UNIVERSITY OF COPENHAGEN
FACULTY OF HEALTH AND MEDICAL SCIENCES

Shining a light on the black cloud of depression

- A study of cognitive markers in Major Depressive Disorder

PhD thesis

Vibeke Naja Høyrup Dam

This thesis has been submitted to the Graduate School of Health and Medical Sciences, University of Copenhagen December 15th 2019

PhD title: Shining a light on the black cloud of depression – A study of cognitive markers in Major Depressive Disorder

Working title: Affective and social cognition as neurocognitive predictors of antidepressant treatment response

Author: Vibeke Naja Høyrup Dam, MSc in psychology

Work place: Neurobiology Research Unit, Department of Neurology, Copenhagen University Hospital Rigshospitalet, Denmark

Academic institution: Faculty of Health and Medical Sciences, University of Copenhagen, Denmark

Principal supervisor: Professor Gitte Moos Knudsen, Department of Clinical Medicine, University of Copenhagen, Denmark

Primary co-supervisor: Senior researcher Vibe Gedsø Frøkjær, Neurobiology Research Unit, Copenhagen University Hospital, Denmark

Co-supervisor: Associate professor, Dea Siggaard Stenbæk, Department of Psychology, University of Copenhagen, Denmark

Date of submission: December 15th 2019

Date of defense: October 23rd 2020 (original data of defense March 20th 2020)

Assessment committee:

Chairperson: Professor Lars Vedel Kessing, Department of Clinical Medicine, University of Copenhagen

Danish assessor: Associate professor Vibeke Fuglsang Bliksted, Department of Clinical Medicine, Aarhus University, Denmark

International assessor: Professor Catherine Harmer, Department of Psychiatry, Oxford University, United Kingdom

TABLE OF CONTENTS

I. Acknowledgements 3
II. Thesis summary 5
III. Dansk resumé 7
IV. List of manuscripts 9
V. Abbreviations 10

INTRODUCTION

1. Major Depressive Disorder 11
   1.1. Depression diagnosis 11
   1.2. Precision medicine 12
   1.3. Antidepressants & the monoamine hypothesis 13
   1.4. Beck's cognitive model of depression 14
   1.5. The cognitive neuropsychological model of depression 14
2. Cognition 15
   2.1. Disturbances of hot cognition in depression 15
   2.2. Disturbances of cold cognition in depression 16
   2.3. Impact of cognitive disturbances in depression 17
   2.4. Cognition as predictor of treatment response 19
   2.5. Antidepressant effect on cognition 20

AIMS & HYPOTHESES

METHODS

3. Study I. EMOTICOM validation study 24
   3.1. Participants 24
   3.2. Study design 25
4. Study II & III. NeuroPharm depression trial 25
   4.1.Participants 25
   4.2. Study design 26

4.3. Clinical outcomes..... 28
4.4. Cognitive outcomes..... 28
5. Statistical analyses..... 31

RESULTS & DISCUSSION

6. Study I..... 34
7. Study II..... 36
8. Study III..... 39
9. Additional cross-study analyses..... 45

DISCUSSION

10. The EMOTICOM test battery..... 48
11. Cognitive disturbances in depression..... 49
12. Cognitive markers of antidepressant response..... 50
13. Cognitive disturbances as treatment target..... 52
14. Dissociation between cognitive and depressive symptoms..... 53
15. Methodological considerations..... 55

CONCLUSION

16. Implications & future directions..... 57

REFERENCES

PAPER I-III

CO-AUTHORSHIP DECLARATIONS

I. ACKNOWLEDGEMENTS

I first want to thank my three supervisors for their support throughout my PhD. In particular, I want to thank Dea Siggaard Stenbæk for inviting me into the world of science and teaching me the ropes; Vibe Gedsø Frøkjær for her unbeatable enthusiasm and razor-sharp insight; and Professor Gitte Moos Knudsen for her continued support and the opportunity to work on such exciting projects.

I also want to give a huge thanks to everybody on Team Depression including our amazing team of (ex)students: Elizabeth Landman, Søren Vinther Larsen, Asbjørn Seenithamby Poulsen, Ida Marie Brandt and Mads Heidemann. You guys rock and we, hands down, could not have done it without you. To my fellow PhDs on the project, Kristin Köhler-Forsberg and Cheng Teng Ip, I just want to say: *I can't believe you both started a year before me and I am still the first to finish my PhD!* (edit: I wrote this before Covid-19 postponed my defense, so it still totally counts even though Kristin *technically* beat me to it). In all seriousness though, I've had the best time working with you and I look forward to many more years of our particular brand of dysfunctional friendship. And of course, thank you to the hardworking PIs including Vibe and Martin Balslev Jørgensen.

A special thanks to all the psych students on Team EMOTICOM who put in a truly Herculean effort to translate, implement, recruit and collect the data. Go team! Apart from being awesome to work with, I also consider each and every one of you a close friend. So thank you Christa Koll Thystrup, Anne-Sofie Schneider, Simone Pleinert and Nanna Hansen. Also a special thanks to Peter Steen Jensen for having the patience of a saint and helping me decipher EMOTICOM's excel outputs.

I wish to express my deep gratitude to Professor Barbara Sahakian for inviting me into her lab and continuing to offer scientific insights and guidance; and to Christelle Langley, George Savulich and the rest of the gang for welcoming me and making my stay at Cambridge University so memorable.

I also want to mention Vincent Beliveau and Brice Ozenne, my first and closest friends at NRU. Thanks for years of boardgame afternoons, obscure statistics discussions in French and way too many hikes up mountains. Thank you, Annette Johansen for our many impromptu discussions about life, the universe and everything and for our epic adventures (and the many more to come). Thank you, Sophia Armand for being the co-worker equivalent of Wonder Woman who always comes to the rescue when needed. Thank you, Lene Lundgaard Donovan for being your awesome, kick-ass self. Thank you,

Martin Korsbak Madsen for your genuine warmth and always amazing hugs. Thank you, Camilla Borgsted Larsen for all our glamorous international culinary adventures. Thank you, Nakul Raval for all the laughs and the NRU drinks nights. In addition, honorary mentions go to Sofie Trolle Pedersen, Sara Kristiansen, Johanna Mariegaard, Sophia Katrine Weber, Gunild Vulpius, Liv Vadskjær Hjordt, Martin Schain, Sanjay Sagar, Melanie Ganz, Claus Svarer, Patrick Fisher, Lars Pinborg, Giske Opheim, Sofi da Cunha-Bang, Sara Marie Larsen, Emily Beaman, Dorte Bonde Zilstorff, Nizar Hamrouni and all the rest of NRU. You are all fantastic human beings and I've thoroughly enjoyed getting to know you both as co-workers and friends.

I also want to acknowledge the unsung heroes of NRU, Peter Steen Jensen, Dorthe Givard, Lone Freyer Ibsgaard and Birgit Tang without whom absolutely nothing would get done. Please promise you will never all go on vacation at the same time! In addition, I want to thank Professor Olaf Paulson for creating an atmosphere of collaboration and good colleagueship when NRU was first founded and to Gitte Moos Knudsen for upholding these values today. It is a rare thing and in the five years I have been at NRU, I can honestly say that there has not been a single day where I did not feel like coming to work.

A special thanks to my friend Michéle Out for the weekly breakfasts and sanity checks and to Anne-Sophie Schwarz for always being there.

Lastly, I owe a big thanks to my family. To my father Mogens for encouraging me to have my head in the clouds and my feet firmly on the ground. To my mother Nina who passed on her love of writing to me. To my brother Lars who showed me that there are many different paths in life, and my brother Mathias for setting impossibly high standards for me to chase after.

II. THESIS SUMMARY

Disturbed cognition is a common but often overlooked symptom in Major Depressive Disorder (MDD). The disturbances are typically expressed across a wide spectrum of cognitive functions and may critically impair the patient's ability to function in everyday life. Recently, researchers have begun to explore the potential of cognitive markers to inform clinical decision-making in the treatment of depression. However, there are still many unanswered questions about the patterns and severity of different types of cognitive deficits during a depressive episode and how they relate to underlying depressive pathology and clinical outcomes. In particular, little is known about the differences and co-occurrence of hot (emotion-dependent) and cold (emotion-independent) cognitive disturbances in depression. The purpose of this PhD was therefore to improve the current knowledge about cognitive disturbances in depression in the hope that this would not only aid our understanding of the disease mechanisms in MDD but also offer clinical insight to help improve patient care.

In Study I, we tested 100 healthy Danish participants with the hot cognitive EMOTICOM test battery to establish the psychometric properties of the battery and create a Danish reference sample. Although select EMOTICOM tasks exhibited problematic test-retest reliability and/or floor and ceiling effects, we found that the test battery as a whole offers a useful set of cognitive tools for measuring hot cognition.

In Study II, we investigated cognitive disturbances in 92 antidepressant-free patients with a moderate to severe depressive episode. We found that performance of the MDD patients was disturbed across both hot and cold cognitive domains relative to healthy controls. However, the severity of cognitive disturbances was not related to the severity of depressive symptoms. We also identified three distinct clusters of cognitive profiles within the MDD cohort: One cluster was characterized by pronounced negative affective biases in emotion processing with minimal disturbances across other cognitive domains; the second cluster was characterized by positive affective biases in emotion processing and moderate deficits in cold cognitive domains; the last cluster was characterized by large global deficits across all domains. The globally impaired cluster had slightly more severe depressive symptoms compared with the other two clusters.

In Study III, we investigated the association between pre-treatment cognitive performance and antidepressant response as well as the effect of 12 weeks of standard treatment with Selective Serotonin Reuptake Inhibitors on cognition in MDD patients. Although we found no association between performance on any single cognitive outcome at baseline and later clinical treatment response, patients from the globally impaired cluster had worse clinical response after 8 but not 12 weeks of treatment compared with the other two clusters. This suggests that severe cognitive disturbances may delay antidepressant treatment effects. Overall cognitive performance improved during the course of treatment, although these improvements were not related to improvement in clinical symptoms indicating a dissociation between cognitive and depressive symptoms in MDD.

In conclusion, we translated and validated the EMOTICOM test battery in Danish, providing new set of hot cognitive tools for future clinical and research use. In addition, the results from this thesis provide new insights into the role of hot and cold cognitive disturbances in depression and emphasize that cognition should be viewed as a distinct symptom and treatment target in MDD. Importantly, we also showed that cognitive profiles may be useful tools for stratifying patients in a precision medicine approach.

III. DANSK RESUMÉ

Kognitive forstyrrelser er et hyppigt forekommende men ofte også overset symptom ved depression. Forstyrrelserne kommer typisk til udtryk på tværs af en lang række kognitive funktioner og kan have store konsekvenser for patientens evne til at fungere i hverdagen. Forskere er for nyligt begyndt at undersøge brugen af kognitive markører til at målrette klinisk behandling af depression. Der er dog stadig mange ubesvarede spørgsmål omkring sværhedgraden og præsentationen af forskellige former for kognitive forstyrrelser under en depressive episode, og hvordan disse relaterer sig til den underliggende patologi i depression samt klinisk behandlingsrespons. Især mangler vi viden om forskellene og sammenfaldene mellem varme (emotions-afhængige) og kolde (emotions-uafhængige) kognitive forstyrrelser i depression. Formålet med denne phd-afhandling er derfor at udbygge den nuværende viden om kognitive forstyrrelser i depression. Håbet er, at dette ikke blot vil bidrage til vores forståelse af sygdomsmekanismerne i depression men også vil give ny klinisk indsigt, som kan være med til at forbedre behandlingen af patienter.

I det første studie testede vi 100 raske danske forsøgspersoner med det varme kognitive testbatteri EMOTICOM for at etablere batteriets psykometriske egenskaber og oprette et dansk reference materiale. Selvom enkelte EMOTICOM tests udvise problematisk test-retest relibilitet og/eller gulv og loftseffekter, så fandt vi at testbatteriet i sin helhed udgør et brugbart sæt af kognitive redskaber til at måle varm kognition med.

I det andet studie undersøgte vi kognitive forstyrrelser hos 92 antidepressiva-fri patienter med moderat til svær depression. Vi fandt, at kognitionen hos de depressive patienter var forstyrret på tværs af både varme og kolde kognitive domæner relativt til raske kontroller. Der var dog ingen sammenhæng mellem sværhedsgraden af kognitive forstyrrelser og sværhedsgraden af depressive symptomer hos patienterne. Vi fandt også, at patienterne kunne inddeles i tre grupper baseret på forskellige kognitive profiler: Én gruppe var karakteriseret ved udtalte negative affektive bias i emotionsprocessering men havde ellers minimale forstyrrelser i andre kognitive mål; den anden gruppe var karatekeriseret ved positive affektive bias i emotionsprocessing og moderate forstyrrelser i kolde kognitive mål; den sidste gruppe var karakteriseret ved svære globale forstyrrelser på tværs af alle kognitive mål. Denne globalt forstyrrede gruppe havde også lidt sværere depressive symptomer sammenlignet med de to andre grupper.

I det tredje studie undersøgte vi sammenhængen mellem kognitive forstyrrelser målt før behandling og antidepressiv respons såvel som effekten af 12 ugers standard behandling med selektive serotoningenoptagshæmmere på kognitive mål hos patienter med depression. Selvom vi ikke så nogen association mellem enkelte kognitive mål ved baseline og senere klinisk behandlingsrespons, så fandt vi, at patienter fra den globalt kognitivt påvirkede gruppe udviste dårligere behandlingsrespons efter 8 men ikke 12 ugers behandling sammenlignet med de to andre grupper. Dette indikerer at svære kognitive forstyrrelser kan forsinke det antidepressive behandlingsrespons. Overordnet set forbedrede patienterne sig kognitivt efter behandlingen. Disse forbedringer var dog ikke relateret til forbedringer i kliniske symptomer, hvilket tyder på en dissociation mellem kognitive og depressive symptomer i depression.

Opsamlingvis har vi med oversættelsen og valideringen af EMOTICOM batteriet stillet et nyt sæt af varme kognitive mål til rådighed for fremtidig forskning og klinisk brug. Envidere har resultaterne fra den foreliggende phd-afhandling givet ny indsigt i relationen mellem varm og kold kognition i depression og understreget, at kognitive forstyrrelser bør ses som et selvstændigt symptom og derfor også som et behandlingsmål i depression. Sidst og måske vigtigst har vi vist, at kognitive profiler kan bruges til at inddele patienter i klinisk relevante subgrupper.

IV. LIST OF MANUSCRIPTS

The thesis is based on the following manuscripts and published articles:

I. **Dam, V.H.**, Thystrup, C.K., Jensen, P.S., Mortensen, E.L., Sahakian, B.J., Knudsen, G.M., Frokjaer, V.G. & Stenbæk, D.S. Psychometric Properties and Validation of the EMOTICOM Test Battery in a Healthy Danish Population. *Frontiers in Psychology 10*.

II. **Dam, V.H.**, Stenbæk, D.S., Köhler-Forsberg, K., Ip, C., Ozenne, B., Sahakian, B.J., Knudsen, G.M., Jørgensen, M.B. & Frokjaer, V.G. Hot and cold cognitive disturbances in antidepressant-free patients with Major Depressive Disorder: A NeuroPharm Study. *Psychological Medicine (in revision)*

III. **Dam, V.H.**, Stenbæk, D.S., Ozenne, B., Ip, C., Köhler-Forsberg, K., Sahakian, B.J., Knudsen, G.M., Jørgensen, M.B. & Frokjaer, V.G. Cognitive disturbances as both treatment targets and predictors of antidepressant action in Major Depressive Disorder: A NeuroPharm study. *In final prep*

Related articles not included in thesis:

IV. Stenbæk, D.S.*, **Dam, V.H.*,** Fisher, P.M., Hansen, N., Hjordt, L.V. & Frokjaer, V.G. (2017). No evidence for a role of the serotonin 4 receptor in five-factor personality traits: A positron emission tomography brain study. *PloS One 12,* e0184403.
*Authors contributed equally

V. Hjordt, L.V., **Dam, V.H.**, Ozenne, B., McMahon, B., Mortensen, E.L., Knudsen, G.M. & Stenbæk, D.S. (2018). Self-perceived personality characteristics in seasonal affective disorder and their implications for severity of depression. *Psychiatry Res 262,* 108-114.

V. ABBREVIATIONS

Abbreviation	*Explicated term*
5-HT	Serotonin (5-hydroxytryptamine)
95% CI	95% Confidence Interval
ANCOVA	Analysis of Covariance
ANOVA	Analysis of Variance
CBT	Cognitive Behavioural Therapy
EEG	Electroencephalogram
ERP	Event-Related Potential
fMRI	Functional Magnetic Resonance Imaging
HDRS	Hamilton Depressive Rating Scale
$HDRS_6$	6-item HDRS
$HDRS_{17}$	17-item HDRS
ICC	Intraclass Correlation Coefficients
ICD-10	International Classification of Diseases 10^{th} edition
MAOIs	Monoamine Oxidase Inhibitors
MDD	Major Depressive Disorder
MINI	Mini-International Neuropsychiatry Interview
MRI	Magnetic Resonance Imaging
PET	Positron Emission Tomography
SNRIs	Serotonin-Norepinephrine Reuptake Inhibitors
SSRIs	Selective Serotonin Reuptake Inhibitors
TCAs	Tricyclic Antidepressants

INTRODUCTION

The following sections provide a brief overview of the MDD diagnosis and the challenges caused by the heterogeneity of the disorder. Next different theories of depression are presented including the monoamine hypothesis, Beck's cognitive model and the cognitive neuropsychological model of depression. Following this, the literature on hot and cold cognition and their role in MDD is outlined and impact of cognitive disturbances on the patient's ability to function in everyday life is explored. Lastly, the literature on cognition as predictors of treatment response and antidepressant effects on cognition is reviewed.

1. Major Depressive Disorder

According to a newly released report from the World Health Organization, MDD is now the leading cause of disability globally (WHO, 2017). The report further estimates that worldwide ~ 322 million people are suffering from a depressive episode; this translates to 1 in 24 people (WHO, 2017). Lifetime incidence of MDD in Denmark is 15-25% for women and 7-12% for men (source: Danish Ministry of Health). Apart from the inestimable human suffering, MDD also comes with a heavy socioeconomic burden in terms of work disability, suicide-related costs and treatment costs (Greenberg *et al.*, 2015). Combined with high recurrence rates of 50-80% (Burcusa and Iacono, 2007), depression thus represents a profound health challenge on a global scale.

1.1. Depression diagnosis

MDD, sometime referred to as unipolar depression, is characterized by persistent negative mood with associated disturbances in thoughts, behaviours and physiological functions such as sleep, appetite and sex drive. The International Classification of Diseases 10th Edition (ICD-10) diagnostic criteria for MDD are shown in Figure 1.

Figure 1. Diagnostic criteria for depression

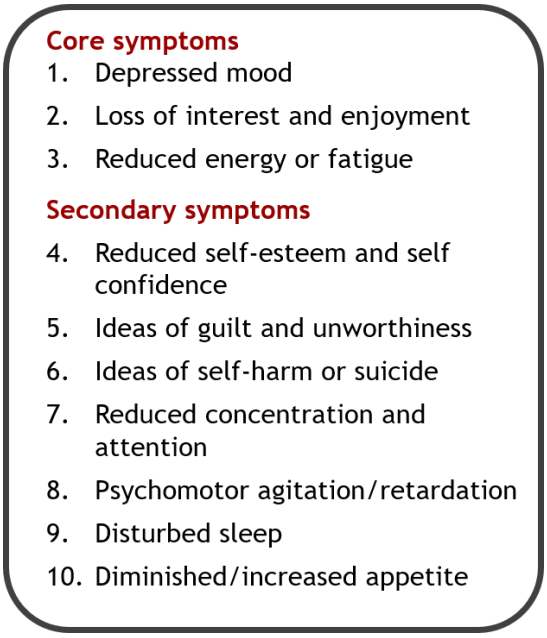


Figure 1. To meet the diagnostic criteria for MDD, the patient must have experienced at least two of the three core symptoms and at least two of the seven secondary symptoms. In addition, the symptoms must have been present 'most of the time' for at least two weeks; other biological cause must have been ruled out; and no history of hypomanic or manic episodes must be present. The severity of the depressive episode is determined by the number of symptoms and is conventionally categorized as mild (2 core plus + 2 secondary symptoms), moderate (2 core + 4 secondary symptoms) or severe (3 core + > 4 secondary symptoms).

1.2. Precision medicine

MDD is not an aetiologically based diagnosis; rather it is a syndrome made up of symptoms clusters. After decades of rigorous scientific efforts and numerous proposed neurobiological models, researchers have mostly abandoned the idea of a single unifying theory of depression (Hasler, 2010). Instead, it has become clear that MDD is an aetiologically heterogeneous construct that may in fact describe several different brain disorders. Adding to this complexity, there are 487 possible ways to meet the ICD-10 diagnostic criteria of MDD. Of course, not all combinations of symptoms are equally likely or common. For example, one study found that 25% of the possible symptom constellations did not occur in a sample of more than 1500 MDD patients (note, this study used MDD criteria from the Diagnostic and Statistical Manual of Mental Disorders, 5th edition) (Zimmerman *et al.*, 2015). The heterogeneity and complexity of MDD pose a serious challenge in patient treatment. For one, there are no available objective tests clinicians can use to verify the presence of a depressive episode. Thus, both diagnosing and treating MDD rely solely on a combination of clinical guidelines and the expertise of mental health

professionals. Secondly, the current treatments offered to patients with MDD are often insufficient. For example, approximately one third of patients fail to respond to Selective Serotonin Reuptake Inhibitors (SSRIs) commonly used as a first-line treatment in MDD (Trivedi *et al.*, 2006). In addition, the remission rates have been shown to worsen for every failed treatment attempt (Rush *et al.*, 2006). Precision medicine has been proposed as a way to overcome this challenge. In precision medicine, the treatment is individually tailored to each patient based on their distinct characteristics (Schumann *et al.*, 2014). Consequently, a great deal of research has gone into identifying biomarkers that can stratify patients into clinically relevant subgroups and/or predict treatment response. Such efforts have traditionally focused on 'classic' biological markers such as neuroimaging with Positron Emission Tomography (PET), functional and structural Magnetic Resonance Imaging (fMRI; MRI) and event-related potential (ERP) and other electroencephalogram (EEG) related measurements; genetic and epigenetic characteristics; and neuroendocrine markers (Hacimusalar and Eşel, 2018), so far with varying success. More recently, cognition has been added to this list of possible markers of antidepressant treatment response although the literature is still scarce (Groves *et al.*, 2018).

1.3. Antidepressants & the monoamine hypothesis

'Antidepressants' is an umbrella term for pharmaceutical drugs used to treat depressive symptoms. There are currently five major classes of antidepressant: Tricyclic Antidepressants (TCAs); Monoamine Oxidase Inhibitors (MAOIs); SSRIs; Serotonin and Norepinephrine Reuptake Inhibitors (SNRIs); and atypical antidepressants. MAOIs and TCAs were discovered serendipitously in the 1950s when pharmaceuticals originally developed for treatment of tuberculosis and schizophrenia showed antidepressant effects. Subsequent research revealed that both classes of drugs act by increasing the amount of available monoamine neurotransmitters in the synaptic cleft: MAOIs by inhibiting the enzyme responsible for breaking down monoamines and TCAs by inhibiting the reuptake of serotonin and norepinephrine from the synaptic cleft to the presynaptic cell. This led to the formulation of the monoamine hypothesis of depression which posits that depressive symptoms are caused by a neurochemical imbalance in the brain and more specifically a lack of serotonin, dopamine and norepinephrine (Pereira and Hiroaki-Sato, 2018).

Over the next decades, the monoamine hypothesis gained traction and in particular serotonin and its role in depressive pathophysiology received a great deal of scientific interest. This ultimately led to the introduction of SSRIs in the 1970s followed by SNRIs in the 1990s. Both classes of drugs were developed to have more targeted neuropharmacological profiles in order to minimize the severe side effects commonly experienced with MAOIs and TCAs and they are currently used as first and second-line treatments in MDD (Cleare *et al.*, 2015). Despite the efficacy of antidepressant drugs in the treatment of MDD (Cipriani *et al.*, 2018), the monoamine hypothesis has been criticized for failing to explain why only 50-70% of patients respond to treatment (Trivedi *et al.*, 2006) and why treatment effects are typically delayed by several weeks when changes in synaptic signalling are observed acutely.

1.4. Beck's cognitive model of depression

One of the earliest cognitive theories of depression was proposed by psychiatrist Aaron Beck in 1976. According to the cognitive model of depression, early adverse life events help shape dysfunctional schemata, i.e. beliefs and expectations, about the self, the world and the future (also termed the negative cognitive triad). When these schemata are activated in daily life, they act by distorting cognitive processes resulting in negative biases in perception, attention, memory, interpretation and mood. If these mechanisms are activated often enough or strongly enough, they enter a feedback loop where a person unconsciously attends and magnifies negative events while ignoring or minimizing positive events. This reinforces the negative schemata, feeding into a spiral ultimately leading to depression (Beck, 2008). Cognitive Behavioural Therapy (CBT), which is an effective and commonly used treatment in MDD (Driessen and Hollon, 2010), focuses on changing this negative feedback loop by challenging the dysfunctional schemata.

1.5. The cognitive neuropsychological model of depression

The newer cognitive neuropsychological model of depression can be viewed as a synthesis of Beck's cognitive model with aspects of the monoamine hypothesis. The model posits that dysfunctional monoamine neurotransmission affects bottom-up processing of emotional information resulting in the

negative affective biases observed in MDD. These bottom-up negative biases form and reinforce top-down negative schemata which in turn reinforce the bottom-up negative biases creating a feedback loop that results in depression (Roiser *et al.*, 2012). Thus, although the primary mechanisms are the same, the main difference between the neuropsychological model of depression and Beck's cognitive model of depression is the role monoamine dysfunction play in the onset of depressive pathology. Furthermore, according to the cognitive neuropsychological model, SSRIs primarily act by acutely attenuating the bottom-up negative affective biases which over time allows a positive restructuring of the negative schemata and ultimately remission of the depressive episode (Harmer and Cowen, 2013). This proposed mechanism of action not only explains the delay between the initiation of antidepressant treatment and clinical improvements but also provides a set of clear and testable predictions. One of the central hypotheses is that antidepressant treatment response is predicted by early changes in affective biases (for a comprehensive list of model predictions see Roiser *et al.* (2012)) which so far has been empirically supported by both behavioural (Browning *et al.*, 2019, Tranter *et al.*, 2009) and neuroimaging studies (Shiroma *et al.*, 2014).

2. Cognition

Cognition can briefly be defined as the mental act of processing information. This includes the acquisition, representation, storage, retrieval, manipulation and/or interpretation of internal (e.g. emotional states) and external (e.g. sensory input) information which ultimately guide thoughts and behaviours.

2.1. Disturbances of hot cognition in depression

Hot cognition describes cognitive processes that have an emotional component in the form of affectively valenced stimuli (e.g. facial expression) or through activation of emotional states (e.g. reward-driven behaviours) (Brand, 1985). Disruptions in hot cognitive functions and the neural substrates underpinning them have been consistently linked to MDD pathology (Elliott *et al.*, 2011) and are believed to play an important role in the onset and maintenance of the depressive episode

(Roiser *et al.*, 2012). One of the strongest and most consistent findings in MDD is negative affective biases in emotion processing. Affective biases describe the tendency to subconsciously prioritize negative information over positive information and have been observed in MDD across a wide range of functions in including perception, memory and attention (Miskowiak and Carvalho, 2014). In particular, emotional face paradigms are well-studied and consistently show that, compared with healthy participants, MDD patients have enhanced recognition and processing of sad faces relative to happy faces (Dalili *et al.*, 2015). This is further supported by findings of abnormal activity in limbic and cortical networks involved in emotion processing (Delaveau *et al.*, 2011). Interestingly, positive affective biases are often observed in healthy individuals, possibly reflecting an inherent optimism or resilience against negative emotional states (Korn *et al.*, 2014). In fact, there is evidence to suggest that in many cases the affective biases observed in MDD describes the *absence* of normal positive biases rather than the presence of objectively negative biases (Moore and Fresco, 2012). Social cognition describes the higher-order mental processes necessary for understanding other people and successfully engaging in social interactions. There is accumulating evidence that many of these functions are at least partially disrupted in MDD including mentalizing (i.e. the ability to understand the behaviour and motivation of other people) (Bora and Berk, 2015), social cooperation (Brendan *et al.*, 2013), and moral emotions such as feelings of guilt and shame (Kim *et al.*, 2011). Lastly, MDD patients have been shown to exhibit aberrant responses to reward and punishment including heightened sensitivity to punishment and/or negative feedback and diminished (anhedonic) sensitivity to reward and/or positive feedback (Miskowiak and Carvalho, 2014).

2.2. Disturbances of cold cognition in depression

Cold cognition describes mental processing of non-affective stimuli (e.g. numbers or letters) independent of emotional states. Overall, deficits in cold cognition have been much more extensively investigated in MDD compared with hot cognitive disturbances. As a result, deficits in psychomotor speed, attention, memory and learning and executive functions have been reliably reported across several large meta-analyses and are by now well-established as a core feature of MDD pathology (Goodall *et al.*, 2018, Lee *et al.*, 2008, Rock *et al.*, 2014). However, some investigators have questioned the accuracy of these reported cold cognitive deficits citing the potential influence of altered

hot cognition on patient performance during cold cognitive tasks. Specifically, they argue that MDD patients are more easily discouraged by negative feedback and may therefore perform poorly in cold cognitive tasks containing explicit feedback (Roiser and Sahakian, 2013). In other words, it is not possible to distinguish fully between cold and hot cognitive tasks in MDD, as motivational factors relating to self-perceived performance may add a 'hot' component to otherwise 'cold' cognitive tasks.

2.3. Impact of cognitive disturbances in depression

According to a recent survey, 99% of patients with MDD reported experiencing one or more symptoms of cognitive dysfunction. In addition, 45% of patients stated that their cognitive and mood symptoms had impacted their everyday life equally, while 10% stated that their cognitive symptoms had impacted their everyday life more than their mood symptoms (Clark Health Communications, 2015).

Compared with other serious psychiatric illnesses, the cognitive disturbances observed in MDD are relatively mild. For example, while the reported effect sizes of both hot and cold cognitive impairments mostly range from small to moderate in MDD (indexed as Cohen's *d* values < 0.8), the cut-off often used to indicate mild cognitive impairments in dementia is much higher (Cohen's *d* values > 1.5). The magnitude of cold cognitive deficits in MDD is comparable to those observed for mild sleep deprivation (24 hours wakefulness) or low-level alcohol intoxication (blood alcohol concentration = 0.05%) (Maruff and Jaeger, 2016). Even though the cognitive disturbances in MDD are relatively mild, they may still impact the patient's ability to function in a work setting as well as their overall quality of life, particularly since patients with MDD often continue work or educational activities throughout their illness.

It is important to note that the reported effect sizes for cognitive disturbances in MDD reflect *group* averages rather than *individual* patient scores. As MDD is a heterogeneous disorder in general, it makes sense that not all patients experience the same degree of impairments. In fact, estimates of how large a proportion of patients experiencing measurable cognitive disturbances range from 21% (Gualtieri and Morgan, 2008) to 28% (Iverson *et al.*, 2011) to 44.4% (McIntyre *et al.*, 2017). Notably, these three studies all used different cognitive measures and different cut-off criteria to define cognitive dysfunction which may help explain the difference between the estimates. For example, Iverson *et al.*

(2011) used a rather strict cut-off criterium in which patients had to score below the 5th percentile on at least two out of five cognitive domains whereas McIntyre *et al.* (2017) used mean scores ≥ 1 SD below healthy controls as criterium. As an alternative to predefined cut-offs, cluster analysis can be used to identify cognitive profiles in a data-driven and hypothesis-free approach. To our knowledge, only two recent studies have used cluster analysis to identify subgroups based on cognition scores in MDD. The first study is from the large iSPOT-D trial where 1008 MDD patients were tested with a comprehensive (mostly cold) cognitive test battery. The researchers found that a subgroup of ~25% of the patients were impaired across most cognitive domains, relative to healthy controls, while the remaining ~75% of the patients were cognitively intact (Etkin *et al.*, 2015). Vicent-Gil *et al.* (2018) identified a similarly sized cluster of patients of 26% with pronounced cognitive impairments across several cold cognitive domains and a larger group of cognitively intact patients of 74% in first-episode depression (N = 50). Thus, there is converging evidence from both criteria-based and data-driven studies that less than half of all MDD patients exhibit substantial (cold) cognitive symptoms. This is of course in direct contrast with the patient survey described previously where 99% of all patients complained of at least one or more cognitive symptom(s) (Clark Health Communications, 2015). In this context it is important to stress that there are large discrepancies between subjectively experienced deficits and objectively measured deficits in MDD patients (Petersen *et al.*, 2019, Srisurapanont *et al.*, 2017). One possible explanation for this is that the emotional symptoms bias the patient's assessment of their own cognitive performance negatively. It is therefore important to not rely solely on self-report or interview-based data when assessing cognitive disturbances in MDD. On the other hand, a major weakness of objectively measures cognition is that the premorbid level of cognitive function is very rarely known. It is therefore possible for patients with above-normal premorbid function to experience a marked decline in cognition functioning without being detected as their scores now range within the normal spectrum: conversely, individuals with poor premorbid function may be wrongly classified as cognitive disturbed without having experienced cognitive decline. Ideally, both subjective and objective measures of cognitive functioning should be used jointly to improve accuracy in identifying and characterizing MDD-related cognitive disturbances.

Lastly, there is accumulating evidence that cognitive dysfunction is directly linked to poor psychosocial functioning in MDD (Cambridge *et al.*, 2018, Weightman *et al.*, 2019). Psychosocial functioning refers to the ability to successfully navigate everyday life in education, work, social and

family domains. Many patients rate functional recovery as equally or even more important than the alleviation of depressive symptoms in MDD emphasizing the importance of understanding the relationship between cognitive symptoms and functional outcomes (Zimmerman *et al.*, 2006).

2.4. Cognition as predictor of treatment response

As described previously, MDD treatment is moving towards an era of precision medicine which requires identification of clinical markers that can be used to stratify patients and predict treatment response. Within the last two decades, there has been a surge of interest in the potential of cognitive performance as a marker in MDD. A recent review by Groves *et al.* (2018) found that baseline performance on executive functions and attention paradigms was consistently predictive of SSRI treatment response but only in elderly patients; in younger adults the evidence was highly conflicted. The evidence for slowed psychomotor speed as a predictor of SSRI treatment response was also conflicted and there was little to no support for an association between learning and memory performance and later treatment response. Meanwhile, the results from the large iSPOT-D trial showed that patient with broad cognitive impairments had poorer clinical outcome than cognitively intact patients. Even more interesting, the researchers found that task performance at baseline predicted remission with 72% accuracy, but only for cognitively impaired patients who received treatment with the SSRI escitalopram (two other groups were treated with sertraline and venlafaxine, respectively) (Etkin *et al.*, 2015). If cold cognitive performance is only predictive of treatment response in subgroup of patients with pronounced cognitive deficits (and if this group only represents 20-40% of the patient population), this may explain the conflicting results from the literature as only studies with large sample sizes would be able to reliably detect the association. Studies have also investigated the predictive utility of early changes in emotion processing. For example, Tranter *et al.* (2009) found that increased accuracy in the recognition of happy faces after 2 weeks of treatment with citalopram or reboxetine correlated with clinical improvement after 6 weeks of treatment. In addition, a recent study showed that changes in facial emotion processing and self-reported depressive symptoms after 1 week of treatment with citalopram predicted treatment response (defined as ≥ 50% reduction in clinical symptoms) with 77% accuracy in a training sample (N = 74) and 60% accuracy in an independent

sample (N = 239) (Browning *et al.*, 2019). Notably, the study also found that baseline scores, including scores on emotional processing tasks, had little to no predictive value.

2.5. Antidepressant effect on cognition

Traditionally, cognitive dysfunction has not been considered a high-priority treatment target in MDD. Instead clinicians tend to focus on the 'classic' mood and somatic symptoms. This was illustrated in a study by Demyttenaere *et al.* (2015) where clinicians and patients were asked to rank the most important criteria for 'being cured of depression'. The study showed clear differences between patients and clinicians with patients prioritizing functional (e.g. enjoying or finding meaning in life) over clinical outcomes (e.g. mood and anhedonia symptom) and vice versa for the clinicians. Notably, while patients ranked the 'ability to concentrate' as the fourth most important criteria, clinicians did not have a single item relating to cognition among their ten most important criteria. In addition, survey data showed that 25% of patients with MDD had never been asked about their cognitive symptoms by their usual healthcare provider (Clark Health Communications, 2015). However, with greater recognition of the importance of functional recovery and increasing knowledge about the role of cognition in MDD, cognitive dysfunction has started to be recognized as an important treatment target (Kaser *et al.*, 2017). Several treatment candidates for cognitive disturbances in MDD are currently being explored (see Miskowiak *et al.* (2016) for a review) but I here focus on pharmacological antidepressant treatments. A recent meta-analysis found that antidepressant drugs, and in particular SSRIs, have a modest positive effect on several cold cognitive domains including processing speed, sustained and divided attention, immediate and recent memory and executive functions (Prado *et al.*, 2018). Importantly, these effects were only present in patient with MDD and not healthy individuals, corroborated by previous reports that antidepressants do not appear to have neuroenhancing effects in healthy individuals (Repantis *et al.*, 2009). In contrast there is evidence that antidepressants have both acute (Browning *et al.*, 2007, Harmer *et al.*, 2003) and sub-acute effects on emotion processing in healthy individuals (Harmer *et al.*, 2004). Meanwhile, the effects of antidepressants on behavioral measures of hot cognition in patients with MDD are generally understudied. Interestingly, one of the few studies to report on this showed that improvements in emotion recognition was detectable after just 2 weeks of antidepressant treatment and remained stable until follow-up after 6 weeks (Tranter *et al.*, 2009). This suggests that the

antidepressants do not have a cumulative effect on emotion processing over time; instead they are characterized by an initial rapid action followed by a plateauing effect. Some neuroimaging studies also indicate that the neural responses involved in emotion processing are normalized by antidepressant drug treatment (see Harmer and Cowen (2013) for an overview).

While there is evidence that antidepressants improve cognition in MDD, it is also well-established that cognitive disturbances do not always resolve with the clinical remission (Hasselbalch *et al.*, 2011). Small to moderate deficits have been shown in remitted patients for several cold cognitive domains including processing speed, learning and memory, attention and executive functions (Semkovska *et al.*, 2019). Similarly, affective biases in processing of emotional faces (Joormann and Gotlib, 2007, LeMoult *et al.*, 2009) and affective memory (Romero *et al.*, 2014) (for a more comprehensive overview, see (Miskowiak and Carvalho, 2014)). Thus, it appears that antidepressants may be effective in treating some but not all cognitive symptoms in MDD.

AIMS & HYPOTHESES

The aim of this thesis was to investigate hot and cold cognitive disturbances as markers of treatment response in MDD. To accomplish this, we first evaluated the psychometric properties of the novel affective and social cognitive test battery EMOTICOM (Bland *et al.*, 2016) in 100 healthy participants to establish a Danish reference sample. Secondly, as part of the large NeuroPharm depression trial, we tested 92 patients suffering from a moderate to severe depressive episode using tasks from the EMOTICOM battery as well as other selected cold cognitive tasks before and after 12 weeks of standard treatment with the SSRI escitalopram. The thesis is made up of three studies described below:

Study I

The aim of the Study I was to two-fold: First, we wished to validate and assess the psychometric properties of the novel EMOTICOM test battery in a Danish version. Secondly, we wished to establish a reference data set of healthy participants that could be used as a control group for the NeuroPharm MDD cohort and future studies. We therefore translated and implemented 11 out of 16 tasks from the English version of the EMOTICOM. Based on data collected from 100 healthy participants, including retest data from 49 participants, we thoroughly assessed the performance of the cognitive tasks across a number of psychometric parameters.

Study II

The aim of Study II was to map baseline disturbances of hot and cold cognitive functioning in patients with MDD compared with healthy controls. In addition, we wanted to determine whether we could identify distinct clusters of cognitive profiles within the patient group. We therefore tested the following hypotheses: 1) that patients with MDD would exhibit disturbed cognition across both hot and cold cognitive domains relative to healthy participants; 2) that severity of cognitive disturbances would be correlated with severity of clinical depression symptoms in the MDD patients and 3) that we could identify clinically meaningful clusters of cognitive profiles within the MDD patient group.

Study III

The aim of Study III was to assess cognitive performance both as a predictive marker of antidepressant treatment response and as a treatment target in MDD. Based on the results from study 2, we also wished to investigate the long-term clinical relevance of baseline cognitive profiles in MDD. We therefore tested the following hypotheses: *1)* that cognitive performance at baseline would be associated with antidepressant treatment response after 12 of weeks SSRI treatment; *2)* that cognitive performance would improve after 12 weeks of SSRI treatment; *3)* that improvement in cognitive performance would be linked to improvement in clinical symptoms; and *4)* that patients with different baseline cognitive profiles would respond differently to SSRI treatment.

METHODS

Study I covers the validation of the EMOTICOM while Study I and II are both based on data from the NeuroPharm depression trial. The EMOTICOM validation study was approved by the Danish Data Protection Agency (protocol RH-2015-255) and the NeuroPharm trial was approved by the National Committee on Health Research Ethics (protocol: H-15017713) and pre-registered at www.clinicaltrial.gov (reg. nr. NCT02869035). All participants received oral and written information and provided informed signed consent prior to study inclusion. The following sections give a brief summary of participants and study designs (for a more detailed description, see the method sections in Paper I-III).

3. Study I – EMOTICOM validation study

3.1. Participants

We recruited 100 healthy Danish participants (50, female; 50, male) through an already established database of healthy volunteers or with online advertisements and flyers distributed around the greater Copenhagen area. Prior to inclusion, potential participants were screened for eligibility based on the following inclusion and exclusion criteria.

Inclusion criteria:

- Age 18-50 years
- Fluency in Danish

Exclusion criteria

- History of severe Axis I psychiatric disorders
- Significant somatic illness
- Moderate to severe brain trauma
- Use of psychotropic medications

- Alcohol dependence (> 7 units of alcohol[1]/week for women; > 14 units of alcohol/week for men)
- Significant lifetime use of illicit drugs (cannabis > 50 times; other drugs > 10 times)
- Pregnancy or breastfeeding

3.2. Study design

All participants completed the 11 EMOTICOM tasks along with cognitive measures of IQ, working memory, verbal affective memory and reaction time. Retest data was collected from 49 participants after 3-5 weeks. In addition, all participants filled in questionnaires indexing education level, NEO personality and mood as well as self-rated motivation and diligence (i.e. how much the participant strove to perform well) for EMOTICOM tasks that included monetary reward incentives.

4. Study II & III – The NeuroPharm depression trial

4.1. Participants

We initially recruited 100 patients with a moderate to severe depressive episode through a central referral centre in the mental health services of the Capital Region of Denmark or through their general practitioner. MDD diagnosis was confirmed with the diagnostic tool Mini International Neuropsychiatric Interview (MINI) by a trained clinician and corroborated by a senior specialist in psychiatry. Prior to inclusion, patients were screened for eligibility based on the following inclusion and exclusion criteria.

Inclusion criteria:

- Age 18-65 years
- Fluency in Danish
- Moderate to severe depressive episode indicated by $HDRS_{17}$ score > 17
- Current depressive episode lasting < 2 years

[1] A unit of alcohol is defined as 12g/15ml alcohol according to Danish standards.

- No use of antidepressants drugs within two months of inclusion

Exclusion criteria

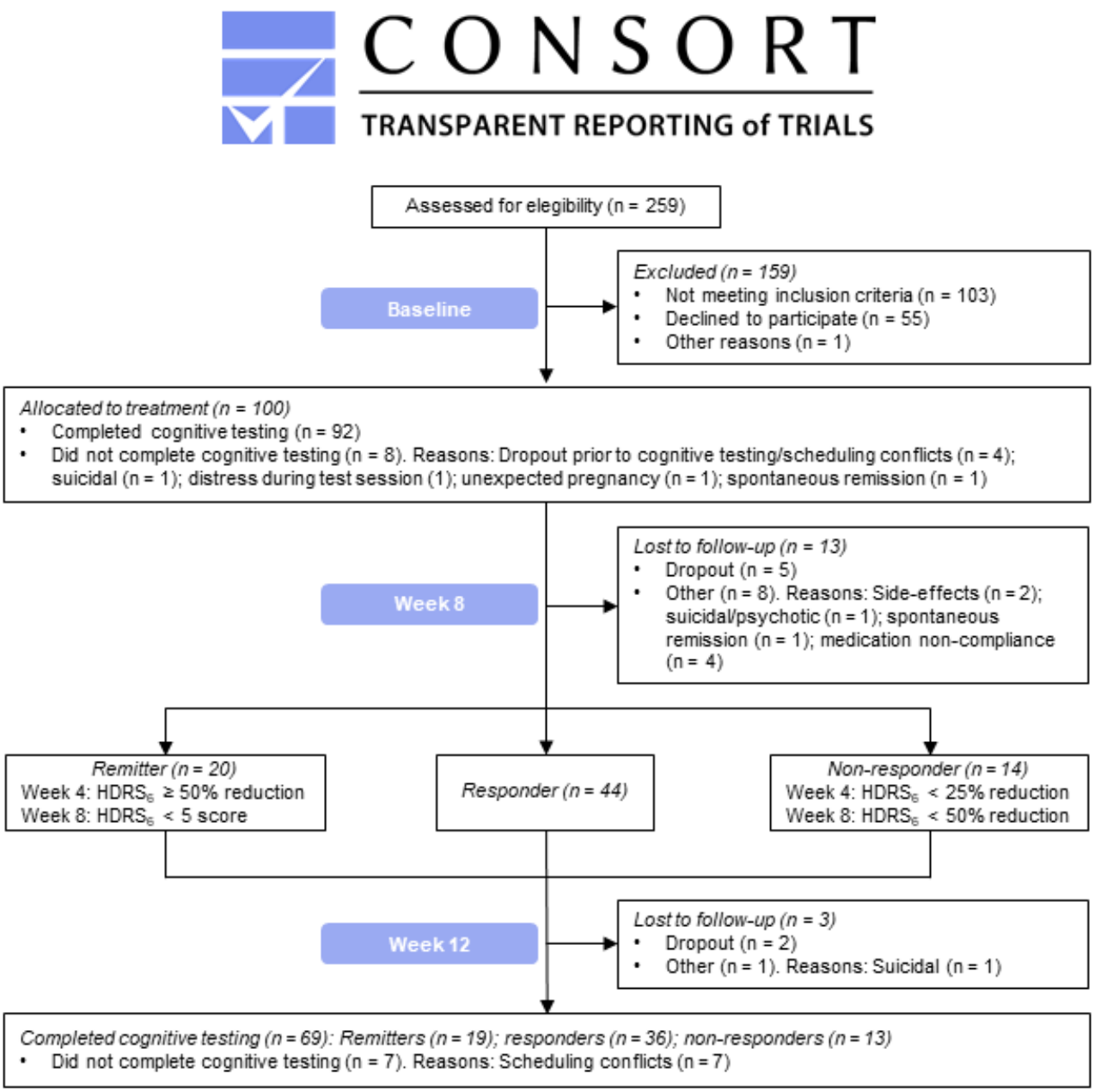
- Acute suicidal ideation or psychosis
- Contraindication or previous non-response to SSRIs
- More than one antidepressant treatment attempt in current depressive episode
- History of primary Axis I psychiatric disorders other than MDD
- Significant somatic illness
- Moderate to severe brain trauma
- Use of psychotropic medications that could not be washed out prior to inclusion (e.g. metoclopramide, ondansetron, clonidine)
- Alcohol dependence (> 7 units of alcohol/week for women; > 14 units of alcohol/week for men)
- Significant lifetime use of illicit drugs (cannabis > 50 times; other drugs > 10 times)
- Pregnancy or breastfeeding

In addition, three healthy control participants were also included (recruitment procedures and inclusion criteria were the same as for the EMOTICOM cohort in Study I).

4.2. Study design

The NeuroPharm trial is a longitudinal, open-label clinical trial investigating potential biomarkers in antidepressant treatment of MDD. The full study program included several neuroimaging modalities as well as biological measures. Here only the cognitive and clinical aspects of the investigate program are presented. Out of the 100 patients included, cognitive data were collected from 92 patients at baseline. Treatment with the SSRI escitalopram was started as soon as patients had completed the baseline program: 5 mg daily for the first 3-5 days followed by flexible doses of 10-20 mg daily. The dose was adjusted based on response and side-effects evaluated by physicians at follow-up visits at week 1, 2, 4, 8 and 12. If patients experienced severe side-effects or showed poor response to escitalopram at week 4, medication was switched to the SNRI duloxetine in accordance with standard practice (n = 16). After 12 weeks of treatment cognitive follow-up data was collected for 69 patients. A detailed overview of enrolment, data collection and patient dropouts are shown in the CONSORT flow diagram (Figure 2).

Figure 2. CONSORT flow diagram of the NeuroPharm trial



4.3. Clinical outcomes

Measures of depressive symptom severity were collected with the clinician rated Hamilton Depressive Rating Scale interview (HDRS) at baseline, week 4, week 8 and week 12. The primary clinical endpoint was categorical classification of treatment status at week 8 (remitter vs non-responder). Remitters were defined as patients who experienced ≥ 50 % reduction in 6-item HDRS ($HDRS_6$) scores at week 4 (early responder) as well as a $HDRS_6$ score < 5 at week 8. Non-responders were defined as patients who experienced < 25% reduction in $HDRS_6$ scores at week 4 (early non-responders) and < 50% reduction in $HDRS_6$ scores at week 8. Patients who did not fit either of these criteria were classified as responders and were not included in primary analyses using categorical treatment status as clinical outcome. Secondary clinical outcome was defined as relative change in $HDRS_6$ scores in percentage calculated as change in $HDRS_6$ from baseline to follow-up at week 8 or week 12 divided by baseline $HDRS_6$ score.

4.4. Cognitive outcomes

Due to time constraints and the limited stamina of the MDD patients, the cognitive test battery had to be relatively short (administration time < 1.5 hours). We therefore selected four of the best performing hot cognitive tasks from the EMOTICOM test battery (two tasks from the affective cognitive domain and two tasks from the social cognitive domain) as well as one hot cognitive and three cold cognitive tasks from the standard neuropsychological test battery used for Cimbi database studies. The Cimbi database is a large multimodal database containing neuroimaging, biological, cognitive and psychometric information on a large number of healthy individuals as well as various patient cohorts (Knudsen *et al.*, 2016). Change in cognitive scores between baseline and follow-up at week 12 was calculated as the absolute difference between baseline and week 12 scores. A brief description of the cognitive tasks used in the NeuroPharm trial and their primary outcomes are outlined in the following sections.

Hot cognition (emotion processing)

Emotion recognition

We used the eyes version of the *Face Emotional Recognition Task* from the EMOTICOM test battery to index the ability to recognize emotions in facial expressions. In this task, participants are asked to identify which emotion (happiness, sadness, anger or fear) is expressed in a pair of eyes briefly shown on a computer screen. As primary outcomes, we chose affective bias for recognition accuracy (hit rate, i.e. percentage of trials where the emotion was correctly identified) and misattribution (false alarm rate, i.e. percentage of trials in which a given emotion was wrongly identified). Affective bias was calculated in percentage as: $Score_{happy} - Score_{sad}$ with a possible range of -100-100%.

Emotion detection threshold

We used the *Emotional Intensity Morphing* task from the EMOTICOM test battery to index the perceptual threshold for detection of emotions in facial expressions. In this task, participants shown a face with a slowly morphing emotional expression and asked to indicate when they are able to detect (increase condition) or no longer detect (decrease condition) a prespecified emotion (happy, sad, angry, fearful, or disgusted). As primary outcome, we chose affective bias averaged across both conditions. Affective bias was calculated in percentage as: $Detection_{sad} - Detection_{happy}$ with a possible range of -100-100%.

Affective verbal memory

We used the *Verbal Affective Memory Test 26* developed by our group (Hjordt *et al.*, in review) to index learning and memory of affective words. In this task, participants are shown a list of words on a screen (10 positive words, 6 neutral words and 10 negative words). Immediately after viewing the list, the participant is asked to recall as many word as possible; this is repeated five times (immediate recall) followed by an interference list after which the participant is prompted to recall the original word list without viewing it again (short-term recall) and again after 30 minutes (long-term recall). As primary outcome, we chose affective bias for total word recall (i.e. average number of words recalled across

immediate, short-term and long-term recall). The affective bias was calculated in percentage as: $Words_{positive}$ - $Words_{negative}$ with a possible range of -100-100%.

