzle J, Iglesias S, Kasper L, Lomakina EI, Mathys C, Muller-Schrader M, Pereira I, Petzschner FH, Raman S, Schobi D, Toussaint B, Weber LA, Yao Y and Stephan KE (2021) TAPAS: An open-source software package for translational neuromodeling and computational psychiatry. *Front Psychiatry* **12,** 680811. https://doi.org/10.3389/fpsyt.2021.680811.

Glover GH, Li TQ and Ress D (2000) Image-based method for retrospective correction of physiological motion effects in fMRI: RETROICOR. *Magn Reson Med* **44**(1)**,** 162-167. https://doi.org/10.1002/1522-2594(200007)44:1<162::aid-mrm23>3.0.co;2-e.

Godlewska BR, Browning M, Norbury R, Igoumenou A, Cowen PJ and Harmer CJ (2018) Predicting treatment response in depression: the role of anterior cingulate cortex. *Int J Neuropsychopharmacol* **21**(11)**,** 988-996. https://doi.org/10.1093/ijnp/pyy069.

Green S, Lambon Ralph MA, Moll J, Deakin JF and Zahn R (2012) Guilt-selective functional disconnection of anterior temporal and subgenual cortices in major depressive disorder. *Arch Gen Psychiatry* **69**(10)**,** 1014-1021.

https://doi.org/10.1001/archgenpsychiatry.2012.135.

Harrison P, Carr E, Goldsmith K, Young AH, Ashworth M, Fennema D, Barrett B and Zahn R (2020) Study protocol for the antidepressant advisor (ADeSS): a decision support system for antidepressant treatment for depression in UK primary care: a feasibility study. *BMJ Open* **10**(5)**,** e035905. https://doi.org/10.1136/bmjopen-2019-035905.

Harrison P, Carr E, Goldsmith K, Young AH, Ashworth M, Fennema D, Duan S, Barrett B and Zahn R (2022) The Antidepressant Advisor (ADeSS): a decision support system for antidepressant treatment for depression in UK primary care - a feasibility study. In.

Harrison P, Walton S, Fennema D, Duan S, Jaeckle T, Goldsmith K, Carr E, Ashworth M, Young AH and Zahn R (2021) Development and validation of the Maudsley Modified Patient Health Questionnaire (MM-PHQ-9). *BJPsych Open* **7**(4)**,** e123. https://doi.org/10.1192/bjo.2021.953.

Harvey AK, Pattinson KT, Brooks JC, Mayhew SD, Jenkinson M and Wise RG (2008) Brainstem functional magnetic resonance imaging: disentangling signal from physiological noise. *J Magn Reson Imaging* **28**(6)**,** 1337-1344. https://doi.org/10.1002/jmri.21623.

Kasper L, Bollmann S, Diaconescu AO, Hutton C, Heinzle J, Iglesias S, Hauser TU, Sebold M, Manjaly ZM, Pruessmann KP and Stephan KE (2017) The PhysIO toolbox for modeling physiological noise in fMRI data. *J Neurosci Methods* **276,** 56-72. https://doi.org/10.1016/j.jneumeth.2016.10.019.

Lythe KE, Moll J, Gethin JA, Workman CI, Green S, Lambon Ralph MA, Deakin JF and Zahn R (2015) Self-blame-selective hyperconnectivity between anterior temporal and subgenual cortices and prediction of recurrent depressive episodes. *JAMA Psychiatry* **72**(11)**,** 1119-1126. https://doi.org/10.1001/jamapsychiatry.2015.1813.

Montgomery SA and Asberg M (1979) A new depression scale designed to be sensitive to change. *Br J Psychiatry* **134,** 382-389. https://doi.org/10.1192/bjp.134.4.382.

Rush AJ, Trivedi MH, Ibrahim HM, Carmody TJ, Arnow B, Klein DN, Markowitz JC, Ninan PT, Kornstein S, Manber R, Thase ME, Kocsis JH and Keller MB (2003) The 16-item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. *Biol Psychiatry* **54**(5)**,** 573-583. https://doi.org/10.1016/s0006-3223(02)01866-8.

Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, Hergueta T, Baker R and Dunbar GC (1998) The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J Clin Psychiatry* **59,** 22-33.

Spitzer RL, Kroenke K, Williams JBW and PHQPCSG (1999) Validation and utility of a self-report version of PRIME-MD: The PHQ primary care study. *JAMA* **282**(18)**,** 1737-1744.

Spitzer RL, Kroenke K, Williams JBW and Lowe B (2006) A brief measure for assessing generalised anxiety disorder: the GAD-7. *Arch Intern Med* **166**(10)**,** 1092-1097. https://doi.org/10.1001/archinte.166.10.1092.

Workman CI, Lythe KE, McKie S, Moll J, Gethin JA, Deakin JF, Elliott R and Zahn R (2016) Subgenual cingulate-amygdala functional disconnection and vulnerability to melancholic depression. *Neuropsychopharmacology* **41**(8)**,** 2082-2090. https://doi.org/10.1038/npp.2016.8.