Hot cognition (social cognition)

Moral emotions

We used the *Moral Emotions* task from the EMOTICOM test battery to index moral emotions in social situations. In this task, participants are shown a series of cartoons in which one character either intentionally or unintentionally harms another character. The participant is asked to imagine themselves as either the character causing harm (i.e. the agent) or the victim and rate how guilty, ashamed, annoyed and good/bad they would feel in the situation. As primary outcomes, we chose ratings of guilt and shame across all conditions (i.e. both agent and victim ratings for both intentional and unintentional harm situations). Possible scores range from 1-7.

Social information preference

We used the *Social Information Preference* task from the EMOTICOM test battery to index information sampling and interpretation of social situations. In this task, participants are shown a series of cartoons depicting socially ambiguous situations in which nine pieces of information is obscured (three facts/objects, three facial expression and three thought bubbles). The participants must choose which four pieces of information they would like to have revealed in order to help them choose between three interpretations of the situations; a positive, a negative or neutral. As primary outcomes, we chose preference for social information over non-social information calculated as: *Thoughts (%) + Faces (%) – Facts (%)* with possible scores ranging from -50-100% as well as affective interpretation bias calculated as: $Outcome_{positive} – Outcome_{negative}$ with possible scores ranging from -100-100%.

Cold cognition

Explicit verbal memory

We also used the *Verbal Affective Memory Test 26* to index overall verbal memory function. As primary outcome, we chose total word recall (i.e. average number of words remembered across immediate, short-term and long-term recall) across all three affective word categories (i.e. positive, negative and neutral). Possible scores range from 0-26.

Working memory

We used the *Letter Number Sequence* task from the Wechsler Adult Intelligence scale III (Wechsler, 1997) to index working memory. In this task, a sequence of jumbled letters and numbers are read out loud (e.g. 8-G-2-D-6) and the participant is asked to remember, mentally sort and then recite them in ordered sequence beginning with the numbers in numerical order followed by the letters in alphabetical order (e.g. 2-6-8-D-G). The sequences gradually increase in length/difficulty. As a primary outcome, we used number of correctly recited sequences. Possible scores range from 0-21.

Reaction time

We used a *Simple Reaction Time* task to index latency in reaction time. In this task, participants are instructed to press a computer key as quickly as possible when a white square appears on the screen. As primary outcome we used mean reaction time latency in milliseconds. There is no pre-specified range for this outcome.

5. Statistical analyses

Demographic and descriptive data were analysed with Student's t-test, Mann-Whitney U test or Analysis of Variance (ANOVA) for continuous variables and Chi-square test (χ^2) for categorical variables. In Study II and III, age and sex were included as co-variates in all linear regression models; linear mixed effect models; and Analysis of Covariance (ANCOVA) models. After careful

consideration, we chose not to include education or IQ as covariates as they may be directly affected by the depressive episode. For example, risk of school drop-out has been linked with depressive symptom severity (Dupéré *et al.*, 2018) and measures of fluid IQ (also known as performance IQ) have been shown to be vulnerable to bias in the depressive state (Sackeim *et al.*, 1992). As a general rule, the independent variable and covariates should be independent of each other, making IQ and education unsuitable as covariates in this case. However, we did investigate the effect of including education and IQ in Study II (reported in the Supplementary Materials) and found that it did not change the overall results. In Study II and III, we employed the Bonferroni-Holm method to correct for familywise error rate for hypotheses which tested all 11 primary cognitive outcomes (corrected *p*-values are denoted $p_{corrected}$). All statistical analyses were carried out in SPSS (v. 25) or R (version 3.5.1).

Study I

We used descriptive statistics to characterize participant scores across EMOTICOM task outcomes including mean, standard deviation, median, interquartile range, range, skewness, floor effect and ceiling effects. To assess test-retest reliability, we used Intraclass Correlation Coefficients and their 95% Confidence Intervals (95% CI) calculated with absolute two-way mixed models. We tested the original test developer's assumption that the EMOTICOM tasks captured three broad cognitive domains (i.e. *Emotion Processing*, *Motivation and Reward* and *Social Cognition*) with Spearman's rank correlation matrices indexing the shared variance between performance on tasks within the same domain. To corroborate the correlation matrices, we also conducted a exploratory factor analysis assessing the underlying factorial structure of the EMOTICOM test battery. Lastly, we used Spearman's rank correlation coefficients to assess the relationship between task performance and demographic and descriptive factors.

Study II

We assessed baseline differences in cognitive performance between patients with MDD and healthy controls with a series of linear regression models in which group, coded as a categorical variable, was entered as the independent variable and cognitive task performance was entered as the dependent

variable. In addition, Cohen's *d* estimates were used to quantify the effect size of the group differences. We further used Spearman's rank correlation coefficients to assess the relationship between cognitive performance at baseline and severity of depressive symptoms within the MDD patient group. Lastly, in an unplanned and exploratory set of analyses we used a combination of Hierarchical Cluster Analysis and K-means cluster analysis to identify subgroups with distinct cognitive profiles within the MDD cohort. This analysis was based on the cognitive outcomes showing significant group differences between healthy controls and MDD patients.

Study III

In this study, we used a series of logistic regression models to test the association between baseline scores on cognitive outcomes and categorical treatment response (remitter vs non-responder) at week 8. Subsequently, we used a series of linear regression models to test the association between baseline scores and relative treatment changes in $HDRS_6$ scores at week 8 and week 12. We also tested whether there were any differences between the three cognitive profile clusters in terms of categorical treatment outcome (remitter vs non-responder) at week 8 using a χ^2 test as well as clinical symptom improvement in $HDRS_6$ scores at week 8 and week 12 using ANCOVA models. To test antidepressant treatment effect on cognitive performance between baseline and week 12 we used a series of linear mixed effect models. Following this, we used Spearman's rank correlation coefficients to test the relationship between improvement on cognitive scores and improvements on clinical $HDRS_6$ scores at week 12. Lastly, we used a series of ANCOVA models to test whether the three cognitive profile clusters differed on how much their cognitive scores had improved between baseline and week 12.

RESULTS

The main results from the thesis work are outlined in the following sections. Additional details and discussion of individual findings are provided in Paper I-III.

6. Study I

We used data collected from 100 healthy participants to assess the psychometric properties of the hot cognitive test battery EMOTICOM in a Danish cohort. The Danish version of the EMOTICOM contained 11 tasks from three cognitive domains: *Emotion Processing*, *Motivation and Reward* and *Social Cognition*. Note, while four of these tasks are described in the Method chapter because of their inclusion in the NeuroPharm trial (section 5.4), a full description of all 11 tasks can be found in Paper I. In addition, while the *tasks* themselves were the same, the specific task outcomes assessed in the EMOTICOM validation differed slightly from those used in the NeuroPharm trial.

We estimated test-retest reliability based on data from 49 participants who had been retested after 3-5 weeks. The majority of EMOTIOM task outcomes exhibited fair to excellent retest reliability; however, three tasks exhibited poor test-retest coefficients across all outcomes (ICC > 0.4): *Face Affective Go/NoGo Task*, *Monetary Incentive Reward Task* and the *Adapted Cambridge Gambling Task*. In addition, almost 40% of the primary EMOTICOM task outcomes showed some degree of either floor or ceiling effects in our sample of healthy individuals. For most tasks, the ceiling/floor effects were mild to moderate; however, four tasks had severe floor/ceiling effects: *Face Affective Go/NoGo Task, Reinforcement Learning Task, Progressive Ratio Task* and *Prisoner's Dilemma* task. Overall, the correlation between tasks from the same broad cognitive domain (e.g. the *Emotion Processing Domain*) were weak, indicating little shared variance and rejecting the original test developers' hypothesis of tasks specific domains (see Figure 3). This was corroborated by an exploratory factor analysis which also failed to detect the proposed three-domain factorial structure in the data. Lastly, with very few exceptions, we found no relationship between cognitive performance on EMOTICOM tasks and demographic factors such as age, sex, education or IQ nor descriptive factors including mood, trait Neuroticism or self-rated motivation and diligence during task completion.

Figure 3. Correlation between EMOTICOM task outcomes

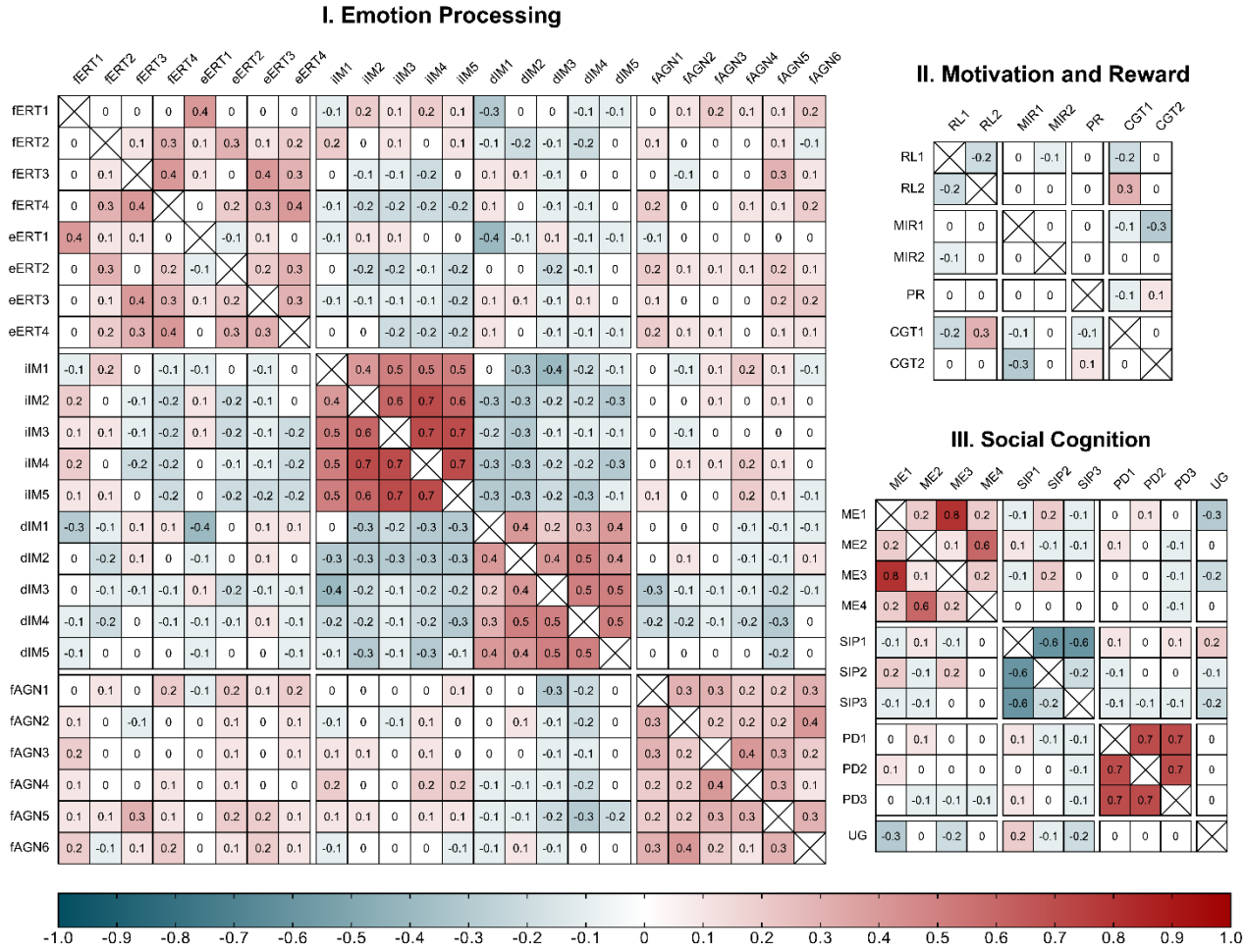


Figure 3. Spearman's Rank Correlations for EMOTICOM outcomes within the three proposed cognitive domains. I. Emotion Processing: fERT = face Emotion Recognition Task; fERT1 = hit rate for happy, fERT2 = hit rate for sad, fERT3 = hit rate for angry, fERT4 = hit rate for fearful. eERT = eyes Emotion Recognition Task; eERT1 = hit rate for happy, eERT2 = hit rate for sad, eERT3 = hit rate for angry, eERT4 = hit rate for fearful. iIM = increase Emotional Intensity Morphing Task; iIM1 = detection threshold for happy, iIM2 = detection threshold for sad, iIM3 = detection threshold for angry, iIM4 = detection threshold for fearful, iIM5 = detection threshold for disgusted. dIM = decrease Intensity Morphing Task; dIM1 = detection threshold for happy, dIM2 = detection threshold for sad, dIM3 = detection threshold for angry, dIM4 = detection threshold for fearful, dIM5 = detection threshold for disgusted. fAGN = Face Affective Go/NoGo Task; fAGN1 = d-prime for 'happy/neutral', fAGN2 = d-prime for 'happy/sad', fAGN3 = d-prime for 'neutral/happy', fAGN4 = d-prime for 'neutral/sad', fAGN5 = d-prime for 'sad/happy', fAGN6 = d-prime for 'sad/neutral'. II. Motivation and Reward: RL = Reinforcement Learning Task; RL1 = learning rate alpha for win condition, RL2 = learning rate alpha for loss condition. MIR = Monetary Incentive Reward Task; MIR1 = reaction time for win condition, MIR2 = reaction time for loss condition. PR = Progressive Ratio Task. aCGT = adapted Cambridge Gambling Task; aCGT1 = risk adjustment for win condition, aCGT2 = risk adjustment for loss condition. III. Social Cognition Domain: ME = Moral Emotions Task; ME1 = guilt for agent, ME2 = guilt for victim, ME3 = shame for agent, ME4 = shame for victim. SIP = Social Information Preference Task; SIP1 = proportion thoughts, SIP2 = proportion faces, SIP3 = proportion facts. UG = Ultimatum Game.

7. Study II

In this study, we used baseline data from the NeuroPharm trial to investigate cognitive disturbances in patients with MDD prior to antidepressant treatment compared with healthy controls. Demographic and descriptive characterization of the MDD patient group and the healthy control group from the EMOTICOM validation is shown in Table 1.

Table 1. Descriptive data

	Depressed patients (n = 92)	Healthy controls (n = 103)	*p*-value
Age in years	27.3 ± 8.1 (18–57)	28.7 ± 7.3 (18–48)	0.2
Male/female	25/68	51/52	<0.001
MDI	34.5 ± 7.2 (16–50)[a]	4.9 ± 3.9 (0–20)[b]	<0.001
$HDRS_6$	12.4 ± 1.6 (7–17)	-	-
$HDRS_{17}$	22.8 ± 3.4 (18–31)	-	-

Table 1. Demographic and descriptive data. MDI = Major Depressive Inventory; $HDRS_6$ = item-6 Hamilton Depressive Rating Scale; $HDRS_{17}$ = Item-17 Hamilton Depressive Rating Scale. Values are presented as mean ± SD with range in brackets. [a]N = 90, and [b]N = 102 due to missing data.

Cognitive disturbances in depression

As expected, we found significant differences between the MDD patients and healthy controls across almost all hot and cold cognitive functions. These differences are summarized as Cohen's *d* effect sizes in Figure 4 (note, the zero-line represents mean scores of the healthy control group). We observed small to medium effect sizes for emotion processing outcomes; medium effect sizes for ratings of guilt and shame; and medium effect sizes for cold cognitive outcomes. The effect sizes for affective memory, information sampling and social interpretation bias were small to negligible, reflected by non-significant *p*-values for these outcomes.

Figure 4. Magnitude of cognitive disturbances in MDD

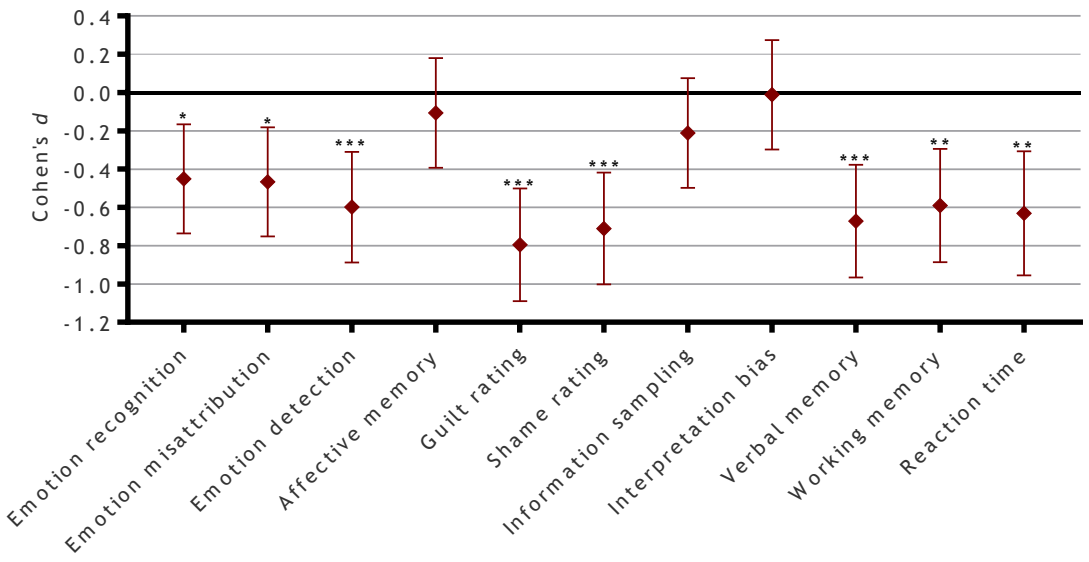


Figure 4. Magnitude of cognitive disturbances for MDD patients (N = 92) as Cohen's *d* effect sizes. The zero-line represents mean scores of the healthy control group (N = 103) and error bars denote 95% Confidence Intervals. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Note, *p*-values have been corrected for 11 tests with the Bonferroni-Holm method.

Correlation between depressive and cognitive symptom severity

We did not observe any significant correlations between depressive symptom severity indexed with $HDRS_6$ and cognitive scores in the patient group (ρ [-0.2; 0.2], all $p_{corrected} > 0.4$), suggesting a dissociation between cognitive symptoms and other depressive symptoms.

Cognitive profiles in depression

We used an exploratory cluster analysis approach to identify subgroups of patients with similar cognitive profiles. Based on performance on the eight hot and cold cognitive outcomes which were disturbed in MDD, we were able to identify three cognitive profile clusters (see Figure 5).

Figure 5. Cognitive profiles in depression

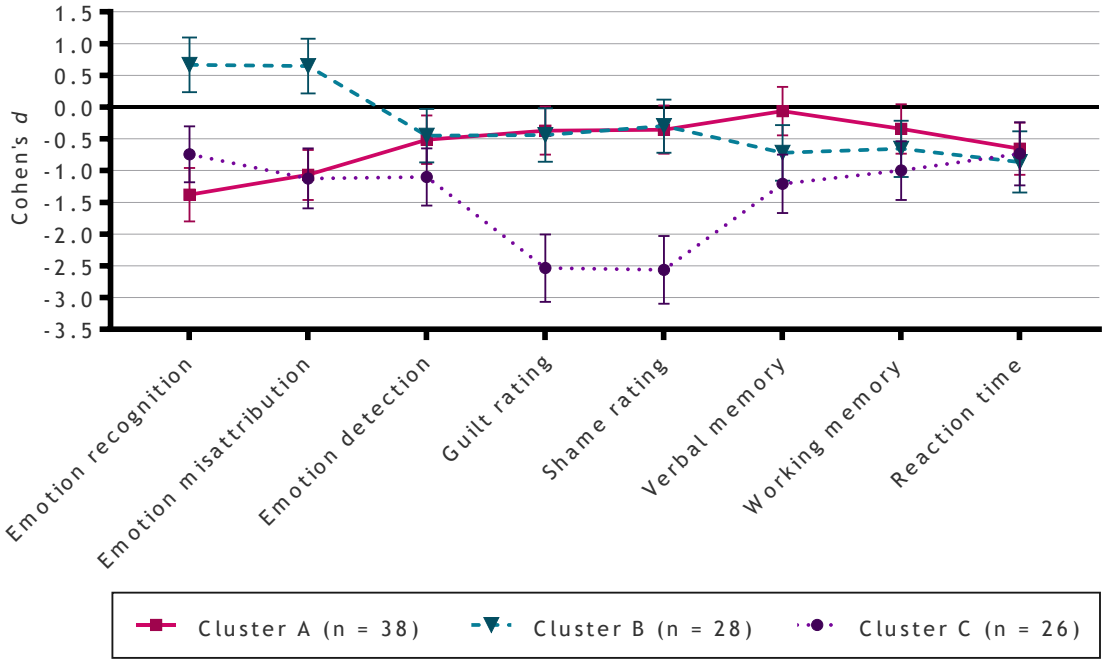


Figure 5. Magnitude of cognitive disturbances for the three cognitive clusters as Cohen's *d* effect sizes. The zero-line represents mean scores of the healthy control group (N = 103) and error bars denote 95% Confidence Intervals.

Cluster A was predominantly characterized by very strong negative affective biases in emotion processing with no other substantial disturbances in cognition except slowed psychomotor speed. Conversely, Cluster B was predominantly characterized by *positive* affective biases in emotion recognition and misattribution as well as moderate to strong deficits in cold cognitive domains. Lastly, Cluster C was characterized by severe global deficits across all cognitive domains including extreme ratings of guilt and shame in the *Moral Emotions* task. Notably, while the three Clusters did not differ in terms of age ($p = 0.6$) or sex ($p = 0.7$), Cluster C had slightly higher depression symptom scores compared with Cluster A and B ($p = 0.02$).

8. Study III

In this study, we used longitudinal cognitive and clinical data from the NeuroPharm trial to investigate cognitive disturbances as markers of antidepressant treatment response in patients with MDD. Figure 6 (not included in papers) provides a descriptive overview of improvements in depressive symptom severity over 12 weeks of treatment with escitalopram (with switch to duloxetine in 16 patients) across the three clinical outcome groups. At week 8, 14 patients (17.9%) fulfilled the criteria for non-responder status; 44 patients (56.4%) fulfilled the criteria for responder status; and 20 patients (25.6%) fulfilled the criteria for remitter status.

Figure 6. Changes in depressive symptoms over time

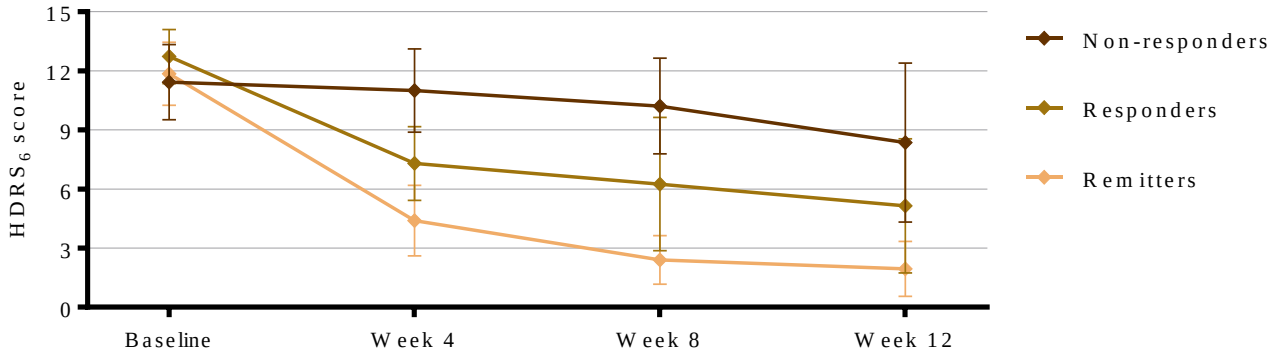


Figure 6. Hamilton Depressive Rating Scale 6 ($HDRS_6$) scores for the three treatment status groups. Error bars denote SD.

Clinical improvement after antidepressant treatment

Contrary to our hypothesis, we found no association between cognitive performance on any of the cognitive domains and treatment status (remitter vs non-responder) at week 8 (all $p_{corrected} = 1.0$) or between cognitive performance and percentage change in $HDRS_6$ scores at week 8 (all $p_{corrected} = 1.0$) or week 12 (all $p_{corrected} = 1.0$).

Cognitive improvements after antidepressant treatment

Figure 7 shows cognitive improvements after 12 weeks of antidepressant treatments.

Figure 7. Antidepressant treatment effects on cognition

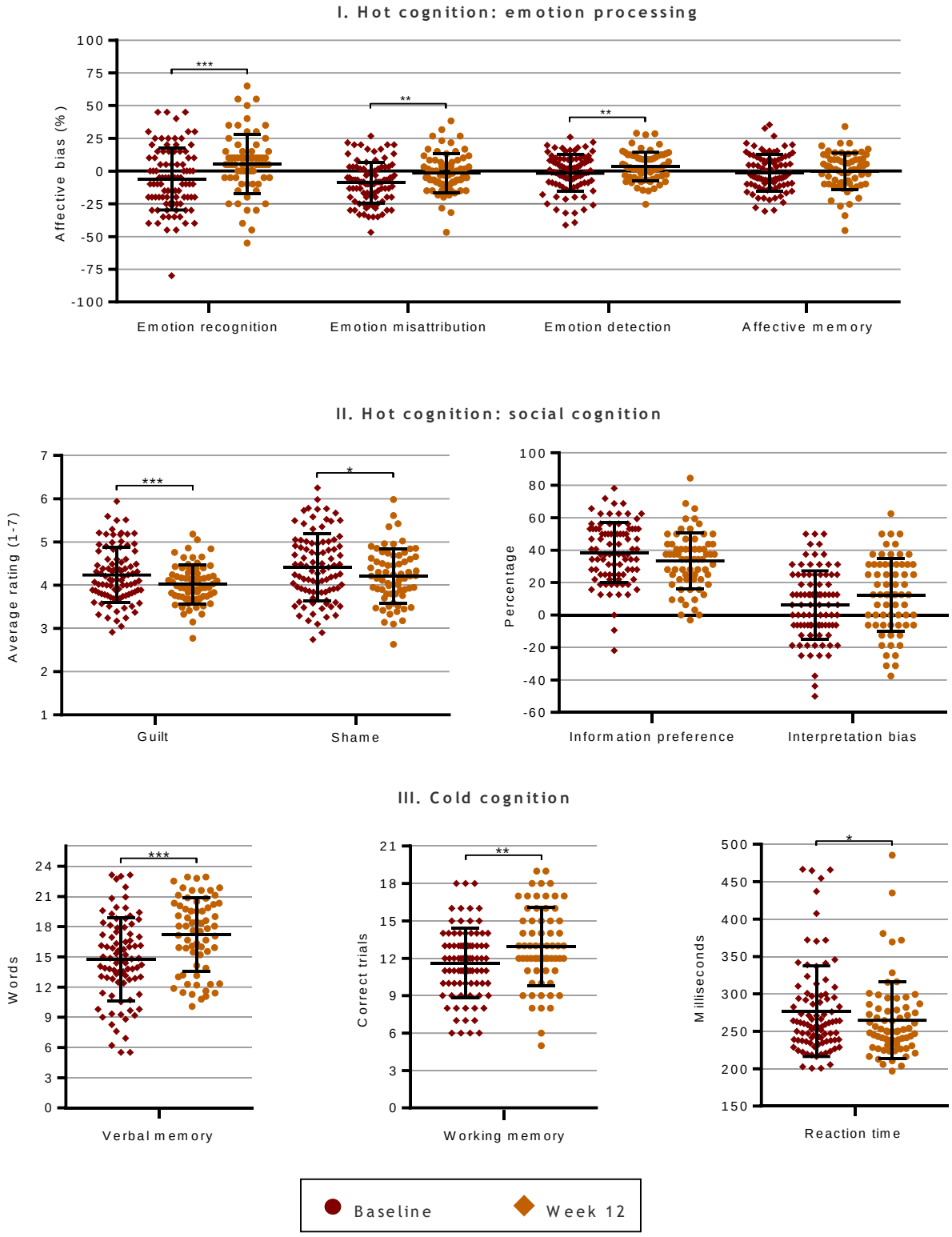


Figure 7. Differences in cognitive performance between baseline and 12 weeks follow-up for MDD patients (Baseline, N = 92; Week 12, N = 69) across hot and cold cognitive domains. Error bars denote SD. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Note, p-values have been corrected for 11 tests with the Bonferroni-Holm method.

As expected, we observed significant improvements in cognitive performance from baseline to week 12 for most of the hot and cold cognitive domains. Only three domains did not improve significantly: affective memory bias ($p_{corrected} = 1.0$), information preference ($p_{corrected} = 0.5$) and social interpretation bias ($p_{corrected} = 0.4$). Figure 8 (not included in papers) shows the magnitude of cognitive disturbances before and after 12 weeks of antidepressant treatment expressed as Cohen's d effect sizes.

Figure 8. Cognitive performance before and after antidepressant treatment

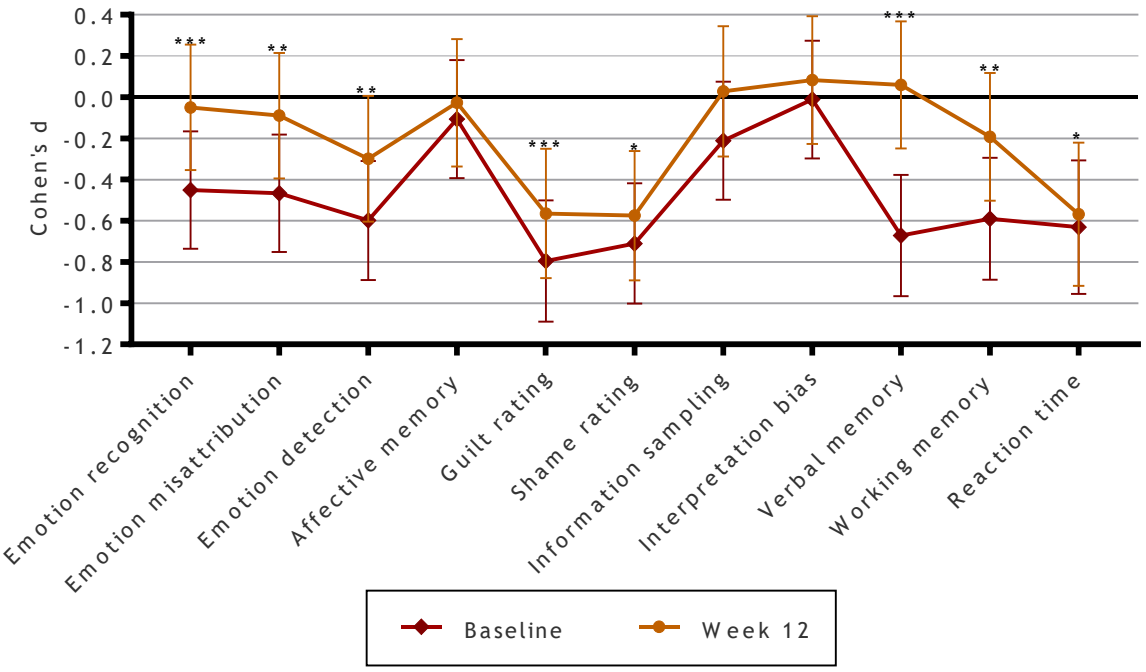


Figure 8. Cohen's d effect sizes of cognitive disturbances for MDD patients (baseline, N = 92; week 12, N = 69) relative to healthy controls scores at baseline (N = 103). The p-values represent differences between baseline and follow-up scores within the MDD patient group. Error bars denote 95% Confidence Intervals. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Note, p-values have been corrected for 11 tests with the Bonferroni-Holm method.

Clinical treatment response across cognitive profile clusters

When we assessed differences in clinical antidepressant treatment response between the three cognitive profile clusters from Study II, we found that Cluster C (characterized by severe global deficits) had poorer treatment response at week 8 compared with Cluster A and Cluster B ($p = 0.03$). However, this difference was no longer detectable at week 12 ($p = 0.8$) (see Figure 9 (not included in papers)). There were also no differences between the three cognitive profiles clusters when looking at remitter vs non-responder status at week 8 ($p = 0.2$), although the reduced sample size for this analysis could potentially have influenced this result.

Figure 9. $HDRS_6$ scores over time for cognitive profile clusters

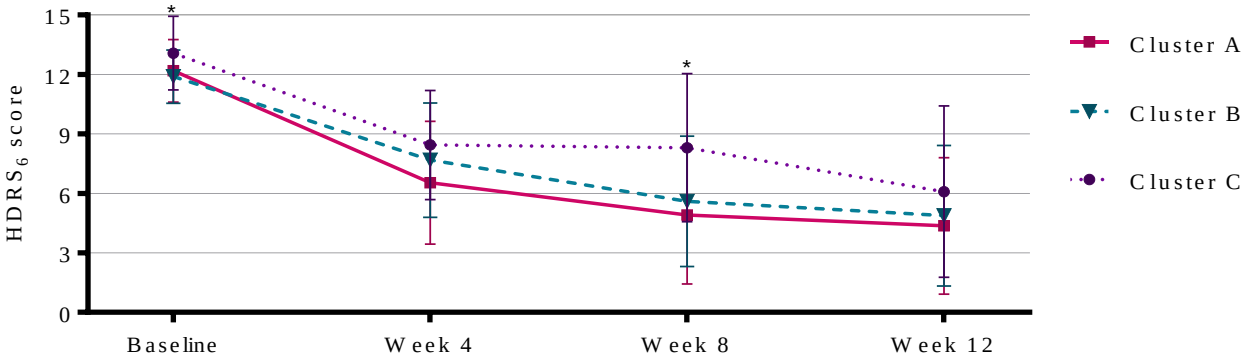


Figure 9. Hamilton Depressive Rating Scale 6 ($HDRS_6$) scores for the three cognitive profile clusters across 12 weeks of antidepressant treatment. There were significant group differences between the three clusters at baseline and week 8 (note, while mean $HDRS_6$ scores are shown here the analysis for week 8 was based on percentage change in $HDRS_6$ score). Error bars denote SD. * $p < 0.05$.

Correlation between clinical and cognitive improvements

We then looked at whether the improvements in cognitive performance were linked to improvements in clinical depressive symptom severity. We found no statistically significant correlations between changes in cognitive performance from baseline to week 12 and changes in $HDRS_6$ symptoms from baseline to week 12 ($\rho = [-0.21;0.16]$, all $p_{corrected} = 1.0$). In conjunction with the findings in Study II, this further supports a dissociation between cognitive symptoms and other core depressive symptoms in MDD.

Cognitive improvement across cognitive profile clusters

We found that the degree of cognitive improvement varied across the three clusters for ratings of guilt ($p_{corrected} < 0.001$) and shame ($p_{corrected} < 0.001$). In addition, the three clusters also differed at trend levels across several other cognitive outcomes including biases in emotion recognition ($p = 0.05$, $p_{corrected} = 0.6$), emotion misattribution bias ($p = 0.03$, $p_{corrected} = 0.3$) and working memory ($p = 0.06$, $p_{corrected} = 0.7$). Figure 10 shows the trajectory of cognitive change from baseline to week 12 follow-up for each of the three clusters. Interestingly, these differences appeared to follow a pattern where cluster(s) with the most impaired scores at baseline experienced the greatest improvement while cluster(s) with the least impaired scores at baseline experienced little to no improvements. We therefore speculate that the antidepressant effect on cognition may be moderated by the severity of the impairment.

Figure 10. Changes in cognitive performance for cognitive profile clusters

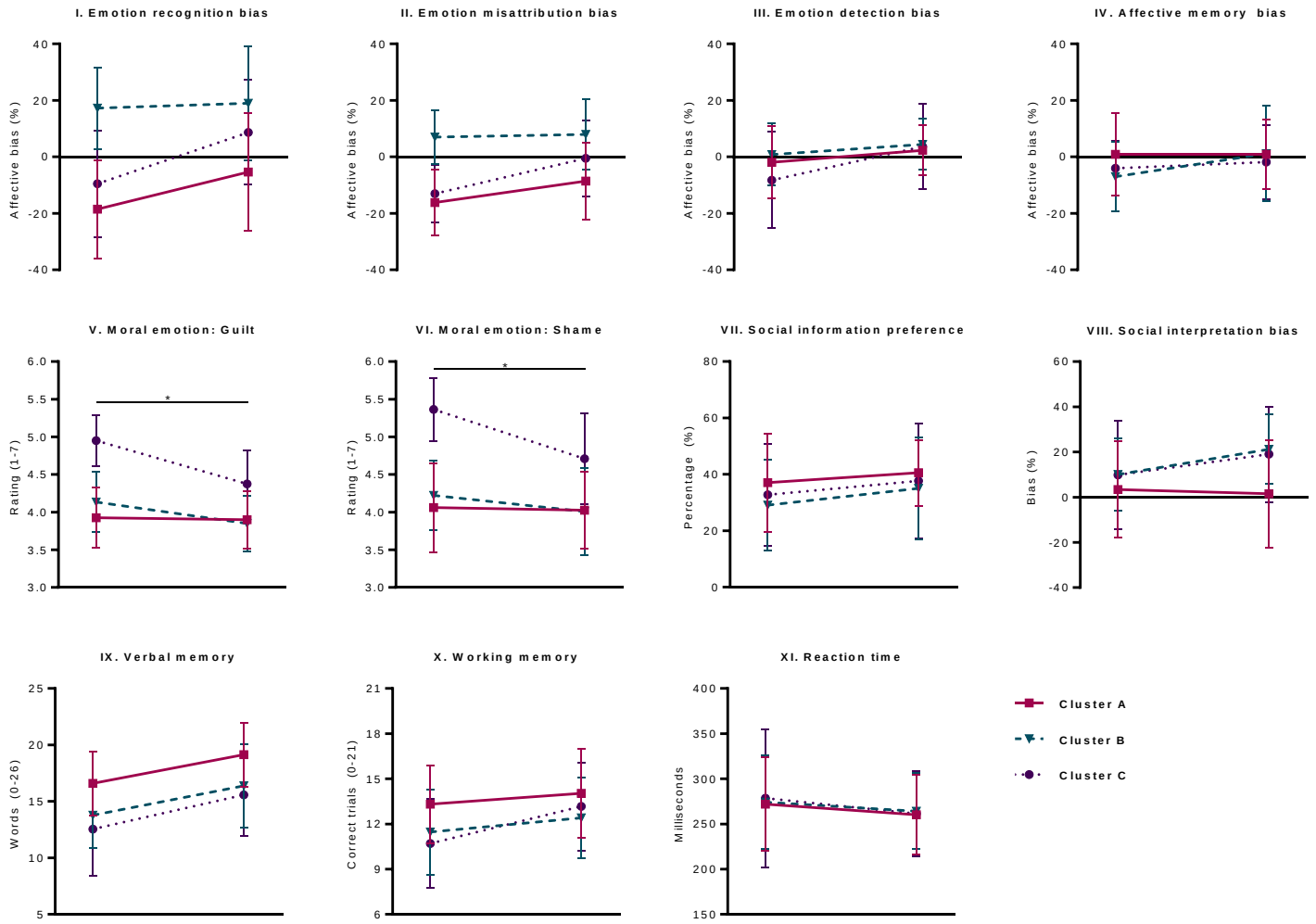


Figure 10. Changes in cognitive performance from baseline to week 12 for the three cognitive clusters. The graphs show mean raw scores at baseline and follow-up at week 12 and the error bars denote SD. We found a significant main effect of changes in cognitive performance at group level for guilt and shame ratings (graph V and VI). Note, *p*-values have been corrected for 11 tests with the Bonferroni-Holm method. * $p < .05$.

9. Additional cross-study analyses

Pre-treatment disturbances and antidepressant effect on cognition

Based on our findings in Study III, we speculated that antidepressant effects on cognitive symptoms in MDD may be moderated by the severity of the pre-treatment disturbances. In order to test this hypothesis more directly, we divided the MDD patients into two groups: A Cognition^{+} group containing patients whose score was equal to or better than the healthy control mean and a Cognition$^{\div}$ group containing patients whose score was worse than the healthy control mean; this was done for each cognitive outcome separately. We then used a series of ANCOVA models, corrected for age and sex, to compare the change in cognitive score (baseline to week 12) between the Cognition^{+} and the Cognition$^{\div}$ groups. As a pseudo-reference group, we also included test-retest data from the 49 EMOTICOM participants who had completed cognitive testing again after 4-6 weeks. Figure 11 shows the the trajectory of cognitive change over time for the Cognition^{+}, Cognition$^{\div}$ and the EMOTICOM retest group.

The ANCOVA models showed significant main effect of group on all hot cognitive outcomes (graph I-VIII; all $p_{corrected} < 0.008$) while the effects did not survive correction for multiple comparisons for the three cold cognitive outcomes (graph IX-XI; all $p < 0.03$, all $p_{corrected} > 0.1$). Just from the graphs, it is clear that the hot cognitive scores of the Cognition$^{\div}$ group improved over time while the scores of the Cognition^{+} group declined slightly, supporting our hypothesis. This was further confirmed with *post hoc* analyses showing not only significant differences between the Cognition^{+} and Cognition$^{\div}$ group (all $p_{corrected} < 0.01$) but also between the Cognitive$^{\div}$ group and the healthy control group for all but two hot cognitive outcomes (affective memory bias, $p_{corrected} = 1.0$; social information bias, $p_{corrected} = 0.9$; all other $p_{corrected} < 0.01$). Perhaps even more convincing, the *post hoc* analyses revealed no substantial differences between the Cognition^{+} group and the healthy control group except for a borderline difference on affective memory bias ($p = 0.005$, $p_{corrected} = 0.053$; all other $p_{corrected} > 0.2$).

Thus, the analyses showed that MDD patients with poor cognitive scores at baseline improved significantly more than both patients with good baseline scores and healthy participants across hot cognitive domains. Meanwhile, there were no statistically detectable differences between the three groups on any of the cold cognitive domains.

Figure 11. Changes in cognitive performance for

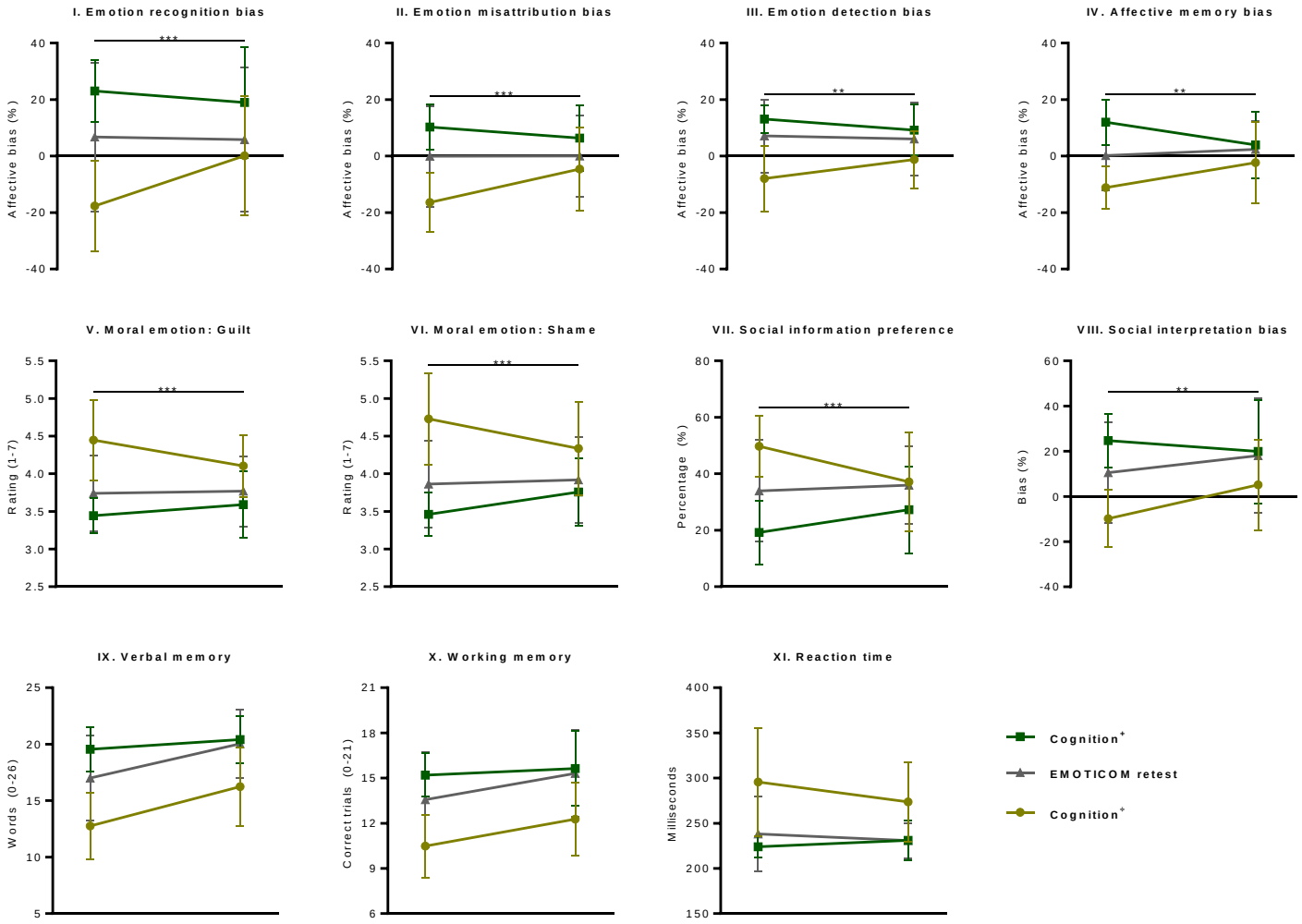


Figure 11. Changes in cognitive performance over time for Cognition + patients and Cognition ÷ patients (12 weeks) and 49 healthy EMOTICOM participants (3-5 weeks). The graphs show mean raw scores at baseline and follow-up time points. Error bars denote SD. We found a significant main effect of changes in cognitive performance at group level for all hot cognitive outcomes (graph I-VIII). Note, *p*-values have been corrected for 11 tests with the Bonferroni-Holm method. ** $p < 0.01$, *** $p < 0.001$.

Scarring effect of previous depressive episodes

Although understudied, there have been reports that number of previous depressive episodes may impact the severity of cognitive impairments, reflecting a so-called 'scar effect' (Hasselbalch *et al.*, 2011, Talarowska *et al.*, 2015). Figure 12 shows the distribution of depressive history in the NeuroPharm cohort. It should be noted that this information was based primarily on patient testimony and may therefore be subject to a certain degree of uncertainty. ANCOVA models corrected for age and sex revealed no significant difference between first-episode depression (n = 41) and recurrent depression (n = 51) on any of the clinical outcomes including baseline $HDRS_6$ scores ($p = 0.7$) or treatment response at week 8 ($p = 0.9$) or week 12 ($p = 0.8$). In addition, there were no differences between the two groups on any cognitive outcome at baseline (all $p_{corrected} > 0.8$) or treatment response at week 12 (all $p_{corrected} > 0.6$). Similarly, using Spearman's rank correlation coefficients, we found no relationship between number of depressive episodes and any clinical (baseline, $p = 0.8$; week 8, $p = 0.5$; week 12, $p = 0.4$) or cognitive (baseline, all $p_{corrected} > 0.5$; week 12 change, all $p_{corrected} > 1.0$) outcomes within the recurrent depression group. In conclusion, we found no evidence of a scarring effect on any pre-treatment or longitudinal clinical or cognitive outcomes in the NeuroPharm MDD cohort.

Figure 12. Previous number of depressive episodes

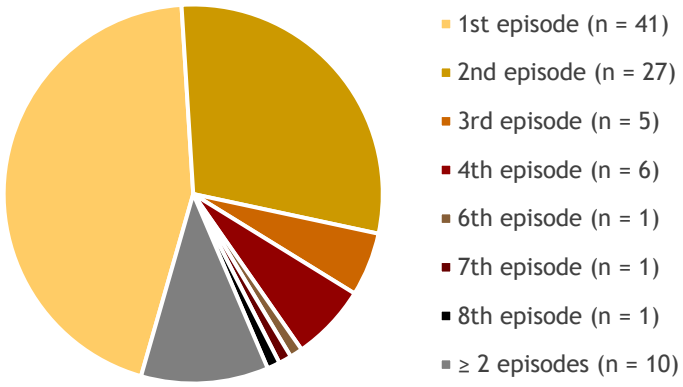


Figure 12. Distribution of MDD patients (N = 92) with first episode depression and recurrent depression (i.e. ≥ 2 episodes). Note for n = 10 patients, the exact number of depressive episodes could not be verified beyond recurrent depression.

DISCUSSION

The main discussion points from the thesis studies are summarized in the following sections. For a more in-depth and detailed discussion of the individual findings, the reader is referred to Paper I-III.

10. The EMOTICOM test battery

Until recently, a major challenge in the field of hot cognition has been the lack of easily accessible standardized psychometric tools capable of reliably assessing affective and social cognitive functions. While many cold cognitive test batteries exist (e.g. the CANTAB (Sandberg, 2011) and D-KEFS (Delis *et al.*, 2001) batteries), EMOTICOM represents the first comprehensive neuropsychological test battery specifically designed to capture hot cognitive functions (Bland *et al.*, 2016). In practice, the use of different, and often non-validated, cognitive tasks makes it difficult to replicate and compare studies because it cannot be determined if contrasting results are caused by true differences in the study populations or by differences in cognitive measures. A major aim of this thesis work was therefore to introduce and validate a new set of hot cognitive tools in Danish in the hope that the EMOTICOM battery will become a useful set of tools for researchers and help standardize future studies of hot cognitive functions. Overall, the EMOTICOM test battery showed satisfactory psychometric properties although we recommend that select tasks should be modified. In particular, tasks from the *Motivation and Reward* domain suffered from poor test-retest reliability and problems with floor and ceiling effects while tasks from the *Emotion Processing* and *Social Cognition* domains generally performed better (with select exceptions, e.g. the *Face Affective Go/NoGo Task*). We did not find evidence to support the presence of a three-domain structure proposed by the original test developers, highlighting that hot cognition describes a highly complex system of diverse cognitive constructs. Apart from establishing a Danish reference material of healthy individuals in Study I, we were also able to show that three out of the four EMOTICOM tasks included in the NeuroPharm were sensitive to MDD pathology (i.e. the *Face Emotional Recognition Task*, the *Intensity Morphing* task and the *Moral Emotions* task). Meanwhile, we did not find any statistically significant difference between patient and healthy control scores at baseline nor detect an effect of antidepressant treatment on the *Social*

Information Preference Task. While this could reflect that preference for social information and social interpretation biases are not affected in MDD, we speculate that it may instead be related to the task design itself. For example, when the participant is asked to choose an interpretation of the social situation, several of the scenarios have options with distinct paranoid components, e.g. the kidnapping of a child or a colleague poisoning a cup of tea. It is likely that these task features are better suited for capturing severe psychopathology associated with e.g. paranoid schizophrenia and not the more subtle interpretation biases associated with MDD.

11. Cognitive disturbances in depression

As expected, the depressed patients exhibited significant disturbances in cognitive performance compared with healthy individuals across the majority of the cognitive outcomes. The only exception was performance on affective memory bias, social information preference and social interpretation bias. As these three outcomes also failed to show significant improvement after 12 weeks of antidepressant treatment (Study III) this may indicate the tasks are not be sensitive to MDD pathology or alternatively that these cognitive functions are not affected to the same degree in MDD. Otherwise, the effect sizes of the observed disturbances all ranged from moderate to small aligning with the existing literature (Maruff and Jaeger, 2016). Additionally, because we included outcomes from both hot and cold cognitive tasks, we were able to directly compare and contrast the effect sizes of the disturbances. Our results showed that the magnitude of disturbance was similar across hot and cold domains, confirming that MDD pathology affects cognitive functioning broadly.

We were able to use the results from the exploratory factor analysis to further nuance the findings from the main group comparisons. As described previously, we identified three patient clusters with distinct cognitive profiles. Cluster A patients were characterized by very pronounced negative affective biases but no other notable cognitive disturbances. Cluster B patients on the other hand had positive affective biases and moderate deficits in the cold cognitive domains. Meanwhile, Cluster C patients showed severe disturbances across all cognitive domains. The presence of these clusters within our MDD cohort provides two key insights: *1)* the severity of cognitive symptoms in MDD is very heterogeneous and the small to moderate effects sizes typically reported in the literature likely obscure the presence of

a small subgroup of patients – Cluster C type patients – who are experiencing severe global impairments and *2)* hot and cold cognitive disturbances do not necessarily co-occur in MDD as evidenced by the different profiles exhibited by Cluster A and Cluster B patients. Interestingly, the proportion of Cluster C patients in our study fits remarkably well with reports from other studies that also used cluster analysis to identify cognitive subgroups in MDD. In the NeuroPharm cohort, Cluster C type patients with global disturbances made up ~28% of the total patient cohort while subgroup of cognitively impaired patients made up ~25% in the large iSPOT-D trial (Etkin *et al.*, 2015) and ~26% in a smaller study on first-episode depression (Vicent-Gil *et al.*, 2018). Together, these three separate cluster analysis studies strongly indicate that the proportion of patients experiencing global cognitive deficits in MDD is approximately one in four. Meanwhile, because the two other studies only included cold cognitive measures (the iSPOT-D trial did include outcomes related to facial emotion processing but did not differentiate between positive and negative emotions) they were not able to detect the presence of Cluster B type patients who had minimal deficits in cold cognitive domains but strong negative affective biases. The clinical implications of the cognitive profile clusters will be discussed in the following section.

12. Cognition as marker of antidepressant treatment response

We did not find any significant associations between baseline scores on any of the cognitive outcomes and later clinical response to antidepressant treatment. This negative finding remained consistent whether we used categorical treatment status (remitter vs non-responder) at week 8 or overall clinical improvement indexed as percentage change in $HDRS_6$ from baseline to week 8 or week 12 as clinical outcome. There are some reports that pre-treatment impairments in cold cognitive functions are associated with worse treatment response in MDD but as a whole the literature is conflicted (Groves *et al.*, 2018). At the same time, the relation between pre-treatment disturbances in hot cognitive functions and treatment response is currently understudied. Despite our negative findings it is therefore not possible to rule out that pre-treatment scores on any single cognitive outcome may have some predictive value. However, we speculate that any such value is too small to be useful for directing treatment choices in MDD. This does not mean that cognition should be discarded as a potentially

useful stratification tool, though. When we looked at clinical differences between the three cognitive profile clusters, we found that the globally impaired Cluster C patients had slightly higher $HDRS_6$ scores at baseline compared with Cluster A and Cluster B, indicating that cognitive profiles mapping performance across several cognitive domains are more sensitive to MDD pathology. Importantly, we found that Cluster A and B patients had better treatment response at week 8 compared with Cluster C patients. Specifically, Cluster A patients showed 58.8% reduction in $HDRS_6$ scores, Cluster B patients showed 53.0% reduction while Cluster C patients showed 36.8% reduction; thus, there was a more than 20 percentage points difference in treatment response between Cluster A and Cluster C patients. This also corroborates one of the main findings from the iSPOT-D trial where the cognitively impaired patient group showed poorer treatment response at after 8 weeks of antidepressant drug treatment compared with cognitive intact patients (Etkin *et al.*, 2015). Unlike the NeuroPharm trial, the iSPOT-D trial did not report clinical follow-up data after 12 weeks of treatment. This is relevant because, intriguingly, we found that the observed difference in treatment between the three clusters was no longer detectable at week 12. It is apparent from Figure 9 that clinical improvements for Cluster A and B do not change much between week 8 and week 12, whereas Cluster C patients experience a much larger improvement over the same time span. Thus, at week 12 Cluster A showed 62.3% reduction in $HDRS_6$ scores, Cluster B showed 59.9% reduction and Cluster C showed 56.7% reduction. Based on this, we speculate that global cognitive disturbances in the depressed state at baseline may delay but not necessarily impair long-term treatment effects of antidepressant drugs in MDD. This interpretation is partially supported by the findings from Vicent-Gil *et al.* (2018) who reported no difference between cognitive intact and cognitive impaired patients after 12 months of SSRI treatment. Current clinical guidelines recommend that patients should be switched to a different antidepressant class of drug if adequate treatment response is not achieved after 4-8 weeks (Cleare *et al.*, 2015). If our findings are correct, this timeframe may be too short for Cluster C type patients and clinicians should therefore consider waiting longer before deciding to switch medication for patients with severe cognitive disturbances.

As a final note, based on the cognitive neuropsychological model of depression, early changes in emotion processing measured after 1-2 weeks of antidepressant treatment have been proposed as a useful predictor of treatment response (Harmer *et al.*, 2017). However, despite representing a promising new line of research (Browning *et al.*, 2019, Kingslake *et al.*, 2017), we were unable to

investigate this aspect of cognition in MDD, as we did not collect follow-up cognitive data during the early phases of treatment.

13. Cognitive disturbances as a treatment target

As expected, cognitive performance improved across both hot and cold domains over the course of 12 weeks of antidepressant treatment; the only exceptions were the three hot cognitive domains which also did not show any group difference between the healthy controls and patients with MDD at baseline, i.e. affective memory bias, social information preference and social interpretation bias. Thus, overall our findings corroborate previous reports that antidepressant treatment has a positive effect on cognition in MDD (Harmer and Cowen, 2013, Prado *et al.*, 2018). We also found significant differences between the three cognitive profile clusters in terms of the magnitude of change in cognition on several cognitive outcomes, i.e. patients from different clusters experienced different rates of cognitive improvements. Notably, it appeared that patient clusters with intact cognitive scores exhibited little to no improvement while the patient clusters with the impaired cognitive scores exhibited the most dramatic improvements. When we investigated this further by dividing the MDD patients into binary groups based on whether their cognitive scores were better (Cognition^{+}) or worse (Cognition$^{\div}$) than the healthy control average, this pattern only became more pronounced. This would seem to indicate that the rate of improvement in cognitive symptoms is moderated by the severity of the pre-treatment disturbances; or to put it simply, the patients with the worst cognitive symptoms also improve the most. It further implies that antidepressants do not act as cognitive enhancers in MDD but instead rescue lost or damaged cognitive functions. This also fits with reports from the literature that antidepressants do not improve (cold) cognitive functions in healthy individuals (Prado *et al.*, 2018). In addition, we included test-retest data collected from 49 healthy individuals in the EMOTICOM validation study in these analyses. Importantly, these data do *not* represent a directly comparable control group as the retest period in the EMOTICOM study was 3-5 weeks while the retest period in the NeuroPharm trial was 12 weeks. Even though this precludes a direct comparison between the two data sets, including the EMOTICOM test-retest data still provided some important insight. First, the shorter test-retest window in the EMOTICOM study should logically result in greater learning or practice effects compared with

the NeuroPharm trial. Therefore, any significant difference in cognitive change between the EMOTICOM retest group and the MDD patients is a strong indication of a treatment effect beyond learning/retest effect. This is exactly what we saw across most of the hot cognitive outcomes; in fact, based on the mostly flat lines representing the change between first and second testing for the EMOTICOM retest group (see Figure 11), the hot cognitive domains show only minimal learning effects. Meanwhile, the EMOTICOM retest group showed clear signs of learning/practice effects on the cold cognitive domains and in particular for verbal memory and working memory. This was further supported by the lack of group differences between the EMOTICOM retest and MDD groups for these outcomes. However, this does not necessarily mean that there were no antidepressant effects. As outlined above, we would expect learning effects to be significantly *greater* for the EMOTICOM retest group compared with the MDD patients due to the short time interval between tests. Since the overall improvements in task scores appear to be close to equal (as seen by the borderline parallel slopes), this actually indicates that the improved scores for the MDD patients are most likely a mix of learning and treatment effects. Unfortunately, because we do not have a properly matched control group (i.e. one tested after 12 weeks) we are not able to disentangle how much each effect is contributing to the overall improvement.

14. Dissociation between cognitive and depressive symptoms

We found no significant correlation between cognitive symptoms and depressive symptoms indexed with the $HDRS_6$ at baseline (Study II) or longitudinally (Study III). This absence of a substantial association was consistent across both hot and cold cognitive domains and was robust as evidenced by the reported effect sizes which were all small to negligible. It highlights not only how the HDRS fails to adequately capture – and therefore also monitor – cognitive disturbances in MDD but also that cognitive symptoms are not simply an epiphenomenon of the classic mood or somatic symptoms. In contrast with our negative findings, a meta-analysis conducted by McDermott and Ebmeier (2009) reported a significant correlation between depressive symptom severity and cognitive performance in the domains of episodic memory, executive functions and processing speed. As we did not include measures of these cognitive domains in our cognitive battery (with the exception of working memory

as an executive function) we were unable to test whether this would also be the case in the NeuroPharm cohort. It should be noted however, that the effect sizes reported by McDermott and Ebmeier (2009) were small (all weighted mean $r > -0.31$) and other recent and large studies found no or very limited associations between longitudinal changes in cognition and depressive symptoms (McIntyre *et al.*, 2014, Shilyansky *et al.*, 2016). Overall, the link between cognition and depressive symptoms appear to be so weak that is more meaningful to view cognitive disturbances as a distinct psychopathological facet of MDD.

As briefly touched upon in a previous section, the NeuroPharm study design did not allow us to address whether early changes in affective biases can be used to predict treatment response as proposed by the cognitive neuropsychological model of depression (Harmer and Cowen, 2013). However, if early changes in affective biases after 1 week of antidepressant treatment are associated with clinical treatment response, we would also expect to see an association after 12 weeks of antidepressant treatment. As we do not see any association between cognitive and clinical changes in $HDRS_6$ at week 12, our data indirectly fail to support the model's assumption. This does not necessarily mean that the cognitive neuropsychological model is wrong; it may simply be that the signal of the early association becomes harder to detect over time due to interreference from other factors, e.g. the quality of social support and overall life situation of the patient. Our findings do, however, emphasize that there are still many things we do not know about the interplay between antidepressants treatment, cognition and clinical treatment response. On a final note, we looked at 'scarring' effect from previous MDD history. The scar hypothesis posits that previous depressive episodes may cause long-term negative changes in cognitive and emotional functioning and that these changes may in turn contribute to make the remitted patient more vulnerable to relapse (Wichers *et al.*, 2010). However, evidence for a cognitive scarring effect in MDD is not only understudied but also conflicted. For example, a recent study using data from two large birth cohorts suggested that the observed diminished cognitive functioning in MDD is related to increased comorbidity with other psychiatric disorders rather than the number, lengths and severity of the depressive episodes (Schaefer *et al.*, 2017). In line with this, we were also unable to detect any effect of previous MDD history on any of our clinical or cognitive outcome measures.

15. Methodological considerations

Study strengths

The three studies included in this thesis had several notable strengths: *1)* By validating the EMOTICOM test battery in a Danish cohort, and combining them with well-established cold cognitive tasks from the Cimbi database setup, we ensured that all the cognitive tasks used in the NeuroPharm trial were psychometrically sound and well-characterized. *2)* In continuation of the previous point, we used a broad cognitive battery containing both hot and cold cognitive tasks to characterize the MDD patients in Study II and III; this allowed us not only to compare and contrast disturbances and treatment response across these domains but also to detect and characterize the three cognitive profile clusters within our depressed study population. *3)* A major strength of Study II and III was the inclusion of MDD patients who had taken no antidepressants for at least two months and whose current depressive episode had not lasted longer than two years. This ensured that the effect of the escitalopram/duloxetine treatment on cognition was not influenced by other antidepressant drugs and that the patient group was relatively homogeneous. *4)* Throughout all three studies, we endeavoured to maintain a high standard for statistical handling of the acquired data including the use of relevant co-variates and rigorous control of the family-wise error rate through the use of Bonferroni-Holm corrections for multiple comparisons in Study II and III.

Study limitations

The findings reported in Paper I-III should be considered in the context of some important methodological limitations: *1)* The sample size of 100 participants, including retest data from 49 participants, in Study I is relatively small for a validation study. This presents some limitations in the interpretations of some of the psychometric analyses; most notably we may not have been able to detect the presence of weak correlations between task performance and demographic and descriptive factors. In addition, we did not have sufficient sample size to reliably estimate the true factorial structure of the EMOTICOM tasks in our exploratory factor analysis beyond rejecting the proposed three-factor solution. *2)* The EMOTICOM validation cohort was not stratified for age or education level. This resulted in a cohort of relatively young and highly educated participants with limited

representability of the general population. *3)* Similarly, the MDD patients included in the NeuroPharm trial were strictly screened for psychiatric and somatic comorbidities and therefore only represent a select subpopulation of MDD patients, lowering the generalizability of findings. *4)* We did not include a control group (either healthy controls or placebo) in Study III. This means we were unable to estimate the effect of learning or practice on the longitudinal cognitive outcomes and as a consequence may be overestimating the effects of antidepressant treatment on cognition in MDD. As shown by the additional cross-study analysis (see subchapter 9), this does not appear to be a serious problem for any of the hot cognitive outcomes. However, the cold cognitive outcomes, and in particular the verbal and learning memory outcomes, showed clear learning effects. *5)* We only investigated the effect of antidepressant treatment of the SSRI escitalopram (and the SNRI duloxetine in a subgroup of patients). Therefore, our findings are not necessarily generalizable to other SSRIs or other classes of antidepressant drugs. *6*) In order to accommodate the limited stamina of the MDD patients, the cognitive test battery we used in Study II and III had to be relatively short. In addition, because we chose to prioritize the inclusion of both hot and cold cognitive domains, we had to leave out several cognitive task domains known to be sensitive to MDD pathology; e.g. reward and motivation tasks, attention paradigms and higher-order executive functions.

CONCLUSION

The main aims of this thesis work were to implement and validate a Danish version of the EMOTICOM test battery and to investigate hot and cold cognitive markers in MDD. Overall, the EMOTICOM test battery exhibited satisfactory psychometric properties although select tasks, primarily from the *Motivation and Reward* domain, might benefit from modification to avoid issues with poor test-retest reliability and floor and ceiling effects in healthy individuals. As part of the EMOTICOM validation, we created a reference sample of healthy individuals and compared them with MDD patients from the NeuroPharm depression trial. We found that patients with MDD exhibited clear cognitive disturbances across both hot and cold cognitive domains relative to the healthy controls. In addition, we detected the presence of three distinct cognitive profile clusters within the MDD patient group: Cluster A patients were predominantly characterized by negative affective biases in emotion

processing tasks with no other notable disturbances in cognition; Cluster B patients were predominantly characterized by positive affective biases in emotion processing tasks coupled with moderate deficits in cold cognitive domains; Cluster C patients were characterized by large global disturbances. We found that Cluster C patients had poorer clinical response after 8 but not 12 weeks of antidepressant treatment compared with the other two clusters. Moreover, we found that antidepressant treatment improved cognitive performance in the MDD patients but that these improvements were independent of improvement in clinical depressive symptoms indexed with $HDRS_6$.

16. Implications and future directions

Most MDD patients today are treated at a primary care centre (i.e. outside of hospital). This means that for an MDD biomarker to be useful it should be easily implemented, cheap and preferably non-invasive. Unlike the majority of MDD biomarkers currently being investigated (e.g. neuroimaging with EEG, fMRI and PET; inflammatory markers; oxidative stress; genetic and epigenetic markers etc. (Hacimusalar and Eşel, 2018)), measures of cognitive performance tick all of the above boxes. Our findings from the NeuroPharm depression trial join a small but rapidly growing literature highlighting the potential use of cognitive markers in predicting treatment response in MDD (Browning *et al.*, 2019, Etkin *et al.*, 2015). We showed that single cognitive measures at baseline have no clinical use as predictors of antidepressant treatment response, cognitive profiles derived from baseline performance on hot and cold cognitive tasks can be used to stratify patients into clinically meaningful groups. We also emphasized the importance of looking at cognitive disturbances as a distinct symptom, and therefore, treatment target in MDD. Lastly, by validating the EMOTICOM in a Danish cohort of healthy individuals and by testing its sensitivity to MDD pathology in a subset of tasks, we have provided a new set of standardized affective and social cognitive tools for future clinical and research use.

The following recommendations for future studies are based on the implications and limitations of Study I-III:

- Because of the paucity of studies investigating hot and cold cognitive domains simultaneously in MDD, the findings reported in Study II and III require replication. In particular, the existence

of not just two (i.e. impaired vs intact cognition) but three distinct cognitive profiles in MDD needs to be verified in independent samples.

- It will be relevant to test if cognitive profiles in MDD correspond to biological phenotypes such as neuroimaging characteristics, oxidative stress levels, genotypes or hormone levels. As a next step in the NeuroPharm trial, we therefore plan to combine cognitive data and PET neuroimaging of the 5-HT 4 receptor also collected from the MDD patients.
- It will also be relevant to test whether there is a link between cognitive profiles and treatment response to different antidepressant drugs or other types of treatment (e.g. neurological or psychotherapeutic). In addition, future studies should investigate whether early (and indeed overall) treatment response can be improved for Cluster C type patients with global cognitive disturbance by augmenting the standard antidepressant treatment with cognitive remediation or cognitive enhancers.
- Future studies should also investigate whether early changes in the EMOTICOM's affective bias outcomes after 1-2 weeks of antidepressant treatment can be used to predict clinical treatment response as predicted by cognitive neuropsychological model of depression. In particular, it would be interesting to investigate whether *1)* these early changes are greater for Cluster A and C type patients who exhibit marked pre-treatment negative biases and *2)* whether early changes only predict treatment outcome for Cluster A and Cluster C. If so, creating cognitive profiles based on baseline cognitive performance could potentially increase the specificity of this putative predictor of treatment response by indicating which subgroup of patients it is be relevant to apply it to.
- In order to account for test-retest effects, future longitudinal studies should aim to include either a placebo group (when ethically feasible) or a matched healthy control group.
- We also encourage future studies to investigate our hypothesis that baseline performance influences antidepressant effects on cognition; both in MDD patients and healthy participants. If confirmed, it would provide novel insight into antidepressant mechanisms of action. In addition, it could also help explain some of the conflicting findings in the literature as the likelihood of detecting antidepressant effect on cognition would depend partly on the composition of high-functioning vs low-functioning individuals in the study cohort.

- Based on our findings in Study III, we also encourage other longitudinal MDD studies to expand the follow-up window to at least 12 weeks.
- As we only included 4 out of 11 EMOTICOM tasks in the NeuroPharm trial, future studies are needed to evaluate the sensitivity of the remaining 7 tasks to MDD (and other psychiatric) pathology.

REFERENCES

Beck, A. T. (2008). The evolution of the cognitive model of depression and its neurobiological correlates. *Am J Psychiatry* **165**, 969-77.

Bland, A. R., Roiser, J. P., Mehta, M. A., Schei, T., Boland, H., Campbell-Meiklejohn, D. K., Emsley, R. A., Munafo, M. R., Penton-Voak, I. S., Seara-Cardoso, A., Viding, E., Voon, V., Sahakian, B. J., Robbins, T. W. & Elliott, R. (2016). EMOTICOM: A Neuropsychological Test Battery to Evaluate Emotion, Motivation, Impulsivity, and Social Cognition. *Front Behav Neurosci* **10**, 25.

Bora, E. & Berk, M. (2015). Theory of mind in major depressive disorder: A meta-analysis. *J Affect Disord* **191**, 49-55.

Brand, A. G. (1985). HOT COGNITION: EMOTIONS AND WRITING BEHAVIOR. *Journal of Advanced Composition* **6**, 5-15.

Brendan, C. C., Thorne, C. B., Hardy, S. & Cropsey, K. L. (2013). Cooperation and depressive symptoms. *J Affect Disord* **150**, 1184-7.

Browning, M., Kingslake, J., Dourish, C. T., Goodwin, G. M., Harmer, C. J. & Dawson, G. R. (2019). Predicting treatment response to antidepressant medication using early changes in emotional processing. *European Neuropsychopharmacology* **29**, 66-75.

Browning, M., Reid, C., Cowen, P. J., Goodwin, G. M. & Harmer, C. J. (2007). A single dose of citalopram increases fear recognition in healthy subjects. *J Psychopharmacol* **21**, 684-90.

Burcusa, S. L. & Iacono, W. G. (2007). Risk for recurrence in depression. *Clinical psychology review* **27**, 959-985.

Cambridge, O. R., Knight, M. J., Mills, N. & Baune, B. T. (2018). The clinical relationship between cognitive impairment and psychosocial functioning in major depressive disorder: A systematic review. *Psychiatry Res* **269**, 157-171.

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

Cleare, A., Pariante, C. M., Young, A. H., Anderson, I. M., Christmas, D., Cowen, P. J., Dickens, C., Ferrier, I. N., Geddes, J., Gilbody, S., Haddad, P. M., Katona, C., Lewis, G., Malizia, A., McAllister-Williams, R. H., Ramchandani, P., Scott, J., Taylor, D. & Uher, R. (2015). Evidence-based guidelines for treating depressive disorders with antidepressants: A revision of the 2008 British Association for Psychopharmacology guidelines. *J Psychopharmacol* **29**, 459-525.

Communications, C. H. (2015). Survey of British adults diagnosed with depression on behalf of Clark HEalth Communications. https://www.comresglobal.com/wp-content/uploads/2015/09/Clark-Health-Communications_Cognitive-Dysfunction-in-Depression.pdf.

Dalili, M. N., Penton-Voak, I. S., Harmer, C. J. & Munafo, M. R. (2015). Meta-analysis of emotion recognition deficits in major depressive disorder. *Psychol Med* **45**, 1135-44.

Delaveau, P., Jabourian, M., Lemogne, C., Guionnet, S., Bergouignan, L. & Fossati, P. (2011). Brain effects of antidepressants in major depression: a meta-analysis of emotional processing studies. *J Affect Disord* **130**, 66-74.

Delis, D. C., Kaplan, E. & Kramer, J. H. (2001). *Delis-Kaplan executive function system: technical manual*. Psychological Corporation.

Demyttenaere, K., Donneau, A. F., Albert, A., Ansseau, M., Constant, E. & van Heeringen, K. (2015). What is important in being cured from depression? Discordance between physicians and patients (1). *J Affect Disord* **174**, 390-6.

Driessen, E. & Hollon, S. D. (2010). Cognitive behavioral therapy for mood disorders: efficacy, moderators and mediators. *The Psychiatric clinics of North America* **33**, 537-555.

Dupéré, V., Dion, E., Nault-Brière, F., Archambault, I., Leventhal, T. & Lesage, A. (2018). Revisiting the Link Between Depression Symptoms and High School Dropout: Timing of Exposure Matters. *Journal of Adolescent Health* **62**, 205-211.

Elliott, R., Zahn, R., Deakin, J. F. & Anderson, I. M. (2011). Affective cognition and its disruption in mood disorders. *Neuropsychopharmacology* **36**, 153-82.
Etkin, A., Patenaude, B., Song, Y. J., Usherwood, T., Rekshan, W., Schatzberg, A. F., Rush, A. J. & Williams, L. M. (2015). A cognitive-emotional biomarker for predicting remission with antidepressant medications: a report from the iSPOT-D trial. *Neuropsychopharmacology* **40**, 1332-42.
Goodall, J., Fisher, C., Hetrick, S., Phillips, L., Parrish, E. & Allott, K. (2018). Neurocognitive Functioning in Depressed Young People: A Systematic Review and Meta-Analysis. *Neuropsychology Review* **28**, 216-231.
Greenberg, P. E., Fournier, A. A., Sisitsky, T., Pike, C. T. & Kessler, R. C. (2015). The economic burden of adults with major depressive disorder in the United States (2005 and 2010). *J Clin Psychiatry* **76**, 155-62.
Groves, S. J., Douglas, K. M. & Porter, R. J. (2018). A Systematic Review of Cognitive Predictors of Treatment Outcome in Major Depression. *Front Psychiatry* **9**, 382.
Gualtieri, C. T. & Morgan, D. W. (2008). The frequency of cognitive impairment in patients with anxiety, depression, and bipolar disorder: an unaccounted source of variance in clinical trials. *J Clin Psychiatry* **69**, 1122-30.
Hacimusalar, Y. & Eşel, E. (2018). Suggested Biomarkers for Major Depressive Disorder. *Noro psikiyatri arsivi* **55**, 280-290.
Harmer, C. J., Bhagwagar, Z., Perrett, D. I., Vollm, B. A., Cowen, P. J. & Goodwin, G. M. (2003). Acute SSRI administration affects the processing of social cues in healthy volunteers. *Neuropsychopharmacology* **28**, 148-52.
Harmer, C. J. & Cowen, P. J. (2013). 'It's the way that you look at it'--a cognitive neuropsychological account of SSRI action in depression. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **368**, 20120407.
Harmer, C. J., Duman, R. S. & Cowen, P. J. (2017). How do antidepressants work? New perspectives for refining future treatment approaches. *The Lancet Psychiatry* **4**, 409-418.
Harmer, C. J., Shelley, N. C., Cowen, P. J. & Goodwin, G. M. (2004). Increased positive versus negative affective perception and memory in healthy volunteers following selective serotonin and norepinephrine reuptake inhibition. *Am J Psychiatry* **161**, 1256-63.
Hasler, G. (2010). Pathophysiology of depression: do we have any solid evidence of interest to clinicians? *World psychiatry : official journal of the World Psychiatric Association (WPA)* **9**, 155-161.
Hasselbalch, B. J., Knorr, U. & Kessing, L. V. (2011). Cognitive impairment in the remitted state of unipolar depressive disorder: a systematic review. *J Affect Disord* **134**, 20-31.
Iverson, G. L., Brooks, B. L., Langenecker, S. A. & Young, A. H. (2011). Identifying a cognitive impairment subgroup in adults with mood disorders. *Journal of Affective Disorders* **132**, 360-367.
Joormann, J. & Gotlib, I. H. (2007). Selective attention to emotional faces following recovery from depression. *J Abnorm Psychol* **116**, 80-5.
Kaser, M., Zaman, R. & Sahakian, B. J. (2017). Cognition as a treatment target in depression. *Psychol Med* **47**, 987-989.
Kim, S., Thibodeau, R. & Jorgensen, R. S. (2011). Shame, guilt, and depressive symptoms: a meta-analytic review. *Psychol Bull* **137**, 68-96.
Kingslake, J., Dias, R., Dawson, G. R., Simon, J., Goodwin, G. M., Harmer, C. J., Morriss, R., Brown, S., Guo, B., Dourish, C. T., Ruhé, H. G., Lever, A. G., Veltman, D. J., van Schaik, A., Deckert, J., Reif, A., Stäblein, M., Menke, A., Gorwood, P., Voegeli, G., Pérez, V. & Browning, M. (2017). The effects of using the PReDicT Test to guide the antidepressant treatment of depressed patients: study protocol for a randomised controlled trial. *Trials* **18**, 558.
Knudsen, G. M., Jensen, P. S., Erritzoe, D., Baare, W. F., Ettrup, A., Fisher, P. M., Gillings, N., Hansen, H. D., Hansen, L. K., Hasselbalch, S. G., Henningsson, S., Herth, M. M., Holst, K. K., Iversen, P., Kessing, L. V., Macoveanu, J., Madsen, K. S., Mortensen, E. L., Nielsen, F. A., Paulson, O. B., Siebner, H. R., Stenbaek, D. S., Svarer, C., Jernigan, T. L., Strother, S. C. & Frokjaer, V. G. (2016). The Center for Integrated Molecular Brain Imaging (Cimbi) database. *Neuroimage* **124**, 1213-9.
Korn, C. W., Sharot, T., Walter, H., Heekeren, H. R. & Dolan, R. J. (2014). Depression is related to an absence of optimistically biased belief updating about future life events. *Psychological medicine* **44**, 579-592.
Lee, R. M., Dean, B. L. & Jung, K.-R. (2008). Social connectedness, extraversion, and subjective well-being: Testing a mediation model.(Report). *Personality and Individual Differences* **45**, 414.

LeMoult, J., Joormann, J., Sherdell, L., Wright, Y. & Gotlib, I. H. (2009). Identification of emotional facial expressions following recovery from depression. *Journal of abnormal psychology* **118**, 828-833.
Maruff, P. & Jaeger, J. (2016). Understanding the importance of cognitive dysfunction and cognitive change in major depressive disorder. In *Cognitive Impairment in Major Depressive Disoder* (ed. S. M. Roger). Cambridge University Press.
McDermott, L. M. & Ebmeier, K. P. (2009). A meta-analysis of depression severity and cognitive function. *J Affect Disord* **119**, 1-8.
McIntyre, R. S., Best, M. W., Bowie, C. R., Carmona, N. E., Cha, D. S., Lee, Y., Subramaniapillai, M., Mansur, R. B., Barry, H., Baune, B. T., Culpepper, L., Fossati, P., Greer, T. L., Harmer, C., Klag, E., Lam, R. W., Wittchen, H. U. & Harrison, J. (2017). The THINC-Integrated Tool (THINC-it) Screening Assessment for Cognitive Dysfunction: Validation in Patients With Major Depressive Disorder. *J Clin Psychiatry* **78**, 873-881.
McIntyre, R. S., Lophaven, S. & Olsen, C. K. (2014). A randomized, double-blind, placebo-controlled study of vortioxetine on cognitive function in depressed adults. *The international journal of neuropsychopharmacology* **17**, 1557-1567.
Miskowiak, K. W. & Carvalho, A. F. (2014). 'Hot' cognition in major depressive disorder: A systematic review. *CNS and Neurological Disorders - Drug Targets* **13**, 1787-1803.
Miskowiak, K. W., Ott, C. V., Petersen, J. Z. & Kessing, L. V. (2016). Systematic review of randomized controlled trials of candidate treatments for cognitive impairment in depression and methodological challenges in the field. *Eur Neuropsychopharmacol* **26**, 1845-1867.
Moore, M. T. & Fresco, D. M. (2012). Depressive realism: a meta-analytic review. *Clin Psychol Rev* **32**, 496-509.
Pereira, V. S. & Hiroaki-Sato, V. A. (2018). A brief history of antidepressant drug development: from tricyclics to beyond ketamine. *Acta Neuropsychiatr* **30**, 307-322.
Petersen, J. Z., Porter, R. J. & Miskowiak, K. W. (2019). Clinical characteristics associated with the discrepancy between subjective and objective cognitive impairment in depression. *J Affect Disord* **246**, 763-774.
Prado, C. E., Watt, S. & Crowe, S. F. (2018). A meta-analysis of the effects of antidepressants on cognitive functioning in depressed and non-depressed samples. **28**, 32-72.
Repantis, D., Schlattmann, P., Laisney, O. & Heuser, I. (2009). Antidepressants for neuroenhancement in healthy individuals: a systematic review. *Poiesis & Praxis* **6**, 139-174.
Rock, P. L., Roiser, J. P., Riedel, W. J. & Blackwell, A. D. (2014). Cognitive impairment in depression: a systematic review and meta-analysis. *Psychol Med* **44**, 2029-40.
Roiser, J. P., Elliott, R. & Sahakian, B. J. (2012). Cognitive mechanisms of treatment in depression. *Neuropsychopharmacology* **37**, 117-36.
Roiser, J. P. & Sahakian, B. J. (2013). Hot and cold cognition in depression. *CNS Spectrums* **18**, 139-149.
Romero, N., Sanchez, A. & Vazquez, C. (2014). Memory biases in remitted depression: the role of negative cognitions at explicit and automatic processing levels. *J Behav Ther Exp Psychiatry* **45**, 128-35.
Rush, A. J., Trivedi, M. H., Wisniewski, S. R., Nierenberg, A. A., Stewart, J. W., Warden, D., Niederehe, G., Thase, M. E., Lavori, P. W., Lebowitz, B. D., McGrath, P. J., Rosenbaum, J. F., Sackeim, H. A., Kupfer, D. J., Luther, J. & Fava, M. (2006). Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. *Am J Psychiatry* **163**, 1905-17.
Sackeim, H. A., Freeman, J., McElhiney, M., Coleman, E., Prudic, J. & Devanand, D. P. (1992). Effects of major depression estimates of intelligence. *Journal of Clinical and Experimental Neuropsychology* **14**, 268-288.
Sandberg, M. A. (2011). Cambridge Neuropsychological Testing Automated Battery. In *Encyclopedia of Clinical Neuropsychology* (ed. J. S. Kreutzer, J. DeLuca and B. Caplan), pp. 480-482. Springer New York: New York, NY.
Schaefer, J. D., Scult, M. A., Caspi, A., Arseneault, L., Belsky, D. W., Hariri, A. R., Harrington, H., Houts, R., Ramrakha, S., Poulton, R. & Moffitt, T. E. (2017). Is low cognitive functioning a predictor or consequence of major depressive disorder? A test in two longitudinal birth cohorts. *Dev Psychopathol*, 1-15.
Schumann, G., Binder, E. B., Holte, A., de Kloet, E. R., Oedegaard, K. J., Robbins, T. W., Walker-Tilley, T. R., Bitter, I., Brown, V. J., Buitelaar, J., Ciccocioppo, R., Cools, R., Escera, C., Fleischhacker, W., Flor, H., Frith, C. D., Heinz, A., Johnsen, E., Kirschbaum, C., Klingberg, T., Lesch, K. P., Lewis, S.,

Maier, W., Mann, K., Martinot, J. L., Meyer-Lindenberg, A., Muller, C. P., Muller, W. E., Nutt, D. J., Persico, A., Perugi, G., Pessiglione, M., Preuss, U. W., Roiser, J. P., Rossini, P. M., Rybakowski, J. K., Sandi, C., Stephan, K. E., Undurraga, J., Vieta, E., van der Wee, N., Wykes, T., Haro, J. M. & Wittchen, H. U. (2014). Stratified medicine for mental disorders. *Eur Neuropsychopharmacol* **24**, 5-50.
Semkovska, M., Quinlivan, L., O'Grady, T., Johnson, R., Collins, A., O'Connor, J., Knittle, H., Ahern, E. & Gload, T. (2019). Cognitive function following a major depressive episode: a systematic review and meta-analysis. *Lancet Psychiatry* **6**, 851-861.
Shilyansky, C., Williams, L. M., Gyurak, A., Harris, A., Usherwood, T. & Etkin, A. (2016). Effect of antidepressant treatment on cognitive impairments associated with depression: a randomised longitudinal study. *Lancet Psychiatry* **3**, 425-35.
Shiroma, P. R., Thuras, P., Johns, B. & Lim, K. O. (2014). Emotion recognition processing as early predictor of response to 8-week citalopram treatment in late-life depression. *Int J Geriatr Psychiatry* **29**, 1132-9.
Srisurapanont, M., Suttajit, S., Eurviriyanukul, K. & Varnado, P. (2017). Discrepancy between objective and subjective cognition in adults with major depressive disorder. *Scientific reports* **7**, 3901-3901.
Talarowska, M., Zajaczkowska, M. & Galecki, P. (2015). Cognitive functions in first-episode depression and recurrent depressive disorder. *Psychiatr Danub* **27**, 38-43.
Tranter, R., Bell, D., Gutting, P., Harmer, C., Healy, D. & Anderson, I. M. (2009). The effect of serotonergic and noradrenergic antidepressants on face emotion processing in depressed patients. *J Affect Disord* **118**, 87-93.
Trivedi, M. H., Rush, A. J., Wisniewski, S. R., Nierenberg, A. A., Warden, D., Ritz, L., Norquist, G., Howland, R. H., Lebowitz, B., McGrath, P. J., Shores-Wilson, K., Biggs, M. M., Balasubramani, G. K. & Fava, M. (2006). Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: implications for clinical practice. *Am J Psychiatry* **163**, 28-40.
Vicent-Gil, M., Keymer-Gausset, A., Serra-Blasco, M., Carceller-Sindreu, M., de Diego-Adelino, J., Trujols, J., Mur, M., Perez, V., Alvarez, E., Cardoner, N. & Portella, M. J. (2018). Cognitive predictors of illness course at 12 months after first-episode of depression. *Eur Neuropsychopharmacol* **28**, 529-537.
Wechsler, D. (1997). *WAIS-III : administration and scoring manual : Wechsler Adult Intelligence Scale*. Psychological Corporation: [San Antonio, Tex].
Weightman, M. J., Knight, M. J. & Baune, B. T. (2019). A systematic review of the impact of social cognitive deficits on psychosocial functioning in major depressive disorder and opportunities for therapeutic intervention. *Psychiatry Res* **274**, 195-212.
WHO (2017). Depression and other common mental disorders: global health estimates. World Health Organization: Geneva.
Wichers, M., Geschwind, N., van Os, J. & Peeters, F. (2010). Scars in depression: is a conceptual shift necessary to solve the puzzle? *Psychol Med* **40**, 359-65.
Zimmerman, M., Ellison, W., Young, D., Chelminski, I. & Dalrymple, K. (2015). How many different ways do patients meet the diagnostic criteria for major depressive disorder? *Compr Psychiatry* **56**, 29-34.
Zimmerman, M., McGlinchey, J. B., Posternak, M. A., Friedman, M., Attiullah, N. & Boerescu, D. (2006). How should remission from depression be defined? The depressed patient's perspective. *Am J Psychiatry* **163**, 148-50.

PAPER I-III

frontiers in Psychology

ORIGINAL RESEARCH
published: 03 December 2019
doi: 10.3389/fpsyg.2019.02660

Psychometric Properties and Validation of the EMOTICOM Test Battery in a Healthy Danish Population

*Vibeke H. Dam[1,2], Christa K. Thystrup[1], Peter S. Jensen[1], Amy R. Bland[3], Erik L. Mortensen[4], Rebecca Elliott[3], Barbara J. Sahakian[5,6], Gitte M. Knudsen[1,2], Vibe G. Frokjaer[1] and Dea S. Stenbæk[1]**

[1] Neurobiology Research Unit, The Neuroscience Centre, Copenhagen University Hospital Rigshospitalet, Copenhagen, Denmark, [2] Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark, [3] Neuroscience and Psychiatric Unit, The University of Manchester, Manchester, United Kingdom, [4] Department of Public Health and Center for Healthy Aging, University of Copenhagen, Copenhagen, Denmark, [5] Department of Psychiatry, University of Cambridge, Cambridge, United Kingdom, [6] Behavioural and Clinical Neuroscience Institute, University of Cambridge, Cambridge, United Kingdom

Disruptions in hot cognition, i.e., the processing of emotionally salient information, are prevalent in most neuropsychiatric disorders and constitute a potential treatment target. EMOTICOM is the first comprehensive neuropsychological test battery developed specifically to assess hot cognition. The aim of the study was to validate and establish a Danish language version and reference data for the EMOTICOM test battery. To evaluate the psychometric properties of 11 EMOTICOM tasks, we collected data from 100 healthy Danish participants (50 males, 50 females) including retest data from 49 participants. We assessed test–retest reliability, floor and ceiling effects, task-intercorrelations, and correlations between task performance and relevant demographic and descriptive factors. We found that test–retest reliability varied from poor to excellent while some tasks exhibited floor or ceiling effects. Intercorrelations among EMOTICOM task outcomes were low, indicating that the tasks capture different cognitive constructs. EMOTICOM task performance was largely independent of age, sex, education, and IQ as well as current mood, personality, and self-reported motivation and diligence during task completion. Overall, many of the EMOTICOM tasks were found to be useful and objective measures of hot cognition although select tasks may benefit from modifications to avoid floor and ceiling effects in healthy individuals.

Keywords: EMOTICOM, affective cognition, social cognition, hot cognition, psychometrics, neuropsychological test battery

OPEN ACCESS

Edited by:
Thomas Kleinsorge, Leibniz Research Centre for Working Environment and Human Factors (IfADo), Germany

Reviewed by:
Xinyang Liu, University of Oldenburg, Germany
Philip D. Harvey, University of Miami, United States

****Correspondence:***
Dea S. Stenbæk
dea@nru.dk

Specialty section:
This article was submitted to Cognition, a section of the journal Frontiers in Psychology

Received: *12 September 2019*
Accepted: *11 November 2019*
Published: *03 December 2019*

Citation:
Dam VH, Thystrup CK, Jensen PS, Bland AR, Mortensen EL, Elliott R, Sahakian BJ, Knudsen GM, Frokjaer VG and Stenbæk DS (2019) Psychometric Properties and Validation of the EMOTICOM Test Battery in a Healthy Danish Population. Front. Psychol. 10:2660. doi: 10.3389/fpsyg.2019.02660

INTRODUCTION

Hot cognition describes cognitive processing of emotionally salient information (Roiser and Sahakian, 2013). Examples of hot cognitive domains include basic emotion processing, motivation and reward driven behaviors as well as social cognition, i.e., the ability to understand and participate in social transactions. Importantly, disruptions in hot cognitive processes have been identified as core features in a wide range of neuropsychiatric disorders such as mood disorders (Elliott et al., 2011), anxiety disorders (Plana et al., 2014), schizophrenia (Ventura et al., 2013), Attention Deficit and Hyperactivity Disorder (ADHD) (Umemoto et al., 2014), and autism (Harms et al., 2010). In particular, negative affective biases, i.e., the preferential processing of negative information over

positive information, have consistently been shown in patients with mood disorders (Elliott et al., 2011; Hjordt et al., 2017), anxiety disorders (Mogg et al., 1995), substance abuse disorders (Ersche and Sahakian, 2007), and eating disorders (Lovell et al., 1997). Notably, one study found mood-congruent attentional biases in bipolar disorder where patients in the depressed state showed enhanced processing of negative information while patients in the manic state showed enhanced processing of positive information (García-Blanco et al., 2013). In contrast, healthy individuals typically show no or a slight positive affective bias (Pool et al., 2016). Meanwhile, impairments in motivation and reward-driven behaviors have been observed in psychopathological conditions including aggression (Kuin et al., 2015), traumatic brain injury (Newcombe et al., 2011), and ADHD (Umemoto et al., 2014) while differences in neural response to rewards and loss and disruptions in reinforcement learning have been linked to schizophrenia and major depressive disorder (MDD) (Chen et al., 2015; Hagele et al., 2015). Disturbances in social cognition including mentalization, i.e., the ability to infer the mental states of others, are central features of disorders such as autism and schizophrenia (Chung et al., 2014) and impairment in moral judgment has been reported for psychopathic individuals (Cardinale and Marsh, 2015), autism (Brewer et al., 2015), and patients suffering from ventromedial prefrontal cortex lesions (Cameron et al., 2018). In addition, self-blaming moral emotions such as guilt and shame have been shown to be exacerbated in MDD (Green et al., 2013) and anxiety disorders (Hedman et al., 2013). In healthy individuals, differences in hot cognitive processes have been linked to pharmacological interventions such as oxytocin (Leppanen et al., 2017) and serotonergic manipulations (Merens et al., 2007). Subclinical symptoms of depression and anxiety (Routledge et al., 2018), as well as natural sex hormone fluctuations in women (Osorio et al., 2018), also produce changes in hot cognition.

In summary, hot cognitive processes are relevant in a wide range of contexts across both normal and disturbed mental functioning. Notably, hot cognition has been proposed as an early predictor for treatment response in MDD (Harmer and Cowen, 2013; Park et al., 2018) as well as a promising target for therapeutic intervention (Roiser et al., 2012). Yet, despite growing recognition of their importance, scientists have so far lacked a validated and comprehensive set of tools capable of assessing hot cognitive processes in a standardized manner. Therefore, a group of researchers from Britain recently developed a novel 3-h computerized neuropsychological test battery called EMOTICOM (Bland et al., 2016). The EMOTICOM battery comprises 16 novel, adapted, and existing tasks designed to capture cognitive functions from four hot cognitive domains; (1) *Emotion Processing*, (2) *Motivation and Reward*, (3) *Impulsivity*, and (4) *Social Cognition*. The British developers validated the EMOTICOM battery in a cohort of 200 healthy participants (Bland et al., 2016). We here assess the psychometric properties of EMOTICOM in a shortened version using a Danish cohort of 100 healthy participants and provide reference data for research and clinical use of the test battery in Danish. In the British validation, test–retest reliability of the EMOTICOM battery was assessed after a relatively short time interval (5–10 days). In the present study we chose to collect retest data after 3–5 weeks in order to provide a reference for longitudinal studies investigating the effects of treatment or interventions over weeks or months. We also supplement the original British study findings by comparing performance on the EMOTICOM tasks in the shortened Danish battery with relevant factors such as personality, mood, and self-reported levels of motivation and diligence during task completion.

MATERIALS AND METHODS

Participants

One hundred healthy Danish participants between 18 and 48 years of age (males, $n = 50$; females, $n = 50$) were recruited from a previously established database of healthy volunteers (Knudsen et al., 2016) or through internet advertisements and flyers posted around the greater Copenhagen area. Exclusion criteria for the study included history of psychiatric disorders, significant somatic illness, brain trauma, use of psychotropic medication, significant lifetime history of drug abuse, pregnancy or breastfeeding, and non-fluency in Danish. The study was approved by the Danish Data Protection Agency (protocol RH-2015-255) and written informed consent was obtained from all participants.

Study Design

Upon inclusion, participants were randomized into single test or retest groups. Three participants originally randomized into the retest group dropped out after completing the first test session; one due to a family emergency and two failed to disclose the reason. To accommodate these dropouts, two unused single-test slots in the randomization system were converted into retest slots while the last dropout happened too late in the data collection process to be recovered. Thus, 51 participants completed a single test session while 49 participants completed retest sessions after 3–5 weeks (time between test–retest: 27.4 ± 4.8 days, mean ± SD)[1]. Intelligence quotient (IQ) was assessed with the Reynolds Intellectual Screening Test (RIST) using the verbal subtest 'Guess What?' and the non-verbal subtest 'Odd-Item Out' (Reynolds, 2011). Level of education was indexed with the Online Stimulant and Family History Assessment Module (OS-FHAM) questionnaire using a five-point Likert scale from 1 (no vocational degree) to 5 (>4 years of higher learning at university level). Personality was assessed with the NEO Personality Inventory Revised (NEO PI-R, $n = 93$) and the NEO Personality Inventory-3 (NEO PI-3, $n = 6$) (Costa and McCrae, 2005). Mood was assessed with the Profile of Mood State (POMS) (McNair and Heuchert, 2007) immediately before each test session. All test sessions took place in standardized testing rooms and were conducted by a team of five trained neuropsychological testers at the Neurobiology Research Unit, Copenhagen University Hospital Rigshospitalet.

[1]Due to scheduling conflicts, one participant completed the retest session after 6 weeks (43 days).

In addition to a flat fee of 200 Danish kroners, participants had the opportunity to win money based on their performance in six EMOTICOM tasks that included monetary reward. For these six tasks, participants were instructed to rate their performance during the task in terms of motivation and diligence, i.e., the degree to which they had 'done their best.' Participants were also encouraged to write down any thoughts or suggestions regarding the overall test experience or any specific task, followed by a brief unstructured interview at the end of each session. The order of tasks within the EMOTICOM battery was randomized to control for any potential effects of test order.

Translation and Implementation of EMOTICOM in Danish

Three native Danish speakers independently translated the full EMOTICOM test battery into Danish. Following a consensus meeting supervised by trained test psychologists, a single version was agreed upon. The consensus version was then back-translated into English by a natural English-Danish bilingual individual and sent for the approval of the original test developers. Implementation of the final Danish translation was done using the open source software PsychoPy. All monetary rewards were converted from British pounds to equivalent sums in Danish kroners.

The EMOTICOM Test Battery

Out of the original 16 tasks in the full EMOTICOM test battery, 11 were selected for translation and implementation in the Danish version. Two tasks, *The Four-choice Serial Reaction Time Task* and *The Discounting Task*, were not translated into Danish because the original test code was unavailable while two others, *The Emotional Memory Recognition Task* and *The Inference Task*, were left out based on the recommendation from the original British test developers who felt these tasks warranted further improvements. Lastly, due to translation concerns (e.g., issues relating to word length, frequency, and translation ambiguity), the *Word Affective Go No/Go* was also not implemented in the Danish validation. Therefore, only three of the original four hot cognitive domains, i.e., *Emotion Processing*, *Reward and Motivation*, and *Social Cognition*, were represented in the present study, while the last domain, *Impulsivity*, was left out. For a brief overview of the selected EMOTICOM tasks and their primary outcomes see **Table 1**. For a description of the full EMOTICOM battery see Bland et al. (2016).

Statistical Analysis

Statistical analyses were performed using SPSS statistical software (version 25.0) and R Studio (version 3.5). Missing data included NEO personality for one participant and self-reported ratings of motivation and diligence for five participants on the *Prisoner's Dilemma* and for one participant on the *Ultimatum Game*. Alpha levels were set at 0.01 for statistical significance in order to account for multiple comparisons.

Task Outcomes and Descriptive Statistics

Primary task outcomes for each EMOTICOM task were selected based on recommendations from the original British test developers and the existing literature. Descriptive and psychometric information on secondary outcomes can be found in the **Supplementary Information**. Mean, SD, median, interquartile range, range, and skewness are reported for all primary task outcomes. Floor and ceiling effects were determined as the percentage of participants who achieved minimum scores (floor effect) or maximum scores (ceiling effects) for a given task outcome. Floor or ceiling effects above 10% were considered moderate while effects above 30% were considered severe/problematic.

Test–Retest Reliability

To assess test–retest reliability, intraclass correlation coefficients (ICCs) and their 95% confidence intervals (95% CI) were calculated based on retest data from 49 participants using an absolute-agreement two-way mixed effect model. ICC values of less than 0.40 were considered poor, values between 0.40 and 0.59 as fair, values between 0.60 and 0.74 as good, and values greater than 0.75 as excellent (Cicchetti, 1994). In addition, test–retest bias, i.e., percent change in scores between first and second test, was calculated as: $Test\text{–}retest\ bias = ((score_{retest} - score_{test})/score_{test}) * 100$.

Task-Intercorrelations and Factor Analysis

To determine EMOTICOM's ability to capture the three proposed underlying cognitive domains, correlation matrices conducted with Spearman's rank correlations were used to index the shared marginal variance between tasks within the same cognitive domain, i.e., *Emotion Processing, Motivation and Reward,* and *Social Cognition*. In addition, we used an exploratory factor analysis to investigate the underlying factorial structure of the EMOTICOM test battery. The analysis was conducted using principal axis factoring with Varimax rotation. We used an eigen-value greater than 1 as criterion for extraction of factors.

Correlations With Demographic and Descriptive Factors

Spearman's rank correlation was used to assess the association between performance on EMOTICOM tasks and relevant demographic and descriptive factors including age, sex, education, IQ, NEO personality trait Neuroticism, and scores for self-reported mood on test days. In addition, correlations between test performance and self-reported motivation and diligence were assessed for the six EMOTICOM tasks containing a monetary reward paradigm, i.e., *Reinforcement Learning Task, Monetary Incentive Reward Task, Progressive Ratio Task, Adapted Cambridge Gambling Task, Prisoner's Dilemma,* and *Ultimatum Game*.

RESULTS

Task Outcomes and Descriptive Statistics

Table 2 shows descriptive data for the 100 healthy Danish participants. Level of education was high with a majority ($n = 74$) of participants currently attending or having completed > 4 years

TABLE 1 | EMOTICOM task overview.

Emotion processing	
Emotional Recognition Task	*Description*
	Assessment of emotion recognition. A series of emotional faces appear briefly (for 250 ms) and the participant is asked to identify the expressed emotion (happy, sad, angry, or fearful). The task has two versions: one using full faces and one showing only eyes.
	Primary outcomes
	Correct identification of each emotion calculated as hit rate (%).
Emotional Intensity Morphing Task	*Description*
	Assessment of perceptual threshold for emotion detection. A face with a slowly morphing emotional expression is shown. The participant must indicate when they can detect the presence of an emotion (increase condition) or no longer perceive an emotion (decrease condition). The emotional expressions include happy, sad, angry, fearful, and disgusted.
	Primary outcomes
	Intensity threshold for detection of each emotion in both the increase and decrease condition.
Face Affective Go/No-Go Task	*Description*
	Assessment of information processing bias in identification of emotional faces. A series of emotional faces (happy, sad, angry, or fearful) is shown and the participant is asked to respond only to a specific emotion while ignoring other emotions.
	Primary outcomes
	Discrimination accuracy of emotional faces indexed as *d*-prime scores for each emotion.
Motivation and reward	
Reinforcement Learning Task	*Description*
	Assessment of learning based on reward and punishment. A series of paired colored circles is shown and the participants is asked to choose one circle. Each color has either a high or low chance of eliciting a monetary reward (win condition) or a high or low risk of eliciting monetary loss (loss condition).
	Primary outcomes
	Learning rate (alpha) calculated with a reinforcement learning rate algorithm for both the no-loss and no-win condition.
Monetary Incentive Reward Task	*Description*
	Assessment of effort to avoid punishment and gain reward. The participant is asked to respond as quickly as possible when a black box appears between two circles each containing two lines. The distance between the lines indicate the size of the loss or gain for each trial. A faster response elicits greater reward/smaller loss.
	Primary outcomes
	Average change in reaction time relative to baseline reaction time for both the win and loss condition.
Progressive Ratio Task	*Description*
	Assessment of motivational breakpoint. Four boxes of varying sizes are shown and the participant is asked to select the odd one out. The frequency and size of monetary reward for successfully completing each trial is gradually decreased. The participant is told they can quit at any time but must still wait passively for the remainder of the task's run time.
	Primary outcomes
	Motivational break-point, i.e., the number of trials the participant completes before quitting the task.
Adapted Cambridge Gambling Task	*Description*
	Assessment of decision making and risk-taking behavior. The participant is shown a roulette wheel divided into two colors; the proportion of each color changes in every trial, representing different odds. The participant is asked to choose the color they wish to bet on as well as the size of their bet. The task consists of a win and a loss condition.
	Primary outcomes
	Risk adjustment score indexing optimizing behavior in both the win and loss condition.
Social cognition	
Moral Emotions Task	*Description*
	Assessment of emotional reactions to moral social situations. The participant is presented with cartoons of moral scenarios in which one character intentionally or unintentionally harms another. The participant must rate how guilty, shameful, annoyed, and bad they would feel if they were either the victim or the agent (i.e., the victimizer).
	Primary outcomes
	Average ratings of guilt and shame for victim and agent scenarios.
Social Information Preference Task	*Description*
	Assessment of preference for different types of information. The participant is shown a socially ambiguous situation in which nine pieces of information (faces, thoughts, and facts/objects) are hidden from view. The participant is instructed to pick four pieces of information to help them decide between three different interpretations of the situations; a positive, neutral, and negative.

(Continued)

TABLE 1 | Continued

	Primary outcomes
	The proportion (%) of thoughts, faces, and facts chosen.
Prisoners' Dilemma	*Description*
	Assessment of cooperative strategy. The participant and a computerized opponent perform a small task to collect money which is pooled. The participant is given the choice to split the money equally with the opponent or steal all the money. If both parties choose to split the money, both get half. If one steals and the other splits, the one who stole wins all the money. If both choose to steal, neither party wins any money. The participant faces three computerized opponents with different strategies: cooperative (opponent always splits), tit-for-two-tats (opponent splits until the participant steals for two consecutive trials), and aggressive (opponent starts with steal and then mirrors the participant's behavior).
	Primary outcomes
	Proportion of trials (%) in which the participant chooses to steal for each type of opponent.
Ultimatum Game	*Description*
	Assessment of sensitivity to fairness. The participant and a computerized opponent perform a small task to collect money which is then pooled. In some trials, the participant decides how the money is split, ranging from fair (50/50) to increasingly unfair (10/90), and in some trials the opponent decides the split, ranging from fair (50/50) to increasingly unfair (10/90). The participants may choose to either accept or decline the offers from the opponent.
	Primary outcomes
	Proportion of accepted offers.

TABLE 2 | Descriptive data.

	Mean	*SD*	*Range*
Age (years)	28.87	7.33	18 to 48
Sex (male/female)		50/50	
Education (1–5)	4.54	0.58	1 to 5
IQ	110.36	6.98	93 to 129
Neuroticism[a]	76.04	27.89	24 to 144
TMD (−32 to 200)	1.56	15.99	−20 to 55

Descriptive data for the Danish validation cohort (N = 100). The table shows age, education score indexed with the Online Stimulant and Family History Assessment Module on a five-point likert scale, IQ score assessed with the Reynolds Intellectual Screening Test, total mood disturbance (TMD) indexed with the Profile of Mood Scale, and trait Neuroticism indexed with the NEO Personality Inventory-Revised (n = 93) and the NEO Personality Inventory 3 (n = 6). [a]*N = 99 due to missing data from one participant.*

of higher learning at university level. The study sample IQ of 110.36 was significantly higher than the population IQ of 100, $t(99) = 14.8$, $p < 0.001$ (Reynolds, 2011). There was no difference in Neuroticism scores between the study sample average of 76.04 and the Danish population average of 77.20, $t(98) = -0.41$, $p = 0.68$ (Skovdahl et al., 2011). Lastly, the study sample exhibited significantly lower levels of self-reported total mood disturbance (TMD) indexed with the POMS (TMD score = 1.56) compared to normative data (TMD score = 18.00), $t(99) = -10.28$, $p < 0.001$ (Nyenhuis et al., 1999).

Task Outcomes and Descriptive Statistics

Table 3 shows the descriptive statistics for the primary outcomes of each EMOTICOM task. A full overview of all secondary EMOTICOM outcomes can be found in **Supplementary Information**.

The majority of EMOTICOM task outcomes were skewed and 32 out of 42 outcomes had non-normal distributions. For these task outcomes, median and IQR should be used as reference instead of mean and SD. We observed small floor effects (<10%) for 4 outcomes; moderate floor effects (≥ 10%) for 1 outcome; and severe floor effects (≥30%) for 5 outcomes. In addition, we observed small ceiling effects for 15 EMOTICOM outcomes; moderate ceiling effects for 7 outcomes; and severe ceiling effects for 3 outcomes.

Test–Retest Reliability

Table 4 shows test–retest reliability and test–retest bias for primary EMOTICOM outcomes.

Intraclass correlation coefficients scores varied across primary EMOTICOM outcomes: 7 task outcomes exhibited excellent test–retest reliability (ICC ≥ 0.75); 21 task outcomes exhibited good test–retest reliability (0.60 ≤ ICC < 0.75); 9 task outcomes exhibited fair test–retest reliability (0.40 ≤ ICC < 0.60); and 10 outcomes exhibited poor test–retest reliability (ICC < 0.40). Test–retest bias ranged from −15.32 to 32.58% across all primary EMOTICOM outcomes.

Task-Intercorrelations and Factor Analysis

Figure 1 shows the results of the correlation matrices conducted for each of the three cognitive domains: *Emotion Processing, Motivation and Reward,* and *Social Cognition*.

Within the *Emotion Processing* domain correlations between tasks were predominantly weak ($-0.2 < \rho < 0.2$) and statistically non-significant at the 0.01 alpha level. Only three pairs of task outcomes showed statistically significant correlations: accuracy for Anger in the *face Emotional Recognition Task* and d-prime for Happy/Neutral in the *Face Affective Go/NoGo task* ($\rho = 0.30$, $p = 0.003$); accuracy for Happy in the *eyes Emotional Recognition Task* and detection threshold for Happy in the decrease condition of the *Emotional Intensity Morphing task* ($\rho = -0.36$, $p < 0.001$); and detection threshold for Anger in the decrease condition

TABLE 3 | Primary outcomes.

	Mean	*SD*	*Median*	*IQR*	*Range*	*Skewness*	*Floor effect*	*Ceiling effect*
Emotional Face Recognition Task: Face version								
Accuracy (%) – Happy	85.45	13.63	90.00	15.00	20 to 100	−2.42***	0%	19%
Accuracy (%) – Sad	84.40	12.07	85.00	15.00	40 to 100	−1.31***	0%	12%
Accuracy (%) – Angry	60.60	13.26	65.00	15.00	15 to 80	−1.27***	0%	0%
Accuracy (%) – Fearful	82.00	11.87	85.00	15.00	45 to 100	−1.19***	0%	6%
Emotional Face Recognition Task: Eyes version								
Accuracy (%) – Happy	78.15	16.46	80.00	20.00	20 to 100	−1.37***	0%	6%
Accuracy (%) – Sad	71.20	19.06	75.00	25.00	10 to 100	−0.46***	0%	5%
Accuracy (%) – Angry	66.20	11.81	65.00	20.00	40 to 90	−0.06**	0%	0%
Accuracy (%) – Fearful	75.35	15.01	77.50	16.25	5 to 100	−0.41***	0%	2%
Emotional Intensity Morphing Task: Increase condition								
Detection threshold – Happy	7.61	2.10	7.50	3.00	2.75 to 13.33	0.21	0%	0%
Detection threshold – Sad	9.46	2.13	9.50	3.00	3.50 to 13.50	−0.45	0%	0%
Detection threshold – Angry	8.79	2.18	8.71	2.31	3.50 to 14.00	0.11	0%	0%
Detection threshold – Fearful	9.58	2.33	9.50	3.25	4.00 to 15.00	−0.12	2%	0%
Detection threshold – Disgusted	9.06	2.06	9.50	2.75	3.50 to 13.50	−0.44	0%	0%
Emotional Intensity Morphing Task: Decrease condition								
Detection threshold – Happy	5.33	2.30	5.00	2.94	1.00 to 11.5	0.51*	0%	6%
Detection threshold – Sad	5.47	1.73	5.50	2.19	1.75 to 10.25	0.29	0%	3%
Detection threshold – Angry	4.53	1.75	4.38	2.44	1.50 to 9.75	0.65**	0%	7%
Detection threshold – Fearful	5.17	1.59	5.00	2.00	1.00 to 10.25	0.37	0%	3%
Detection threshold – Disgusted	4.04	1.75	3.75	2.31	1.00 to 10.50	0.85**	0%	11%
Face Affective Go/NoGo								
d-prime – Happy/Neutral	2.85	0.67	2.93	0.73	−0.80 to 3.29	−2.70***	0%	47%
d-prime – Happy/Sad	2.77	0.63	2.93	0.80	0 to 3.29	−1.60***	0%	38%
d-prime – Neutral/Happy	2.50	0.81	2.93	0.76	0 to 3.29	−1.32***	0%	19%
d-prime – Neutral/Sad	2.15	0.86	2.17	1.28	0 to 3.29	−0.63***	0%	11%
d-prime – Sad/Happy	2.69	0.62	2.93	0.80	0.78 to 3.29	−1.23***	0%	29%
d-prime – Sad/Neutral	2.05	1.01	2.17	1.28	−2.49 to 3.29	−1.61***	0%	6%
Reinforcement Learning Task[a]								
Alpha – Win condition	0.23	0.33	0.02	0.40	0.00 to 1.00	1.33***	32%	0%
Alpha – Loss condition	0.43	0.38	0.29	0.73	0.00 to 1.00	0.41***	32%	0%
Monetary Incentive Reward Task								
Reaction time (ms) – Win condition	17.41	18.94	16.13	26.15	−30.3 to 72.87	0.05	–	–
Reaction time (ms) – Loss condition	18.73	18.45	16.67	25.88	−27.52 to 84.65	1.38	–	–
Progressive Ratio Task								
Breakpoint (trials)	316.77	148.33	424.50	251.00	1 to 436	−0.83***	2%	48%
Adapted Cambridge Gambling Task								
Risk adjustment – Win condition	1.72	1.09	1.93	1.40	−0.56 to 3.56	−0.60**	0%	0%
Risk adjustment – Loss condition	2.21	0.92	2.43	1.26	−0.71 to 3.64	−0.84***	0%	0%
Moral Emotions Task								
Guilt – Agent	5.86	0.78	6.04	0.66	4.58 to 7.00	−2.08***	0%	1%
Guilt – Victim	1.59	0.53	1.42	0.61	1.00 to 3.39	1.48***	10%	0%
Shame – Agent	5.74	0.80	5.87	1.00	2.42 to 7.00	−1.35***	0%	1%
Shame – Victim	1.97	0.70	1.91	1.00	1.00 to 4.42	0.78**	8%	0%
Social Information Preference Task								
Information (%) – Thoughts	55.17	13.01	56.25	12.50	0.00 to 75.00	−1.64***	1%	2%
Information (%) – Faces	11.52	11.38	7.81	10.16	0.00 to 57.81	1.83***	5%	0%
Information (%) – Facts	33.31	9.34	32.81	10.94	7.81 to 57.81	−0.09	0%	0%

(Continued)

TABLE 3 | Continued

	Mean	SD	Median	IQR	Range	Skewness	Floor effect	Ceiling effect
Prisoner's Dilemma								
Proportion steals (%) – Cooperative	20.56	29.00	0.00	33.33	0 to 100	1.35***	55%	5%
Proportion steals (%) – Tit-for-two-tat	25.56	32.84	0.00	52.78	0 to 100	0.89***	54%	4%
Proportion steals (%) – Aggressive	35.00	32.03	33.33	66.67	0 to 100	0.3***	33%	3%
Ultimatum Game								
Average acceptance rate (%)	61.07	26.64	59.52	42.26	14.29 to 100	0.01***	0%	14%

Mean, standard deviation (SD), median, interquartile range (IQR), range, and skewness are reported for the primary outcomes of the 11 EMOTICOM tasks. Shapiro–Wilks tests were used to assess normality of data; non-normal distribution of data is denoted with asterisks next to skewness ($p < 0.05$, **$p < 0.01$, ***$p < 0.001$). Note, mean and SD should be used as reference for normally distributed outcomes while median and IQR should be used as reference for non-normally distributed outcomes. Floor and ceiling effects are presented as percentage of test subjects who achieved the minimum score (floor effect) or maximum score (ceiling effect).* [a]*N = 68, as 32 participants performed below chance level, violating the assumptions of the reinforcement learning algorithm used to determine the alpha value.*

of the *Emotional Intensity Morphing task* and d-prime for Happy/Neutral in the *Face Affective Go/NoGo task* ($\rho = -0.31$, $p = 0.002$). Meanwhile correlations between outcomes within the same task ranged from week to moderate for the *Emotional Recognition Task* ($\rho = [-0.12;0.45]$); from weak to strong for the *Emotional Intensity Morphing task* ($\rho = [-0.35;0.70]$); and from weak to moderate for the *Face Affective Go/NoGo task* ($\rho = [0.13;0.35]$). Within the *Motivation and Reward* domain correlations between tasks were predominantly weak ($-0.2 < \rho < 0.2$) and statistically non-significant at the 0.01 alpha level. Only one pair of outcomes showed a statistically significant correlation: reaction time for the win condition in the *Monetary Incentive Reward task* and risk adjustment for the win condition in the *Adapted Cambridge Gambling Task* ($\rho = -0.28$, $p = 0.005$). Correlations between outcomes within the same task was moderate for the *Reinforcement Learning Task* ($\rho = -0.22$); weak for the *Monetary Incentive Reward task* ($\rho = 0.05$); and weak for the *Adapted Cambridge Gambling Task* ($\rho = 0.04$). Within the *Social Cognition Domain* correlations between tasks were predominantly weak ($-0.2 < \rho < 0.2$) and statistically non-significant. Only one pair of outcomes showed a statistically significant correlation: Agent Guilt rating from the *Moral Emotions* task and average acceptance rate from the *Ultimatum Game* ($\rho = -0.28$, $p = 0.006$). Correlations between outcomes within the same task ranged from weak to strong for the *Moral Emotions task* ($\rho = [0.13;0.76]$); from weak to strong for the *Social Information Preference task* ($\rho = [-0.61; -0.17]$); and were strong for the *Prisoner's Dilemma* task ($\rho = [0.67;0.71]$).

The exploratory factor analysis indicated a 13-factor solution with a majority of factors loading onto a single task (see **Supplementary Information** for summary of factor loadings). The 13 factors cumulatively accounted for 70.4% of the total variance. The Kaiser-Meyer-Olkin measure of sampling adequacy was low but acceptable (KMO = 0.53) and Bartlett's test of sphericity was significant [$\chi^2(820) = 1807.0$, $p < 0.001$], indicating that the data was suitable for structure detection.

Correlations With Demographic and Descriptive Factors

Table 5 shows correlations between primary EMOTICOM outcomes and various demographic and descriptive factors. A full overview of correlation between demographic and descriptive factors and all EMOTICOM outcomes can be found in **Supplementary Information**.

Age was negatively correlated with accuracy in recognizing angry and fearful emotions in the eyes version of the *Emotional Face Recognition Task* while differences in sex were correlated with risk adjustment in the win condition in the *Adapted Cambridge Gambling Task* (men performed better); ratings of shame in the *Moral Emotions task* (women rated higher); and proportion of steals against and aggressive opponent in the *Prisoner's Dilemma* (men stole more). Education level showed a negative correlation with detection threshold of fearful emotions in the decrease condition of *Intensity Morphing task* while IQ and Neuroticism scores were not statistically correlated with performance on any primary outcome. Negative mood was positively correlated with accuracy in recognizing sad emotions in the face version of the *Emotional Face Recognition Task* and self-rated motivation and diligence during task completion was positively correlated with breakpoint in the *Progressive Ratio Task*.

DISCUSSION

We here present data collected from 100 healthy participants in order to validate the EMOTICOM test battery and provide reference material for future clinical and research use in Danish populations. Overall the shortened EMOTICOM test battery exhibited mostly acceptable test–retest reliability, low task-intercorrelations indicating limited redundancy between the tasks, and independence between task performance and demographic factors. Therefore, many of the EMOTICOM tasks provide a useful objective method for measuring hot cognition. Below we discuss some task-specific considerations regarding the use of the EMOTICOM test battery in research or clinical practice.

Skewness of Data

A majority of primary EMOTICOM outcomes (76%) exhibited non-normal distributions. One explanation for this could be that our study sample is biased or that the tasks contain threshold constraints such as floor or ceiling effects which skew the distribution. The observed non-normal distributions may also reflect that the construct being assessed is not normally

TABLE 4 | Test–retest reliability.

	Baseline (*n* = 49)		Retest (*n* = 49)		Test–retest bias (%)	*ICC*	*95% CI*
	Mean	**SD**	**Mean**	**SD**			
Emotional Face Recognition Task: Face version							
Accuracy (%) – Happy	85.92	13.76	90.20	13.38	4.98	0.83	0.66 to 0.91
Accuracy (%) – Sad	84.80	13.03	86.12	9.42	1.56	0.67	0.42 to 0.82
Accuracy (%) – Angry	63.27	12.73	70.00	12.42	10.64	0.60	0.25 to 0.78
Accuracy (%) – Fearful	83.47	10.62	83.27	9.77	−0.24	0.50	0.10 to 0.72
Emotional Face Recognition Task: Eyes version							
Accuracy (%) – Happy	80.41	14.21	80.51	14.44	0.12	0.50	0.10 to 0.72
Accuracy (%) – Sad	73.78	15.33	74.69	17.27	1.23	0.74	0.54 to 0.85
Accuracy (%) – Angry	69.49	10.96	74.29	11.90	6.91	0.65	0.36 to 0.80
Accuracy (%) – Fearful	77.86	12.20	79.08	11.02	1.57	0.64	0.36 to 0.80
Emotional Intensity Morphing Task: Increase condition							
Detection threshold – Happy	7.78	2.08	7.55	1.99	−2.96	0.67	0.48 to 0.80
Detection threshold – Sad	9.46	2.02	9.19	1.88	−2.85	0.57	0.35 to 0.73
Detection threshold – Angry	8.57	1.96	8.11	1.79	−5.37	0.66	0.41 to 0.81
Detection threshold – Fearful	9.33	1.98	9.04	2.2	−3.11	0.74	0.54 to 0.85
Detection threshold – Disgusted	9.05	2.04	8.21	1.78	−9.28	0.71	0.45 to 0.85
Emotional Intensity Morphing Task: Decrease condition							
Detection threshold – Happy	5.52	2.24	5.12	1.78	−7.25	0.75	0.56 to 0.86
Detection threshold – Sad	5.36	1.4	5.18	1.53	−3.36	0.50	0.11 to 0.72
Detection threshold – Angry	4.42	1.51	4.76	1.82	7.69	0.29	−0.25 to 0.60
Detection threshold – Fearful	5.03	1.25	4.65	1.42	−7.55	0.34	−0.14 to 0.63
Detection threshold – Disgusted	3.88	1.48	4.18	1.46	7.73	0.53	0.17 to 0.73
Face Affective Go/NoGo							
d-prime – Happy/Neutral	2.94	0.45	3.03	0.34	3.06	0.62	0.34 to 0.79
d-prime – Happy/Sad	2.79	0.69	2.88	0.44	3.23	0.12	−0.47 to 0.54
d-prime – Neutral/Happy	2.48	0.75	2.80	0.66	12.90	0.45	0.06 to 0.68
d-prime – Neutral/Sad	2.00	0.86	2.34	0.83	17.00	0.42	0 to 0.66
d-prime – Sad/Happy	2.79	0.55	2.73	0.57	−2.15	0.15	−0.52 to 0.52
d-prime – Sad/Neutral	2.15	0.84	2.43	0.81	13.02	0.44	0.03 to 0.68
Reinforcement Learning Task[a]							
Alpha – Win condition	0.23	0.33	0.21	0.37	−11.83	−0.04	−0.63 to 0.37
Alpha – Loss condition	0.46	0.36	0.43	0.41	−6.61	−0.18	−1.11 to 0.34
Monetary Incentive Reward Task							
Reaction time (ms) – Win condition	18.48	19.68	16.94	20.51	−8.34	−0.26	−1.07 to 0.25
Reaction time (ms) – Loss condition	19.05	20.86	18.98	21.33	−0.38	−1.47	−3.52 to −0.37
Progressive Ratio Task							
Breakpoint (trials)	309.76	153.64	350.69	130.61	13.21	0.56	0.24 to 0.75
Adapted Cambridge Gambling Task							
Risk adjustment – Win condition	1.78	1.17	2.36	0.82	32.58	0.20	−0.29 to 0.52
Risk adjustment – Loss condition	2.24	0.92	2.54	0.77	13.39	0.18	−0.40 to 0.53
Moral Emotions Task							
Guilt – Agent	5.88	0.79	5.85	0.69	−0.54	0.85	0.73 to 0.91
Guilt – Victim	1.63	0.54	1.69	0.51	3.49	0.83	0.70 to 0.90
Shame – Agent	5.80	0.82	5.68	0.73	−2.03	0.85	0.73 to 0.91
Shame – Victim	2.05	0.67	2.16	0.67	5.22	0.81	0.66 to 0.89
Social Information Preference Task							
Information (%) – Thoughts	51.5	15.43	53.99	14.73	4.83	0.71	0.48 to 0.83
Information (%) – Faces	13.52	13.19	14.00	13.55	3.55	0.74	0.54 to 0.85
Information (%) – Facts	34.98	9.90	32.02	6.90	−8.46	0.62	0.34 to 0.79

(Continued)

TABLE 4 | Continued

	Baseline (n = 49)		Retest (n = 49)		Test–retest bias (%)	*ICC*	*95% CI*
	Mean	**SD**	**Mean**	**SD**			
Prisoner's Dilemma							
Proportion steals (%) – Cooperative	14.74	22.84	17.69	26.64	20.00	0.67	0.40 to 0.81
Proportion steals (%) – Tit-for-two-tat	18.82	28.62	16.55	23.69	−12.05	0.65	0.38 to 0.80
Proportion steals (%) – Aggressive	29.48	32.04	28.34	29.66	−3.85	0.67	0.42 to 0.82
Ultimatum Game							
Average acceptance rate (%)	59.14	25.05	71.96	27.68	21.69	0.77	0.46 to 0.89

*Test–retest reliability for the EMOTICOM tasks indexed with Intraclass Correlation Coefficients (ICC) and their 95% Confidence Intervals (95% CI). In addition, test–retest bias, i.e., percent change in scores between first and second test, was calculated as: Test–retest bias = 100 * ((retest – test)/test).* [a]*N = 35, as 14 participants performed below chance level, violating the assumptions of the reinforcement learning algorithm used to determine the alpha value.*

distributed within the general population. For example, norm data reported for emotion recognition paradigms similar to those included in the EMOTICOM test battery indicate that the performance of healthy individuals is not normally distributed within this cognitive domain (Kessels et al., 2014). Due to the skewness observed in some of the EMOTICOM tasks, we recommend using the median and interquartile ranges to gauge task performance instead of mean and SD.

Floor and Ceiling Effects

Floor and ceiling effects occur when a task is either too difficult (floor effect) or too easy (ceiling effect). It represents a serious psychometric issue because it limits the variability of the collected data and therefore the amount of useful information obtained. Several EMOTICOM tasks exhibited floor or ceiling effects: out of the 42 primary task outcomes, 16 outcomes exhibited either floor or ceiling effects above 10% (i.e., at least 10% of all participants achieved either minimum or maximum scores), including eight outcomes that exhibited severe floor or ceiling effects of 30–55%. In particular, the *Face Affective Go/NoGo Task* had severe ceiling effects while the *Reinforcement Learning Task* had severe floor effects. For the *Face Affective Go/NoGo Task*, this issue could potentially be helped by using reaction time instead of *d*-prime as the primary outcome as reaction time is less vulnerable to floor and ceiling effects. Meanwhile, the presence of floor effects was particularly problematic for the *Reinforcement Learning Task* as a basic assumption in the algorithm used to determine the main outcome (learning rate, alpha) is that the participant performs better than chance level, i.e., that they learn the rules for choosing the best option and stop guessing randomly. In the present sample this meant that the learning rate could not be computed for 32 of the 100 participants. The difficulty of the task was corroborated by the unstructured interviews in which many participants reported they were unable to detect any patterns and kept randomly guessing throughout the task. We therefore suggest that the *Reinforcement Learning Task* may benefit from modifications or at least careful consideration before being applied in clinical practice or research. Other tasks including the *Prisoner's Dilemma Task* and the *Progressive Ratio Task* also had a large proportion of participants who met our criteria for ceiling effects. However, as the purpose of these tasks is to assess different behavioral strategies (e.g., aggressive vs. cooperative) we argue that it is not meaningful to use the terms floor and ceiling effects in the conventional sense for these types of tasks even though they contain optimal strategies for maximizing monetary reward (e.g., not quitting in the *Progressive Ratio Task*).

Test–Retest Reliability

In the original British validation study, test–retest reliability was assessed over a time-period of 5–10 days while we chose a retest span of 3–5 weeks. This longer timeframe is suited to inform studies that include long-term interventions or follow clinical progress over time. However, life events and mood may change considerably more over periods of weeks, as compared with days, which may influence test–retest reliability. The majority of EMOTICOM task outcomes exhibited fair to excellent test–retest reliability although notably only two tasks, the *Moral Emotions task* and the *Ultimatum Game,* had excellent test–retest coefficients of $\geq$ 0.75 for all primary outcomes. In addition, several tasks showed very poor reliability including the *Face Affective Go/NoGo Task*, *Monetary Incentive Reward Task*, and the *Adapted Cambridge Gambling Task*. It should be noted that low ICC scores can be caused by limited variance in the data which in turn may occur as a result of ceiling or floor effects (Koo and Li, 2016). For example, the low ICC scores reported for the *Face Affective Go/NoGo Task* may in part be explained by the severe ceiling effects exhibited by this task. Overall, tasks from the *Social Cognition* domain appeared to have the highest degree of reliability followed by tasks from the *Emotional Processing* domain, while tasks from the *Motivation and Reward* domain had poorer reliability. These observations were largely in accordance with the reports from the original British validation study for related outcomes from the same tasks (Bland et al., 2016). However, what may appear as poor reliability for *Motivation and Reward* tasks could instead reflect learning effects or adaptation in playing strategy. For instance, several participants reported deliberately prioritizing optimizing their winnings during their second session rather than 'playing fair' against the computer opponent. Furthermore, the reported test–retest biases were predominantly positive across most tasks, supporting the presence of a slight behavioral learning effect. It should be noted that for tasks without right/wrong answers (e.g., *Moral Emotions Task* and *Prisoner's Dilemma*), the test–retest bias cannot be interpreted as a learning

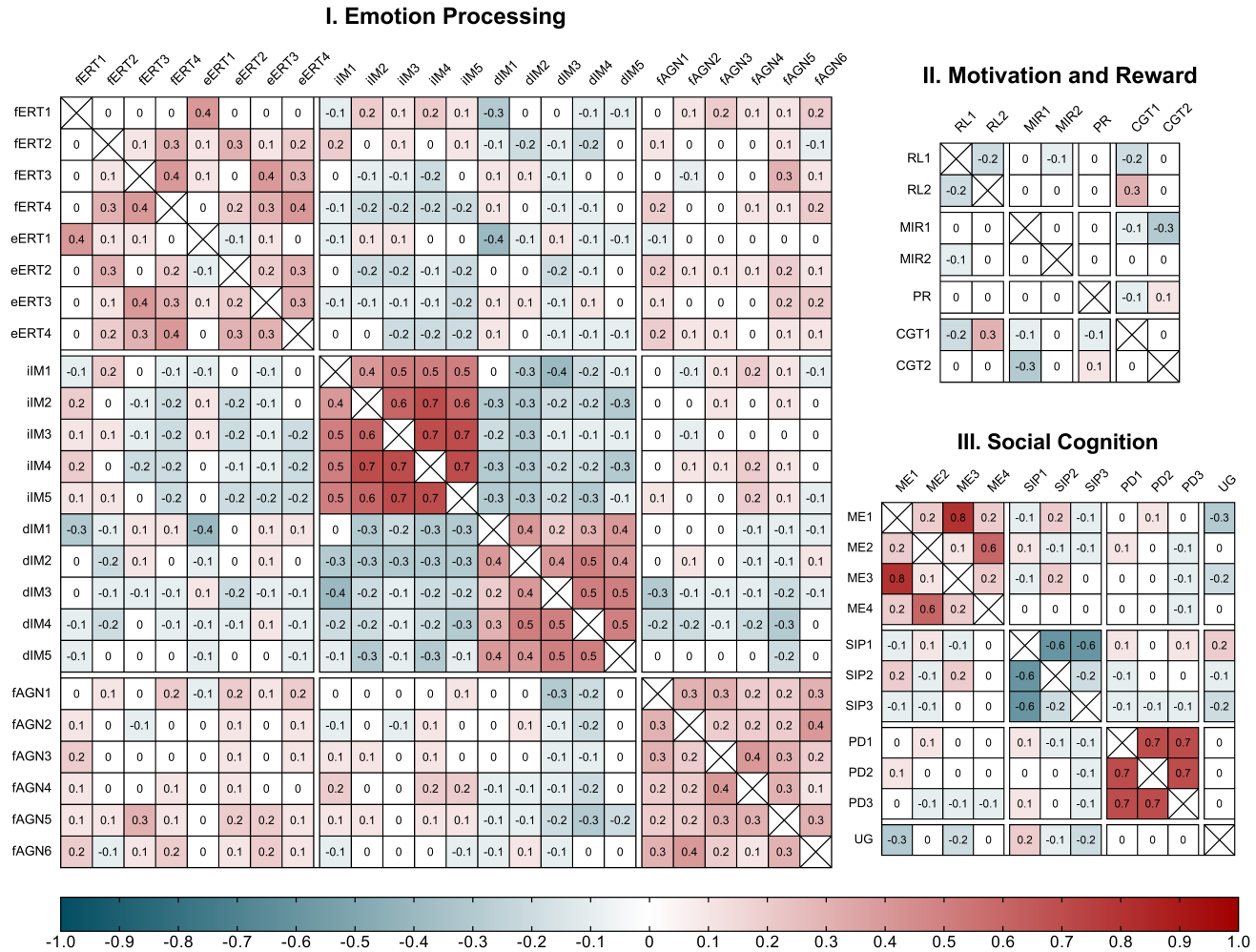


FIGURE 1 | Spearman's Rank Correlations for EMOTICOM outcomes within the three proposed cognitive domains. (I) Emotion Processing: fERT, face Emotion Recognition Task; fERT1, hit rate for happy; fERT2, hit rate for sad; fERT3, hit rate for angry; fERT4, hit rate for fearful. eERT, eyes Emotion Recognition Task; eERT1, hit rate for happy; eERT2, hit rate for sad; eERT3, hit rate for angry; eERT4, hit rate for fearful. iIM, increase Emotional Intensity Morphing Task; iIM1, detection threshold for happy; iIM2, detection threshold for sad; iIM3, detection threshold for angry; iIM4, detection threshold for fearful; iIM5, detection threshold for disgusted. dIM, decrease Intensity Morphing Task; dIM1, detection threshold for happy; dIM2, detection threshold for sad; dIM3, detection threshold for angry; dIM4, detection threshold for fearful; dIM5, detection threshold for disgusted. fAGN, Face Affective Go/NoGo Task; fAGN1, d-prime for 'happy/neutral'; fAGN2, d-prime for 'happy/sad'; fAGN3, d-prime for 'neutral/happy'; fAGN4, d-prime for 'neutral/sad'; fAGN5, d-prime for 'sad/happy'; fAGN6, d-prime for 'sad/neutral.' (II) Motivation and Reward: RL, Reinforcement Learning Task; RL1, learning rate alpha for win condition; RL2, learning rate alpha for loss condition. MIR, Monetary Incentive Reward Task; MIR1, reaction time for win condition; MIR2, reaction time for loss condition. PR, Progressive Ratio Task. aCGT, adapted Cambridge Gambling Task; aCGT1, risk adjustment for win condition; aCGT2, risk adjustment for loss condition. (III) Social Cognition Domain: ME, Moral Emotions Task; ME1, guilt for agent; ME2, guilt for victim; ME3, shame for agent; ME4, shame for victim. SIP, Social Information Preference Task; SIP1, proportion thoughts; SIP2, proportion faces; SIP3, proportion facts. UG, Ultimatum Game.

effect but could instead reflect a shift in response style or choice of strategy.

Construct Validity

The tasks in the EMOTICOM test battery were originally chosen to capture distinct hot cognitive domains including *Emotion Processing*, *Motivation and Reward*, and *Social Cognition*. In order to test the extent to which each individual task loaded onto their respective domains, we mapped the shared variance for the task outcomes within the same domain in three correlation matrices. We found that there were little to no correlation between tasks from the same hot cognitive domain indicating that the original hypothesis of task specific domains could not be supported. This was further corroborated by the results of the exploratory factor analysis which indicated a 13-factor solution and thus did not support the proposed three-domain factorial structure. These results align with the findings from the original British validation which also failed to detect the proposed domain-specific pattern across EMOTICOM tasks (Bland et al., 2016). A possible explanation is that the proposed hot cognitive domains do not represent a single unitary cognitive construct; instead they should be seen as umbrella-terms for multiple inter-related cognitive

TABLE 5 | Correlations.

	Age	*Sex*[$]	*Education*	*IQ*	*Neuroticism*[a]	*TMD*	*Motivation*	*Diligence*
Emotional Face Recognition Task: Face version								
Accuracy (%) – Happy	−0.04	0.01	−0.10	0.04	0.00	0.17	–	–
Accuracy (%) – Sad	−0.19	−0.15	0.05	0.15	0.19	0.28**	–	–
Accuracy (%) – Angry	−0.32***	−0.02	−0.05	0.16	0.14	−0.05	–	–
Accuracy (%) – Fearful	−0.38***	0.09	−0.01	0.16	0.14	0.19	–	–
Emotional Face Recognition Task: Eyes version								
Accuracy (%) – Happy	−0.01	0.06	−0.21	0.09	0.04	0.14	–	–
Accuracy (%) – Sad	−0.17	0.18	−0.02	−0.01	0.23	0.21	–	–
Accuracy (%) – Angry	−0.03	0.07	−0.04	−0.004	0.14	0.07	–	–
Accuracy (%) – Fearful	−0.24	0.02	−0.04	−0.003	0.14	0.04	–	–
Intensity Morphing Task: Increase condition								
Detection threshold – Happy	−0.03	−0.01	0.07	0.06	0.06	0.05	–	–
Detection threshold – Sad	0.04	−0.09	0.08	0.12	−0.05	−0.05	–	–
Detection threshold – Angry	−0.02	−0.12	0.12	0.07	0.03	0.09	–	–
Detection threshold – Fearful	0.15	−0.12	0.15	0.06	−0.13	0.04	–	–
Detection threshold – Disgusted	0.12	−0.19	0.14	0.09	−0.05	−0.04	–	–
Intensity Morphing Task: Decrease condition								
Detection threshold – Happy	−0.08	0.00	0.02	−0.10	−0.02	−0.06	–	–
Detection threshold – Sad	−0.03	0.03	−0.15	−0.21	−0.05	−0.13	–	–
Detection threshold – Angry	−0.03	−0.18	−0.12	0.02	−0.01	0.00	–	–
Detection threshold – Fearful	0.08	0.05	−0.27**	−0.19	−0.03	0.00	–	–
Detection threshold – Disgusted	−0.02	−0.09	−0.06	0.01	−0.07	−0.11	–	–
Face Affective Go/NoGo								
d-prime – Happy/Neutral	0.02	0.10	0.04	0.05	−0.05	−0.07	–	–
d-prime – Happy/Sad	0.05	0.08	−0.08	0.04	−0.01	0.03	–	–
d-prime – Neutral/Happy	0.04	0.001	0.05	−0.07	0.05	0.05	–	–
d-prime – Neutral/Sad	0.02	−0.09	0.14	0.11	0.06	0.22	–	–
d-prime – Sad/Happy	0.00	−0.10	0.09	0.05	−0.04	0.01	–	–
d-prime – Sad/Neutral	−0.08	0.21	−0.19	−0.03	−0.02	0.14	–	–
Reinforcement Learning Task[b]								
Alpha – Win condition	−0.30	0.13	−0.04	−0.23	0.29	0.06	0.05	0.01
Alpha – Loss condition	0.23	−0.16	0.14	0.05	−0.31	0.03	−0.13	−0.20
Monetary Incentive Reward Task								
Reaction time (ms) – Win	−0.08	−0.01	0.14	−0.11	0.15	−0.18	−0.08	0.09
Reaction time (ms) – Loss	0.02	0.06	−0.07	−0.10	−0.02	−0.14	−0.06	0.09
Progressive Ratio Task								
Breakpoint (trials)	−0.23	0.12	0.05	−0.09	−0.07	−0.07	0.39***	0.29**
Adapted Cambridge Gambling Task								
Risk adjustment – Win condition	0.12	−0.28**	0.11	0.19	0.05	0.05	−0.16	−0.05
Risk adjustment – Loss condition	−0.21	0.03	0.17	0.14	−0.01	−0.01	0.02	0.06
Moral Emotions Task								
Guilt – Agent	0.14	0.17	−0.05	−0.11	0.01	0.01	–	–
Guilt – Victim	0.08	0.17	−0.07	−0.03	0.15	0.15	–	–
Shame – Agent	0.02	0.28**	−0.02	−0.17	0.1	0.10	–	–
Shame – Victim	−0.03	0.16	−0.03	−0.17	0.23	0.23	–	–
Social Information Preference Task								
Information (%) – Thoughts	−0.02	0.03	0.02	−0.05	−0.1	−0.10	–	–
Information (%) – Faces	0.13	0.03	−0.10	0.03	0.06	0.06	–	–
Information (%) – Facts	−0.08	−0.09	0.10	0.10	0.07	0.07	–	–

(Continued)

TABLE 5 | Continued

	Age	*Sex*[$]	*Education*	*IQ*	*Neuroticism*[a]	*TMD*	*Motivation*	*Diligence*
Prisoner's Dilemma[c]								
Proportion steals (%) – Cooperative	−0.13	−0.14	−0.02	−0.07	−0.12	−0.12	0.01	0.06
Proportion steals (%) – Tit-for-two-tat	−0.08	−0.23	−0.01	−0.11	−0.01	−0.01	0.04	0.10
Proportion steals (%) – Aggressive	−0.06	−0.26**	0.03	0.005	−0.03	−0.03	0.11	0.14
Ultimatum Game[d]								
Average acceptance rate (%)	−0.16	0.07	0.16	−0.08	0.06	0.17	−0.02	−0.22

*Correlations between EMOTICOM primary outcomes and age, sex, education indexed with the Family History Assessment Module on a five-point Likert scale, IQ score assessed with the Reynolds Intellectual Screening Test, total mood disturbance (TMD) indexed with the Profile of Mood Scale, and trait Neuroticism indexed with the NEO Personality Inventory-Revised (n = 93) and the NEO Personality Inventory 3 (n = 6). Correlations between self-reported motivation and diligence and outcomes from the six EMOTICOM tasks containing monetary reward are also shown. Correlation coefficients are reported as Spearman's ρ; only p-values < 0.01 are considered significant. **p < 0.01, ***p < 0.001. [$]A negative ρ value indicates males score higher while a positive ρ value indicates females score higher. [a]N = 99 due to missing data from one participant. [b]N = 68 as 32 participants performed below chance level, violating the assumptions of the reinforcement learning algorithm used to determine the alpha value. [c]N = 95 due to missing data from five participants. [d]N = 99 due to missing data from one participant.*

processes. In addition, while previous studies have indicated the existence of an underlying facial expression decoding construct in the *Emotion Processing domain* (Hildebrandt et al., 2015), we speculate that the EMOTICOM tasks within this domain are too heterogeneous both in terms of task design and outcome scales to capture this single construct. Overall, these findings emphasize that hot cognition is a complex phenomenon made up of multifaceted cognitive constructs. As a consequence, we recommend that researchers aiming to investigate hot cognition using EMOTICOM should view the battery as a tool box and carefully consider the exact target of their investigation before choosing the appropriate task.

Lastly, some EMOTICOM tasks exhibited very low within-task correlation, suggesting that (a) the task itself does not measure a single construct or (b) the outcomes are unreliable. This was particularly pronounced for tasks from the *Motivation and Reward* domain and indicates that these tasks may benefit from modifications.

Demographic Factors

With few exceptions, performance on EMOTICOM tasks was not strongly influenced by demographic factors. Age was negatively correlated to recognition of anger and fear in the face version of the *Emotional Face Recognition Task* but not in the eye version. Age effects on emotion recognition have previously been reported in the literature and in particular for recognition of negative emotions (Ruffman et al., 2008). Therefore, it may be advantageous to use the eye version of the *Emotional Face Recognition Task* in study cohorts containing middle-aged and older adults as this version appears to be less sensitive to age effects. Corroborating the original British validation study, we did not observe sex effect on tasks from the *Emotion Processing* domain (Bland et al., 2016), but women exhibited higher ratings of shame in the *Moral Emotions Task*. This fits with previous reports of sex differences in proneness to experience shame and guilt (O'Connor et al., 1994; Else-Quest et al., 2012). Women were also less likely to steal from their opponent in the *Prisoner's Dilemma* task while men exhibited better risk adjustment in the *Adapted Cambridge Gambling Task*. Performance on EMOTICOM appeared to be largely independent of IQ and education with the single exception of a negative correlation between education level and detection of fear in the *Intensity Morphing task*'s decrease condition. However, it should be emphasized that the included participants were not stratified for education. This resulted in a cohort with very high education levels as well as high IQ which limits our ability to accurately assess the potential effect of these factors on task performance. Overall, it is a strength of the EMOTICOM test battery that demographic factors do not seem to influence task performance. However, given the stratification issues described above, other studies are needed to investigate the impact of demographic factors on test performance in older as well as less well-educated cohorts.

Mood, Personality, Motivation Factors

In addition to demographic characteristics, we also looked at how other relevant factors such as trait Neuroticism and self-reported mood might influence responses on EMOTICOM tasks. Trait Neuroticism is used to index the tendency to experience negative emotions and is strongly linked to risk of developing psychopathology (Malouff et al., 2005; Ormel et al., 2013). Trait Neuroticism did not correlate significantly with any EMOTICOM outcomes while mood was positively correlated with recognition of sad faces in the face version of the *Emotional Face Recognition Task* only. The latter finding is in line with previous reports showing that mood can influence recognition of emotional faces. However, the effect appears to be relatively small and in most studies requires the active evocation of emotion in the participant prior to the presentation of the stimuli (Schmid and Mast, 2010). Lastly, the correlation between self-reported motivation and diligence during the six tasks containing the possibility of winning an extra sum of money was also assessed. We found that self-reported motivation and diligence had little effect on performance except for motivation on the *Progressive Ratio Task*. This provides further validation for the Progressive Ratio Task as an objective measure of motivation. Overall, the general lack of correlations between performance on EMOTICOM tasks

and trait Neuroticism, mood disturbance, and self-reported motivation and diligence indicates that EMOTICOM is not sensitive to differences in emotion fluctuations or personality characteristics in healthy participants.

Comparison With British Validation Study

There are several differences between the original British validation study and the present work. For example, we chose a longer test–retest interval and included measures of mood, Neuroticism and motivation and diligence to characterize potential influences on task performance. In addition, many of the reported task outcomes differ. We based our choice of primary outcomes for each task on consultation with the original test developers as well as standard practice in the literature. However, as most cognitive tasks do not have a single, clearly defined outcome, the 'optimal' choice of primary outcome may vary from study to study depending on the research question. For example, recognition of angry faces may be especially relevant in studies investigating aggression whereas recognition of fearful faces may be especially relevant for studying anxiety. We therefore endeavored to pick outcomes that we believe best capture the core cognitive function of each task and, when possible, limit the use of composite outcomes (i.e., complex outcomes created from two or more outcomes). While these choices make a direct one-to-one comparison between the two studies difficult, overall our findings align with those from the British validation study. We observed similar patterns of test–retest reliability at both task and domain level and were able to replicate the report that EMOTICOM is largely independent of demographic factors. In addition, we corroborate the original study's rejection of a three-domain structure. As information on floor and ceiling effects were not reported in the British validation study, we cannot compare our results to the British study.

Methodological Limitations

EMOTICOM was initially validated in 200 volunteers by the British test developers. The purpose of this study was to replicate the original study with a smaller sample of 100 Danish participants. This is a used practice for psychometric studies comparing populations with large biological, environmental, and cultural overlaps; e.g., the Danish version of the Delis-Kaplan Executive Function System (D-KEFS) test battery was validated against American norms based on data collected from 111 Danish individuals. However, the relatively small sample size of the present study does present some limitations. In particular the reported correlations between task performance and demographic and descriptive factors should be interpreted with caution as the study may not have had sufficient power to detect weaker correlations. In addition, as the present study likely does not have a sufficiently large sample size to accurately estimate the true factorial structure of the EMOTICOM task outcomes (Beavers et al., 2013), we refrain from interpreting the meaning of individual factors derived from the analysis. Importantly, our study sample does not represent a normative sample but rather a reference sample based on well-educated individuals with high IQ. In addition, due to the high level of ethnic and cultural homogeneity in the Danish population, the present study sample could not provide any insight into potential effects of ethnicity or cultural differences on task performance. Therefore, caution should be taken when comparing the findings to other types of study groups or the general population. Also, based on the current study it cannot be ascertained whether the observed ceiling effects in healthy participants would also be present in clinical samples nor how sensitive the tasks may be to psychological or pharmacological interventions. So far, one study has used the EMOTICOM battery to investigate the association between paranoid thinking in healthy participants and social cognition, reporting a link between increased paranoia and likelihood of stealing from the cooperative opponent in the *Prisoner's Dilemma* task (Savulich et al., 2018).

As a final note, we caution against using the rating of 'annoyance' from the *Moral Emotions* task. Based on the qualitative interviews, we discovered that some participants reported high levels of annoyance in moral scenarios where they were the agent (i.e., when they caused harm to others) because they 'felt annoyed with themselves' while some participants reported low levels of annoyance because they 'did not feel annoyed with the victim or the situation.' Since this ambiguity of interpretation was not seen in the original publication of a healthy United Kingdom sample, it may reflect cultural differences. We therefore recommend that the task instructions be modified to eliminate this ambiguity.

CONCLUSION

We here present reference material for performance on the hot cognitive test battery EMOTICOM from a Danish cohort of healthy participants. While most tasks exhibited acceptable psychometric properties, select tasks may not be appropriate for use in healthy individuals due to issues relating to floor and ceiling effects, low test–retest reliability and lack of within-task correlations. While these issues may be ameliorated by choosing alternate task outcomes in some cases (e.g., for the *Face Affective Go/NoGo* task) other tasks, in particular those from the *Motivation and Reward* domain, may benefit from modifications. We observed overall weak correlations between tasks within the same domain, indicating that the proposed structure of an *Emotion Processing* domain, *Reward and Motivation* domain and *Social Cognition* domain cannot be substantiated. EMOTICOM tasks were largely independent of demographic factors such as age, sex, education as well as IQ, personality, mood, and self-reported motivation and diligence during task completion. The present study may help guide future study designs by indicating which EMOTICOM tasks may be most appropriate for the study population planned. In conclusion, many EMOTICOM tasks provide useful, objective methods for measuring social and emotional cognition; however, future studies are needed to investigate the performance of EMOTICOM tasks in patient groups as well as their performance in intervention trials.

DATA AVAILABILITY STATEMENT

For legal reasons we are not allowed to upload and share our data. The data from the study is available upon request from the CIMBI database (http://www.cimbi.dk/db).

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

DS, VF, and GK conceived and designed the study. VD and CT collected the data. PJ organized the database. VD, PJ, and AB defined and implemented the outcomes used. VD wrote the first draft of the manuscript. EM consulted on the statistical analysis which was performed by VD. CT wrote the sections of the manuscript. RE and BS consulted on the analysis and interpretation of the findings. All authors contributed to the manuscript revision, and read and approved the submitted version.

FUNDING

This study was supported by the Augustinus Foundation (Grant 16-0058), Rigshospitalet's Research Council (Grant R149-A6325), and the Innovation Fund Denmark (Grant 4108-00004B). The financial supporters were not involved in the study design, collection, analysis, interpretation, or publication of data.

ACKNOWLEDGMENTS

We would like to thank Nanna Hansen and Vincent Beliveau for their invaluable assistance with the implementation of the EMOTICOM test battery in Danish as well as Anne-Sofie Schneider, Sophia Armand, and Simone Pleinert for their assistance with recruitment and data collection. We are grateful to all participants who donated their time to this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2019.02660/full#supplementary-material

REFERENCES

Beavers, A. S., Lounsbury, J. W., Richards, J. K., Huck, S. W., Skolits, G., and Esquivel, S. L. (2013). Practical considerations for using exploratory factor analysis in educational research. *Pract. Assess. Res. Eval.* 18, 1–13. doi: 10.1080/15389588.2018.1476689

Bland, A. R., Roiser, J. P., Mehta, M. A., Schei, T., Boland, H., Campbell-Meiklejohn, D. K., et al. (2016). EMOTICOM: a neuropsychological test battery to evaluate emotion, motivation, impulsivity, and social cognition. *Front. Behav. Neurosci.* 10:25. doi: 10.3389/fnbeh.2016.00025

Brewer, R., Marsh, A. A., Catmur, C., Cardinale, E. M., Stoycos, S., Cook, R., et al. (2015). The impact of autism spectrum disorder and alexithymia on judgments of moral acceptability. *J. Abnorm. Psychol.* 124, 589–595. doi: 10.1037/abn0000076

Cameron, C. D., Reber, J., Spring, V. L., and Tranel, D. (2018). Damage to the ventromedial prefrontal cortex is associated with impairments in both spontaneous and deliberative moral judgments. *Neuropsychologia* 111, 261–268. doi: 10.1016/j.neuropsychologia.2018.01.038

Cardinale, E. M., and Marsh, A. A. (2015). Impact of psychopathy on moral judgments about causing fear and physical harm. *PLoS One* 10:e0125708. doi: 10.1371/journal.pone.0125708

Chen, C., Takahashi, T., Nakagawa, S., Inoue, T., and Kusumi, I. (2015). Reinforcement learning in depression: a review of computational research. *Neurosci. Biobehav. Rev.* 55, 247–267. doi: 10.1016/j.neubiorev.2015.05.005

Chung, Y. S., Barch, D., and Strube, M. (2014). A meta-analysis of mentalizing impairments in adults with schizophrenia and autism spectrum disorder. *Schizophrenia Bull.* 40, 602–616. doi: 10.1093/schbul/sbt048

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol. Assess.* 6, 284–290. doi: 10.1037//1040-3590.6.4.284

Costa, J. P. T., and McCrae, R. R. (2005). *Personality in Adulthood, A Five-Factor Theory Perspective*, 2nd Edn. New York, NY: Guilford Press.

Elliott, R., Zahn, R., Deakin, J. F., and Anderson, I. M. (2011). Affective cognition and its disruption in mood disorders. *Neuropsychopharmacology* 36, 153–182. doi: 10.1038/npp.2010.77

Else-Quest, N. M., Higgins, A., Allison, C., and Morton, L. C. (2012). Gender differences in self-conscious emotional experience: a meta-analysis. *Psychol. Bull.* 138, 947–981. doi: 10.1037/a0027930

Ersche, K. D., and Sahakian, B. J. (2007). The neuropsychology of amphetamine and opiate dependence: implications for treatment. *Neuropsychol. Rev.* 17, 317–336. doi: 10.1007/s11065-007-9033-y

García-Blanco, A. C., Perea, M., and Livianos, L. (2013). Mood-congruent bias and attention shifts in the different episodes of bipolar disorder. *Cogn. Emot.* 27, 1114–1121. doi: 10.1080/02699931.2013.764281

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

Hagele, C., Schlagenhauf, F., Rapp, M., Sterzer, P., Beck, A., Bermpohl, F., et al. (2015). Dimensional psychiatry: reward dysfunction and depressive mood across psychiatric disorders. *Psychopharmacology* 232, 331–341. doi: 10.1007/s00213-014-3662-7

Harmer, C. J., and Cowen, P. J. (2013). 'It's the way that you look at it'–a cognitive neuropsychological account of SSRI action in depression. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 368:20120407. doi: 10.1098/rstb.2012.0407

Harms, M. B., Martin, A., and Wallace, G. L. (2010). Facial emotion recognition in autism spectrum disorders: a review of behavioral and neuroimaging studies. *Neuropsychol. Rev.* 20, 290–322. doi: 10.1007/s11065-010-9138-6

Hedman, E., Ström, P., Stünkel, A., and Mörtberg, E. (2013). Shame and guilt in social anxiety disorder: effects of cognitive behavior therapy and association with social anxiety and depressive symptoms. *PLoS One* 8:e61713. doi: 10.1371/journal.pone.0061713

Hildebrandt, A., Sommer, W., Schacht, A., and Wilhelm, O. (2015). Perceiving and remembering emotional facial expressions — A basic facet of emotional intelligence. *Intelligence* 50, 52–67. doi: 10.1016/j.intell.2015.02.003

Hjordt, L. V., Stenbaek, D. S., Ozenne, B., Mc Mahon, B., Hageman, I., Hasselbalch, S. G., et al. (2017). Season-independent cognitive deficits in seasonal affective

disorder and their relation to depressive symptoms. (Report). *Psychiatry Res.* 257:219. doi: 10.1016/j.psychres.2017.07.056

Kessels, R. P., Montagne, B., Hendriks, A. W., Perrett, D. I., and de Haan, E. H. (2014). Assessment of perception of morphed facial expressions using the emotion recognition task: normative data from healthy participants aged 8-75. *J. Neuropsychol.* 8, 75–93. doi: 10.1111/jnp.12009

Knudsen, G. M., Jensen, P. S., Erritzoe, D., Baare, W. F., Ettrup, A., Fisher, P. M., et al. (2016). The center for integrated molecular brain imaging (Cimbi) database. *Neuroimage* 124(Pt B), 1213–1219. doi: 10.1016/j.neuroimage.2015.04.025

Koo, T. K., and Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J. Chiropractic Med.* 15, 155–163. doi: 10.1016/j.jcm.2016.02.012

Kuin, N., Masthoff, E., Kramer, M., and Scherder, E. (2015). The role of risky decision-making in aggression: a systematic review. *Aggres. Violent Behav.* 25, 159–172. doi: 10.1016/j.avb.2015.07.018

Leppanen, J., Ng, K. W., Tchanturia, K., and Treasure, J. (2017). Meta-analysis of the effects of intranasal oxytocin on interpretation and expression of emotions. *Neurosci. Biobehav. Rev.* 78, 125–144. doi: 10.1016/j.neubiorev.2017.04.010

Lovell, D. M., Williams, J. M., and Hill, A. B. (1997). Selective processing of shape-related words in women with eating disorders, and those who have recovered. *Br. J. Clin. Psychol.* 36(Pt 3), 421–432. doi: 10.1111/j.2044-8260.1997.tb01249.x

Malouff, J., Thorsteinsson, E., and Schutte, N. (2005). The relationship between the five-factor model of personality and symptoms of clinical disorders: a meta-analysis. *J. Psychopathol. Behav. Assess.* 27, 101–114. doi: 10.1007/s10862-005-5384-y

McNair, D. M., and Heuchert, J. P. (2007). *Profile of Mood States - Technical Update*. Canada: Multi Health System Inc.

Merens, W., Willem Van der Does, A. J., and Spinhoven, P. (2007). The effects of serotonin manipulations on emotional information processing and mood. *J. Affect. Disord.* 103, 43–62. doi: 10.1016/j.jad.2007.01.032

Mogg, K., Bradley, B. P., and Williams, R. (1995). Attentional bias in anxiety and depression: the role of awareness. *Br. J. Clin. Psychol.* 34(Pt 1), 17–36. doi: 10.1111/j.2044-8260.1995.tb01434.x

Newcombe, V. F., Outtrim, J. G., Chatfield, D. A., Manktelow, A., Hutchinson, P. J., Coles, J. P., et al. (2011). Parcellating the neuroanatomical basis of impaired decision-making in traumatic brain injury. *Brain* 134(Pt 3), 759–768. doi: 10.1093/brain/awq388

Nyenhuis, D. L., Yamamoto, C., Luchetta, T., Terrien, A., and Parmentier, A. (1999). Adult and geriatric normative data and validation of the profile of mood states. *J. Clin. Psychol.* 55, 79–86. doi: 10.1002/(sici)1097-4679(199901)55:1<79::aid-jclp8>3.0.co;2-7

O'Connor, L. E., Berry, J. W., Inaba, D., Weiss, J., and Morrison, A. (1994). Shame, guilt, and depression in men and women in recovery from addiction. *J. Subst. Abuse Treat.* 11, 503–510. doi: 10.1016/0740-5472(94)90001-9

Ormel, J., Jeronimus, B. F., Kotov, R., Riese, H., Bos, E. H., Hankin, B., et al. (2013). Neuroticism and common mental disorders: meaning and utility of a complex relationship. *Clin. Psychol. Rev.* 33, 686–697. doi: 10.1016/j.cpr.2013.04.003

Osorio, F. L., de Paula Cassis, J. M., Machado de Sousa, J. P., Poli-Neto, O., and Martin-Santos, R. (2018). Sex hormones and processing of facial expressions of emotion: a systematic literature review. *Front. Psychol.* 9:529. doi: 10.3389/fpsyg.2018.00529

Park, C., Pan, Z., Brietzke, E., Subramaniapillai, M., Rosenblat, J. D., Zuckerman, H., et al. (2018). Predicting antidepressant response using early changes in cognition: a systematic review. *Behav. Brain Res.* 353, 154–160. doi: 10.1016/j.bbr.2018.07.011

Plana, I., Lavoie, M. A., Battaglia, M., and Achim, A. M. (2014). A meta-analysis and scoping review of social cognition performance in social phobia, posttraumatic stress disorder and other anxiety disorders. *J. Anxiety Disord.* 28, 169–177. doi: 10.1016/j.janxdis.2013.09.005

Pool, E., Brosch, T., Delplanque, S., and Sander, D. (2016). Attentional bias for positive emotional stimuli: a meta-analytic investigation. *Psychol. Bull.* 142, 79–106. doi: 10.1037/bul0000026

Reynolds, C. R. (2011). *RIAS/RIST, Reynolds Intellectual Assessment Scales, Reynolds Intellectual Screening Test, vejledning, 1. oplag Ed*. Virum: Hogrefe Psykologisk Forlag.

Roiser, J. P., Elliott, R., and Sahakian, B. J. (2012). Cognitive mechanisms of treatment in depression. *Neuropsychopharmacology* 37, 117–136. doi: 10.1038/npp.2011.183

Roiser, J. P., and Sahakian, B. J. (2013). Hot and cold cognition in depression. *CNS Spectr.* 18, 139–149. doi: 10.1017/S1092852913000072

Routledge, K. M., Williams, L. M., Harris, A. W. F., Schofield, P. R., Clark, C. R., and Gatt, J. M. (2018). Genetic correlations between wellbeing, depression and anxiety symptoms and behavioral responses to the emotional faces task in healthy twins. *Psychiatry Res.* 264, 385–393. doi: 10.1016/j.psychres.2018.03.042

Ruffman, T., Henry, J. D., Livingstone, V., and Phillips, L. H. (2008). A meta-analytic review of emotion recognition and aging: implications for neuropsychological models of aging. *Neurosci. Biobehav. Rev.* 32, 863–881. doi: 10.1016/j.neubiorev.2008.01.001

Savulich, G., Jeanes, H., Rossides, N., Kaur, S., Zacharia, A., Robbins, T. W., et al. (2018). Moral emotions and social economic games in paranoia. *Front. Psychiatry* 9:615. doi: 10.3389/fpsyt.2018.00615

Schmid, P. C., and Mast, M. S. (2010). Mood effects on emotion recognition. *Motivat. Emot.* 34, 288–292. doi: 10.1007/s11031-010-9170-0

Skovdahl, H. H., Mortensen, E. L., and Schiøtz, H. K. (2011). *NEO PI-R, manual - klinisk. 1. udgave, 5. oplag edn*. Copenhagen: Hogrefe Psykologisk Forlag.

Umemoto, A., Lukie, C., Kerns, K., Müller, U., and Holroyd, C. (2014). Impaired reward processing by anterior cingulate cortex in children with attention deficit hyperactivity disorder. *Cogn. Affect. Behav. Neurosc.* 14, 698–714. doi: 10.3758/s13415-014-0298-3

Ventura, J., Wood, R. C., Jimenez, A. M., and Hellemann, G. S. (2013). Neurocognition and symptoms identify links between facial recognition and emotion processing in schizophrenia: meta-analytic findings. *Schizophr Res.* 151, 78–84. doi: 10.1016/j.schres.2013.10.015

Conflict of Interest: AB, RE, and BS are co-inventors of the EMOTICOM test battery and BS consults for Cambridge Cognition.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Appendix

Emotion Processing

1. Emotional Recognition Task…… p. 2
2. Emotional Intensity Morphing Task…..……... p. 6
3. Face Affective Go/NoGo Task……..... p. 7

Motivation and Reward

4. Reinforcement Learning Task…… p. 11
5. Monetary Incentive Reward Task…….. p. 14
6. Progressive Ratio Task…….. p. 17
7. Adapted Cambridge Gambling Task……. p. 20

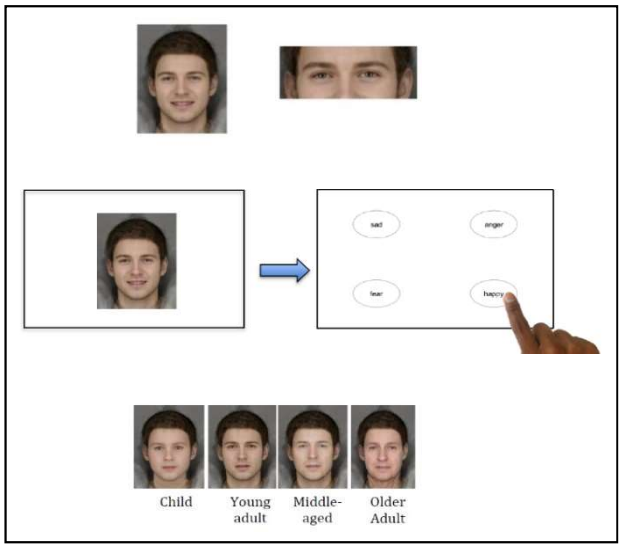
Social Cognition

8. Moral Emotions Task……. p. 24
9. Social Information Preference Task…….. p. 28
10. Prisoner's Dilemma……... p. 31
11. Ultimatum Game……... p. 35

Exploratory factor analysis…… p. 39

1. Emotional Recognition Task

Task description



The face and eyes Emotional Recognition Task (fERT & eERT) measures the ability to identify emotions in facial expressions. The EMOTICOM test battery contains two separate ERT task versions: a full-face version and an eyes-only version. The faces, or eyes, are briefly presented on a computer screen (250ms) after which the test subject is asked to identify the emotion expressed (happy, sad, angry, or fearful). Each emotion is shown twice at ten different intensity levels (10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%) equaling 20 trials for each emotion and 80 trials in total. In addition, a control condition displaying faces/eyes at different ages (child, young adult, middle aged, and older adult) is also included in the task.

Main outcomes (reported in the main article)

- *Hit rate (H)* – Percentage of trials in which a given emotion is correctly identified

Secondary outcomes

- *False alarm rate (FA)* – Percentage of trials in which a given emotion is incorrectly identified
- *d-prime (d')* – Index of discrimination sensitivity calculated as: $d' = z(H) - z(FA)$.

Emotional Recognition Task – Descriptive data

	Mean	*SD*	*Median*	*IQR*	*Range*	*Skewness*	*Floor effect*	*Ceiling effect*
Face version								
False alarm rate (%) - Happy	*8.17*	*8.82*	5.00	10.00	0.00–40.00	1.49	0%	15%
False alarm rate (%) - Sad	*15.42*	*8.48*	15.00	8.75	0.00–46.67	0.89	0%	2%
False alarm rate (%) - Angry	*1.83*	*3.59*	0.00	1.67	0.00–21.67	3.65	0%	53%
False alarm rate (%) - Fearful	*3.77*	*4.15*	1.67	6.67	0.00–20.00	1.55	0%	28%
d-prime - Happy	*2.76*	*0.52*	2.80	0.61	1.09–3.92	-0.53	0%	0%
d-prime - Sad	2.21	0.56	2.24	0.71	0.42–3.46	-0.56	0%	0%
d-prime - Angry	*2.16*	*0.42*	2.21	0.52	0.46–2.8	-1.02	0%	0%
d-prime - Fearful	*2.76*	*0.52*	2.80	0.61	1.09–3.79	-0.48	0%	0%
Eyes version								
False alarm rate (%) - Happy	*15.00*	*12.06*	11.67	13.33	0–63.33	1.43	0%	3%
False alarm rate (%) - Sad	*15.18*	*8.49*	13.33	11.67	1.67–41.67	0.87	0%	0%
False alarm rate (%) - Angry	*3.63*	*5.94*	1.67	5.00	0–50	5.24	0%	32%
False alarm rate (%) - Fearful	*2.55*	*3.90*	1.67	3.33	0–26.67	3.14	0%	45%
d-prime - Happy	2.59	0.57	2.63	0.66	0.32–3.92	-0.49	0%	0%
d-prime - Sad	1.75	0.59	1.77	0.75	0.48–3.29	-0.12	0%	0%
d-prime - Angry	2.22	0.40	2.21	0.46	1.04–3.24	-0.09	0%	0%
d-prime - Fearful	*2.59*	*0.57*	2.63	0.66	0.32–3.92	-0.89	0%	0%

Legend: Mean, SD, median, interquartile range, range, and skewness are shown. Note, for outcomes with non-parametric data distribution, mean and SD are denoted in cursive gray indicating the median and interquartile range should be used as reference. Floor and ceiling effects are shown as percentage of test subjects who achieved minimum (floor) or maximum (ceiling) scores. Note, for False alarm rate maximum score = 0% and minimum score = 100%.

Emotional Recognition Task –Test-retest reliability

	Baseline (n = 49)		Retest (n = 49)		*Test-retest bias (%)*	*ICC*	*95% CI*
	Mean	*SD*	*Mean*	*SD*			
Face version							
False alarm rate (%) - Happy	8.27	8.69	7.96	8.00	-3.75	0.69	0.45–0.83
False alarm rate (%) - Sad	14.25	6.79	11.46	9.06	-19.58	0.57	0.25–0.75
False alarm rate (%) - Angry	1.67	2.85	1.46	2.25	-12.57	0.32	-0.22–0.62
False alarm rate (%) - Fearful	3.33	3.32	2.59	3.79	-22.22	0.66	0.40–0.81
d-prime - Happy	2.82	0.47	2.86	0.45	1.42	0.65	0.37–0.80
d-prime - Sad	2.28	0.57	2.48	0.54	8.77	0.59	0.29–0.77
d-prime - Angry	2.24	0.44	2.46	0.39	9.82	0.59	0.25–0.78
d-prime - Fearful	2.82	0.47	2.86	0.45	1.42	0.32	-0.22–0.62
Eyes version							
False alarm rate (%) - Happy	13.81	8.80	12.21	9.22	-11.59	0.62	0.33–0.79
False alarm rate (%) - Sad	13.50	7.03	12.18	8.15	-9.78	0.65	0.38–0.80
False alarm rate (%) - Angry	3.50	3.82	3.27	3.86	-6.57	0.45	0.01–0.69
False alarm rate (%) - Fearful	2.01	2.50	2.82	3.03	40.30	0.67	0.42–0.81
d-prime - Happy	2.71	0.49	2.69	0.45	-0.74	0.45	0.03–0.69
d-prime - Sad	1.88	0.50	2.01	0.54	6.91	0.61	0.31–0.78
d-prime - Angry	2.31	0.40	2.49	0.46	7.79	0.48	0.11–0.70
d-prime - Fearful	2.71	0.49	2.69	0.45	-0.74	0.43	-0.02–0.68

Legend: Mean and SD are reported for baseline and retest sessions (3-5 weeks) along with Intraclass Correlation Coefficient (ICC) and their 95% Confidence Interval (95% CI). Test-retest bias is calculated as percentage change between first and second test: *Test-retest bias = ((retest-test)/test)*100.*

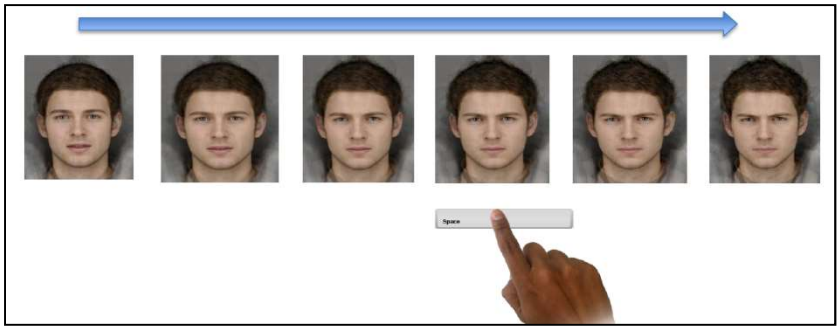
Emotional Recognition Task – Correlations

	Age	*Sex*[$]	*Education*	*IQ*	*Neuroticism*[a]	*TMD*
Face version						
False alarm rate (%) - Happy	0.21	0.04	0.01	-0.02	-0.11	-0.04
False alarm rate (%) - Sad	0.21	-0.04	0.02	-0.14	0.01	-0.04
False alarm rate (%) - Angry	0.16	-0.01	-0.14	**-0.26****	-0.07	-0.22
False alarm rate (%) - Fearful	-0.17	0.01	-0.06	-0.13	0.09	-0.08
d-prime - Happy	**-0.26****	-0.04	-0.11	0.06	0.15	0.24
d-prime - Sad	**-0.32****	-0.12	0.05	0.21	0.17	**0.28****
d-prime - Angry	**-0.34*****	-0.09	-0.09	0.23	0.11	0.04
d-prime - Fearful	**-0.33*****	0.04	0.03	0.19	0.09	**0.26****
Eyes version						
False alarm rate (%) - Happy	0.06	-0.07	0.08	0.15	-0.15	0.03
False alarm rate (%) - Sad	-0.03	0.01	0.07	-0.13	0.05	-0.03
False alarm rate (%) - Angry	0.22	-0.12	0.06	-0.01	0.02	-0.15
False alarm rate (%) - Fearful	0.04	-0.16	0.05	0.09	-0.17	**-0.31****
d-prime - Happy	-0.02	0.17	-0.26	-0.06	0.15	0.06
d-prime - Sad	-0.17	0.17	-0.11	0.07	0.23	0.24
d-prime - Angry	-0.16	0.08	-0.11	0.03	0.08	0.17
d-prime - Fearful	**-0.29****	0.01	-0.02	0.03	0.25	0.22

Legend: Correlations between Emotional Recognition Task (ERT) outcomes and age, sex, education indexed with the Family History Assessment Module (OS-FHAM) on a five-point likert scale, IQ score assessed with the Reynolds Intellectual Screening Test (RIST), total mood disturbance (TMD) indexed with the Profile of Mood Scale (POMS), and trait Neuroticism indexed with the NEO Personality Inventory-Revised (NEO PI-R, n = 93) and the NEO Personality Inventory 3 (NEO PI-3, n = 6). Correlation coefficients are reported as Spearman's rho; only *p*-values < .01 are considered significant and are marked in bold. ** = $p < .01$, *** = $p < .001$. [$] A negative rho value indicates males score higher while a positive rho value indicates females score higher. [a] N = 99 due to missing data from one participant.

2. Emotional Intensity Morphing Task

Task description



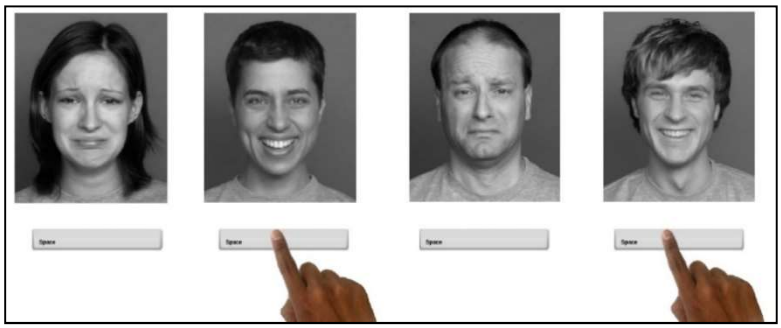
The Emotional Intensity Morphing task (IM) measures the perceptual threshold for the detection of emotions in facial expressions at different intensity levels. The task contains two conditions; an increase and a decrease condition. At the start of each trial, the participant is told which emotion to look for (happiness, sadness, anger, fear or disgust) before a face is shown whose emotional expression slowly morphs through 15 different intensity levels (1 = neutral, 15 = maximum intensity). For the increase condition, the emotional expression starts at neutral and morphs towards maximum intensity and the participant is instructed to press a button when they think they can see the emotion. For the decrease condition, the emotional expression starts at maximum intensity and morphs towards neutral and the participant is instructed to press a button when they think they can no longer see the emotion. For both conditions, each emotion is shown four times, equaling 40 trials in total.

Main outcomes (reported in main article)

- *Detection threshold for increase condition* – Average perceptual threshold for each emotion
- *Detection threshold for decrease condition* – Average perceptual threshold for each emotion

3. Face Affective Go/NoGo Task

Task description



The Face Affective Go/NoGo Task (fAGN) measures attentional bias and behavioural inhibitory control. A series of emotional faces are shown and the participant is instructed to react to one specific type of emotion (e.g., happy faces) by pressing a button while refraining from reacting to another type of emotion (e.g., sad faces). The ratio of 'Go' to 'NoGo' stimuli is 50/50. The task contains six blocks: happy targets with neutral distractors (happy/neutral), happy targets with sad distractors (happy/sad), neutral targets with happy distractors (neutral/happy), neutral targets with sad distractors (neutral/sad), sad targets with happy distractors (sad/happy), and sad targets with neutral distractors (sad/neutral). Each block consists of 20 trials equaling 120 trials in total.

Main outcomes (reported in main article)

- *d-prime (d')* – Index of discrimination sensitivity calculated as: *d' = z(hits) - z(false alarms)*

Secondary outcomes[1]

- *Hits (H)* – Percentage of correct responses during Go-trials
- *False alarm (FA)* – Percentage of incorrect responses during NoGo-trials
- *Reaction time* – Reaction time (ms) for correct responses during Go-trials

[1]Note other possible secondary outcomes include Misses (i.e., failure to respond to Go-trials), and Correct rejections (i.e., correctly withholding response during NoGo-trials), but as these outcomes are the inverse of H and FA respectively and can be inferred by their counterparts they are not reported here.

Face Affective Go/NoGo Task – Descriptive data

	Mean	*SD*	*Median*	*IQR*	*Range*	*Skewness*	*Floor effect*	*Ceiling effect*
Accuracy (%) - Happy/Neutral	*96.00*	*12.79*	100.00	0.00	0–100	-5.26	1%	82%
Accuracy (%) - Happy/Sad	*95.20*	*13.14*	100.00	2.50	0–100	-4.99	1%	75%
Accuracy (%) - Neutral/Happy	*86.60*	*20.16*	90.00	20.00	10–100	-1.95	0%	48%
Accuracy (%) - Neutral/Sad	*80.70*	*22.17*	90.00	30.00	10–100	-1.35	0%	33%
Accuracy (%) - Sad/Happy	*93.10*	*12.37*	100.00	10.00	20–100	-2.97	0%	62%
Accuracy (%) - Sad/Neutral	*91.50*	*14.93*	100.00	10.00	0–100	-3.14	1%	59%
False alarms (%) - Happy/Neutral	*7.80*	*9.91*	0.00	10.00	0–40	5.26	0%	51%
False alarms (%) - Happy/Sad	*9.30*	*10.47*	10.00	10.00	0–40	4.99	0%	43%
False alarms (%) - Neutral/Happy	*8.90*	*11.71*	10.00	10.00	0–60	1.95	0%	46%
False alarms (%) - Neutral/Sad	*12.70*	*13.40*	10.00	20.00	0–50	1.35	0%	34%
False alarms (%) - Sad/Happy	*9.30*	*10.47*	10.00	10.00	0–50	2.97	0%	41%
False alarms (%) - Sad/Neutral	*26.80*	*24.28*	20.00	30.00	0–100	3.14	4%	12%
Accuracy (ms) - Happy/Neutral	*420.30*	*56.52*	406.52	76.20	307.29–619.02	0.88	-	-
Accuracy (ms) - Happy/Sad	438.30	56.99	435.56	72.37	296.12–577.24	0.28	-	-
Accuracy (ms) - Neutral/Happy	484.76	89.29	480.48	107.87	263.64–728.27	0.19	-	-
Accuracy (ms) - Neutral/Sad	526.40	89.41	516.30	110.12	297.8–736.81	0.13	-	-
Accuracy (ms) - Sad/Happy	476.95	71.51	480.80	113.40	306.61–639.37	0.12	-	-
Accuracy (ms) - Sad/Neutral	460.82	66.54	457.36	96.99	297.88–613.92	0.10	-	-

Legend: Mean, SD, median, interquartile range, range, and skewness are shown. Note, for outcomes with non-parametric data distribution, mean and SD are denoted in cursive gray indicating the median and interquartile range should be used as reference. Floor and ceiling effects are shown as percentage of test subjects who achieved minimum (floor) or maximum (ceiling) scores. Note, for False alarm rate maximum score = 0% and minimum score = 100%.

Face Affective Go/NoGo Task – Test-retest reliability

	Baseline (n = 49)		Retest (n = 49)		*Test-retest bias (%)*	*ICC*	*95% CI*
	Mean	*SD*	*Mean*	*SD*			
Accuracy (%) - Happy/Neutral	97.55	6.30	98.78	3.89	1.26	0.66	0.40–0.81
Accuracy (%) - Happy/Sad	94.08	16.82	97.96	6.12	4.12	-0.02	-0.78–0.42
Accuracy (%) - Neutral/Happy	86.33	19.33	92.45	16.40	7.09	0.26	-0.27–0.58
Accuracy (%) - Neutral/Sad	78.57	22.27	85.31	18.27	8.58	0.02	-0.70–0.44
Accuracy (%) - Sad/Happy	93.47	10.52	94.90	9.60	1.53	0.30	-0.24–0.61
Accuracy (%) - Sad/Neutral	92.86	10.61	96.33	7.27	3.74	0.33	-0.15–0.61
False alarms (%) - Happy/Neutral	6.73	8.75	5.71	7.36	-15.16	0.33	-0.19–0.62
False alarms (%) - Happy/Sad	7.96	10.99	8.78	8.81	10.30	0.22	-0.40–0.56
False alarms (%) - Neutral/Happy	8.98	9.84	5.92	8.14	-34.08	0.56	0.24–0.75
False alarms (%) - Neutral/Sad	14.69	14.16	11.84	13.95	-19.40	0.39	-0.08–0.65
False alarms (%) - Sad/Happy	6.73	7.47	9.80	11.08	45.62	0.33	-0.16–0.61
False alarms (%) - Sad/Neutral	24.90	23.99	20.00	21.21	-19.68	0.45	0.03–0.69
Accuracy (ms) - Happy/Neutral	413.82	54.45	423.20	56.72	2.27	0.59	0.27–0.77
Accuracy (ms) - Happy/Sad	437.27	61.64	442.57	61.80	1.21	0.63	0.34–0.79
Accuracy (ms) - Neutral/Happy	490.24	97.54	491.45	75.22	0.25	0.68	0.42–0.82
Accuracy (ms) - Neutral/Sad	533.64	82.67	538.30	87.19	0.87	0.56	0.22–0.75
Accuracy (ms) - Sad/Happy	487.26	68.28	492.43	66.59	1.06	0.48	0.08–0.71
Accuracy (ms) - Sad/Neutral	465.23	62.04	481.79	70.20	3.56	0.59	0.28–0.77

Legend: Mean and SD are reported for baseline and retest sessions (3-5 weeks) along with Intraclass Correlation Coefficient (ICC) and their 95% Confidence Interval (95% CI). Test-retest bias is calculated as percentage change between first and second test: *Test-retest bias = ((retest-test)/test)*100.*

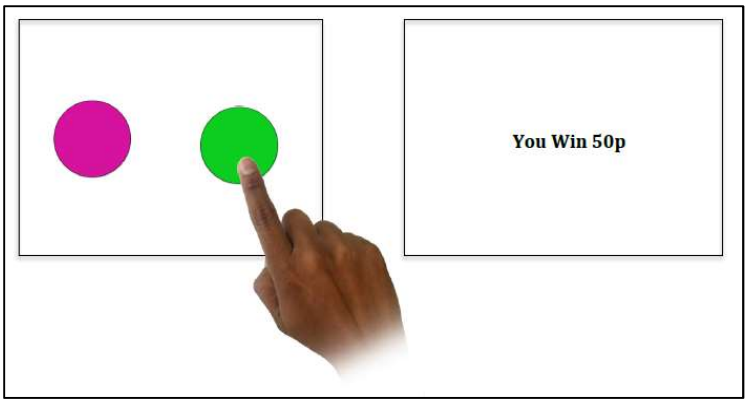
Face Affective Go/NoGo Task – Correlations

	Age	*Sex*[$]	*Education*	*IQ*	*Neuroticism*[a]	*TMD*
Accuracy (%) - Happy/Neutral	-0.15	0.13	-0.07	-0.02	-0.02	-0.04
Accuracy (%) - Happy/Sad	-0.07	0.03	0.02	0.02	0.02	-0.03
Accuracy (%) - Neutral/Happy	-0.01	0.004	0.02	-0.17	0.06	0.01
Accuracy (%) - Neutral/Sad	-0.17	-0.03	0.12	0.13	0.03	0.24
Accuracy (%) - Sad/Happy	-0.08	-0.06	0.09	0.003	0.11	0.08
Accuracy (%) - Sad/Neutral	-0.14	0.18	0.01	0.00	0.08	0.24
False alarms (%) - Happy/Neutral	-0.08	-0.08	-0.05	-0.08	0.03	0.03
False alarms (%) - Happy/Sad	-0.10	-0.13	0.08	-0.03	-0.03	-0.05
False alarms (%) - Neutral/Happy	-0.04	0.06	-0.11	-0.18	0.01	0.05
False alarms (%) - Neutral/Sad	**-0.27****	0.13	-0.18	0.02	-0.10	-0.10
False alarms (%) - Sad/Happy	-0.09	0.13	-0.02	-0.09	0.15	0.08
False alarms (%) - Sad/Neutral	0.03	-0.16	0.17	-0.01	0.07	-0.01
Accuracy (ms) - Happy/Neutral	0.13	-0.01	-0.03	-0.18	0.02	-0.02
Accuracy (ms) - Happy/Sad	0.21	-0.04	-0.02	-0.20	0.10	-0.02
Accuracy (ms) - Neutral/Happy	0.15	-0.053	-0.01	0.03	0.05	-0.11
Accuracy (ms) - Neutral/Sad	0.22	-0.04	0.20	0.14	0.05	-0.03
Accuracy (ms) - Sad/Happy	0.13	-0.11	-0.08	0.203	0.06	-0.06
Accuracy (ms) - Sad/Neutral	**0.29****	-0.03	-0.05	0.06	-0.05	-0.02

Legend: Correlations between Face Affective Go/NoGo task (fAGN) outcomes and age, sex, education indexed with the Family History Assessment Module (OS-FHAM) on a five-point likert scale, IQ score assessed with the Reynolds Intellectual Screening Test (RIST), total mood disturbance (TMD) indexed with the Profile of Mood Scale (POMS), and trait Neuroticism indexed with the NEO Personality Inventory-Revised (NEO PI-R, n = 93) and the NEO Personality Inventory 3 (NEO PI-3, n = 6). Correlation coefficients are reported as Spearman's rho; only *p*-values < .01 are considered significant and are marked in bold. ** = $p < .01$. [$] A negative rho value indicates males score higher while a positive rho value indicates females score higher. [a] N = 99 due to missing data from one participant.

4. Reinforcement Learning Task

Task description



The Reinforcement Learning Task assesses learning based on reward and punishment feedback. In every trial, two colored circles appear and the participant is asked to make a choice between the two. Within each color pair, the ratio between favorable and unfavorable outcomes for one color is 70/30 while the ratio between favorable and unfavorable outcome for the other color is 30/70. In total, there are four pairs of colors (grey/red, yellow/black, orange/blue, and purple/green). Half the color-pairs represent a win condition in which the favorable outcome is a monetary reward while the unfavorable outcome is no reward and the other half of the color-pairs represent a loss condition in which the favorable outcome is no loss and the unfavorable outcome is monetary loss. The task contains a learning phase made up of 120 trials and a transfer phase in which the colors are paired randomly made up of 48 trials, equaling a total of 168 trials.

Main outcomes (reported in main article)

- *Learning rate alpha for win and loss conditions*– Index of learning speed in the context of positive and negative feedback calculated using a reinforcement learning algorithm

Secondary outcomes

- *Temperature for win and loss conditions* – Index of exploration vs static behavior calculated using a reinforcement learning algorithm

Reinforcement Learning Task – Descriptive data

	Mean	*SD*	*Median*	*IQR*	*Range*	*Skewness*	*Floor effect*	*Ceiling effect*
Temperature - Win condition	*0.75*	*1.27*	0.05	0.82	0.01–4.58	1.84	32%	0%
Temperature - Loss condition	*0.92*	*0.99*	0.61	1.05	0.01–4.17	1.56	32%	0%

Legend: N = 68, as 32 participants performed below chance level, violating the assumptions of the reinforcement learning algorithm used to determine the temperature outcome. Mean, SD, median, interquartile range, range, and skewness are shown. Note, for outcomes with non-parametric data distribution, mean and SD are denoted in cursive gray indicating the median and interquartile range should be used as reference. Floor and ceiling effects are shown as percentage of test subjects who achieved minimum (floor) or maximum (ceiling) scores. Note, floor and ceiling effects are here based on the percentage of trials in which the correct (i.e., most favorable) colored circle is chosen; this data is not shown but should be inspected prior to running the reinforcement learning algorithm. Thus, to reach the criteria for floor effect the most favorable colored circled must be chosen in 50% > of trials (indicating a hit rate of below chance level).

Reinforcement Learning Task – Test-retest reliability

	Baseline (n = 35)		**Retest (n = 35)**		*Test-retest bias (%)*	*ICC*	*95% CI*
	Mean	*SD*	*Mean*	*SD*			
Temperature - Win condition	0.91	1.38	0.80	1.27	-11.87	0.18	-0.48–0.54
Temperature - Loss condition	0.97	1.02	1.13	1.46	17.09	-0.07	-0.94–0.41

Legend: N = 35, as 14 participants performed below chance level, violating the assumptions of the reinforcement learning algorithm used to determine the temperature outcome. Mean and SD are reported for baseline and retest sessions (3-5 weeks) along with Intraclass Correlation Coefficient (ICC) and their 95% Confidence Interval (95% CI). Test-retest bias is calculated as percentage change between first and second test: *Test-retest bias = ((retest-test)/test)*100.*

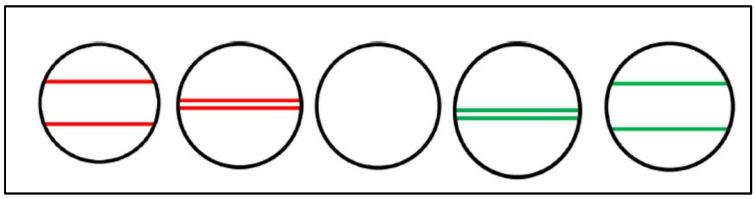
Reinforcement Learning Task – Correlations

	Age	*Sex*[$]	*Education*	*IQ*	*Neuroticism*[a]	*TMD*	*Motivation*	*Diligence*
Temperature - Win condition	-0.01	0.04	0.08	0.09	-0.09	-0.19	**-0.34****	**-0.35****
Temperature - Loss condition	0.12	-0.11	-0.12	0.07	-0.17	-0.16	-0.24	**-0.31****

Legend: N = 68, as 32 participants performed below chance level, violating the assumptions of the reinforcement learning algorithm used to determine the temperature outcome. Correlations between Reinforcement Learning outcomes and age, sex, education indexed with the Family History Assessment Module (OS-FHAM) on a five-point likert scale, IQ score assessed with the Reynolds Intellectual Screening Test (RIST), total mood disturbance (TMD) indexed with the Profile of Mood Scale (POMS), and trait Neuroticism indexed with the NEO Personality Inventory-Revised (NEO PI-R, n = 62) and the NEO Personality Inventory 3 (NEO PI-3, n = 5). Correlations between self-reported motivation and diligence are also shown. Correlation coefficients are reported as Spearman's rho; only *p*-values < .01 are considered significant and are marked in bold. ** = $p < .01$. [$] A negative rho value indicates males score higher while a positive rho value indicates females score higher. [a] N = 67 due to missing data from one participant.

5. Monetary Incentive Reward Task

Task description



The Monetary Incentive Reward task (MIR) assesses sensitivity to reward and punishment. Each trial, a pair of circles appear on the screen followed by a black box. Participants are instructed to press a key as soon as they see the black box appear. The difficulty level of the task is tailored to the individual reaction time of each participant, calculated from 30 baseline trials. In the main task, two lines appear in the circles indicating the potential to win money (green lines) or lose money (red lines). The distance between the lines indicate the size of the gain/loss, with large distances signaling a large gain/loss and small distances signaling a small gain/loss. Participants must respond faster than their baseline reaction time in order to gain money or avoid loss and are given feedback after each individual trial. Each condition (high win, low win, high loss, low loss, and neutral) are shown 20 times, equaling 100 trials in total as well as 30 neutral baseline trials.

Main outcomes (reported in the main article)

- *Reaction time (ms) for win* – Average reaction time in milliseconds across the two win conditions relative to baseline calculated as: *Win = baseline - (high win + low win)/2*
- *Reaction time (ms) for loss* – Average reaction time in milliseconds across the two loss conditions relative to baseline calculated as: *Loss = baseline - (high loss + low loss)/2*

Secondary outcomes

- *Reaction time for each condition* – Average reaction time in milliseconds for high win, low win, high loss, low loss relative to baseline, e.g., calculated as: *High win = baseline - high win*

Monetary Incentive Reward Task – Descriptive data

	Mean	*SD*	*Median*	*IQR*	*Range*	*Skewness*	*Floor effect*	*Ceiling effect*
Reaction time (ms) - High win	17.02	20.31	17.96	27.39	-26.53–78.93	0.41	-	-
Reaction time (ms) - Low win	17.79	20.33	17.59	26.00	-34.04–67.5	0.22	-	-
Reaction time (ms) - High loss	19.60	21.15	18.67	20.32	-38.95–94.53	0.26	-	-
Reaction time (ms) - Low loss	*17.87*	*18.34*	15.90	23.65	-50.08–74.76	0.19	-	-

Legend: Mean, SD, median, interquartile range, range, and skewness are shown. Note, for outcomes with non-parametric data distribution, mean and SD are denoted in cursive gray indicating the median and interquartile range should be used as reference. Note floor and ceiling effects cannot be estimated for reaction time outcomes.

Monetary Incentive Reward Task – Test-retest reliability

	Baseline (n = 49)		**Retest (n = 49)**		*Test-retest bias (%)*	*ICC*	*95% CI*
	Mean	*SD*	*Mean*	*SD*			
Reaction time (ms) - High win	17.10	22.16	20.76	25.58	21.40	-0.64	-1.96–0.08
Reaction time (ms) - Low win	19.85	20.15	13.12	19.64	-33.90	0.02	-0.68–0.44
Reaction time (ms) - High loss	21.01	23.50	22.12	24.47	5.28	-0.63	-1.96–0.10
Reaction time (ms) - Low loss	17.10	21.17	15.85	20.10	-7.31	0.03	-0.75–0.46

Legend: Mean and SD are reported for baseline and retest sessions (3-5 weeks) along with Intraclass Correlation Coefficient (ICC) and their 95% Confidence Interval (95% CI). Test-retest bias is calculated as percentage change between first and second test: *Test-retest bias = ((retest-test)/test)*100.*

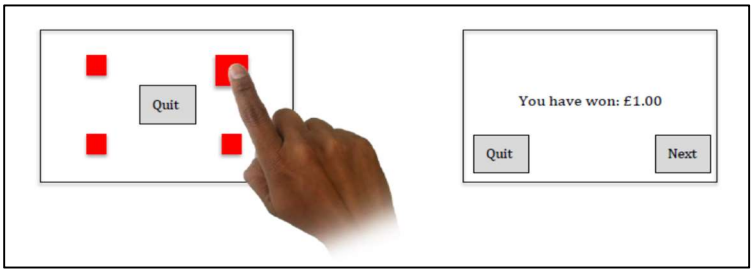
Monetary Incentive Reward Task – Correlations

	Age	*Sex*[$]	*Education*	*IQ*	*Neuroticism*[a]	*TMD*	*Motivation*	*Diligence*
Reaction time (ms) - High win	-0.03	-0.01	0.01	-0.12	0.02	-0.14	-0.08	0.09
Reaction time (ms) - Low win	-0.10	0.00	0.22	-0.10	-0.11	-0.19	-0.09	0.08
Reaction time (ms) - High loss	0.01	0.07	-0.04	-0.12	-0.07	-0.19	-0.14	0.06
Reaction time (ms) - Low loss	0.02	0.05	-0.09	-0.09	-0.04	-0.11	0.01	0.09

Legend: Correlations between Monetary Incentive Reward Task outcomes and age, sex, education indexed with the Family History Assessment Module (OS-FHAM) on a five-point likert scale, IQ score assessed with the Reynolds Intellectual Screening Test (RIST), total mood disturbance (TMD) indexed with the Profile of Mood Scale (POMS), and trait Neuroticism indexed with the NEO Personality Inventory-Revised (NEO PI-R, n = 93) and the NEO Personality Inventory 3 (NEO PI-3, n = 6). Correlations between self-reported motivation and diligence are also shown. Correlation coefficients are reported as Spearman's rho; only *p*-values < .01 are considered significant and are marked in bold. [$] A negative rho value indicates males score higher while a positive rho value indicates females score higher. [a] N = 99 due to missing data from one participant.

6. Progressive Ratio Task

Task description



The Progressive Ratio Task assesses self-control and motivational breakpoint. Four squares are shown on a screen and the participant is instructed to pick the odd one out. The task is made up of a high-yield block (reward 10 Danish kroners), a medium-yield block (reward 2 Danish kroners), and a low-yield block (reward ½ Danish kroners). Within each block, the number of trials needed to obtain a monetary reward continually doubles (4, 8, 16, 32, etc.). The participant is told that they can quit the task at any time, however they must remain in front of the computer until the full run-time of the task is complete (~20 minutes). The high-yield block contains 60 trials, the medium-yield block contains 124 trials, and the low-yield block contains 252 trials, equaling 436 trials in total.

Main outcomes (reported in the main article)

- *Breakpoint* – Number of trials completed

Secondary outcomes

- *Post reinforcement break* – Average time in seconds between reward and decision to continue task calculated for each block

Progressive Ratio Task – Descriptive data

	N	*Mean*	*SD*	*Median*	*IQR*	*Range*	*Skewness*	*Floor effect*	*Ceiling effect*
Post reinforcement pause (s) – High-yield	98	*1.11*	*1.07*	0.79	0.63	0.28–8.66	4.53	-	-
Post reinforcement pause (s) – Medium-yield	90	*0.68*	*0.69*	0.49	0.29	0.16–4.85	4.32	-	-
Post reinforcement pause (s) - Low-yield	73	*0.55*	*0.42*	0.44	0.34	0.15–2.77	3.00	-	-

Legend: N indicates number of participants who completed each block. Mean, SD, median, interquartile range, range, and skewness are shown. Note, for outcomes with non-parametric data distribution, mean and SD are denoted in cursive gray indicating the median and interquartile range should be used as reference. Note floor and ceiling effects cannot be estimated for outcomes measured in time.

Progressive Ratio Task – Test-retest reliability

	Baseline (n = 49)			**Retest (n = 49)**			*Test-retest bias (%)*	*ICC*	*95% CI*
	N	*Mean*	*SD*	*N*	*Mean*	*SD*			
Post reinforcement pause (s) – High-yield	48	1.14	0.71	49	0.63	0.23	-44.74	0.26	-0.16–0.56
Post reinforcement pause (s) - Medium-yield	43	0.78	0.71	48	0.69	1.72	-11.54	0.09	-0.71–0.51
Post reinforcement pause (s) - Low-yield	35	0.68	0.56	38	0.67	1.01	-1.47	-0.13	-1.56–0.49

Legend: N indicates number of participants who completed each block. Mean and SD are reported for baseline and retest sessions (3-5 weeks) along with Intraclass Correlation Coefficient (ICC) and their 95% Confidence Interval (95% CI). Test-retest bias is calculated as percentage change between first and second test: *Test-retest bias = ((retest-test)/test)*100.*

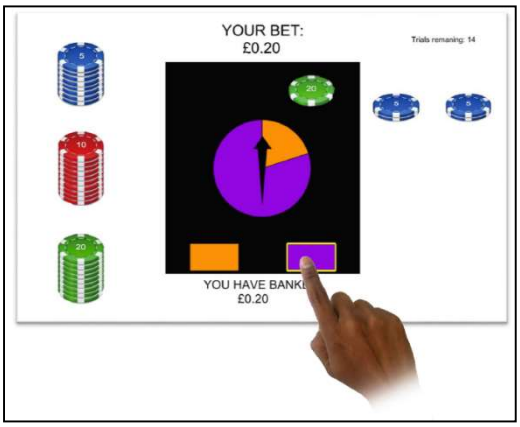
Progressive Ratio Task – correlations

	N	*Age*	*Sex*[$]	*Education*	*IQ*	*Neuroticism*[a]	*TMD*	*Motivation*	*Diligence*
Post reinforcement pause (s) - High-yield	98	0.09	0.10	0.09	-0.12	-0.06	0.03	-0.02	0.03
Post reinforcement pause (s) - Medium-yield	90	0.07	0.12	-0.03	-0.09	0.07	0.06	0.05	0.06
Post reinforcement pause (s) - Low-yield	73	0.11	0.16	0.01	0.02	0.12	0.05	-0.03	0.10

Legend: N indicates number of participants who completed each block. Correlations between Progressive Ratio Task outcomes and age, sex, education indexed with the Family History Assessment Module (OS-FHAM) on a five-point likert scale, IQ score assessed with the Reynolds Intellectual Screening Test (RIST), total mood disturbance (TMD) indexed with the Profile of Mood Scale (POMS), and trait Neuroticism indexed with the NEO Personality Inventory-Revised (NEO PI-R, n = 93) and the NEO Personality Inventory 3 (NEO PI-3, n = 6). Correlations between self-reported motivation and diligence are also shown. Correlation coefficients are reported as Spearman's rho; only *p*-values < .01 are considered significant and are marked in bold. [$] A negative rho value indicates males score higher while a positive rho value indicates females score higher. [a] N = 99 due to missing data from one participant.

7. Adapted Cambridge Gambling Task

Task description



The Adapted Cambridge Gambling Task (aCGT) assesses risk-taking and decision-making. At the beginning of the task, the participant is given three stacks of chips; 10 chips worth 1.5 Danish kroners, 10 chips worth 2 Danish kroners, and 10 chips worth 4 Danish kroners. In each trial, a roulette wheel made up of two colors is presented; the participant must pick a color and place two chips on their bet. The task contains two conditions; a win and a loss condition. In the win condition, the participant will either double or keep the money they bet and in the loss condition, the participant will either keep or lose the money they bet. Each condition contains 15 trials (2 x 90% odds, 2 x 80% odds, 4 x 70% odds, 4 x 60% odds, and 3 x 50% odds), equaling 30 trials in total.

Main outcomes (reported in the main article)

- *Risk adjustment for win and loss conditions* – Index of association between odds of winning and the size of bet calculated as: *Risk adjustment = ((2 x mean bet 90%) + (1 x mean bet 80%) + (0 x mean bet 70%) - (1 x mean bet 60%) - (2 x mean bet 50%)) / average bet*.

Secondary outcomes

- *Mean bets for win and loss conditions* – Mean bet in kroners for individual odds (90%, 80%, 70%, 60%, and 50%) for both win and loss conditions.
- *Quality of decision making* – Percentage of trials in which the best odds were picked

Adapted Cambridge Gambling Task – Descriptive data

	Mean	*SD*	*Median*	*IQR*	*Range*	*Skewness*	*Floor effect*	*Ceiling effect*
Mean bet 50% (kroners) - Win condition	*1.57*	*0.54*	1.42	1.00	1–3	0.72	-	-
Mean bet 60% (kroners) - Win condition	*2.04*	*0.54*	2.00	0.69	1–3.50	0.49	-	-
Mean bet 70% (kroners) - Win condition	2.54	0.51	2.50	0.62	1.38–4	-0.07	-	-
Mean bet 80% (kroners) - Win condition	*2.90*	*0.77*	3.00	1.00	1–4	-0.31	-	-
Mean bet 90% (kroners) - Win condition	*3.15*	*0.85*	3.00	1.50	1.25–4	-0.57	-	-
Mean bet 50% (kroners) - Loss condition	*1.47*	*0.50*	1.33	0.71	1–3.17	1.13	-	-
Mean bet 60% (kroners) - Loss condition	*1.81*	*0.45*	1.75	0.50	1–3.50	1.13	-	-
Mean bet 70% (kroners) - Loss condition	*2.58*	*0.49*	2.56	0.53	1.33–4	0.36	-	-
Mean bet 80% (kroners) - Loss condition	*3.21*	*0.72*	3.00	1.25	1–4	-0.68	-	-
Mean bet 90% (kroners) - Loss condition	*3.37*	*0.69*	3.50	1.00	1.50–4	-0.79	-	-
Quality of decision making	*97.53*	*5.62*	100.00	3.33	60–100	-4.26	0%	66%

Legend: Mean, SD, median, interquartile range, range, and skewness are shown. Note, for outcomes with non-parametric data distribution, mean and SD are denoted in cursive gray indicating the median and interquartile range should be used as reference. Floor and ceiling effects are shown as percentage of test subjects who achieved minimum (floor) or maximum (ceiling) scores. Note, floor and ceiling effects cannot be estimated for monetary bets.

Adapted Cambridge Gambling Task – Test-retest reliability

	Baseline (n = 49)		Retest (n = 49)		*Test-retest bias (%)*	*ICC*	*95% CI*
	Mean	*SD*	*Mean*	*SD*			
Mean bet 50% (kroners) - Win condition	1.59	0.54	1.41	0.44	-11.32	0.44	0.04–0.68
Mean bet 60% (kroners) - Win condition	1.96	0.53	1.84	0.40	-6.12	0.28	-0.26–0.59
Mean bet 70% (kroners) - Win condition	2.55	0.55	2.52	0.51	-1.18	0.36	-0.16–0.64
Mean bet 80% (kroners) - Win condition	2.98	0.78	3.16	0.71	6.04	-0.02	-0.81–0.43
Mean bet 90% (kroners) - Win condition	3.17	0.87	3.51	0.70	10.73	0.38	-0.05–0.65
Mean bet 50% (kroners) - Loss condition	1.46	0.50	1.39	0.39	-4.79	-0.06	-0.90–0.40
Mean bet 60% (kroners) - Loss condition	1.83	0.50	1.76	0.49	-3.83	0.48	0.08–0.71
Mean bet 70% (kroners) - Loss condition	2.56	0.43	2.49	0.44	-2.73	0.37	-0.11–0.65
Mean bet 80% (kroners) - Loss condition	3.26	0.72	3.44	0.66	5.52	0.09	-0.59–0.48
Mean bet 90% (kroners) - Loss condition	3.37	0.69	3.52	0.69	4.45	0.10	-0.58–0.49
Quality of decision making	97.96	6.04	99.18	1.45	1.25	0.22	-0.36–0.56

Legend: Mean and SD are reported for baseline and retest sessions (3-5 weeks) along with Intraclass Correlation Coefficient (ICC) and their 95% Confidence Interval (95% CI). Test-retest bias is calculated as percentage change between first and second test: *Test-retest bias = ((retest-test)/test)*100.*

Adapted Cambridge Gambling Task – Correlations

	Age	*Sex*[$]	*Education*	*IQ*	*Neuroticism*[a]	*TMD*	*Motivation*	*Diligence*
Mean bet 50% (kroners) - Win condition	-0.11	0.16	-0.05	-0.05	0.16	0.09	0.03	0.03
Mean bet 60% (kroners) - Win condition	0.03	0.20	-0.03	-0.25	0.06	-0.08	0.15	0.09
Mean bet 70% (kroners) - Win condition	-0.09	-0.01	-0.01	-0.04	-0.03	-0.09	0.06	0.04
Mean bet 80% (kroners) - Win condition	0.18	-0.26	-0.12	0.11	-0.04	0.03	0.08	0.07
Mean bet 90% (kroners) - Win condition	0.00	-0.16	0.22	0.18	-0.17	0.07	-0.17	-0.06
Mean bet 50% (kroners) - Loss condition	0.16	0.16	0.00	-0.06	0.12	0.03	0.00	-0.05
Mean bet 60% (kroners) - Loss condition	0.15	-0.07	-0.14	-0.02	-0.04	-0.08	-0.09	-0.09
Mean bet 70% (kroners) - Loss condition	-0.03	-0.09	-0.06	-0.15	0.12	0.04	0.03	0.05
Mean bet 80% (kroners) - Loss condition	-0.20	-0.11	0.16	0.12	-0.22	-0.02	0.02	0.03
Mean bet 90% (kroners) - Loss condition	-0.09	0.15	0.19	0.12	-0.06	0.00	0.03	0.05
Quality of decision making	-0.05	0.32	0.01	0.19	-0.02	-0.01	-0.18	-0.04

Legend: Correlations between Adapted Cambridge Gambling Task outcomes and age, sex, education indexed with the Family History Assessment Module (OS-FHAM) on a five-point likert scale, IQ score assessed with the Reynolds Intellectual Screening Test (RIST), total mood disturbance (TMD) indexed with the Profile of Mood Scale (POMS), and trait Neuroticism indexed with the NEO Personality Inventory-Revised (NEO PI-R, n = 93) and the NEO Personality Inventory 3 (NEO PI-3, n = 6). Correlations between self-reported motivation and diligence are also shown. Correlation coefficients are reported as Spearman's rho; only *p*-values < .01 are considered significant and are marked in bold. [$] A negative rho value indicates males score higher while a positive rho value indicates females score higher. [a] N = 99 due to missing data from one participant.

8. Moral Emotions Task

Task description

The Moral Emotions Task (ME) assesses the effect of intention on experience of moral emotions.

A series of moral scenarios are shown in which one character either intentionally or unintentionally causes another character physical or emotional harm. The participant is instructed to imagine themselves as either the agent (i.e., the person causing the harm) or the victim and to rate on a scale from 1 (not at all) to 7 (extremely) how 'guilty', 'ashamed', and 'annoyed' they would feel as well as how overall 'bad' to 'good' they would feel. The task consists of 12 moral scenarios and each scenario is shown twice so the participant can take on the role of both agent and victim, equaling a total of 24 trials.

Main outcomes (reported in the main article)

- *Agent and victim ratings of guilt and shame* – Average ratings of guilt and shame for agent and victim scenarios e.g., calculated as: *Agent guilt = (agent guilt intentional + agent guilt unintentional)/2.*

Secondary outcomes

- *Ratings for each condition (role x intentionality x emotion)* – Ratings for each of the 16 possible combinations of agent vs victim, intentional vs unintentional, and 'guilt', 'shame', 'annoyance'[2], and 'feeling bad'.

[2] As discussed in the main article, we observed a serious issue with the ratings for 'annoyance'. As a consequence, this outcome will not be presented here.

Moral Emotions Task – Descriptive data

	Mean	*SD*	*Median*	*IQR*	*Range*	*Skewness*	*Floor effect*	*Ceiling effect*
Guilt - Agent intentional	*6.14*	*0.85*	6.33	0.83	2.50–7.00	-1.88	0%	13%
Guilt - Agent unintentional	*5.58*	*0.82*	5.71	0.86	2.00–7.00	-1.61	0%	1%
Guilt - Victim intentional	*1.63*	*0.63*	1.50	0.83	1.00–3.83	1.35	19%	0%
Guilt - Victim unintentional	*1.54*	*0.59*	1.43	0.86	1.00–4.43	1.88	26%	0%
Shame - Agent intentional	*5.99*	*0.87*	6.17	1.04	2.50–7.00	-1.46	0%	10%
Shame - Agent unintentional	*5.48*	*0.88*	5.57	1.14	2.00–7.00	-1.12	0%	1%
Shame - Victim intentional	*2.46*	*0.96*	2.25	1.67	1.00–4.83	0.26	9%	0%
Shame - Victim unintentional	*1.47*	*0.63*	1.14	0.75	1.00–4.00	1.57	38%	0%
Bad - Agent intentional	*1.92*	*0.70*	1.83	1.00	1.00–4.83	1.11	0%	8%
Bad - Agent unintentional	*2.30*	*0.72*	2.29	0.75	1.00–5.57	1.38	0%	3%
Bad - Victim intentional	*2.54*	*0.69*	2.50	0.67	1.00–5.17	1.17	0%	1%
Bad - Victim unintentional	*3.06*	*0.61*	3.00	0.75	1.57–5.43	0.65	0%	0%

Legend: Mean, SD, median, interquartile range, range, and skewness are shown. Note, for outcomes with non-parametric data distribution, mean and SD are denoted in cursive gray indicating the median and interquartile range should be used as reference. Floor and ceiling effects are shown as percentage of test subjects who achieved minimum (floor) or maximum (ceiling) scores. Note, ratings for bad are reversed so a score of 1 is high (i.e., feeling 'bad') and a score of 7 is low (i.e., feeling 'good').

Moral Emotions Task – Test-retest reliability

	Baseline (n = 49)		Retest (n = 49)		*Test-retest bias (%)*	*ICC*	*95% CI*
	Mean	*SD*	*Mean*	*SD*			
Guilt - Agent intentional	6.18	0.83	6.12	0.69	-0.97	0.73	0.51–0.85
Guilt - Agent unintentional	5.59	0.86	5.59	0.84	0	0.85	0.74–0.92
Guilt - Victim intentional	1.69	0.67	1.71	0.57	1.18	0.75	0.56–0.86
Guilt - Victim unintentional	1.57	0.55	1.66	0.55	5.73	0.80	0.65–0.89
Shame - Agent intentional	6.06	0.86	5.92	0.78	-2.31	0.80	0.65–0.89
Shame - Agent unintentional	5.54	0.92	5.44	0.82	-1.81	0.81	0.67–0.89
Shame - Victim intentional	2.52	0.88	2.69	0.92	6.75	0.78	0.60–0.87
Shame - Victim unintentional	1.57	0.61	1.62	0.62	3.18	0.79	0.63–0.88
Bad - Agent intentional	1.84	0.63	1.89	0.56	2.72	0.82	0.67–0.90
Bad - Agent unintentional	2.28	0.71	2.29	0.63	0.44	0.85	0.74–0.92
Bad - Victim intentional	2.48	0.55	2.41	0.61	-2.82	0.80	0.65–0.89
Bad - Victim unintentional	3.06	0.60	3.00	0.59	-1.96	0.83	0.71–0.91

Legend: Mean and SD are reported for baseline and retest sessions (3-5 weeks) along with Intraclass Correlation Coefficient (ICC) and their 95% Confidence Interval (95% CI). Test-retest bias is calculated as percentage change between first and second test: *Test-retest bias = ((retest-test)/test)*100.*

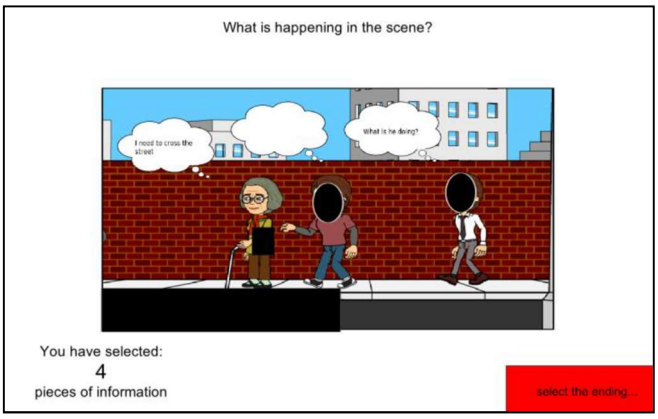
Moral Emotions Task - Correlations

	Age	*Sex*[$]	*Education*	*IQ*	*Neuroticism*[a]	*TMD*
Guilt - Agent intentional	0.21	0.08	-0.09	-0.08	0.06	-0.07
Guilt - Agent unintentional	0.05	0.24	0.01	-0.13	0.24	0.10
Guilt - Victim intentional	0.05	0.20	-0.15	-0.01	**0.28****	0.22
Guilt - Victim unintentional	0.09	0.05	0.09	0.01	0.12	0.07
Shame - Agent intentional	0.08	0.19	-0.07	-0.07	0.14	-0.02
Shame - Agent unintentional	-0.04	**0.29****	0.02	-0.20	**0.22****	0.17
Shame - Victim intentional	-0.07	0.19	-0.07	-0.18	**0.29****	0.18
Shame - Victim unintentional	0.01	0.06	0.10	-0.03	**0.28****	0.22
Bad - Agent intentional	-0.21	-0.13	0.05	0.04	0.01	0.10
Bad - Agent unintentional	0.01	-0.23	0.08	0.18	-0.08	0.07
Bad - Victim intentional	0.02	-0.09	0.12	0.04	-0.13	-0.06
Bad - Victim unintentional	0.05	-0.02	0.19	0.04	-0.26	-0.09

Legend: Correlations between Moral Emotions Task (ME) outcomes and age, sex, education indexed with the Family History Assessment Module (OS-FHAM) on a five-point likert scale, IQ score assessed with the Reynolds Intellectual Screening Test (RIST), total mood disturbance (TMD) indexed with the Profile of Mood Scale (POMS), and trait Neuroticism indexed with the NEO Personality Inventory-Revised (NEO PI-R, n = 93) and the NEO Personality Inventory 3 (NEO PI-3, n = 6). Correlation coefficients are reported as Spearman's rho; only *p*-values < .01 are considered significant and are marked in bold. ** = $p < .01$. [$] A negative rho value indicates males score higher while a positive rho value indicates females score higher. [a] N = 99 due to missing data from one participant.

9. Social Information Preference Task

Task description



The Social Information Preference Task (SIP) assesses information sampling for interpretation of social situations. A series of social scenarios are shown in which nine pieces of information are hidden; three thought bubbles, three faces, and three object/facts. The participant is allowed to choose four pieces of information to help them decide what is happening in the situation. They are then asked to choose between three different interpretations (a positive, a neutral, and a negative) as well as rate how confident they are in their choice on a scale from 1 ('not at all') to 7 ('very much so'). All outcomes are equally plausible and there are no right or wrong answers. The task consists of 18 scenarios.

Main outcomes (reported in the main article)

- *Type of information chosen* – The percentage of thoughts, faces, and facts chosen

Secondary outcomes

- *Type of outcomes chosen* – The percentage of positive, neutral, and negative outcomes chosen
- *Confidence in choice of outcome* – Average confidence rating for positive, neutral, and negative outcome choices

Social Information Preference Task – Descriptive data

	Mean	*SD*	*Median*	*IQR*	*Range*	*Skewness*	*Floor effect*	*Ceiling effect*
Outcome (%) - Positive	*38.44*	*12.09*	40.62	12.50	6.25–75.00	-0.14	0%	0%
Outcome (%) - Neutral	*33.56*	*10.90*	31.25	14.06	12.50–62.50	0.44	0%	0%
Outcome (%) - Negative	*28.00*	*12.67*	28.12	18.75	0.00–56.25	0.32	1%	0%
Confidence in outcome (%) - Positive	4.90	0.83	4.88	1.16	2.38–6.50	-0.48	0%	0%
Confidence in outcome (%) - Neutral	4.74	0.93	4.80	1.18	2.86–7.00	0.04	0%	1%
Confidence in outcome (%) - Negative	4.85	1.03	4.83	1.23	2.00–7.00	-0.26	0%	3%

Legend: Mean, SD, median, interquartile range, range, and skewness are shown. Note, for outcomes with non-parametric data distribution, mean and SD are denoted in cursive gray indicating the median and interquartile range should be used as reference. Floor and ceiling effects are shown as percentage of test subjects who achieved minimum (floor) or maximum (ceiling) scores.

Social Information Preference Task – Test-retest reliability

	Baseline (n = 49)		**Retest (n = 49)**		*Test-retest bias (%)*	*ICC*	*95% CI*
	Mean	*SD*	*Mean*	*SD*			
Outcome (%) - Positive	37.24	13.26	40.31	14.66	8.24	0.65	0.38–0.80
Outcome (%) - Neutral	35.84	11.47	37.50	11.76	4.63	0.33	-0.20–0.62
Outcome (%) - Negative	26.91	12.51	22.19	13.32	-17.54	0.77	0.55–0.87
Confidence in outcome (%) - Positive	4.85	0.86	4.86	0.84	0.21	0.58	0.26–0.77
Confidence in outcome (%) - Neutral	4.59	0.92	4.75	0.83	3.49	0.58	0.27–0.76
Confidence in outcome (%) - Negative	4.80	1.10	4.68	1.11	-2.50	0.34	-0.18–0.63

Legend: Mean and SD are reported for baseline and retest sessions (3-5 weeks) along with Intraclass Correlation Coefficient (ICC) and their 95% Confidence Interval (95% CI). Test-retest bias is calculated as percentage change between first and second test: *Test-retest bias = ((retest-test)/test)*100.*

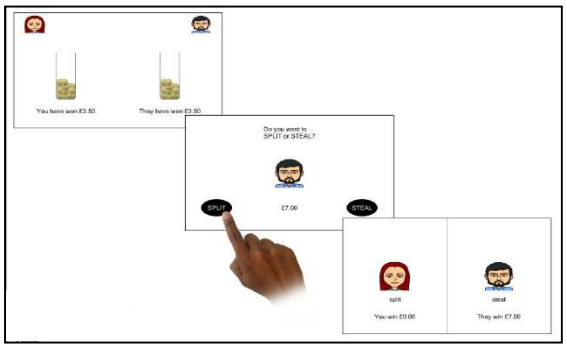
Social Information Preference Task – Correlations

	Age	*Sex*[$]	*Education*	*IQ*	*Neuroticism*[a]	*TMD*
Outcome (%) - Positive	0.19	-0.16	-0.06	-0.11	0.01	-0.08
Outcome (%) - Neutral	**0.29****	-0.07	-0.04	0.24	-0.11	-0.08
Outcome (%) - Negative	**-0.42*****	0.23	0.07	-0.08	0.12	0.17
Confidence in outcome (%) - Positive	0.06	-0.06	-0.07	-0.02	-0.10	-0.01
Confidence in outcome (%) - Neutral	0.05	-0.10	0.04	-0.03	-0.14	-0.10
Confidence in outcome (%) - Negative	0.01	-0.06	0.06	0.09	-0.16	-0.21

Legend: Correlations between Social Information Preference task (SIP) outcomes and age, sex, education indexed with the Family History Assessment Module (OS-FHAM) on a five-point likert scale, IQ score assessed with the Reynolds Intellectual Screening Test (RIST), total mood disturbance (TMD) indexed with the Profile of Mood Scale (POMS), and trait Neuroticism indexed with the NEO Personality Inventory-Revised (NEO PI-R, n = 93) and the NEO Personality Inventory 3 (NEO PI-3, n = 6). Correlation coefficients are reported as Spearman's rho; only *p*-values <.01 are considered significant and are marked in bold. ** = $p < .01$, *** = $p < .001$. [$] A negative rho value indicates males score higher while a positive rho value indicates females score higher. [a] N = 99 due to missing data from one participant.

10. Prisoner's Dilemma

Task description



The Prisoner's Dilemma task (PD) assesses cooperative behavior. At the beginning of each trial, the participant and a computer opponent complete a small task to collect money which is then pooled; a third of the time, the participant contributes more, a third of the time both parties contribute equally, and a third of the time the opponent contributes more. The participant is asked whether they wish to split or steal the pooled winnings. If both the participant and the opponent decide to split, they each get half, if they both decide to steal, neither of them get anything. If one player decides to steal and the other to split, the player who stole gets the whole sum. The participant is faced with three different opponents exhibiting different strategies: Cooperative (always splits), tit-for-two-tat (starts with a split and the changes behavior if the participant steals two times consecutively), and aggressive (tit-for-tat, starting with a steal). There are nine trials for each opponent type, equaling 27 trials in total.

Main outcomes (reported in the main article)

- *Proportion of steals* – Proportion of trials (%) in which the participant chooses to steal for each type of opponent (cooperative, tit-for-two-tat, aggressive)

Secondary outcomes

- *Proportion of steals based on contribution* – Proportion of trials (%) in which the participant chooses to steal from each type of opponent (cooperative, tit-for-tat, aggressive) for each level of contribution (more, equal, less)

Prisoner's Dilemma – Descriptive data

	Mean	*SD*	*Median*	*IQR*	*Range*	*Skewness*	*Floor effect*	*Ceiling effect*
Steals cooperative (%) - Player contributes more	*23.33*	*34.98*	0.00	33.33	0–100	1.22	63%	11%
Steals cooperative (%) - Equal contribution	*13.33*	*28.43*	0.00	0.00	0–100	2.11	78%	6%
Steals cooperative (%) - Player contributes less	*25.00*	*37.12*	0.00	33.33	0–100	1.14	63%	14%
Steals tit-for-two-tat (%) - Player contributes more	*28.00*	*38.99*	0.00	66.67	0–100	0.94	61%	16%
Steals tit-for-two-tat (%) - Equal contribution	*21.00*	*34.38*	0.00	33.33	0–100	1.36	68%	10%
Steals tit-for-two-tat (%) - Player contributes less	*27.67*	*37.32*	0.00	66.67	0–100	0.92	59%	13%
Steals aggressive (%) - Player contributes more	*38.33*	*37.72*	33.33	66.67	0–100	0.38	41%	16%
Steals aggressive (%) - Equal contribution	*30.33*	*35.17*	33.33	66.67	0–100	0.76	49%	11%
Steals aggressive (%) - Player contributes less	*36.33*	*39.10*	33.33	66.67	0–100	0.44	48%	16%

Legend: Mean, SD, median, interquartile range, range, and skewness are shown. Note, for outcomes with non-parametric data distribution, mean and SD are denoted in cursive gray indicating the median and interquartile range should be used as reference. Floor and ceiling effects are shown as percentage of test subjects who achieved minimum (floor) or maximum (ceiling) scores. Note, floor effects are defined as players who never steals from the opponent while ceiling effects are defined as players who always steals from opponent.

Prisoner's Dilemma – Test-retest reliability

	Baseline (n = 49)		Retest (n = 49)		*Test-retest bias (%)*	*ICC*	*95% CI*
	Mean	*SD*	*Mean*	*SD*			
Steals cooperative (%) - Player contributes more	17.69	28.95	21.77	33.02	23.06	0.50	0.12–0.72
Steals cooperative (%) - Equal contribution	8.84	24.32	13.61	27.99	53.96	0.37	-0.12–0.64
Steals cooperative (%) - Player contributes less	17.69	31.26	17.69	32.70	0.00	0.71	0.49–0.84
Steals tit-for-two-tat (%) - Player contributes more	23.13	36.77	19.05	29.66	-17.64	0.66	0.40–0.81
Steals tit-for-two-tat (%) - Equal contribution	13.61	27.99	14.29	28.05	5.00	0.63	0.34–0.79
Steals tit-for-two-tat (%) - Player contributes less	19.73	30.37	16.33	28.16	-17.23	0.39	-0.09–0.66
Steals aggressive (%) - Player contributes more	31.97	38.47	29.25	32.37	-8.51	0.63	0.34–0.79
Steals aggressive (%) - Equal contribution	25.85	32.11	21.09	31.69	-18.41	0.58	0.26–0.76
Steals aggressive (%) - Player contributes less	30.61	38.99	34.69	39.06	13.33	0.54	0.18–0.74

Legend: Mean and SD are reported for baseline and retest sessions (3-5 weeks) along with Intraclass Correlation Coefficient (ICC) and their 95% Confidence Interval (95% CI). Test-retest bias is calculated as percentage change between first and second test: *Test-retest bias = ((retest-test)/test)*100.*

Prisoner's Dilemma - Correlations

	Age	*Sex*[$]	*Education*	*IQ*	*Neuroticism*[a]	*TMD*	*Motivation*	*Diligence*
Steals cooperative (%) - Player contributes more	-0.08	-0.17	-0.08	-0.02	0.08	-0.05	-0.10	0.01
Steals cooperative (%) - Equal contribution	-0.20	-0.20	0.17	0.03	-0.03	-0.09	0.03	0.01
Steals cooperative (%) - Player contributes less	-0.15	-0.08	-0.08	-0.13	0.04	-0.11	0.00	0.06
Steals tit-for-two-tat (%) - Player contributes more	-0.06	-0.25	-0.09	-0.08	0.11	-0.01	0.01	0.09
Steals tit-for-two-tat (%) - Equal contribution	-0.13	-0.24	0.03	-0.07	0.08	0.02	-0.03	0.02
Steals tit-for-two-tat (%) - Player contributes less	-0.04	-0.20	0.00	-0.15	0.05	-0.04	0.06	0.12
Steals aggressive (%) - Player contributes more	-0.05	-0.18	0.07	-0.06	0.06	-0.03	0.06	0.12
Steals aggressive (%) - Equal contribution	-0.01	**-0.32****	0.03	0.08	0.07	-0.02	0.08	0.08
Steals aggressive (%) - Player contributes less	-0.07	-0.20	0.01	-0.02	0.10	0.01	0.12	0.13

Legend: Correlations between Prisoner's Dilemma task outcomes and age, sex, education indexed with the Family History Assessment Module (OS-FHAM) on a five-point likert scale, IQ score assessed with the Reynolds Intellectual Screening Test (RIST), total mood disturbance (TMD) indexed with the Profile of Mood Scale (POMS), and trait Neuroticism indexed with the NEO Personality Inventory-Revised (NEO PI-R, n = 93) and the NEO Personality Inventory 3 (NEO PI-3, n = 6). Correlations between self-reported motivation and diligence are also shown. Correlation coefficients are reported as Spearman's rho; only p-values < 0.01 are considered significant and are marked in bold. ** = $p < .01$. [$] A negative rho value indicates males score higher while a positive rho value indicates females score higher. [a] N = 99 due to missing data from one participant.

11. Ultimatum Game

Task description

The Ultimatum Game (UG) assesses sensitivity to fairness. At the beginning of each trial, the participant and a computer opponent complete a small task to collect money which is then pooled; a third of the time, the participant contributes more, a third of the time both parties contribute equally, and a third of the time the opponent contributes more. Some of the time, the participant decides how the money is split and other times the opponent decides how the money is split. The opponent's offer ranges in fairness (50% offer, 40% offer, 35% offer, 30% offer, 25% offer, 20% offer, 10% offer) and the participant can decide to accept the offer and get the proposed amount or reject the offer and get no money. The participant decides the split in 15 trials and the opponent decides the split in 36 trials, equaling a total of 51 trials.

Main outcomes (reported in the main article)

- *Proportion of accepted offers* – Proportion of trials (%) in which the participant accepts the proposed offer

Secondary outcomes

- *Proportion of accepted offers at different fairness levels* – Proportion of offers (%) accepted at different fairness levels (50%, 40%, 35%, 30%, 25%, 20%, 10%).
- *Average offer proposed* – The average offer proposed by the participant

Ultimatum Game – Descriptive data

	Mean	*SD*	*Median*	*IQR*	*Range*	*Skewness*	*Floor effect*	*Ceiling effect*
Acceptance rate (%) - 50% offer	*98.33*	*7.30*	100.00	0.00	66.67–100	-4.19	0%	95%
Acceptance rate (%) - 40% offer	*79.83*	*29.23*	100.00	33.33	0–100	-1.30	3%	57%
Acceptance rate (%) - 35% offer	*74.00*	*33.28*	91.67	50.00	0–100	-0.99	7%	50%
Acceptance rate (%) - 30% offer	*61.00*	*35.31*	66.67	66.67	0–100	-0.35	11%	31%
Acceptance rate (%) - 25% offer	*49.00*	*37.44*	41.67	66.67	0–100	0.11	19%	22%
Acceptance rate (%) - 20% offer	*39.67*	*39.91*	33.33	87.50	0–100	0.57	32%	25%
Acceptance rate (%) - 10% offer	*25.67*	*39.03*	0.00	33.33	0–100	1.16	64%	18%
Average offer proposed (%)	*37.93*	*9.42*	39.67	14.50	20–50	-0.47	4%	12%

Legend: Mean, SD, median, interquartile range, range, and skewness are shown. Note, for outcomes with non-parametric data distribution, mean and SD are denoted in cursive gray indicating the median and interquartile range should be used as reference. Floor and ceiling effects are shown as percentage of test subjects who achieved minimum (floor) or maximum (ceiling) scores. Note, for acceptance rate outcomes floor effect describes players who rejected all offers (0%) while ceiling effect describes players who accepted all offers (100%). For the average offer proposed outcome, floor effect describes players who made only the lowest offers (20%) while ceiling effects describes players who made only the highest offer (50%).

Ultimatum Game – Test-retest reliability

	Baseline (n = 49)		Retest (n = 49)		*Test-retest bias (%)*	*ICC*	*95% CI*
	Mean	*SD*	*Mean*	*SD*			
Acceptance rate (%) - 50% offer	98.64	6.66	98.64	6.66	0.00	0.65	0.38–0.80
Acceptance rate (%) - 40% offer	77.55	30.72	85.03	29.71	9.65	0.81	0.66–0.89
Acceptance rate (%) - 35% offer	74.15	34.70	81.63	31.41	10.09	0.86	0.74–0.92
Acceptance rate (%) - 30% offer	60.20	34.33	74.83	34.55	24.30	0.75	0.50–0.87
Acceptance rate (%) - 25% offer	47.62	35.19	67.69	36.07	42.15	0.68	0.32–0.84
Acceptance rate (%) - 20% offer	35.37	36.90	60.54	41.48	71.16	0.59	0.18–0.79
Acceptance rate (%) - 10% offer	20.41	34.57	35.37	43.78	73.30	0.64	0.35–0.79
Average offer proposed (%)	38.78	8.70	32.84	11.39	-15.32	0.72	0.32–0.87

Legend: Mean and SD are reported for baseline and retest sessions (3-5 weeks) along with Intraclass Correlation Coefficient (ICC) and their 95% Confidence Interval (95% CI). Test-retest bias is calculated as percentage change between first and second test: *Test-retest bias = ((retest-test)/test)*100.*

Ultimatum Game - Correlations

	Age	*Sex*[$]	*Education*	*IQ*	*Neuroticism*[a]	*TMD*	*Motivation*	*Diligence*
Acceptance rate (%) - 50% offer	-0.15	-0.05	-0.13	-0.12	0.06	0.04	0.25	0.03
Acceptance rate (%) - 40% offer	-0.20	0.03	0.19	-0.09	0.08	0.08	-0.02	-0.18
Acceptance rate (%) - 35% offer	-0.18	0.08	0.22	-0.11	0.14	0.16	0.08	-0.05
Acceptance rate (%) - 30% offer	-0.12	0.08	0.12	-0.02	0.07	0.13	-0.02	-0.19
Acceptance rate (%) - 25% offer	-0.15	0.04	0.16	-0.01	0.03	0.19	-0.04	-0.24
Acceptance rate (%) - 20% offer	-0.12	0.16	0.08	-0.10	0.01	0.19	-0.03	-0.21
Acceptance rate (%) - 10% offer	-0.12	0.04	0.08	-0.09	0.01	0.13	-0.10	**-0.27****
Average offer proposed (%)	0.23	0.04	-0.04	0.20	-0.11	-0.06	0.00	0.18

Legend: Correlations between Prisoner's Dilemma task outcomes and age, sex, education indexed with the Family History Assessment Module (OS-FHAM) on a five-point likert scale, IQ score assessed with the Reynolds Intellectual Screening Test (RIST), total mood disturbance (TMD) indexed with the Profile of Mood Scale (POMS), and trait Neuroticism indexed with the NEO Personality Inventory-Revised (NEO PI-R, n = 93) and the NEO Personality Inventory 3 (NEO PI-3, n = 6). Correlations between self-reported motivation and diligence are also shown. Correlation coefficients are reported as Spearman's rho; only *p*-values < .01 are considered significant and are marked in bold. ** = $p < .01$. [$] A negative rho value indicates males score higher while a positive rho value indicates females score higher. [a] N = 99 due to missing data from one participant.

Exploratory Factor Analysis

Table A shows the factors and item loadings of an exploratory factor analysis. The analysis was conducted using principal axis factoring with Varimax rotation; only items with loadings > 0.3 are reported. Measures of sampling adequacy were acceptable: the Kaiser-Meyer-Olkin value was sufficient (KMO = 0.53) and Bartlett's test of sphericity was significant ($p < .001$) indicating that the data was suitable for structure detection. Using an eigen-value cutoff of 1, the factor analysis suggested 13 factors (see Figure A for scree plot).

Table A. Factor loadings for EMOTICOM tasks on factors 1-13

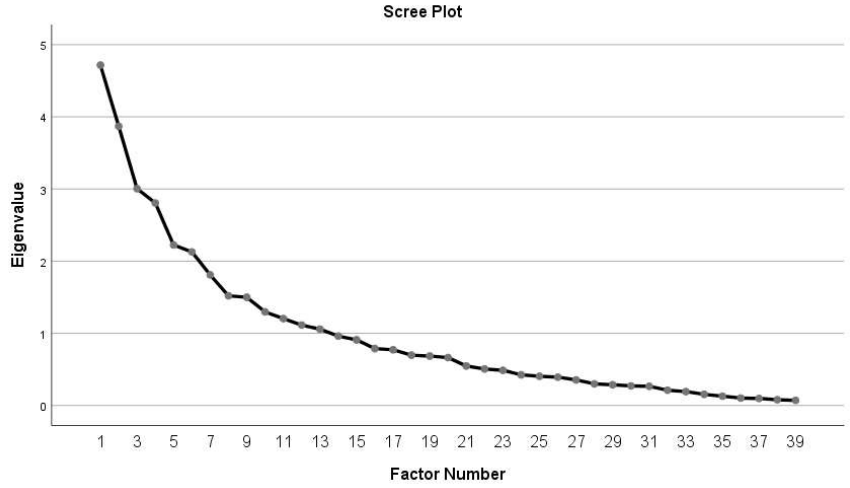
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*
Emotional Intensity Morphing Task: Increase condition													
Detection threshold - Happy	0.62					0.57							
Detection threshold - Sad	0.76												
Detection threshold - Angry	0.79												
Detection threshold - Fearful	0.83												
Detection threshold - Disgusted	0.85												
Emotional Intensity Morphing Task: Decrease condition													
Detection threshold - Happy		0.42						-0.57					-0.31
Detection threshold - Sad		0.64											
Detection threshold - Angry		0.68											
Detection threshold - Fearful		0.71											
Detection threshold - Disgusted		0.80											
Prisoner's Dilemma													
Proportion steals (%) - Cooperative			0.77										
Proportion steals (%) - Tit-for-two-tat			0.84										
Proportion steals (%) - Aggressive			0.86										
Affective Go/NoGo													

d-prime - Happy/Neutral				0.33									
d-prime - Happy/Sad				0.49									
d-prime - Neutral/Happy				0.57									
d-prime - Neutral/Sad				0.63									0.31
d-prime - Sad/Happy				0.64									
d-prime - Sad/Neutral				0.39									
Moral Emotions Task													
Guilt - Agent					0.88								
Guilt - Victim									0.80				
Shame - Agent					0.84								
Shame - Victim									0.83				
Emotional Face Recognition Task: Face version								0.61					
Accuracy (%) - Happy													
Accuracy (%) - Sad						0.31				0.41			
Accuracy (%) - Angry										0.72			
Accuracy (%) - Fearful						0.40				0.51			
Emotional Face Recognition Task: Eyes version													
Accuracy (%) - Happy								0.73					
Accuracy (%) - Sad						0.70							
Accuracy (%) - Angry										0.40			
Accuracy (%) - Fearful						0.57				0.33			
Social Information Preference Task													
Information (%) - Thoughts							-0.88						
Information (%) - Faces							0.77						
Information (%) – Facts*													
Reinforcement Learning Task[a]													
Alpha - Win condition													
Alpha - Loss condition													
Monetary Incentive Reward Task													
Reaction time (ms) - Win											-0.32		0.34

Reaction time (ms) - Loss												0.73		
Progressive Ratio Task														
Breakpoint (trials)														
Adapted Cambridge Gambling Task														
Risk adjustment - Win condition														
Risk adjustment - Loss condition											0.83			
Ultimatum Game														
Average acceptance rate (%)											0.36		0.30	

* We removed one of the three outcomes ('Facts') from the *Social Information Preference task* from the analysis as it is redundant; the sum of the three outcomes always equals 100%, meaning the value of one can always be inferred from the other two.

Figure A. Scree plot



Psychological Medicine

cambridge.org/psm

Hot and cold cognitive disturbances in antidepressant-free patients with major depressive disorder: a NeuroPharm study

V. H. Dam[1,2], D. S. Stenbæk[1], K. Köhler-Forsberg[1,2,3], C. Ip[1,2,4], B. Ozenne[1,5], B. J. Sahakian[6,7], G. M. Knudsen[1,2], M. B. Jørgensen[3] and V. G. Frokjaer[1,3]

[1]Neurobiology Research Unit, the Neuroscience Centre, Copenhagen University Hospital Rigshospitalet, Denmark; [2]Faculty of Health and Medical Sciences, University of Copenhagen, Denmark; [3]Psychiatric Center Copenhagen, Copenhagen University Hospital Rigshospitalet, Denmark; [4]Department of Clinical Pharmacology, H. Lundbeck A/S, Valby, Denmark; [5]Department of Public Health, Section of Biostatistics, University of Copenhagen, Denmark; [6]Department of Psychiatry, University of Cambridge, Cambridge, UK and [7]Behavioral and Clinical Neuroscience Institute, University of Cambridge, Cambridge, UK

Original Article

Cite this article: Dam VH, Stenbæk DS, Köhler-Forsberg K, Ip C, Ozenne B, Sahakian BJ, Knudsen GM, Jørgensen MB, Frokjaer VG (2020). Hot and cold cognitive disturbances in antidepressant-free patients with major depressive disorder: a NeuroPharm study. *Psychological Medicine* 1–10. https://doi.org/10.1017/S0033291720000938

Received: 17 September 2019
Revised: 5 February 2020
Accepted: 24 March 2020

Key words:
Affective biases; cognitive profiles; cold cognition; EMOTICOM; hot cognition; major depressive disorder; social cognition

Author for correspondence:
Vibe G. Frokjaer, E-mail: Vibe.Frokjaer@nru.dk

Abstract

Background. Cognitive disturbances are common and disabling features of major depressive disorder (MDD). Previous studies provide limited insight into the co-occurrence of hot (emotion-dependent) and cold (emotion-independent) cognitive disturbances in MDD. Therefore, we here map both hot and cold cognition in depressed patients compared to healthy individuals.
Methods. We collected neuropsychological data from 92 antidepressant-free MDD patients and 103 healthy controls. All participants completed a comprehensive neuropsychological test battery assessing hot cognition including emotion processing, affective verbal memory and social cognition as well as cold cognition including verbal and working memory and reaction time.
Results. The depressed patients showed small to moderate negative affective biases on emotion processing outcomes, moderate increases in ratings of guilt and shame and moderate deficits in verbal and working memory as well as moderately slowed reaction time compared to healthy controls. We observed no correlations between individual cognitive tasks and depression severity in the depressed patients. Lastly, an exploratory cluster analysis suggested the presence of three cognitive profiles in MDD: one characterised predominantly by disturbed hot cognitive functions, one characterised predominantly by disturbed cold cognitive functions and one characterised by global impairment across all cognitive domains. Notably, the three cognitive profiles differed in depression severity.
Conclusion. We identified a pattern of small to moderate disturbances in both hot and cold cognition in MDD. While none of the individual cognitive outcomes mapped onto depression severity, cognitive profile clusters did. Overall cognition-based stratification tools may be useful in precision medicine approaches to MDD.

Introduction

Disturbance of cognitive functioning is a common feature of major depressive disorder (MDD) and has been proposed as an important treatment target (Collins et al., 2011). Cognitive symptoms including inability to concentrate or difficulty making decisions are listed among the diagnostic criteria for MDD (APA, 2013; WHO, 2007). Investigations into cognitive disturbances in MDD have typically focused on either so-called 'hot' or 'cold' cognitive functions (Roiser & Sahakian, 2013). Hot cognition describes mental functions that involve the processing of emotionally salient information (e.g. identifying emotional facial expressions) or emotional responses (e.g. reward-driven behaviours). In particular negative affective biases, i.e. the subconscious allocation of more attention and mental resources to the processing of negative information over positive information, have been associated with MDD psychopathology (Elliott, Zahn, Deakin, & Anderson, 2011; Miskowiak & Carvalho, 2014) and may play a key role in the onset and maintenance of depressive symptoms (Roiser, Elliott, & Sahakian, 2012). Another hot cognitive domain which may be impaired in MDD is social cognition which includes functions such as seeing oneself in the 'other', i.e. Theory of Mind (Bora & Berk, 2015; Wolkenstein, Schonenberg, Schirm, & Hautzinger, 2011), interpretation of social situations and excessive experiences of negative social emotions such as shame and guilt (Kim et al., 2015).

Cold cognition describes mental processes that include emotionally neutral information and do not directly involve activation of emotional states (Roiser & Sahakian, 2013). Recent meta-analyses suggest that cold cognitive deficits in MDD are predominantly found in domains of attention, learning and memory and executive functions (Goodall et al., 2018; Lee, Hermens, Porter, & Redoblado-Hodge, 2012; Rock, Roiser, Riedel, & Blackwell, 2014). Slowed reaction time

has also been reported for depressed patients and is, along with agitated psychomotor function, considered a distinct symptom in MDD (Bennabi, Vandel, Papaxanthis, Pozzo, & Haffen, 2013).

The effects sizes reported for cognitive disturbances in MDD are typically small to moderate which is relatively modest compared to those reported for other serious neuropsychiatric disorders such as Alzheimer's Disease and schizophrenia (Maruff & Jaeger, 2016). Nevertheless, their impact on daily life may be very disruptive. Both hot and cold cognitive disturbances in MDD have been found to be detrimental to the patient's ability to engage successfully in work or educational activities as well as overall psychosocial functioning (Cambridge, Knight, Mills, & Baune, 2018; Weightman, Knight, & Baune, 2019). This is especially relevant as cognitive disturbances do not always resolve with the remission of core depressive symptoms (Hernaus, Gold, Waltz, & Frank, 2018).

Despite a growing body of data on specific cognitive deficits in MDD, we currently know little about the co-occurrence and magnitude of impairments across different types of cognitive domains. Few studies on MDD have included both hot and cold cognitive tasks and comparisons between studies are often hampered by differences in cohort characteristics such as medication/treatment status, comorbidity, age-range, chronicity and severity of current depressive episode. To address this, we therefore applied a broad range of both hot and cold cognitive tasks in a large cohort of well-characterised and antidepressant-free depressed patients.

Methods

Participants and study design

One hundred non-psychotic antidepressant-free patients suffering from a moderate to severe depressive episode lasting less than two years [Hamilton Depression Rating Scale-17 ($HDRS_{17}$) ⩾18] were included in a large multimodal neuroimaging clinical trial (NeuroPharm 1). Patients were eligible for inclusion if they had been antidepressant free for >2 months; had not previously exhibited non-response to SSRIs; and had not undergone more than one antidepressant treatment attempt in the current depressive episode. Patients were recruited through their primary care centre or a central referral site for 'depression treatment packages' at the Mental Health Services of the Capital Region of Copenhagen. MDD diagnosis was confirmed by a certified psychiatrist and corroborated by a Mini-International Neuropsychiatric Interview (MINI). Out of the 100 patients who entered the study, neuropsychological data was available from 92 patients (67 females); out of these, 41 patients had first-episode depression while 51 patients had recurrent depression. In addition, data from 100 healthy participants were collected as part of a validation study of the EMOTICOM test battery, a novel neuropsychological battery specifically designed to assess hot cognitive functions (Dam et al., 2019), and additionally three healthy controls were recruited via internet advertisements and flyers posted around the greater Copenhagen area (52 females). Exclusion criteria for the study were history of psychiatric disorders for healthy controls and prior or present history of other primary axis I psychiatric disorders for depressed patients; significant somatic illness, brain trauma; use of psychotropic medication within 4 weeks of inclusion; significant lifetime history of drug abuse and pregnancy or breastfeeding. Neuropsychological testing was conducted by trained testers in standardised test rooms. Hamilton Depression Rating Scale-6 ($HDRS_6$), a subscale of the $HDRS_{17}$ that indexes core MDD symptom, was chosen as the primary clinical outcome with $HDRS_{17}$ as a secondary clinical outcome.

The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation (protocol: H- 15017713) and with the Helsinki Declaration of 1975, as revised in 2008. The present study is based on baseline data from a longitudinal clinical trial registered at https://www.clinicaltrial.gov (protocol: NCT02869035).

Hot cognition

Affective biases

We used to two tasks from the EMOTICOM test battery to asses biases in emotion processing: the eyes version of the *Emotional* Recognition *Task* (ERT) was used to index biases in basic emotion recognition (i.e. hit rates) as well as misattribution (i.e. false alarm rates) and the *Intensity Morphing* (IM) task was used to assess biases for perceptual detection threshold of emotions in facial expressions (Dam et al., 2019). A modified version of the Verbal Affective Memory Task 24 (Jensen et al., 2016), the *Verbal Affective Memory Task 26* (VAMT-26), was used to assess affective memory biases. Biases were calculated by subtracting negative information scores from positive information scores (e.g. hit rate for recognition of happy faces minus hit rate for recognition of sad faces).

Social cognition

We used two tasks from the EMOTICOM test battery to assess social cognition: the *Moral Emotions* (ME) task was used to index moral emotions (guilt and shame) in social situations and the *Social Information Preference* (SIP) task was used to assess preference for social information over non-social information and bias in interpretation of social situations.

Cold cognition

We used three tasks to assess cold cognition: total word recall in the VAMT-26 was used to assess explicit non-affective verbal memory function; the *Letter* Number *Sequence* (LNS) task was used to assess working memory capacity; and the *Simple Reaction Time* (SRT) task was used to assess reaction time.

Note, a full description of all task including both primary and secondary outcomes can be found in online Supplementary Materials.

Statistical analysis

Group differences

Differences on cognitive performance between depressed patients and healthy controls were assessed with linear regression models with primary cognitive outcome as the dependent variable and age, sex and group coded as a categorical variable as independent variables. The reported p values were corrected for 11 tests using the Bonferroni–Holm method. The standardised effect size was estimated for each primary cognitive outcome by computing the Cohen's d on the partial residuals relative to the group variable (i.e. after removing the age and sex effects from the cognitive outcome). Normality assumptions were assessed and found to be acceptable for all models except the SRT task outcome. *Post hoc* linear regression analyses were used to investigate secondary outcomes for tasks showing statistically significant group differences. The reported p values for the *post hoc* analyses were adjusted to control the family-wise error rate within each task: first the

Bonferroni–Holm method was applied (e.g. to eight tests for the ERT secondary outcomes). If the resulting p value was smaller than the adjusted p value for the primary task outcome, then the adjusted p value of the primary task outcome was used for this secondary outcome. This was done to ensure that the adjusted p values were coherent between the main and secondary outcomes when using significance threshold below 0.05.

Correlation with symptom severity

In MDD patients, Spearman's rank correlation coefficient was used to assess the relationship between depressive symptom severity assessed with $HDRS_6$ and $HDRS_{17}$ and performance on primary cognitive outcomes. The reported p values were corrected for 11 tests using the Bonferroni–Holm method.

Clustering of cognitive profiles

In an exploratory *post hoc* analysis, we used a K-means cluster analysis to delineate potential groups of cognitive profiles within the depressed cohort (Clatworthy, Buick, Hankins, Weinman, & Horne, 2005). The input into the cluster analysis was restricted to the primary cognitive outcomes found to characterise the depressed state, i.e. those outcomes that differed with statistical significance between the depressed patients and healthy controls. The cognitive outcomes were standardised to z-scores and the number of clusters and the position used to initialise the K-means algorithm were obtained using a hierarchical clustering algorithm (see online Supplementary Material) (Milligan, 1980). The groups derived from the K-means analysis were compared on clinical and demographic factors using Analysis of Variance (ANOVA) for continuous outcomes and the χ^2 test for binary outcomes (i.e. sex).

Outliers and missing data

Outliers were defined as observations 1.5 Interquartile Range (IQR) above or below the 1st quartile or 3rd quartile respectively. Each outlier was qualitatively evaluated based on notes from the testing session and congruency with scores on outcomes from the same task as well as outcomes from similar tasks. In total, 38 outlying data points were detected (representing 1.8% of all observations across the 11 primary cognitive outcomes); 37 were deemed to be 'true outliers' and kept in the analysis while one patient outlier from the LNS task was excluded as the patient had misunderstood the test instructions. Importantly, we found that none of the reported estimates changed critically when all outliers were removed. Note, in the SRT task two patient scores were so extreme (8.1 and 16.1 IQR above the 1st quartile respectively) that independent of their potential neurobiological meaningfulness they were capped to one and two units above the third highest score respectively; this allowed the data to be included without losing their rank or impacting the group estimates unduly. Missing data included IM data from one patient, VAMT-26 data from six patients, SIP data from three controls, LNS data from nine patients and one control and SRT data from one patient and 37 controls.

Results

Descriptive

While there was no significant age difference between the two groups, the proportion of females was significantly higher in the patient group (73.9%) compared to the control group (50.4%) (Table 1). This reflects the well-documented higher prevalence of depression in females in the general population; although notably the proportion of females in the present depressed sample was higher than the ~60% reported for European countries in a recent report by the World Health Organization (WHO, 2017). In accordance with the inclusion criteria, all depressed patients had a $HDRS_{17}$ score above 17, indicating moderate to severe depression. IQ was assessed with the Reynolds Intellectual Screening Test (Reynolds, 2011) and all study participants scored within the normal range (see Supplementary Materials). Cognitive performance did not differ between patients with first-episode ($N = 41$) and recurrent depression ($N = 51$) on any task outcome (all $p_{corrected} > 0.75$; see online Supplementary Materials for a full overview).

Table 1. Descriptive data

	Depressed patients ($n = 92$)	Healthy controls ($n = 103$)	p value
Age in years	27.3 ± 8.1 (18–57)	28.7 ± 7.3 (18–48)	0.19
Male/female	25/68	51/52	<0.001
MDI	34.5 ± 7.2 (16–50)[a]	4.9 ± 3.9 (0–20)[b]	<0.001
$HDRS_6$	12.4 ± 1.6 (7–17)	–	–
$HDRS_{17}$	22.8 ± 3.4 (18–31)	–	–

Age, sex and self-rated depressive symptoms indexed with the Major Depressive Inventory (MDI) are reported for both depressed patients and healthy controls.For depressed patients, clinically rated depressive symptoms indexed with the Hamilton Depression Rating Scale 6 and 17 ($HDRS_6$ and $HDRS_{17}$) are also reported. Values are presented as mean ± SD with range in brackets. Group differences were assessed with an independent t test for age; χ^2 test for sex; and Mann–Whitney U test for MDI.
[a]$N = 90$ due to missing questionnaire data.
[b]$N = 102$ due to missing questionnaire data.

Group differences

Hot cognition: In the ERT task, the affective bias expressed by the depressed patients was 11.1 percentage points more negative for recognition rates ($p_{corrected} = 0.03$) and 7.8 percentage points more negative for misattribution rates ($p_{corrected} = 0.02$) compared to healthy controls. Likewise, the affective bias for emotion detection threshold in the IM task was 7.9 percentage points more negative for the depressed patients ($p_{corrected} < 0.001$) while no substantial difference in bias was observed for affective verbal memory in the VAMT-26 ($p_{corrected} = 1.00$). When asked to identify with cartoon characters in negative social situations in the ME task, the depressed patients also reported stronger experiences of negative moral emotions equivalent to 0.5 points on a seven-point Likert scale for both guilt ($p_{corrected} < 0.001$) and shame ($p_{corrected} < 0.001$) compared to healthy controls. We observed no substantial group differences in choice of social information ($p_{corrected} = 1.00$) nor bias in interpretation of social situations ($p_{corrected} = 1.00$) in the SIP task.

Cold cognition: The depressed patients recalled a total of 2.6 fewer words in the VAMT-26 ($p_{corrected} < 0.001$) independent on affective valence; successfully sorted 1.7 fewer sequences on the working memory task (LNS, $p_{corrected} = 0.002$); and exhibited 30.7 ms slower reaction time (SRT, $p_{corrected} = 0.006$) compared to the healthy controls. Note, as model assumptions for normality were not met for the SRT outcome, we used bootstrapping to determine the reported p value. In addition, we also conducted a quantile regression analysis to assess whether the reported results were robust to outliers and found a similar effect (estimated group effect = 20.7 ms, p value = 0.003) (Fig. 1).

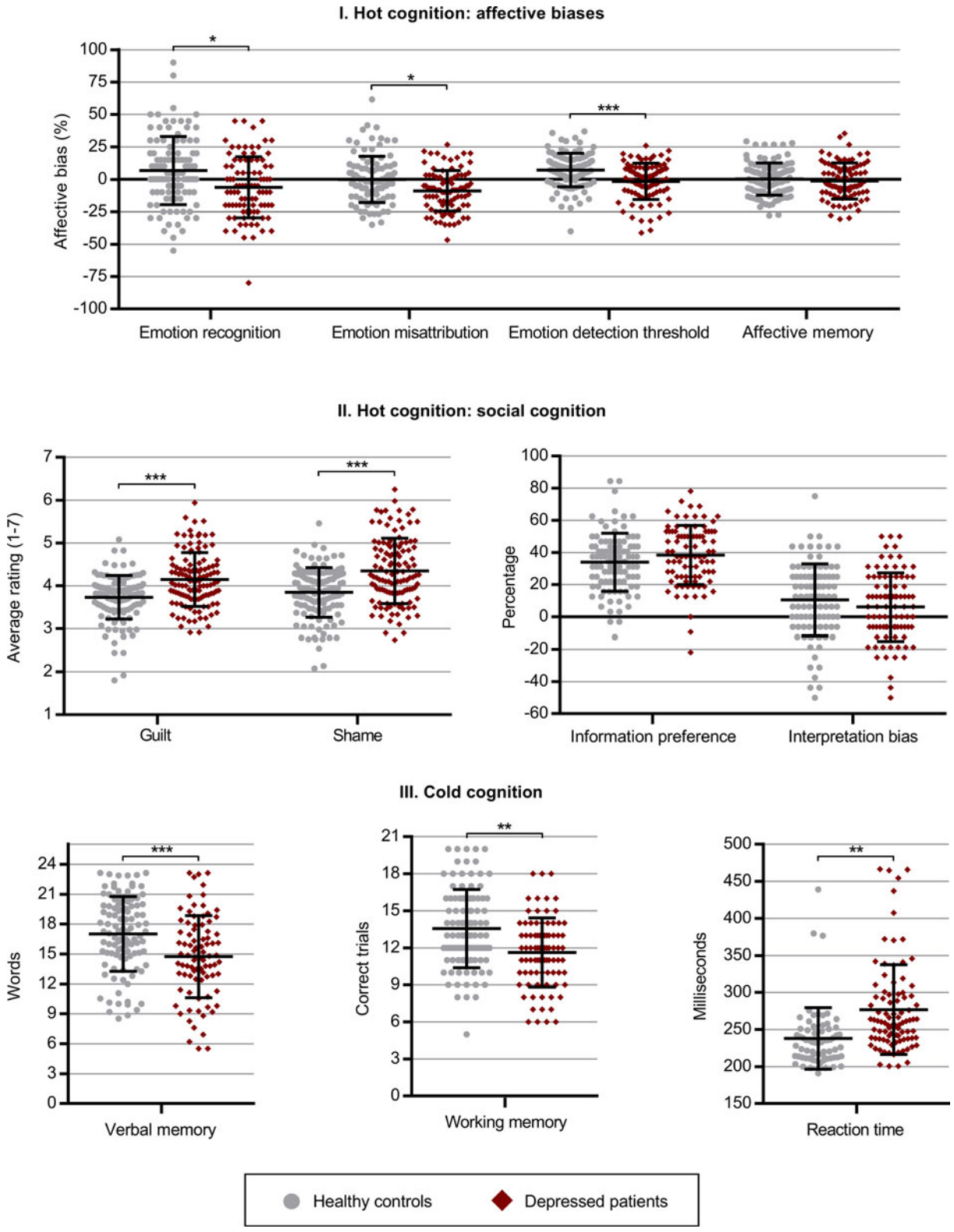


Fig. 1. Group differences on affective, social and cold cognitive outcomes between depressed patients and healthy controls. (I) Affective cognition: Recognition = affective bias for hit rate in the Emotional Recognition Task (patients $n = 92$, controls $n = 103$); Misattribution = affective bias for false alarm rate in the Emotional Recognition Task (patients $n = 92$, controls $n = 103$); Detection threshold = affective bias for the Intensity Morphing Task (patients $n = 91$, controls $n = 103$); Affective memory = affective bias for the Verbal Affective Memory Task 26 (patients $n = 86$, controls $n = 103$). (II) Social cognition: Guilt = average ratings of guilt in the Moral Emotions task (patients $n = 91$, controls $n = 103$); Shame = average ratings of shame in the Moral Emotions task (patients $n = 91$, controls $n = 103$); Information preference = choice of theory of mind-related information relative to facts in the Social Information Preference task (patients $n = 89$, controls $n = 100$); Interpretation bias = affective bias in choice of outcome in the Social Information Preference task (patients $n = 89$, controls $n = 100$). (III) Cold cognition: Verbal memory = Total recall score for the Verbal Affective Memory Task (patients $n = 85$, controls $n = 103$); Working memory = Letter-Number Sequence task (patients $n = 83$, controls $n = 103$); Reaction time = Simple Reaction Time (patients $n = 91$, controls $n = 66$). All models were corrected for age and sex. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

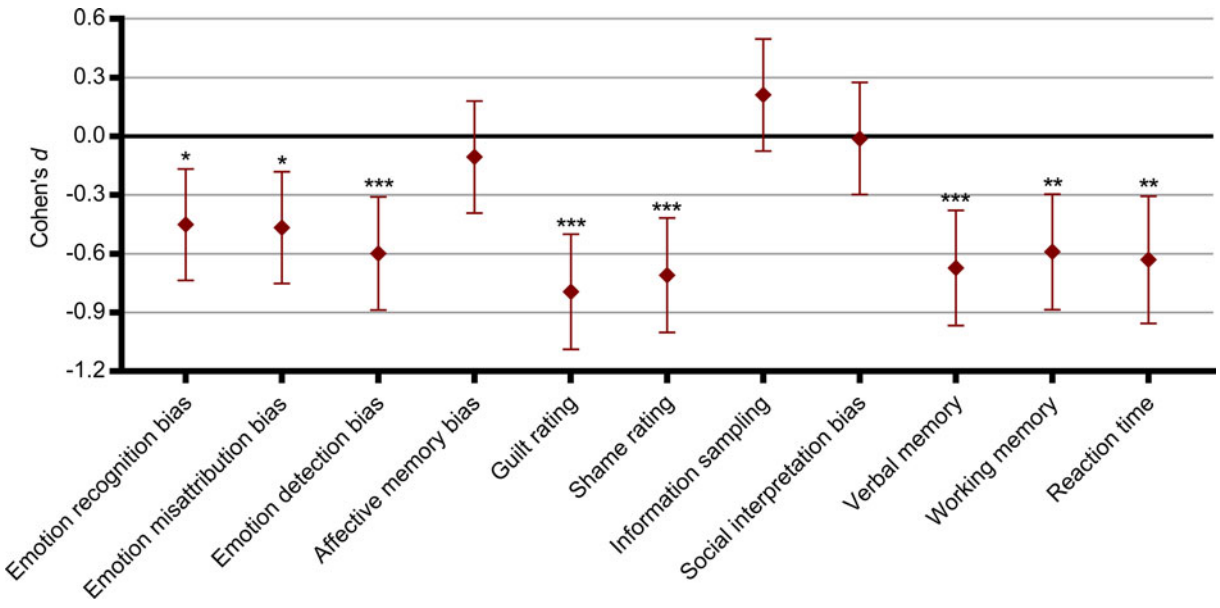


Fig. 2. Summary of differences in performance across cognitive domains for depressed patients relative to healthy controls. Zero represents the healthy control group and differences are expressed as Cohen's *d* effect sizes. Error bars denote 95% confidence intervals (95% CI). Recognition bias = affective bias for hit rate in the Emotional Recognition Task; Misattribution bias = affective bias for false alarm rate in the Emotional Recognition Task; Detection bias = affective bias for the Intensity Morphing task; Affective memory bias = affective bias for the Verbal Affective Memory Task 26; Guilt rating = average guilt rating from the Moral Emotions task; Shame rating = average shame rating from the Moral Emotions task; Information sampling = choice of theory of mind-related information relative to facts in the Social Information Preference task; Interpretation bias = affective bias in choice of outcome in the Social Information Preference task; Verbal memory = total recall from the Verbal Affective Memory Task 26; Working memory = Letter-Number Sequence task; Reaction time = Simple Reaction Time task. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Compared to healthy controls, MDD patients were more likely to incorrectly identify other emotions as sadness in the ERT task and continued to perceive sadness at lower intensity levels (decrease condition) in the IM task. In the ME task, patients also reported higher levels of guilt and shame in social scenarios where they identified with characters who accidentally harmed another person and increased guilt and scenarios where they identified with the victim of either accidental or intentional harm. Lastly, patients exhibited deficits for both immediate, short-term and long-term non-affective verbal memory compared to controls in the VAMT-26 (Fig. 2).

Correlation with depression severity

Correlations between cognitive performance and clinically rated depressive symptoms within the depressed group ranged from weak to negligible on all tasks and were statistically non-significant [$HDRS_6$, ρ (−0.2; 0.2), all $p_{corrected} > 0.42$; $HDRS_{17}$, ρ (−0.2; 0.2), all $p_{corrected} > 0.44$] (see online Supplementary Materials for a full overview).

Clustering of cognitive profiles

Based on the eight cognitive outcomes which showed a significant group difference, an initial hierarchical cluster analysis was run that indicated a three-cluster solution for cognitive profiles within the depressed group. The clustering centroids from the hierarchical cluster analysis were subsequently used to initialise a K-means analysis that converged within six iterations (Fig. 3).

There were no statistically significant differences between the clusters on age ($p = 0.58$) or sex ($p = 0.71$). The three clusters differed significantly on severity of core depressive symptoms indexed with $HDRS_6$ [$F(2, 90) = 4.1$, $p = 0.02$] with patients from Cluster C (13.1 ± 1.9, mean ± S.D.) having higher scores than Cluster A (12.2 ± 1.6, mean ± S.D.) and B (11.9 ± 1.3, mean ± S.D.). The same pattern was present in severity of broad depressive symptoms indexed with $HDRS_{17}$ (Cluster A, $HDRS_{17}$ scores = 22.1 ± 3.2, mean ± S.D.; Cluster B, $HDRS_{17}$ scores = 23.0 ± 3.1, mean ± S.D.; Cluster C, $HDRS_{17}$ scores = 23.6 ± 3.8, mean ± S.D.), but did not reach statistical significance ($p = 0.19$).

Discussion

We here map the presence and magnitude of both hot and cold cognitive disturbances in a large cohort of antidepressant-free patients with a moderate to severe depressive episode. We found small to moderately sized negative biases in emotion processing but not in explicit verbal memory, large increases in experience of negative social emotions but no detectable differences in preference between social and non-social information or interpretation of ambiguous social situations. We also observed moderate impairment of cold cognitive functions including working and verbal memory and moderate slowing of reaction time. We found no direct link between depressive symptom severity and patient performance on any of the single task domains. Using an exploratory and data-driven approach, we identified three clusters with distinct cognitive profiles within the cohort of depressed patients: Cluster A was characterised by disturbances in hot but not cold cognition; Cluster B was characterised by positive biases and moderate deficits in cold cognitive domains; and Cluster C was characterised by large deficits across both hot and cold cognitive domains including extreme scores of guilt and shame.

Affective biases

As expected, the depressed patients exhibited negative affective biases across all emotion processing outcomes including

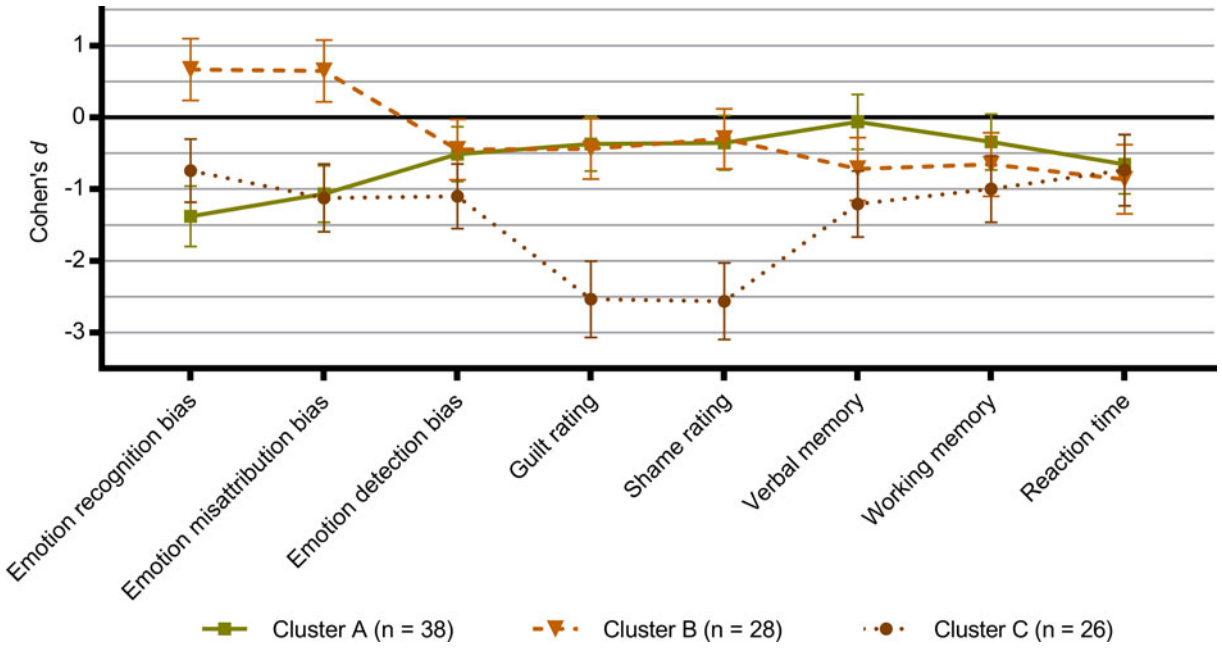


Fig. 3. Clusters of cognitive profiles within the cohort of depressed patients ($N = 92$) based on the eight cognitive outcomes that showed a significant group difference between depressed patients and healthy controls. Zero represents the healthy control group and differences are expressed as Cohen's *d* effect sizes. Error bars denote 95% confidence intervals (95% CI). Recognition bias = affective bias for hit rate in the Emotional Recognition Task; False alarm bias = affective bias for false alarm rate in the Emotional Recognition Task; Detection bias = affective bias for the Intensity Morphing task; Affective memory bias = affective bias for the Verbal Affective Memory Task 26; Guilt rating = average guilt rating from the Moral Emotions task; Shame rating = average shame rating from the Moral Emotions task; Information sampling = choice of theory of mind-related information relative to facts in the Social Information Preference task; Interpretation bias = affective bias in choice of outcome in the Social Information Preference task; Verbal memory = total recall from the Verbal Affective Memory Task 26; Working memory = Letter-Number Sequence task; Reaction time = Simple Reaction Time task.

recognition, misattribution and perceptual detection threshold. While abnormal processing of facial expressions is well-established in the MDD literature, the underlying cognitive mechanisms are still unclear (Elliott et al., 2011). Indeed, the findings from our study emphasise that the negative biases exhibited by the depressed patients must be understood *relative* to healthy controls. For example, while the depressed patients exhibited a clear negative affective bias in emotion misattribution, the negative bias observed for emotion detection threshold predominantly reflected the loss of positive bias exhibited by the healthy controls (see Fig. 1). Another line of research from attention paradigms suggests that affective bias in emotion processing is related to reduced orientation towards positive stimuli combined with an inability to disengage from negative stimuli (Armstrong & Olatunji, 2012). This aligns with our findings from the IM task as patients continued to perceive sadness at much lower intensity levels in the decrease condition compared to healthy controls.

Notably, we did not observe any negative affective bias in verbal memory performance. The concept of a mood-congruent memory bias was first proposed by Bower (1981) and posits that individuals will remember information that matches their current emotional state better than information that is not mood-congruent. Although this theory is relatively well supported by empirical studies investigating autobiographical (Köhler et al., 2015) and implicit (Gaddy & Ingram, 2014) types of memory, evidence for a bias in explicit, non-self-referential memory remains inconclusive. In fact, only a handful of studies have specifically investigated this type of memory (typically using valanced word-lists) in depression with some reporting a negative bias (Bradley, Mogg, & Williams, 1995; Neshat-Doost, Taghavi, Moradi, Yule, & Dalgleish, 1998; Watkins, Mathews, Williamson, & Fuller, 1992) while others suggest a positive bias (Calev, 1996; Danion, Kauffmann-Muller, Grange, Zimmermann, & Greth, 1995; Zupan, Žeželj, & Andjelković, 2017). In contrast to word memory, there have been attentional biases to negative words reported for unmedicated depressed subject (Beavers et al., 2013). The majority of memory studies had very small sample sizes and used different cognitive tasks and different depression criteria, which may contribute to the lack of consensus. We here present data from one of the largest study populations to date of well-characterised depressed patients which suggest that MDD symptomatology is not related to mood-congruent memory bias in explicit non-self-referential affective memory. Future studies should therefore consider using cognitive tasks assessing autobiographical and implicit memory as they may be more sensitive to affective memory disturbances in MDD.

Social cognition

Feelings of excessive shame and guilt are common in MDD and critically contribute to low self-esteem and social withdrawal (Mills et al., 2015). In the most severe cases they can even reach the threshold of psychosis (Lake, 2008). In particular contextual-maladaptive guilt (i.e. exaggerated guilt related to uncontrollable events) and generalised guilt (i.e. guilt divorced from concrete contexts) as well as external shame (i.e. shame based on beliefs about other people's opinions) are strongly associated with depressive symptoms (Kim, Thibodeau, & Jorgensen, 2011). Notably, in the ME task we observed most group differences between the depressed patients and the healthy controls in moral scenarios where the participants were asked to identify

Table 2. Group differences between depressed patients and healthy controls on secondary cognitive outcomes

	Depressed patients		Healthy controls				
	Mean ± SD	Range	Mean ± SD	Range	*B*	*p* value	$p_{corrected}$
Emotion recognition task							
Recognition – Happy	72.4 ± 17.8	15.0–100.0	78.0 ± 16.7	20.0–100.0	−5.0	5.25×10^{-2}	0.21
Recognition – Sad	78.6 ± 16.6	30.0–100.0	71.3 ± 18.9	10.0–100.0	6.2	2.01×10^{-2}	0.14
Recognition – Angry	62.5 ± 12.8	30.0–95.0	66.0 ± 11.4	40.0–90.0	−3.7	4.15×10^{-2}	0.21
Recognition – Fearful	73.5 ± 11.7	35.0–95.0	75.4 ± 14.8	5.0–100.0	−2.4	2.18×10^{-1}	0.44
Misattribution – Happy	10.9 ± 8.4	0.0–36.7	15.0 ± 12.0	0.0–63.3	−3.5	2.37×10^{-2}	0.14
Misattribution – Sad	19.8 ± 9.3	0.0–48.3	15.1 ± 8.6	0.0–41.7	4.3	1.52×10^{-3}	0.03
Misattribution – Angry	3.0 ± 3.7	0.0–18.3	3.7 ± 5.8	0.0–50.0	−0.5	4.58×10^{-1}	0.46
Misattribution – Fearful	3.9 ± 6.0	0.0–28.3	2.6 ± 3.8	0.0–26.7	1.4	5.13×10^{-2}	0.21
Intensity morphing task[a]							
Increase – Happy	51.1 ± 16.1	23.2–92.9	46.9 ± 14.9	12.5–88.1	4.5	5.09×10^{-2}	0.31
Increase – Sad	53.9 ± 14.0	23.2–82.1	60.2 ± 15.2	17.9–89.3	−5.0	2.08×10^{-2}	0.15
Increase – Angry	52.8 ± 13.2	21.4–85.7	55.6 ± 15.6	17.9–92.9	−1.9	3.68×10^{-1}	0.82
Increase – Fearful	56.1 ± 16.2	21.4–100.0	61.0 ± 16.7	21.4–100.0	−2.9	2.31×10^{-1}	0.82
Increase – Disgusted	55.5 ± 15.3	19.6–92.9	57.2 ± 14.7	17.9–89.3	0.06	9.80×10^{-1}	0.98
Decrease – Happy	29.0 ± 13.2	2.4–66.1	30.9 ± 16.2	0.0–75.0	−2.8	2.06×10^{-1}	0.82
Decrease – Sad	22.9 ± 12.2	0.0–69.6	31.9 ± 12.3	5.4–66.1	−9.2	1.00×10^{-6}	<0.001
Decrease – Angry	21.9 ± 13.5	0.0–100.0	25.5 ± 12.2	3.6–62.5	−3.5	7.29×10^{-2}	0.36
Decrease – Fearful	24.4 ± 13.6	0.0–92.9	29.7 ± 11.4	0.0–66.1	−5.0	7.37×10^{-3}	0.06
Decrease – Disgusted	17.2 ± 9.9	0.0–44.6	22.0 ± 12.4	0.0–67.9	−4.7	6.29×10^{-3}	0.06
Moral emotions task							
Agent/intentional – Guilt	6.3 ± 0.6	4.2–7.0	6.1 ± 0.8	2.5–7.0	0.1	2.00×10^{-1}	0.40
Agent/accident – Guilt	6.1 ± 0.8	3.6–7.0	5.6 ± 0.8	2.0–7.0	0.4	4.53×10^{-4}	0.002
Victim/intentional – Guilt	2.5 ± 1.1	1.0–5.3	1.7 ± 0.7	1.0–3.8	0.8	1.67×10^{-8}	<0.001
Victim/accident – Guilt	2.0 ± 0.9	1.0–5.0	1.5 ± 0.6	1.0–4.4	0.5	8.00×10^{-6}	<0.001
Agent/intentional – Shame	6.1 ± 0.8	4.3–7.0	6.0 ± 0.9	2.5–7.0	0.03	8.00×10^{-1}	0.80
Agent/accident – Shame	5.9 ± 0.8	3.6–7.0	5.5 ± 0.9	2.0–7.0	0.3	7.68×10^{-3}	0.02
Victim/intentional – Shame	3.5 ± 1.3	1.0–6.5	2.5 ± 1.0	1.0–4.8	0.9	2.48×10^{-7}	<0.001
Victim/accident – Shame	2.1 ± 1.1	1.0–6.1	1.5 ± 0.6	1.0–4.0	0.6	7.72×10^{-7}	<0.001
Verbal affective memory task 26[b]							
Immediate recall	14.6 ± 3.3	6.6–23.4	16.0 ± 3.0	8.2–21.0	−1.6	7.17×10^{-4}	0.001
Short-term recall	14.7 ± 4.6	4.0–24.0	17.2 ± 4.5	7.0–25.0	−2.8	3.40×10^{-5}	<0.001
Delayed recall	15.0 ± 4.9	4.0–26.0	17.9 ± 4.2	9.0–26.0	−3.2	3.00×10^{-6}	<0.001

Raw *p* values as well as corrected *p* values are reported (see Method sections for description). *β*-Values represent difference in scores between patients and healthy controls once age and sex has been accounted for.
[a]Depressed patients $n = 91$.
[b]Depressed patients $n = 86$.

with the victim of harm (see Table 2). In fact, the level of shame and guilt reported by the patients appeared most pronounced when they were the victim of intentional rather than accidental harm, likely reflecting a disengaging of other-blaming schemata in favour of maladaptive self-blaming and internalising schemata. Clearly, this represents a critical target to address in psychotherapy, e.g. cognitive behavioural therapy (CBT).

Recent evidence suggests that ToM, i.e. the ability to attribute mental states to other people, is impaired in MDD and linked to depressive symptom severity (Bora & Berk, 2015). While we did not have a direct measure ToM in the present study, the SIP task indexes the preference for choosing social information (thoughts or facial expression) over non-social information (facts) when interpreting socially ambiguous situations. We did

not observe any differences between patients and healthy controls, suggesting that it is not a lack of attention towards or a preference away from social information that is causing the reported ToM deficits in MDD. This aligns with reports that dysphoria is associated with slightly *increased* sensitivity to social cues required for ToM (Harkness, Sabbagh, Jacobson, Chowdrey, & Chen, 2005). Our depressed patients did not exhibit negative bias in the interpretation of social situations. We speculate that this may partially be related to the task design; in several scenarios the negative interpretation had paranoid components (e.g. believing a colleague is poisoning a cup of tea) and might be too extreme to capture the more subtle negative biases in MDD.

Cold cognition

We were able to replicate previously reported impairments in verbal memory, working memory and reaction time (Bennabi et al., 2013; Goodall et al., 2018; Lee et al., 2012; Marazziti, Consoli, Picchetti, Carlini, & Faravelli, 2010; Rock et al., 2014) showing moderate effect sizes. The difference in number of remembered words in the VAMT-26 between patients and healthy controls also appeared to become progressively larger across the three time points (immediate recall, 1.6 words; short-term recall 2.8 words, and long-term recall, 3.2 words). This could potentially indicate that initial learning is less affected than long-term memory or alternatively reflect effects of fatigue and/or apathy in the depressed patients (Marazziti et al., 2010).

Correlation with depression severity

We were unable to identify a clear association between any of the individual cognitive task domains and depression severity indexed with $HDRS_6$ or $HDRS_{17}$ scores. This suggests that cognitive disturbances may not simply be an extension of the 'classic' core mood and somatic symptoms in MDD but rather represent distinct characteristics of the depressive pathology. Furthermore, while cognitive disturbances may be largely independent from symptom severity during the depressive episode, they are still associated with long-term clinical and functional outcomes (Cicchetti, 1994; Collins et al., 2011) and thus hold promise as a relevant stratification tool for the identification of clinically meaningful subgroups in MDD. Indeed, empirical trials are currently investigating the benefits of cognition-based tools for optimising treatment in MDD (Kingslake et al., 2017). However, more work is still needed to evaluate the clinical value of such approaches.

Clustering of cognitive profiles

In an exploratory analysis, we identified three clusters of cognitive profiles in the depressed patients. Cluster A was the largest group ($n = 38$) and characterised patients with strong negative biases in emotion recognition and misattribution but no substantial deficits in cold cognitive domains apart from slowed reaction time. Cluster B ($n = 28$) conversely characterised patients with positive biases in emotion processing and moderate deficits across all cold cognition domains. Lastly, Cluster C ($n = 26$) characterised patients who had large deficits across both hot and cold cognitive domains and in particular extremely high ratings of shame and guilt. These findings not only suggest a dissociation between the presence of hot and cold cognitive deficits in MDD, as illustrated by the differences between Cluster A and B, but also the existence of a subgroup of patients with severe global cognitive deficits represented by Cluster C.

To our knowledge, only two other studies have used cluster analysis to identify cognitive profiles in MDD: the large iSPOT-D trial ($N = 1008$) (Etkin et al., 2015) and a smaller study ($N = 50$) in patients with first-episode depression (Vicent-Gil et al., 2018). Both studies identified two clusters based on performance on cold cognitive tasks: a large cluster of cognitive intact patients and a smaller cluster of cognitive impaired patients. Notably, the proportion of cognitive impaired patients reported in both studies was 25–26%, closely matching the size of the globally impaired Cluster C in the present cohort (~28%). This further aligns with previous reports that only a small proportion of patients experience pronounced impairments in cognitive performance with estimates ranging between 21 and 44% depending on the cognitive measures and cut-off criteria used (Gualtieri & Morgan, 2008; Iverson, Brooks, Langenecker, & Young, 2011; McIntyre et al., 2017). Importantly, none the above studies included measures of affective biases or social cognition and may therefore have overlooked the presence of Cluster A type patients who exhibit strong negative biases in emotion processing but little to no deficits in cold cognitive domains. This highlights the importance of characterising *both* cold and hot cognitive disturbances in MDD concurrently.

Interestingly, the degree of cognitive disturbances across the three clusters partly mirrored the severity of depressive symptoms within the clusters, i.e. Cluster C had overall higher levels of depressive symptoms compared to Cluster A and B. This indicates that these cognitive profiles are able to capture MDD characteristics not captured by any individual task domain. Future studies should evaluate whether such cluster labelling, in addition to single cognitive domain information, may be useful for guiding antidepressants treatment choices and/or identify patients who will benefit from augmentation with e.g. cognitive remediation (Maruff & Jaeger, 2016) or cognitive enhancers (Bowie, Gupta, & Holshausen, 2013).

Methodological considerations

Some methodological limitations should be considered: (1) The depressed patients and healthy controls were unevenly matched on sex and because of the recruitment and inclusion procedures, the healthy controls likely represent very high-functioning individuals which may have inflated the observed differences on cognitive outcomes between the two groups. (2) We did not correct for the effect of IQ or education in the analyses as previous reports indicate that IQ measurement (Goss, Kaser, Costafreda, Sahakian, & Fu, 2013; Miskowiak et al., 2014) and education dropout rates (Marazziti et al., 2010) are affected by depressive symptoms. None of the reported estimates changed critically when IQ or education were included in the models (for corrected estimates see online Supplementary Materials). (3) Because the wordlist in the VAMT-26 contained both positively and negatively valanced words, the total word recall score does not represent a 'pure' cold measure of explicit memory. (4) Due to the limited stamina of the MDD patients, we had to restrict the number of cognitive domains tested; as a consequence, we did not collect data on e.g. attention or higher-executive functions despite their relevance in MDD pathology.

Conclusion

The current study represents one of the most comprehensive investigations into hot and cold cognitive impairment in a

large, well-characterised and antidepressant-free cohort of depressed patients to date. This allowed us to assess and directly compare the magnitude and patterns of impairment across a broad range of cognitive domains as well as investigate the presence of clusters of distinct cognitive profiles in depression. It is also the first time tasks from the EMOTICOM test battery have been applied and shown to be sensitive to MDD pathology in a patient cohort. Overall, our findings highlight the importance of including both hot and cold cognitive domains in investigations into MDD and further suggest that cognitive measures capture features beyond those reflected by depression severity. While cognitive disturbances are not present in all patients, they do represent significant impairments in identifiable and large subgroups of patients that may benefit from augmentation with cognition targeted treatments. Thus, we argue that cognition-based tools hold promise as clinically useful stratification aids in the care of depressed patients.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S0033291720000938.

Acknowledgements. We would like to thank the patients and healthy participants for their time. The data from the study is available upon request from the CIMBI database (http://www.cimbi.dk/db).

Financial support. This study was supported by the Augustinus Foundation (VGF, grant 16-0058), Rigshospitalet's Research Council (VHD, grant R149-A6325) and the Innovation Fund Denmark (grant 4108-00004B). The financial supporters were not involved in the study design, collection, analysis, interpretation or publication of data.

Conflict of interest. BJS is a co-inventor of CANTAB and EMOTICOM test battery and consults for Cambridge Cognition, Greenfield BioVentures and Cassava Sciences. VGF and GMK have received honorarium as consultant for Sage Therapeutics. The remaining authors all declare no conflict of interest.

References

APA. (2013). *Diagnostic and statistical manual of mental disorders, DSM-5* (5th ed.). Washington, DC: American Psychiatric Association.

Armstrong, T., & Olatunji, B. O. (2012). Eye tracking of attention in the affective disorders: A meta-analytic review and synthesis. *Clinical Psychology Review*, *32*(8), 704–723. doi:10.1016/j.cpr.2012.09.004.

Beavers, A. S., Lounsbury, J. W., Richards, J. K., Huck, S. W., Skolits, G., & Esquivel, S. L. (2013). Practical considerations for using exploratory factor analysis in educational research. *Practical Assessment, Research and Evaluation*, *18*, 1–13.

Bennabi, D., Vandel, P., Papaxanthis, C., Pozzo, T., & Haffen, E. (2013). Psychomotor retardation in depression: A systematic review of diagnostic, pathophysiologic, and therapeutic implications. *BioMed Research International*, *2013*, 158746–158746. doi:10.1155/2013/158746.

Bora, E., & Berk, M. (2015). Theory of mind in major depressive disorder: A meta-analysis. *Journal of Affective Disorders*, *191*, 49–55. doi:10.1016/j.jad.2015.11.023.

Bower, G. H. (1981). Mood and memory. *American Psychologist*, *36*(2), 129–148. doi:10.1037/0003-066X.36.2.129.

Bowie, C. R., Gupta, M., & Holshausen, K. (2013). Cognitive remediation therapy for mood disorders: Rationale, early evidence, and future directions. *Canadian Journal of Psychiatry*, *58*(6), 319–325. doi:10.1177/070674371305800603.

Bradley, B. P., Mogg, K., & Williams, R. (1995). Implicit and explicit memory for emotion-congruent information in clinical depression and anxiety. *Behavior Research and Therapy*, *33*(7), 755–770. doi:https://doi.org/10.1016/0005-7967(95)00029-W.

Calev, A. (1996). Affect and memory in depression: Evidence of better delayed recall of positive than negative affect words. *Psychopathology*, *29*(2), 71–76. doi:10.1159/000284974.

Cambridge, O. R., Knight, M. J., Mills, N., & Baune, B. T. (2018). The clinical relationship between cognitive impairment and psychosocial functioning in major depressive disorder: A systematic review. *Psychiatry Research*, *269*, 157–171. doi:10.1016/j.psychres.2018.08.033.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6*(4), 284–290. doi:10.1037/1040-3590.6.4.284.

Clatworthy, J., Buick, D., Hankins, M., Weinman, J., & Horne, R. (2005). The use and reporting of cluster analysis in health psychology: A review. *British Journal of Health Psychology*, *10*(Pt 3), 329–358. doi:10.1348/135910705x25697.

Collins, P. Y., Patel, V., Joestl, S. S., March, D., Insel, T. R., Daar, A. S., … Stein, D. J. (2011). Grand challenges in global mental health. *Nature*, *475*(7354), 27–30. doi:10.1038/475027a.

Dam, V. H., Thystrup, C. K., Jensen, P. S., Bland, A. R., Mortensen, E. L., Elliott, R., … Stenbæk, D. S. (2019). Psychometric properties and validation of the EMOTICOM test battery in a healthy Danish population. *Frontiers in Psychology*, *10*(2660). doi:10.3389/fpsyg.2019.02660.

Danion, J. M., Kauffmann-Muller, F., Grange, D., Zimmermann, M. A., & Greth, P. (1995). Affective valence of words, explicit and implicit memory in clinical depression. *Journal of Affective Disorders*, *34*(3), 227–234.

Elliott, R., Zahn, R., Deakin, J. F., & Anderson, I. M. (2011). Affective cognition and its disruption in mood disorders. *Neuropsychopharmacology*, *36*(1), 153–182. doi:10.1038/npp.2010.77.

Etkin, A., Patenaude, B., Song, Y. J., Usherwood, T., Rekshan, W., Schatzberg, A. F., … Williams, L. M. (2015). A cognitive-emotional biomarker for predicting remission with antidepressant medications: A report from the iSPOT-D trial. *Neuropsychopharmacology*, *40*(6), 1332–1342. doi:10.1038/npp.2014.333.

Gaddy, M. A., & Ingram, R. E. (2014). A meta-analytic review of mood-congruent implicit memory in depressed mood. *Clinical Psychology Review*, *34*(5), 402–416. doi:10.1016/j.cpr.2014.06.001.

Goodall, J., Fisher, C., Hetrick, S., Phillips, L., Parrish, E., & Allott, K. (2018). Neurocognitive functioning in depressed young people: A systematic review and meta-analysis. *Neuropsychology Review*, *28*(2), 216–231. doi:10.1007/s11065-018-9373-9.

Goss, A. J., Kaser, M., Costafreda, S. G., Sahakian, B. J., & Fu, C. H. (2013). Modafinil augmentation therapy in unipolar and bipolar depression: A systematic review and meta-analysis of randomized controlled trials. *The Journal of Clinical Psychiatry*, *74*(11), 1101–1107. doi:10.4088/JCP.13r08560.

Gualtieri, C. T., & Morgan, D. W. (2008). The frequency of cognitive impairment in patients with anxiety, depression, and bipolar disorder: An unaccounted source of variance in clinical trials. *The Journal of Clinical Psychiatry*, *69*(7), 1122–1130. doi:10.4088/jcp.v69n0712.

Harkness, K., Sabbagh, M., Jacobson, J., Chowdrey, N., & Chen, T. (2005). Enhanced accuracy of mental state decoding in dysphoric college students. *Cognition and Emotion*, *19*(7), 999–1025. doi:10.1080/02699930541000110.

Hernaus, D., Gold, J. M., Waltz, J. A., & Frank, M. J. (2018). Impaired expected value computations coupled with overreliance on stimulus-response learning in schizophrenia. *Biological Psychiatry. Cognitive Neuroscience and Neuroimaging*, *3*(11), 916-926. doi:10.1016/j.bpsc.2018.03.014.

Iverson, G. L., Brooks, B. L., Langenecker, S. A., & Young, A. H. (2011). Identifying a cognitive impairment subgroup in adults with mood disorders. *Journal of Affective Disorders*, *132*(3), 360–367. doi:https://doi.org/10.1016/j.jad.2011.03.001.

Jensen, C. G., Hjordt, L. V., Stenbaek, D. S., Andersen, E., Back, S. K., Lansner, J., … Hasselbalch, S. G. (2016). Development and psychometric validation of the verbal affective memory test. *Memory (Hove, England)*, *24*(9), 1208–1223. doi:10.1080/09658211.2015.1087573.

Kim, J. H., Son, Y. D., Kim, J. H., Choi, E. J., Lee, S. Y., Joo, Y. H., … Cho, Z. H. (2015). Self-transcendence trait and its relationship with in vivo serotonin transporter availability in brainstem raphe nuclei: An ultra-high resolution PET-MRI study. *Brain Research*, *1629*, 63–71. doi:10.1016/j.brainres.2015.10.006.

Kim, S., Thibodeau, R., & Jorgensen, R. S. (2011). Shame, guilt, and depressive symptoms: A meta-analytic review. *Psychological Bulletin*, *137*(1), 68–96. doi:10.1037/a0021466.

Kingslake, J., Dias, R., Dawson, G. R., Simon, J., Goodwin, G. M., Harmer, C. J. … Browning, M. (2017). The effects of using the PReDicT test to guide the

antidepressant treatment of depressed patients: Study protocol for a randomised controlled trial. *Trials*, *18*(1), 558. doi:10.1186/s13063-017-2247-2.

Köhler, C. A., Carvalho, A. F., Alves, G. S., McIntyre, R. S., Hyphantis, T. N., & Cammarota, M. (2015). Autobiographical memory disturbances in depression: A novel therapeutic target? *Neural Plasticity*, *2015*, 759139–759139. doi:10.1155/2015/759139.

Lake, C. R. (2008). Hypothesis: Grandiosity and guilt cause paranoia; paranoid schizophrenia is a psychotic mood disorder; a review. *Schizophrenia Bulletin*, *34*(6), 1151–1162. doi:10.1093/schbul/sbm132.

Lee, R. S., Hermens, D. F., Porter, M. A., & Redoblado-Hodge, M. A. (2012). A meta-analysis of cognitive deficits in first-episode major depressive disorder. *Journal of Affective Disorders*, *140*(2), 113–124. doi:10.1016/j.jad.2011.10.023.

Marazziti, D., Consoli, G., Picchetti, M., Carlini, M., & Faravelli, L. (2010). Cognitive impairment in major depression. *European Journal of Pharmacology*, *626*(1), 83–86. doi:https://doi.org/10.1016/j.ejphar.2009.08.046.

Maruff, P. & Jaeger, J. (2016). Understanding the importance of cognitive dysfunction and cognitive change in major depressive disorder. In R. S. McIntyre (Ed.), *Cognitive impairment in Major Depressive Disorder* (pp. 15–29). Cambridge, U.K: Cambridge University Press.

McIntyre, R. S., Best, M. W., Bowie, C. R., Carmona, N. E., Cha, D. S., Lee, Y., … Harrison, J. (2017). The THINC-integrated tool (THINC-it) screening assessment for cognitive dysfunction: Validation in patients with major depressive disorder. *The Journal of Clinical Psychiatry*, *78*(7), 873–881. doi:10.4088/JCP.16m11329.

Milligan, G. W. (1980). An examination of the effect of six types of error perturbation on fifteen clustering algorithms. *Psychometrika*, *45*(3), 325–342. doi:10.1007/BF02293907.

Mills, R. S., Hastings, P. D., Serbin, L. A., Stack, D. M., Abela, J. R., Arbeau, K. A., & Lall, D. I. (2015). Depressogenic thinking and shame proneness in the development of internalizing problems. *Child Psychiatry and Human Development*, *46*(2), 194–208. doi:10.1007/s10578-013-0416-4.

Miskowiak, K. W., & Carvalho, A. F. (2014). 'Hot' cognition in major depressive disorder: A systematic review. *CNS and Neurological Disorders – Drug Targets*, *13*(10), 1787–1803.

Miskowiak, K. W., Vinberg, M., Christensen, E. M., Bukh, J. D., Harmer, C. J., Ehrenreich, H., & Kessing, L. V. (2014). Recombinant human erythropoietin for treating treatment-resistant depression: A double-blind, randomized, placebo-controlled phase 2 trial. *Neuropsychopharmacology*, *39*(6), 1399–1408. doi:10.1038/npp.2013.335.

Neshat-Doost, H. T., Taghavi, M. R., Moradi, A. R., Yule, W., & Dalgleish, T. (1998). Memory for emotional trait adjectives in clinically depressed youth. *Journal of Abnormal Psychology*, *107*(4), 642–650.

Reynolds, C. R. (2011). RIAS/RIST, Reynolds Intellectual Assessment Scales, Reynolds Intellectual Screening Test, vejledning (1. oplag ed.). Virum: Hogrefe Psykologisk Forlag.

Rock, P. L., Roiser, J. P., Riedel, W. J., & Blackwell, A. D. (2014). Cognitive impairment in depression: A systematic review and meta-analysis. *Psychological Medicine*, *44*(10), 2029–2040. doi:10.1017/s0033291713002535.

Roiser, J. P., Elliott, R., & Sahakian, B. J. (2012). Cognitive mechanisms of treatment in depression. *Neuropsychopharmacology*, *37*(1), 117–136. doi:10.1038/npp.2011.183.

Roiser, J. P., & Sahakian, B. J. (2013). Hot and cold cognition in depression. *CNS Spectrums*, *18*(3), 139–149. doi:10.1017/S1092852913000072.

Vicent-Gil, M., Keymer-Gausset, A., Serra-Blasco, M., Carceller-Sindreu, M., de Diego-Adelino, J., Trujols, J., … Portella, M. J. (2018). Cognitive predictors of illness course at 12 months after first-episode of depression. *European Neuropsychopharmacology : the Journal of the European College of Neuropsychopharmacology*, *28*(4), 529–537. doi:10.1016/j.euroneuro.2018.02.001.

Watkins, P. C., Mathews, A., Williamson, D. A., & Fuller, R. D. (1992). Mood-congruent memory in depression: Emotional priming or elaboration? *Journal of Abnormal Psychology*, *101*(3), 581–586.

Weightman, M. J., Knight, M. J., & Baune, B. T. (2019). A systematic review of the impact of social cognitive deficits on psychosocial functioning in major depressive disorder and opportunities for therapeutic intervention. *Psychiatry Research*, *274*, 195–212. doi:10.1016/j.psychres.2019.02.035.

WHO. (2007). ICD-10, international statistical classification of diseases and related health problems (10. revision, version for 2007 ed.). Geneva: World Health Organization.

WHO. (2017). *Depression and other common mental disorders: Global health estimates*. Geneva: World Health Organization.

Wolkenstein, L., Schonenberg, M., Schirm, E., & Hautzinger, M. (2011). I can see what you feel, but I can't deal with it: Impaired theory of mind in depression. *Journal of Affective Disorders*, *132*(1–2), 104–111. doi:10.1016/j.jad.2011.02.010.

Zupan, Z., Žeželj, I., & Andjelković, I. (2017). Memory bias in depression: Effects of self-reference and age. *Journal of Social and Clinical Psychology*, *36*(4), 300–315. doi:10.1521/jscp.2017.36.4.300.

Supplementary Materials

1. Description of cognitive tasks and outcomes……. p. 2

2. Cluster analysis…….. p. 6

3. First-episode vs recurrent depression…….…… p. 9

4. Correlation between cognitive scores and clinical symptoms……. p. 11

5. Cognitive task scores across groups……. p. 12

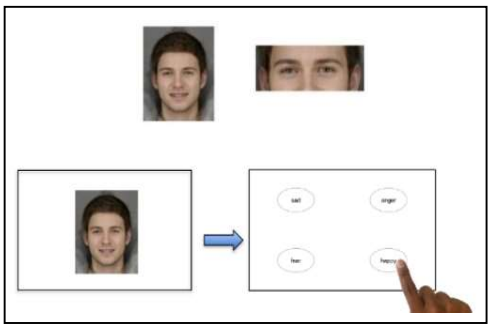
6. Effect of IQ and education on group difference estimates…….. p. 14

1. Description of cognitive tasks and outcomes

Hot cognition

Emotion recognition

The eyes version of the Emotional Recognition Task (ERT) from the EMOTICOM test battery was used to assess recognition of facial expressions (Bland et al., 2016; Dam et al., 2019). Participants were asked to determine which emotion (happy, sad, angry or fearful) was being expressed by a pair of eyes shown briefly (250ms) on a computer screen.



Main outcomes: Affective bias for recognition (hit rate i.e. percentage of trials in which a given emotion was correctly identified) and misattribution (false alarm rate i.e. percentage of trials in which a given emotion was wrongly identified). Affective bias was calculated as: Hit_{happy} - Hit_{sad} and $FalseAlarm_{happy}$ - $FalseAlarm_{sad}$

Secondary outcomes: Hit rate and false alarm rate for each emotion: happy, sad, angry and fearful.

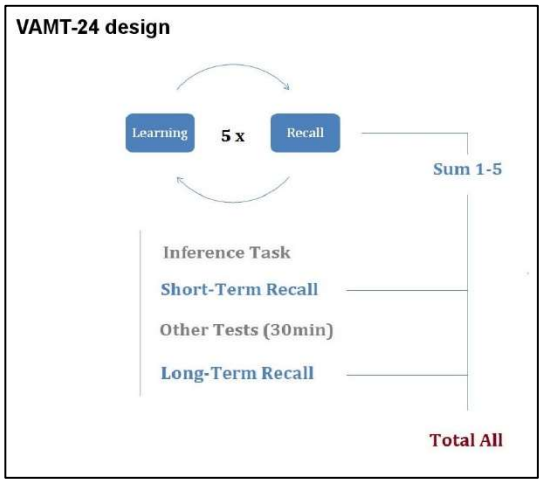
Emotion detection threshold

The Emotional Intensity Morphing task (IM) from the EMOTICOM test battery was used to assess the perceptual threshold for detection of emotions in facial expressions (Bland et al., 2016). Participants were asked to indicate when they were either able to detect (increase condition) or no longer detect (decrease condition) a given emotion (happy, sad, angry, fearful, or disgusted) on a face with a slowly morphing expression.

Main outcomes: Affective bias in percentage averaged across both increase and decrease condition calculated as: $DetectionThreshold_{sad}$ - $DetectionThreshold_{happy}$. To ensure that a positive score indicate a positive bias, 'Happy' was subtracted from 'Sad' for the IM task because low scores indicate a low detection threshold, i.e., higher perceptual sensitivity to the presence of the emotion.

Secondary outcomes: Detection threshold for each of the five emotions for both the increase and decrease condition.

Affective verbal memory



A modified version of the Verbal Affective Memory Task 24 with 26 words (Verbal Affective Memory Task 26; VAMT-26) was used to assess learning and memory of affective words (Jensen et al., 2016). A list of words (10 positive, 10 negative and 6 neutral) were presented briefly one by one in a pseudo-randomized order on a computer screen. After viewing the list, the participants were asked to verbally recall as many words as possible in no specific order. Immediate recall was determined as average number of words remembered across five viewings of the list, short-term recall as number of words remembered after viewing an interference list, and delayed recall as number of words remembered after a span of 30 minutes.

Main outcome: Affective bias for total word recall (i.e. average words remembered across immediate, short-term, and delayed recall) calculated as: $WordScore_{positive} - WordScore_{negative}$

Moral emotions

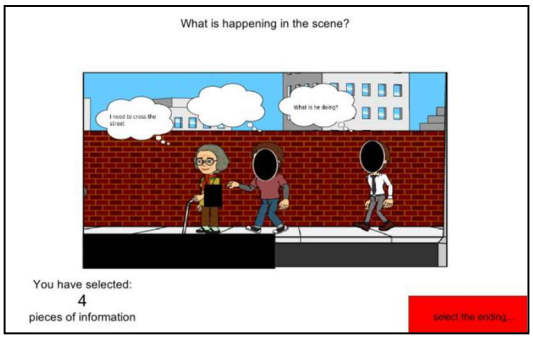
The Moral Emotions Task (ME) was used to assess moral emotions in social situations. Participants were shown cartoons in which one person either intentionally or accidentally causes another person harm. Participants were instructed to imagine themselves as either the agent (i.e. the person causing the harm) or the victim and rate how guilty and ashamed they would feel.

Main outcomes: Average rating across all conditions for guilt and shame.

Secondary outcomes: Ratings of guilt and shame for each of the four conditions: Agent intentionally causing harm; agent accidentally causing harm; victim of intentional harm; and victim of accidental harm.

Social information preference

The Social Information Preference task (SIP) was used to assess information sampling and interpretation of social situations. Participants were shown cartoons depicting social interactions in which several pieces of information were hidden (thoughts, facial expression, and facts/items). Participants were instructed to pick four pieces of information to help them interpret the situation and to choose between a positive, neutral, and negative outcome.



Main outcomes: Preference for social information over non-social information calculated as: $Choicet_{houghts} + Choice_{faces} - Choice_{facts}$ and affective bias in choice of scenario outcome calculated as: $Outcome_{positive} - Outcome_{negative}$

Cold cognition

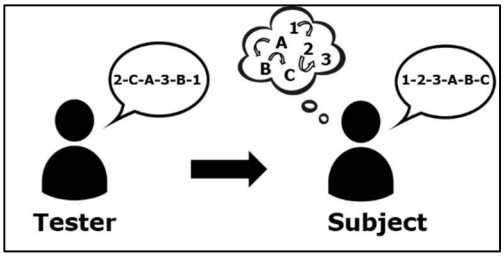
Explicit verbal memory

VAMT-26 was used to assess overall verbal memory function.

Main outcome: Total word recall calculated as average number of words (positive, negative and neutral) recalled across immediate, short-term and long-term recall.

Working memory

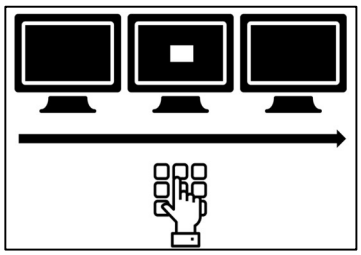
The Letter Number Sequence task (LNS) was used to assess working memory capacity. Participants were asked to remember and mentally sort a sequence of jumbled letters and numbers of increasing length.



Main outcome: Total number of correctly recited sequences with scores ranging from 0-21.

Reaction time

The Simple Reaction Time task (SRT) was used to assess reaction time. Participants were instructed to press a button as fast as possible at the appearance of a white square on a screen. The location of the square did not change but the interval between appearances varied.



Main outcome(s): Reaction time latency in milliseconds.

References

Bland, A. R., Roiser, J. P., Mehta, M. A., Schei, T., Boland, H., Campbell-Meiklejohn, D. K., . . . Elliott, R. (2016). EMOTICOM: A Neuropsychological Test Battery to Evaluate Emotion, Motivation, Impulsivity, and Social Cognition. *Front Behav Neurosci, 10*, 25. doi: 10.3389/fnbeh.2016.00025

Dam, V. H., Thystrup, C. K., Jensen, P. S., Bland, A. R., Mortensen, E. L., Elliott, R., . . . Stenbæk, D. S. (2019). Psychometric Properties and Validation of the EMOTICOM Test Battery in a Healthy Danish Population. *Frontiers in Psychology, 10*(2660). doi: 10.3389/fpsyg.2019.02660

Jensen, C. G., Hjordt, L. V., Stenbaek, D. S., Andersen, E., Back, S. K., Lansner, J., . . . Hasselbalch, S. G. (2016). Development and psychometric validation of the verbal affective memory test. *Memory, 24*(9), 1208-1223. doi: 10.1080/09658211.2015.1087573

2. Cluster analyses

To ascertain the optimal number of clusters for the K-means cluster analysis, we conducted a Hierarchical Cluster Analysis (HCA) using Ward's method with squared Euclidian distance as the similarity measure (Ward, 1963). Input into the HCA was *z*-transformed scores from the eight primary cognitive outcomes which was found to significantly differ between depressed patients and healthy controls (i.e., recognition and misattribution rates from the *Emotional Recognition Task*-eyes version, detection threshold from the *Intensity Morphing* task, guilt and shame ratings from the *Moral Emotions* task, verbal memory from the *Verbal Affective Memory Task-26*, working memory from the *Letter Number Sequence* task, and reaction time from the *Simple Reaction Time* task). From the HCA a dissimilarity measure coefficient was obtained (Table S2.1). The change in dissimilarity was displayed against the number of clusters and the optimal number of clusters was determined as the cluster just prior to the largest jump in change in dissimilarity measure coefficient (Figure S2.1). Thus, a three-cluster solution was indicated supported by the produced dendrogram (Figure S2.2). Lastly, the clustering centroids from the three clusters (i.e. group means for each cognitive outcome) were used to initialize the K-means analysis (Milligan, 1980).

Table S2.1. Agglomeration schedule

Number of clusters	Dissimilarity measure coefficient	Change in dissimilarity measure coefficient	Jump in change of dissimilarity measure coefficient
1	643.9	96.3	18.5
2	547.7	77.8	31.2
3	469.8	46.6	3.9
4	423.2	42.7	7.3
5	380.5	35.4	5.8
6	345.1	29.6	3.7
7	315.5	26.0	8.3
8	289.5	17.6	3.0

9	271.9	14.7	1.9
10	257.2	12.8	0.7
11	244.5	12.0	0.4
12	232.5	11.6	0.8
13	220.9	10.8	1.2
14	210.1	9.7	0.2
15	200.4	9.4	1.0

Table S2.1. The table shows the dissimilarity measure coefficients from the agglomeration schedule. In addition, the calculated change in dissimilarity measure coefficient as well as the jump in change of dissimilarity measure coefficient is shown. Note we here only show the coefficients for cluster solutions 1-15 as a larger number of clusters would not be meaningful in the present study.

Figure S2.1. Changes in dissimilarity measure coefficients

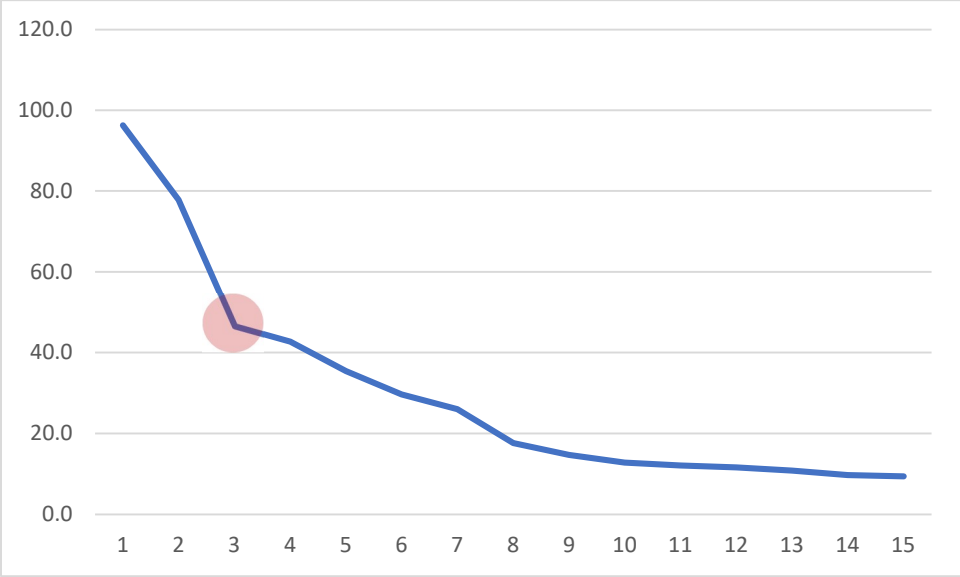


Figure S2.1. The graph shows changes in dissimilarity measure coefficient (y-axis) across different numbers of clusters (x-axis). The changes were calculated as the difference in dissimilarity measure coefficient for each added cluster. The red circle indicates the largest and most inconsistent change, suggesting that the clustering process should be stopped at a three-cluster solution.

Figure S2.2. Dendrogram

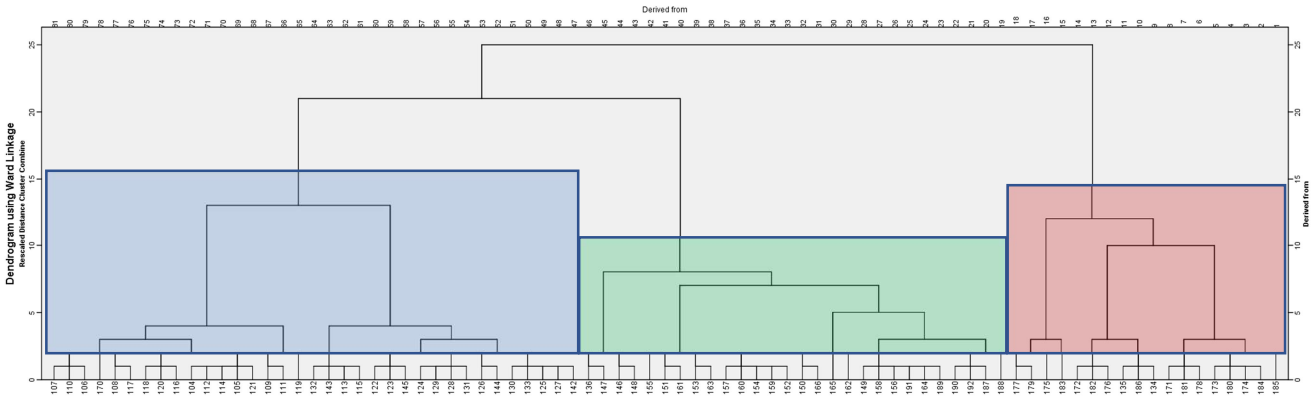


Figure S2.2. Dendrogram produced from Hierarchical Cluster analysis based on the scores of depressed patients on eight primary cognitive outcomes. The colored boxes indicate the three clusters suggested by the dendrogram.

References

Milligan, G. W. (1980). An examination of the effect of six types of error perturbation on fifteen clustering algorithms. *Psychometrika, 45*(3), 325-342. doi:10.1007/BF02293907

Ward, J. H. (1963). Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association, 58*(301), 236-244. doi:10.1080/01621459.1963.10500845

3. First-episode vs recurrent depression

History of previous depressive episodes was collected based on patient testimony supplemented with information from medical records when possible. Figure S3.1. provides an overview of the number of patients with first-episode depression (N = 41) and recurrent depression (N = 51), including number of previous depressive episodes in the latter group.

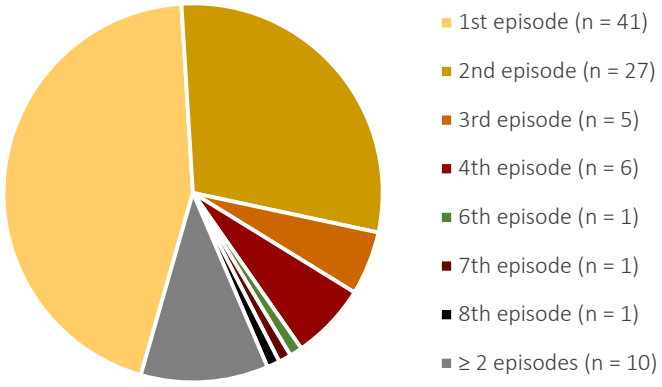


Figure S3.1. Distribution of MDD patients (N = 92) with first episode depression and recurrent depression (i.e. ≥ 2 episodes). Note for = 10 patients, the exact number of depressive episodes could not be verified beyond recurrent depression (i.e. more than one episode).

Table S3.1. shows descriptive characteristics of first-episode and recurrent depression patients.

Table S3.1 History of depressive episodes

	First-episode (N = 41)		Recurrent (N = 51)		*p*
	Mean ± SD	Range	Mean ± SD	Range	
Age	28.8 ± 10.6	18–57	26.0 ± 5.3	18–43	0.73
Male/female	13/28		12/39		0.40
HDRS6	12.4 ± 1.7	8–17	12.3 ± 1.6	7–17	0.89
$HDRS_{17}$	22.7 ± 3.2	18–31	22.9 ± 3.5	18–31	0.82

Table S3.1. Age, sex distribution and depressive symptoms severity (Hamilton Depressive Rating Scale 6, $HDRS_6$) between patients with first-episode depression and patients with recurrent depression. Group differences were assessed using Mann Whitney U-tests and χ^2 test (for sex).

There here was no statistically significant difference between the number of patients with first-episode vs recurrent depression across the three cognitive profile clusters, $\chi^2(2, N = 92) = 1.4$, $p = 0.5$.

Table S3.2 shows differences in cognitive performance between patients with first-episode depression and recurrent depression. We detected no statistically significant differences in performance on any of the cognitive tasks.

Table S3.2. Depressive history and cognitive performance

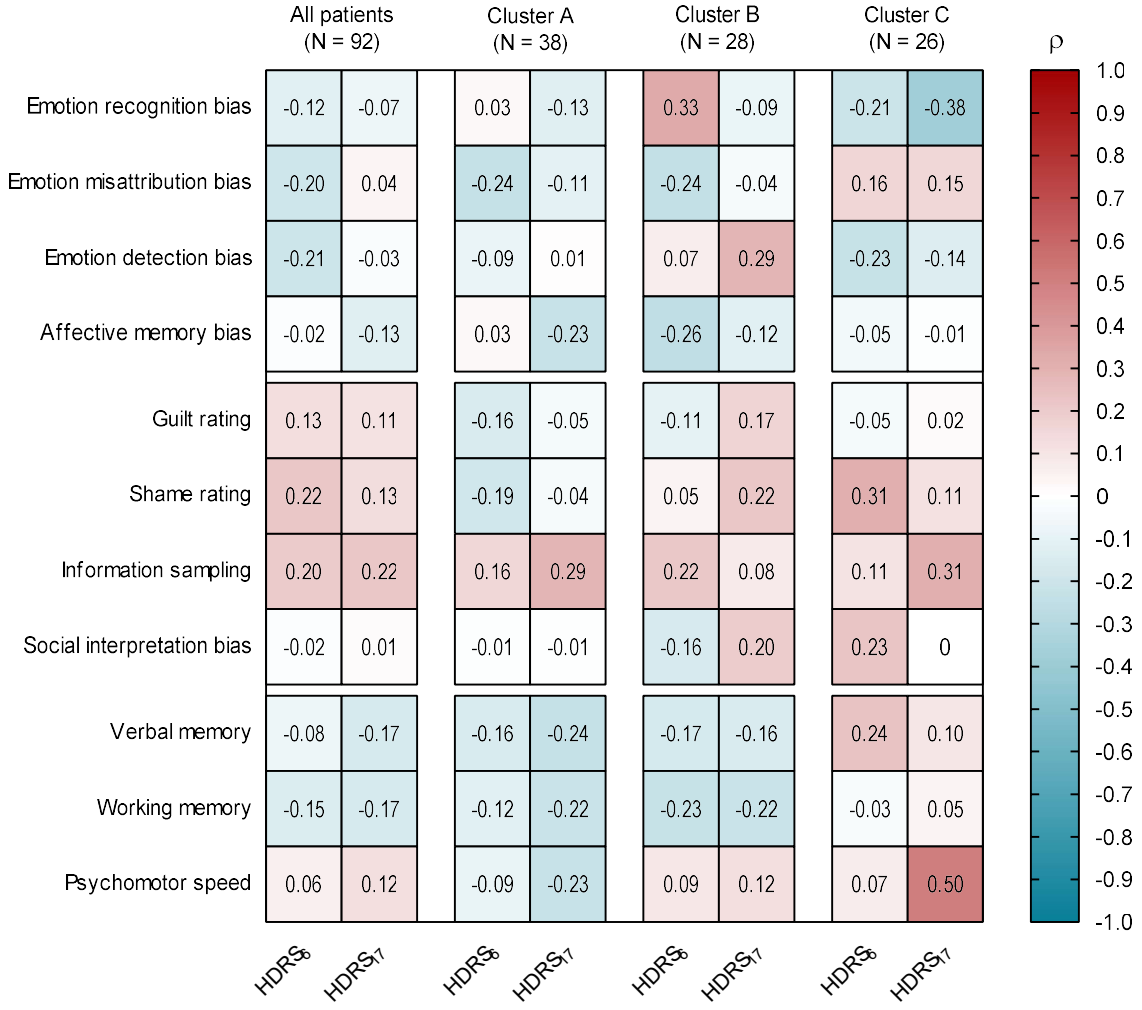
	First-episode (N = 41)		Recurrent (N = 51)		β	p	$p_{corrected}$
	Mean ± SD	Range	Mean ± SD	Range			
Hot cognition I: affective biases							
Emotion recognition	-9.5 ± 24.2	-80.0–45.0	-3.4 ± 22.8	-40.0–45.0	7.3	0.15	1.00
Emotion misattribution	-12.0 ± 15.9	-46.7–20.0	-6.3 ± 15.0	-35.0–26.7	6.1	0.07	0.75
Emotion detection	0.03 ± 14.7	-41.4–25.9	-3.0 ± 13.5	-39.3–17.9	-2.2	0.46	1.00
Affective memory	1.8 ± 13.0	-20.7–35.3	-3.8 ± 14.3	-30.7–32.7	-4.5	0.13	1.00
Hot cognition II: social cognition							
Guilt ratings	4.2 ± 0.7	2.9–5.9	4.2 ± 0.6	3.2–5.5	-0.01	0.93	1.00
Shame ratings	4.4 ± 0.8	2.7–5.8	4.4 ± 0.7	3.2–6.3	0.03	0.84	1.00
Information sampling	36.5 ± 17.7	-21.9–71.9	39.9 ± 19	-9.4–78.1	3.1	0.44	1.00
Social interpretation bias	9.9 ± 23.6	-50.0–50.0	3.3 ± 18.9	-43.8–50.0	-4.3	0.30	1.00
Cold cognition							
Verbal memory	14.3 ± 3.6	6.9–21.9	15.1 ± 4.5	5.5–23.1	0.8	0.41	1.00
Working memory	11.2 ± 2.6	6–16	12.0 ± 3.0	6–18	0.8	0.15	1.00
Reaction time	285 ± 57.6	205.4–466.7	270.1 ± 62.6	200.5–465.7	-17.2	0.17	1.00

Figure S3.2. Group differences between patients with first episode-depression and recurrent depression. The group differences were assessed with linear regression models with cognitive task score as the dependent variable and depression history (first-episode depression was coded as 0 and recurrent depression as 1), age and sex as independent variables. Uncorrected *p*-values and *p*-values corrected for 11 tests using the Bonferroni-Holm method are shown.

4. Correlation between cognitive scores and clinical symptoms

Figure S4.1 shows Spearman's ranked order correlations between clinical depression symptoms indexed with Hamilton Depressive Ratings Scale-6 ($HDRS_6$) and Hamilton Depressive Ratings Scale-17 ($HDRS_{17}$) and scores on cognitive tasks for all depressed patients (N = 92) and the three cognitive profile clusters.

Figure S4.1 Correlation between cognitive scores and clinical symptoms



After correction for 11 tests, none of the *p*-values were statistically significant for either $HDRS_6$ scores (all $p_{corrected} > 0.42$) or $HDRS_{17}$ (all $p_{corrected} > 0.11$).

5. Cognitive task scores across groups

Table S5.1. shows the average scores on cognitive tasks for the all healthy controls (N = 103) and patients (N = 92).

Table S5.1. Performance on cognitive tasks

	Healthy controls (N = 103)		Depressed patients (N = 92)	
	Mean ± SD	Range	Mean ± SD	Range
Hot cognition: affective biases				
Emotion recognition	6.7 ± 26.3	-55.0–90.0	-6.1 ± 23.5	-80.0–45.0
Emotion misattribution	-0.05 ± 17.9	-35.0–61.7	-8.8 ± 15.6	-46.7–26.7
Emotion detection	7.1 ± 13.0	-40.2–36.9	-1.7 ± 14.0	-41.4–25.9
Affective memory	0.2 ± 12.5	-28.0–29.3	-1.2 ± 13.9	-30.7–35.3
Hot cognition: social cognition				
Guilt ratings	3.7 ± 0.5	1.8–5.1	4.2 ± 0.6	2.9–5.9
Shame ratings	3.9 ± 0.6	2.1–5.5	4.4 ± 0.8	2.7–6.3
Information sampling	74.1 ± 16.9	0.0–100.0	77.3 ± 18.8	0.0–100.0
Social interpretation bias	10.6 ± 22.3	-50.0–75.0	6.1 ± 21.2	-50.0–50.0
Cold cognition				
Verbal memory	17.0 ± 3.7	8.5–23.1	14.7 ± 4.1	5.5–23.1
Working memory	13.6 ± 3.2	5.0–20.0	11.6 ± 2.8	6.0–18.0
Reaction time	238.2 ± 41.8	191.0–439.0	276.8 ± 60.5	200.5–466.7

Table S5.2. shows the average scores on cognitive tasks for the three cognitive profiles clusters.

Table S5.2. Cognitive task scores for profile clusters

	Cluster A (N = 38)		Cluster B (N = 28)		Cluster C (N = 26)	
	Mean ± SD	Range	Mean ± SD	Range	Mean ± SD	Range
Hot cognition: affective biases						
Emotion recognition	-20.7 ± 16.9	-80.0–15.0	18.8 ± 14.9	-10.0–45.0	-11.7 ± 17.4	-40.0–20.0
Emotion misattribution	-17.6 ± 11.6	-46.7–3.3	7.9 ± 10.2	-13.3–26.7	-14.0 ± 10.7	-35.0–6.7
Emotion detection	0.5 ± 12.6	-32.1–19.6	1.5 ± 10.4	-17.9–17.9	-8.2 ± 17.3	-41.4–25.9
Affective memory	1.6 ± 14.8	-24.0–32.7	-4.9 ± 14.9	-30.7–35.3	-1.3 ± 10.9	-18–19.3
Hot cognition: social cognition						
Guilt ratings	3.9 ± 0.4	3.1–4.9	4.0 ± 0.5	2.9–5.0	5.0 ± 0.4	4.0–5.9
Shame ratings	4.1 ± 0.6	2.7–5.0	4.0 ± 0.5	2.9–5.1	5.3 ± 0.5	4.1–6.3
Information sampling	38.5 ± 18.1	-21.9–78.1	36.7 ± 18.3	0.0–68.8	40.3 ± 19.6	-9.4–71.9
Social interpretation bias	4.0 ± 22.2	-50.0–37.5	10.3 ± 16.9	-18.8–50	4.5 ± 24.0	-37.5–50.0
Cold cognition						
Verbal memory	16.8 ± 3.2	9.8–23.1	14.3 ± 3.6	6.2–23.0	12.3 ± 4.4	5.5–23.1
Working memory	12.5 ± 2.8	8.0–18.0	11.6 ± 2.5	6.0–15.0	10.5 ± 2.8	6.0–16.0
Reaction time	267.4 ± 49.5	200.5–454.7	278.8 ± 57.8	219.3–436.9	288.9 ± 76.9	205.4–466.7

6. Effect of IQ and education on group difference estimates

Table S6.1. shows IQ scores indexed with the Reynolds Intellectual Screening Test (RIST) and education levels indexed with the Online Stimulant and Family History Assessment Module (OS-FHAM) questionnaire for the depressed patients (N = 92) and healthy controls (N = 103). Table S6.2. shows group difference estimates after correction with IQ and education respectively.

Table S6.1. Education and IQ

	Depressed patients (n = 92)	Healthy controls (n = 103)	*p*-value
IQ score	102.9 ± 8.4 (86–124)[a]	110.2 ± 7.0 (93–129)	<0.001
Education	16.4 ± 1.3 (11–17)[c]	14.9 ± 2.2 (8–17)[b]	<0.001

Table S6.1. The table shows IQ indexed with the Reynolds Intellectual Screening Test (RIST) and education score indexed with the Online Stimulant and Family History Assessment Module (OS-FHAM) as completed number of school years added to an education score between 1 (no vocational degree) and 5 (> 4 years of higher learning at university level). Group differences were assessed with an independent *t*-test. [a]N = 87, [b]N = 102, and [c]N = 74 due to missing data.

Table S6.2. Group differences on cognitive performance corrected for IQ and education

	Corrected for IQ		Corrected for education	
	β	*p*	*β*	*p*
Emotion processing				
Recognition bias	-12.76	0.002	-10.85	0.01
Misattribution bias	-7.42	0.01	-7.39	0.01
Detection bias	-8.419	< 0.001	-9.76	< 0.001
Affective memory	-1.326	0.54	-2.08	0.35
Social cognition				
Guilt ratings	0.49	< 0.001	0.48	< 0.001
Shame ratings	0.54	< 0.001	0.54	< 0.001
Information sampling	2.029	0.50	1.82	0.56
Social interpretation bias	-2.12	0.58	-0.03	1.00
Cold cognition				
Verbal memory	-1.60	0.009	-1.84	0.004
Working memory	-0.81	0.07	-1.20	0.02
Reaction time	23.5194	0.01	28.29	0.005

Table S7.2. Group difference between patients (N = 92) and healthy controls (N = 103) on primary cognitive outcomes after correction for IQ and correction for education. Note *p*-values are reported uncorrected; all models were corrected for age and sex.

Title: Cognitive disturbances as both treatment targets and predictors of antidepressant action in Major Depressive Disorder: A NeuroPharm study

Authors: Dam, V. H.[1,2], Stenbæk, D. S.[1], Köhler-Forsberg, K.[1,2,3], Ip, C.[1,2,4], Ozenne, B.[1,5], Sahakian, B. J.[6,7], Knudsen, G. M. [1,2], Jørgensen, M. B.[2,3], Frokjaer, V. G.[1, 2, 3]

Affiliations:

[1] Neurobiology Research Unit, the Neuroscience Centre, Copenhagen University Hospital Rigshospitalet, Denmark

[2] Faculty of Health and Medical Sciences, University of Copenhagen, Denmark

[3]Psychiatric Center Copenhagen, Copenhagen University Hospital Rigshospitalet, Denmark

[4]Department of Clinical Pharmacology, H. Lundbeck A/S, Valby, Denmark

[5]Department of Public Health, Section of Biostatistics, University of Copenhagen, Denmark

[6]Department of Psychiatry, University of Cambridge, Cambridge, UK

[7] Behavioral and Clinical Neuroscience Institute, University of Cambridge, Cambridge, UK

Corresponding author:

Vibe G. Frokjaer, MD, PhD

Neurobiology Research Unit, Rigshospitalet, Section 6932, 9 Blegdamsvej, DK-2100 Copenhagen, Denmark

Phone: (+45) 35456714, Email : Vibe.Frokjaer@nru.dk

Abstract: 227

Manuscript: 4,329

Figures: 2

Tables: 3

Target journal : *1. Neuropsychopharmacology or 2. Psychological Medicine*

Keywords: Cognitive predictors; neuropsychological predictors; depression; Major Depressive Disorder; EMOTICOM; Treatment response; SSRI; Antidepressant; Escitalopram

Abstract

Cognitive disturbances in Major Depressive Disorder (MDD) are critical treatment targets and also holds promise as an early predictor of antidepressant treatment response; yet their clinical relevance as predictors is not fully established. We therefore tested 92 non-psychotic and antidepressant-free patients with a moderate to severe depressive episode with a comprehensive cognitive test battery. Patients were tested before and after 12 weeks (N = 69) of standard antidepressant treatment with escitalopram in flexible doses of 5-20mg (in case of poor response patients were offered to switch to duloxetine). We found no evidence that performance on any single cognitive measure at baseline was associated with clinical response to antidepressant treatment. However, a small cluster of patients with globally disturbed cognition at baseline exhibited poorer clinical response after 8 but not 12 weeks of pharmacological treatment, suggesting that severe cognitive disturbances may delay treatment response. Thus, while individual cognitive outcomes may not be useful as clinical markers of treatment response, our data indicate that cognitive profiles capturing performance across different domains may be useful for stratification of clinically meaningful groups in MDD. Importantly, antidepressant treatment improved cognitive performance across almost all domains. The improvements were not associated with improvement of depressive symptom severity, as captured by the 6-item Hamilton Depressive Rating Scale, emphasizing that cognitive disturbances are a distinct symptom in MDD and therefore constitute an important treatment target.

1. Introduction

Major Depressive Disorder (MDD) is a highly heterogeneous disorder. Despite decades of effort, researchers have yet to identify the etiologies behind MDD and it has been suggested that the diagnosis covers several brain pathologies [1]. This may help explain why 30-50% of patients do not respond adequately to Selective Serotonin-Reuptake Inhibitors (SSRIs) which is the standard first-line treatment for moderate to severe MDD [2]. Importantly, for every failed treatment attempt chances of remission decrease [3], and we therefore face an urgent need for new strategies to optimize antidepressant treatment. Recognizing the complexity and heterogeneity of MDD, many research efforts are now directed towards a shift from a one-size-fit-all treatment approach towards precision medicine. This requires identification of biomarkers that can help stratify patients into clinically meaningful groups [4]. In recent years, disturbances in cognitive functions have been highlighted as a promising candidate for monitoring and even predicting treatment response to antidepressant drugs [5, 6]. Cognitive disturbances are well-documented in MDD and include impairments in cold (emotion-independent) cognitive functions such as processing speed, attention, memory and executive functions [7, 8]. Negative affective biases in hot (emotion-dependent) cognitive functions are also closely associated with depressive psychopathology [9, 10] and may play a key role in antidepressant drug actions [11]. A recent review highlights that impairments in executive functions, and to a lesser extent slowed psychomotor speed, are associated with poor antidepressant treatment outcome [6]. Additionally, results from the large iSPOT-D trial showed that cognitive performance in a small cluster of patients with pronounced cognitive deficits was predictive of treatment response after eight weeks of treatment with the SSRI escitalopram [12]. Meanwhile, early changes in emotion processing biases after administration of antidepressant drugs have shown promise as a tool for guiding clinical decision-making in MDD treatment [13].

Apart from their potential as biomarkers, cognitive disturbances are also a critical treatment target in MDD [14, 15] as impaired cognition negatively impacts patients' everyday functioning and contributes to work presenteeism and absenteeism [16, 17]. Antidepressants appear to have a modest positive effect on cold cognition in patients with MDD [18] and hot cognitive processes in both patients and healthy individuals [19]. However, cognitive disturbances do not always fully resolve with the remission of traditional core symptoms after a depressive episode [20, 21]. For example, a recent meta-analysis reported small to medium impairments in processing speed, learning and memory, attention, and

executive functions in remitted patients relative to healthy individuals [22]. Likewise, disturbances in hot cognitive processes have also been reported to remain in remitted patients [10]. Together, these findings indicate that treatment with antidepressants may alleviate some but not all cognitive symptoms in MDD and further point to a dissociation of core depressive symptoms and cognitive symptoms. So far it, is not clear to what extent cognitive biomarkers can be used to guide clinical decision making in MDD or to what extent cognitive disturbances can be rescued by antidepressant treatment. Further, studies investigating both hot and cold cognitive disturbances in MDD are scarce, making it difficult to map and contrast the effect of antidepressant treatment on different types of cognition. Therefore, we here investigate both hot and cold cognitive functioning in a large cohort of non-psychotic and antidepressant-free patients before and after 12 weeks of standard treatment with SSRIs. We previously described three clusters of distinct cognitive profiles in the present study cohort prior to treatment (Dam *et al.* in revision). Here, we explore the clinical relevance of these clusters in terms of treatment outcome as well as antidepressant effect on cognition.

2. Methods

We here report findings from the cognitive part of the NeuroPharm study; a longitudinal, open-label clinical trial investigating potential biomarkers in antidepressant treatment of MDD. The NeuroPharm study was approved by the National Committee on Health Research Ethics (protocol: H-15017713) and pre-registered at www.clinicaltrial.gov (reg. nr. NCT02869035). For a detailed description of the full trial protocol see Köhler-Forsberg *et al.* (in review).

2.1 Participants

A total of 100 patients with single-episode or recurrent MDD were recruited through a central referral site part of the Mental Health Services in the Capital Region of Denmark or through their general practitioner (see supplementary material for CONSORT flow diagram). Patients with MDD were diagnosis by a trained clinician in accordance with ICD-10 (International Statistical Classification of Disease and Related Health Problems-10) criteria and confirmed with a Mini-International Neuropsychiatric Interview (MINI). To be eligible for inclusion, patients had to be between 18 and 65 years old and have a 17-item Hamilton Depression Rating ($HDRS_{17}$) score > 17 indicating a moderate to severe depressive episode. Patients were only included if the current depressive episode had lasted

less than 2 years and no more than a single antidepressant treatment attempt had been made during the episode. Other exclusion criteria were: use of antidepressants within two months of inclusion; previous non-response to SSRIs; severe somatic illness; history of other primary Axis I psychiatric disorders; substance or alcohol abuse; acute suicidal ideation or psychosis; use of psychotropic medication; history of brain trauma; pregnancy or breastfeeding; and insufficient fluency in Danish. All patients gave written informed consent prior to start of the study.

2.2 Study program

Out of the 100 patients who entered the study, cognitive data was collected from 92 patients at baseline (67 females). After completing the baseline investigative program, patients started standard antidepressant treatment with the SSRI escitalopram at flexible doses of 10-20mg/day. Dosages were adjusted based on effects and side effects evaluated by trained physicians at follow-up visits at week 1, 2, 4, 8 and 12. In accordance with standard practice, patients experiencing severe side effects or showing poor response to escitalopram after 4 weeks of treatment were switched to the Serotonin-Norepinephrine Reuptake Inhibitor (SNRI) duloxetine (n = 16). Follow-up cognitive data was collected from 69 patients after 12 weeks of antidepressant treatment (49 females). Cognitive testing took place in standardized test rooms by trained neuropsychological testers.

Clinical outcomes

Depression symptom severity was assessed with the $HDRS_{17}$ interview at baseline and week 4, 8 and 12. Although $HDRS_{17}$ scores are more widely reported in the literature, we choose to use the 6-item Hamilton Depressive Rating scale ($HDRS_6$) which is a subscale of the $HDRS_{17}$ as it specifically captures core depressive symptoms and has been shown to be more sensitive to antidepressant treatment response [23].

The primary clinical outcome was categorical treatment status at week 8 classified as either 'remitter' or 'non-responder'. Remitter status were defined as early response (≥ 50 % reduction in $HDRS_6$) at week 4 and 5 > $HDRS_6$ score at week 8. Non-responder status was defined as early non-response (< 25% reduction in $HDRS_6$ score) at weak 4 and < 50% reduction in $HDSR_6$ score at week 8. Patients who did not meet criteria for either remitter or non-responder status were classified as 'responders' and were not included in analyses using categorical treatment status as outcome. Secondary clinical

outcomes were relative change in $HDRS_6$ score ($\Delta HDRS_6$ in %) calculated as change in $HDRS_6$ from baseline to follow-up at week 8 or week 12 divided by baseline $HDRS_6$ score.

Cognitive measures

Eleven primary cognitive outcomes were derived from seven neuropsychological tasks capturing emotion processing biases, social cognition and cold cognition. Emotion processing bias outcomes included emotion recognition (hit rate) and misattribution (false alarm rate) from the *Emotional Recognition Task* (eyes version); emotion detection threshold from the *Emotional Intensity Morphing* task [24]; and affective memory from the *Affective Verbal Memory Task 26* (VAMT-26; Hjordt *et al.* in review). The social cognition domain included ratings of guilt and shame from the *Moral Emotions* task and social information preference and social interpretation bias from the *Social Information Preference* task [24]. Lastly, verbal memory was assessed with the VAMT-26; working memory was assessed with the *Letter Number Sequence* task; and reaction time was assessed with a *Simple Reaction Time* task. A full description of cognitive tasks and task outcomes are described elsewhere (Dam *et al.*, in revision).

Cognitive profiles clusters

Using a data-driven cluster analysis approach, we previously identified three distinct cognitive profiles based on cognitive performance at baseline in the present MDD cohort (see Dam *et al,* in revision). Cluster A (n = 38) was characterized by strong negative affective biases in emotion processing; Cluster B (n = 28) was characterized by positive affective biases in emotion processing and moderate deficits in cold cognitive domains; lastly Cluster C (n = 26) was characterized by large global disturbances across all cognitive domains.

2.3 Statistical analyses

2.3.1 Cognitive performance at baseline and antidepressant treatment response

We used a series of logistic regression models to test the association between cognitive performance at baseline and the primary clinical outcome of treatment status (remitter vs non-responder) at week 8. In addition, we used a series of linear regression models to test the association between cognitive

performance at baseline and the secondary clinical outcomes of week 8 and week 12 $\Delta HDRS_6$ scores. Sex and age were included in all models.

2.3.2 Antidepressant treatment response in cognitive profile clusters

We used a χ^2 test to assess whether the three cognitive profile clusters differed in terms of treatment status (remitter vs non-responder) at week 8. Additionally, one-way ANCOVA analyses including sex and age as co-variates were used to assess group difference between the cognitive profile clusters on week 8 and week 12 $\Delta HDRS_6$ scores.

2.3.2 Antidepressant treatment effects on cognition

We used linear mixed-models to assess changes in cognitive performance between baseline and week 12. In addition, we used one-way ANCOVA analyses to test whether these changes differed between the three cognitive profile clusters. Sex and age were included in all models.

2.3.3 Link between change in cognition and change in $HDRS_6$ core depressive symptom severity

We used Spearman's rank correlation coefficients to assess whether change in cognitive performance was correlated with change in clinical symptoms at week 12.

2.3.4 Missing data, outliers and correction for multiple comparisons

Missing values including patient data lost to follow-up were not imputed but instead treated as missing in analyses. Outliers were defined as datapoints > 1.5 interquartile range above or below the 1st and 3rd quartile respectively. All outliers were individually inspected and cross-checked against notes from the testing session to ensure they were not caused by testing-related errors. As a result, one baseline score from the *Letter Number Sequence* task was excluded as the patient had misunderstood the task instruction. In addition, two scores on the *Simple Reaction Time* task were so extreme (8.1 and 16.1 IQR above the 1st quartile) that they were capped to one and two units above the third highest score respectively, allowing the scores to keep their ranking without skewing the estimates unduly. As exclusion of other outliers did not alter the results substantially, they were kept in all analyses. In addition, removing the 16 patients who were switched from escitalopram to duloxetine from the analysis did not alter the results substantially and consequently they were included in all analyses. We

used the Bonferroni-Holm method to control the family-wise error rate when testing hypotheses that included the 11 cognitive outcomes. Corrected *p*-values are denoted $p_{corrected}$ and have been corrected for 11 tests. Statistical analyses were conducted in SPSS (v 25).

3. Results

Patients were between 18 and 57 years old (mean = 27.3; SD = 8.1). Baseline $HDRS_6$ scores ranged from 7 to 17 (mean = 12.3; SD = 1.6; N = 92); week 8 $HDRS_6$ scores ranged from 0 to 16 (mean = 6.0; SD = 3.8; N = 79); and week 12 $HDRS_6$ scores ranged from 0 to 14 (mean = 4.8; SD = 3.7; N = 69). At week 8, 20 patients (25.6%) fulfilled the criteria for remitter status; 44 patients (56.4%) fulfilled the criteria for responder status; and 14 patients (17.9%) fulfilled the criteria for non-responder status.

3.1 Cognitive performance at baseline and antidepressant treatment response

Table 1 shows associations between cognitive performance at baseline and treatment status at week 8 as well as week 8 and week 12 $\Delta HDRS_6$ scores.

Table 1 here

We observed no statistically significant association between baseline cognitive performance and treatment status (all $p_{corrected} = 1.0$). Nor was there an association between baseline cognitive performance and change in clinical symptom severity at week 8 (all $p_{corrected} = 1.0$) or week 12 (all $p_{corrected} = 1.0$).

3.2 Antidepressant treatment response in cognitive profile clusters

We observed no statistically significant difference between the three cognitive profile clusters in terms of remitters vs non-responder status after 8 weeks of antidepressant treatment ($\chi^2(2, N = 34) = 3.3$, $p = 0.2$). In Cluster A, 11 patients (28.9%) were remitters and 5 patients (13.2%) were non-responders; in Cluster B, 6 patients (21.4%) were remitters and 3 patients (10.1%) were non-responders; in Cluster C, 3 patients (11.5%) were remitters and 6 patients (23.1%) were non-responders.

Table 2 shows baseline $HDRS_6$ scores and week 8 and week 12 $\Delta HDRS_6$ for the three clusters.

Table 2 here

At group level, there was a statistically significant difference between the three clusters on week 8 $\Delta HDRS_6$ scores ($F(2, 75) = 3.7$, $p = 0.03$) but not on week 12 $\Delta HDRS_6$ scores ($F(2, 72) = 0.3$, $p = 0.8$).

Post hoc analyses showed that Cluster C had worse response to antidepressant treatment at week 8 than Cluster A ($p = 0.009$).

3.3 Antidepressant treatment effect on cognition

Figure 2 shows difference in cognitive performance at baseline and after 12 weeks of antidepressant treatment.

Figure 1 here

At group level, affective bias for emotion recognition increased (i.e. became more positive) by 11.1 percentage points (95% CI = [6.0;16.2], $p_{\text{corrected}} < 0.001$); affective bias for emotion misattribution increased by 7.0 percentage points (95% CI = [3.5;10.6], $p_{corrected} = 0.002$); and affective bias for emotion detection threshold increased by 5.5 percentage points (95% CI = [2.4;8.6], $p_{corrected} = 0.007$). Meanwhile, a weak, statistically non-significant increase of 1.7 percentage point was observed for affective memory bias (95% CI = [-2.3;5.7], $p_{corrected} = 1.0$). Ratings of negative moral emotions decreased by 0.2 points on a seven-point Likert scale for both guilt (95% CI = [-0.3;-0.1], $p_{corrected} < 0.001$) and shame (95% CI = [-0.4;-0.1], $p_{corrected} = 0.01$). Preference for social information showed a weak, statistically non-significant decrease of 4.9 percentage points (95% CI = [-9.7;-0.1], $p_{corrected} = 0.5$) while social interpretation bias showed a weak, statistically non-significant increase of 5.8 percentage point (95% CI = [0.3;11.2], $p_{corrected} = 0.4$). Lastly, overall memory capacity increased by 2.6 words (max score 26 words; 95% CI = [2.0;3.2], $p_{corrected} < 0.001$); working memory increased by 1.3 points (max score 21 points; 95% CI = [0.6;1.9], $p_{corrected} = 0.001$); and reaction time decreased by 14.1 milliseconds (95% CI = [-22.2;-5.9], $p_{corrected} = 0.01$).

Figure 2 shows differences in cognitive changes from baseline to week 12 for the three cognitive profile clusters.

Figure 2 here

At group level, the three cognitive profile clusters differed on changes in ratings of guilt ($F(2, 65) = 13.2$, $p_{corrected} < 0.001$) and shame ($F(2, 65) = 8.6$, $p_{corrected} < 0.001$). *Post hoc* analysis indicated that Cluster C experienced the biggest decrease in guilt ratings compared with both Cluster A ($p < 0.001$) and Cluster B ($p = 0.03$) while Cluster B experienced a bigger decrease than Cluster A ($p = 0.02$). Similarly, Cluster C experienced a bigger decrease in shame ratings than both Cluster A ($p < 0.001$) and Cluster B ($p = 0.01$). Meanwhile, group level trends were observed for several other cognitive

outcomes including emotion recognition bias ($p = 0.05$, $p_{corrected} = 0.6$), emotion misattribution bias ($p = 0.03$, $p_{corrected} = 0.3$), and working memory ($p = 0.06$, $p_{corrected} = 0.7$).

3.4 Link between change in cognition and change in $HDRS_6$ core depressive symptom severity

Table 3 shows correlations between changes in cognition between performance from baseline to week 12 and week 12 $\Delta HDRS_6$ scores.

Table 3 here

There was no statistically significant correlation between changes in cognitive performance and week 12 $\Delta HDRS_6$ scores (rho = [-0.21;0.16], all $p_{corrected} = 1.0$).

4. Discussion

We here present results from the NeuroPharm trial investigating associations between cognitive disturbances and antidepressant treatment effects in patients with a moderate to severe depressive episode. Antidepressant treatment significantly improved cognitive performance across a range of both hot and cold cognitive functions. Our findings did not, however, support that pre-treatment performance on single cognitive tests are clinically useful as markers of antidepressant treatment response in MDD. Patients characterized by global cognitive disturbances at baseline exhibited less improvement in clinical symptoms at week 8 compared with other cognitive profile clusters, though this difference was no longer detectable after 12 weeks of treatment. While the improvements in cognitive performance were not correlated with improvement in clinical symptoms ($HDRS_6$) the degree of cognitive improvement did differ between the three cognitive profile clusters on several outcomes.

4.1 Cognitive disturbances as marker of treatment response

We found no evidence that performance on any single cognitive outcome at baseline was associated with remission or non-responder status at week 8. Nor did we observe any associations between pre-treatment cognitive performance and changes in depressive symptoms after 8 or 12 weeks of antidepressant treatment. Most previous studies investigating disturbances in emotion processing and antidepressant treatment response have focused on early changes (e.g., change in performance from baseline to week 1) as a predictive marker [25]. Thus, the present study represents the first comprehensive investigation showing that pre-treatment disturbances in a range of hot cognitive

outcomes are *not* associated with later escitalopram treatment response in moderate to severe MDD. A recent review found that deficits in cold cognitive domains including executive and psychomotor functions are predictive of antidepressant treatment response in MDD [6]. However, the evidence was only robust for elderly patients whereas the literature on younger patients was highly conflicted. Together with our negative finding, this suggests that even if a single cognitive function has some predictive value, it is likely too limited to be clinically relevant. Instead, cognitive profiles capturing patterns of performance across several cognitive domains may provide a much stronger predictive construct. In the large iSPOT-D trial (baseline, N = 1008; completers, N = 665), Etkin, Patenaude (12) identified two patient subgroups using cluster analysis: one with intact cognitive functions (~75% of patients) and one with broadly impaired cognitive functions (~ 25% of patients). Importantly, the study found that the impaired patient group had poorer clinical response after 8 weeks of antidepressant treatment and that treatment response could be predicted by baseline performance within the impaired group for patients who received treatment with escitalopram [12]. Using a similar data-driven clustering approach, we previously identified three clusters with distinct cognitive profiles in the NeuroPharm cohort (Dam *et al.* in revision). Notably, our patients from Cluster C (~28% of patients), who were characterized by severe global deficits across all cognitive domains, had poorer clinical treatment response to 8 weeks of SSRI treatment, mirroring the findings from the iSPOT-D trial. However, this difference in clinical response was no longer detectable after 12 weeks with Cluster C patients 'catching up' to both Cluster A and Cluster B patients (see Table 2). This observation aligns with another smaller study which also identified a subgroup (26%) of patients with impaired cognitive functions and found no difference in clinical scores between the intact and impaired groups at 12 months follow-up [26]. Together, this suggests that while global cognitive impairments may slow or delay treatment response, it does not necessarily affect longer-term treatment outcome. One implication of this is that clinicians may consider waiting longer before switching medication if a patient with global cognitive disturbances do not respond to the first-line SSRI treatment within 4-8 weeks.

4.2 Antidepressant treatment effect on cognition

We found significant improvements in both hot and cold cognitive domains after 12 weeks of escitalopram treatment including biases in emotion recognition, misattribution and detection; ratings of guilt and shame; verbal and working memory; and reaction time. Meanwhile, there was no significant

change in performance for affective bias in verbal memory; social information preference; or social interpretation bias. This is perhaps not surprising, as we did not observe any disturbances on these task outcomes when we compared the same cohort of MDD patients with healthy controls at baseline (see Dam et al., in revision), suggesting that they may not be sensitive and/or relevant to MDD pathology. Overall, our findings align with previous reports that antidepressant treatment improve cognition across both hot [19] and cold domains [18]. Interestingly, the three cognitive clusters differed significantly on changes in cognitive performance over the course of treatment on ratings of guilt and shame in the *Moral Emotions* tasks. Graphically, the largest decrease in self-reported guilt and shame was exhibited by Cluster C patients who also had the highest ratings at baseline. A similar pattern could also be seen for the other cognitive outcomes showing group differences at a trend level (i.e. emotion recognition, emotion detection and working memory) where the most severely disturbed cluster(s) appear to exhibit the greatest improvement over time (see Figure 2). We therefore speculate that improvements in cognitive performance may be mediated by the magnitude of pre-treatment disturbances. This in turn suggests that antidepressants are not cognitive enhancers *per se* but rather act by normalizing impaired cognitive functions. It also aligns with reports of little to no effect of antidepressant intervention on (cold) cognition in healthy individuals [18]. To our knowledge, the presence of a severity-dependent antidepressant effect on cognitive disturbances in MDD has not previously been documented empirically. If confirmed, it could help explain the large heterogeneity in the literature since the ability to detect antidepressant effects on cognition would largely depend on the distribution of cognitively impaired vs cognitive intact patients in the study cohort.

4.3 Dissociation between cognitive disturbances and core depressive symptoms in MDD

Any correlation between improvements in core depressive symptoms ($HDRS_6$) and changes in cognitive performance over the course of antidepressant treatment was weak and statistically non-significant. Together with our previous observation of no association between cognitive performance at baseline and depression severity (Dam et al. in revision), this strongly suggests a dissociation between cognitive disturbances and core depressive symptoms in MDD. We therefore argue that disturbed cognition should be seen as a distinct and independent symptom in depression and not merely as an epiphenomenon (i.e. extension) of mood, anhedonia, decreased energy, and somatic symptoms. This claim is supported by other large clinical studies which also found no or only partial overlap between

treatment effects on (cold) cognition and core depressive symptoms [27, 28]. Consequently, this interpretation raises the intriguing possibility of parallel mechanisms of drug action for cognitive and mood modalities in MDD, which would need to be verified in future studies. Meanwhile, a contrasting view of antidepressant drug action in MDD is offered by the cognitive neuropsychological model of depression. The model posits that antidepressant drugs act by acutely remediating negative affective biases via neuromodulatory mechanisms in limbic and frontal regions. Over time, the changes in affective biases enable positive restructuring of dysfunctional cognitive and psychological processes, which ultimately leads to alleviation of mood symptoms [29]. While not in direct conflict with the prediction that early improvements in affective biases predict later treatment response, our observation that longer-term improvements in affective biases are *not* related to clinical improvement does not lend support to the cognitive neuropsychological model of depression. Rather, our findings suggest that the interaction between cognition and core depressive symptoms over the course of antidepressant treatment may be more complex than previously thought.

4.4 Methodological limitations

First, the present study did not include a healthy control group and/or placebo group. This means that we cannot account for any potential learning effect in cognitive performance and may consequently overestimate the antidepressant effects on cognition. While this is likely not an issue for the hot cognitive domains, as none of the task outcomes contain learning or practice aspects [30], verbal memory and working memory domains are more vulnerable to effects of repeated testing. Second, as we investigated the effects of escitalopram (and duloxetine as a second line treatment), our findings may not be generalizable to treatments with other antidepressants drugs. Third, although the inclusion of both hot and cold cognitive tasks is a notable strength of the study, it also limited the number of tasks we could include. Thus, we were unable to collect data on several cognitive domains known to be relevant in MDD pathology, e.g. attention, reward and motivation processing, and higher-order executive functions [31].

4.5 Implications and future perspectives

Our findings emphasize not only the complexity of cognitive disturbances in depression but also their importance as a distinct symptom and therefore treatment target in MDD. They also raise concern that

the clinical scales used to measure and monitor depressive symptoms fail to adequately capture cognitive symptoms in MDD. Lastly, our findings show that while individual cognitive outcomes may not be clinically useful as markers of treatment response, cognitive profiles that map performance across a wide range of cognitive domains may be useful stratification tools in MDD. For example, our findings suggest that antidepressant treatment response is delayed in patients with global cognitive disturbances which could impact clinical treatment choices. Importantly, future studies should investigate whether such cognitive profiles relate to biological phenotypes (e.g. neuroimaging characteristics) and whether patients with a certain profile may respond better to specific antidepressant drugs or non-pharmacological treatments. It would also be relevant to investigate whether early treatment response may be improved in the subgroup of patients with global dysfunction through antidepressant treatment augmented with cognitive remediation training or cognition-enhancing drugs [32].

Funding and Disclosure

This study was supported by the Augustinus Foundation (grant 16-0058), Rigshospitalet's Research Council (grant R149-A6325), and the Innovation Fund Denmark (grant 4108-00004B). The financial supporters were not involved in the study design, collection, analysis, interpretation, or publication of data. BJS is a co-inventor of CANTAB and EMOTICOM test batteries and consults for Cambridge Cognition, Greenfield BioVentures, and Cassava Sciences. VGF and GMK have received speakers honorarium from Janssen Pharmaceuticals.

Acknowledgments

We would like to thank the patients and healthy participants for their time. The data from the study is available upon request from the CIMBI database (http://www.cimbi.dk/db).

References

1. Pitsillou E, Bresnehan S, Kagarakis E, Wijoyo S, Liang J, Hung A, et al. The cellular and molecular basis of major depressive disorder: towards a unified model for understanding clinical depression. Molecular Biology Reports. 2019 10/14.

2. Trivedi MH, Rush AJ, Wisniewski SR, Nierenberg AA, Warden D, Ritz L, et al. Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: implications for clinical practice. Am J Psychiatry. 2006 Jan;163(1):28-40. PubMed PMID: 16390886. Epub 2006/01/05. eng.
3. Rush AJ, Trivedi MH, Wisniewski SR, Nierenberg AA, Stewart JW, Warden D, et al. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. Am J Psychiatry. 2006 Nov;163(11):1905-17. PubMed PMID: 17074942. Epub 2006/11/01. eng.
4. Collins PY, Patel V, Joestl SS, March D, Insel TR, Daar AS, et al. Grand challenges in global mental health. Nature. 2011 Jul 6;475(7354):27-30. PubMed PMID: 21734685. Pubmed Central PMCID: Pmc3173804. Epub 2011/07/08. eng.
5. Park C, Pan Z, Brietzke E, Subramaniapillai M, Rosenblat JD, Zuckerman H, et al. Predicting antidepressant response using early changes in cognition: A systematic review. Behav Brain Res. 2018 Nov 1;353:154-60. PubMed PMID: 30031025. Epub 2018/07/22. eng.
6. Groves SJ, Douglas KM, Porter RJ. A Systematic Review of Cognitive Predictors of Treatment Outcome in Major Depression. Front Psychiatry. 2018;9:382. PubMed PMID: 30210368. Pubmed Central PMCID: Pmc6121150. Epub 2018/09/14. eng.
7. Rock PL, Roiser JP, Riedel WJ, Blackwell AD. Cognitive impairment in depression: a systematic review and meta-analysis. Psychol Med. 2014 Jul;44(10):2029-40. PubMed PMID: 24168753. Epub 2013/10/31. eng.
8. Ahern E, Semkovska M. Cognitive functioning in the first-episode of major depressive disorder: A systematic review and meta-analysis. Neuropsychology. 2017 Jan;31(1):52-72. PubMed PMID: 27732039. Epub 2016/10/28. eng.
9. Elliott R, Zahn R, Deakin JF, Anderson IM. Affective cognition and its disruption in mood disorders. Neuropsychopharmacology. 2011 Jan;36(1):153-82. PubMed PMID: 20571485. Pubmed Central PMCID: Pmc3055516. Epub 2010/06/24. eng.
10. Miskowiak KW, Carvalho AF. 'Hot' cognition in major depressive disorder: A systematic review. CNS and Neurological Disorders - Drug Targets. 2014;13(10):1787-803.
11. Harmer CJ, Duman RS, Cowen PJ. How do antidepressants work? New perspectives for refining future treatment approaches. The Lancet Psychiatry. 2017 2017/05/01/;4(5):409-18.
12. Etkin A, Patenaude B, Song YJ, Usherwood T, Rekshan W, Schatzberg AF, et al. A cognitive-emotional biomarker for predicting remission with antidepressant medications: a report from the iSPOT-D trial. Neuropsychopharmacology. 2015 May;40(6):1332-42. PubMed PMID: 25547711. Pubmed Central PMCID: Pmc4397406. Epub 2014/12/31. eng.
13. Browning M, Kingslake J, Dourish CT, Goodwin GM, Harmer CJ, Dawson GR. Predicting treatment response to antidepressant medication using early changes in emotional processing. European Neuropsychopharmacology. 2019 2019/01/01/;29(1):66-75.
14. Kaser M, Zaman R, Sahakian BJ. Cognition as a treatment target in depression. Psychol Med. 2017 Apr;47(6):987-9. PubMed PMID: 27938430. Epub 2016/12/13. eng.
15. Medicine Io, National Academies of Sciences E, Medicine. Enabling Discovery, Development, and Translation of Treatments for Cognitive Dysfunction in Depression: Workshop Summary. Bain L, Stroud C, editors. Washington, DC: The National Academies Press; 2015. 90 p.
16. McIntyre RS, Cha DS, Soczynska JK, Woldeyohannes HO, Gallaugher LA, Kudlow P, et al. Cognitive deficits and functional outcomes in major depressive disorder: determinants, substrates, and treatment interventions. Depress Anxiety. 2013 Jun;30(6):515-27. PubMed PMID: 23468126. Epub 2013/03/08. eng.
17. Knight MJ, Baune BT. Cognitive dysfunction in major depressive disorder. Current Opinion in Psychiatry. 2018;31(1):26-31. PubMed PMID: 00001504-201801000-00006.
18. Prado CE, Watt S, Crowe SF. A meta-analysis of the effects of antidepressants on cognitive functioning in depressed and non-depressed samples. 2018 Mar;28(1):32-72. PubMed PMID: 29446012.

19. Harmer CJ, Cowen PJ. 'It's the way that you look at it'--a cognitive neuropsychological account of SSRI action in depression. Philosophical transactions of the Royal Society of London Series B, Biological sciences. 2013;368(1615):20120407.
20. Bora E, Harrison BJ, Yucel M, Pantelis C. Cognitive impairment in euthymic major depressive disorder: a meta-analysis. Psychol Med. 2013 Oct;43(10):2017-26. PubMed PMID: 23098294. Epub 2012/10/27. eng.
21. Hasselbalch BJ, Knorr U, Kessing LV. Cognitive impairment in the remitted state of unipolar depressive disorder: a systematic review. J Affect Disord. 2011 Nov;134(1-3):20-31. PubMed PMID: 21163534. Epub 2010/12/18. eng.
22. Semkovska M, Quinlivan L, O'Grady T, Johnson R, Collins A, O'Connor J, et al. Cognitive function following a major depressive episode: a systematic review and meta-analysis. Lancet Psychiatry. 2019 Oct;6(10):851-61. PubMed PMID: 31422920. Epub 2019/08/20. eng.
23. Ostergaard SD, Bech P, Miskowiak KW. Fewer study participants needed to demonstrate superior antidepressant efficacy when using the Hamilton melancholia subscale (HAM-D(6)) as outcome measure. J Affect Disord. 2016 Jan 15;190:842-5. PubMed PMID: 25487682. Epub 2014/12/10. eng.
24. Bland AR, Roiser JP, Mehta MA, Schei T, Boland H, Campbell-Meiklejohn DK, et al. EMOTICOM: A Neuropsychological Test Battery to Evaluate Emotion, Motivation, Impulsivity, and Social Cognition. Front Behav Neurosci. 2016;10:25. PubMed PMID: 26941628. Pubmed Central PMCID: PMC4764711. Epub 2016/03/05. eng.
25. Harmer CJ, O'Sullivan U, Favaron E, Massey-Chase R, Ayres R, Reinecke A, et al. Effect of acute antidepressant administration on negative affective bias in depressed patients. Am J Psychiatry. 2009 Oct;166(10):1178-84. PubMed PMID: 19755572. Epub 2009/09/17. eng.
26. Vicent-Gil M, Keymer-Gausset A, Serra-Blasco M, Carceller-Sindreu M, de Diego-Adelino J, Trujols J, et al. Cognitive predictors of illness course at 12 months after first-episode of depression. European neuropsychopharmacology : the journal of the European College of Neuropsychopharmacology. 2018 Apr;28(4):529-37. PubMed PMID: 29482974. Epub 2018/02/28. eng.
27. McIntyre RS, Lophaven S, Olsen CK. A randomized, double-blind, placebo-controlled study of vortioxetine on cognitive function in depressed adults. Int J Neuropsychopharmacol. 2014;17(10):1557-67. PubMed PMID: 24787143. Epub 04/30. eng.
28. Shilyansky C, Williams LM, Gyurak A, Harris A, Usherwood T, Etkin A. Effect of antidepressant treatment on cognitive impairments associated with depression: a randomised longitudinal study. Lancet Psychiatry. 2016 May;3(5):425-35. PubMed PMID: 26995298. Pubmed Central PMCID: Pmc4860142. Epub 2016/03/21. eng.
29. Roiser JP, Elliott R, Sahakian BJ. Cognitive mechanisms of treatment in depression. Neuropsychopharmacology. 2012 Jan;37(1):117-36. PubMed PMID: 21976044. Pubmed Central PMCID: Pmc3238070. Epub 2011/10/07. eng.
30. Dam VH, Thystrup CK, Jensen PS, Bland AR, Mortensen EL, Elliott R, et al. Psychometric Properties and Validation of the EMOTICOM Test Battery in a Healthy Danish Population. Frontiers in Psychology. 2019 2019-December-03;10(2660). English.
31. Roiser JP, Sahakian BJ. Hot and cold cognition in depression. CNS spectrums. 2013;18(3):139-49. Epub 03/12.
32. Goss AJ, Kaser M, Costafreda SG, Sahakian BJ, Fu CH. Modafinil augmentation therapy in unipolar and bipolar depression: a systematic review and meta-analysis of randomized controlled trials. J Clin Psychiatry. 2013 Nov;74(11):1101-7. PubMed PMID: 24330897. Epub 2013/12/18. eng.

Table 1. Association between baseline cognition and treatment response at week 8

	Treatment status: remitters vs non-responders					Week 8 $\Delta HDRS_6$			Week 12 $\Delta HDRS_6$		
	β	Exp(β)	Exp(β) 95% CI	p	$p_{corrected}$	β	p	$p_{corrected}$	β	p	$p_{corrected}$
Emotion processing											
Recognition bias	0.01	1.01	[0.97–1.05]	0.7	1.0	0.02	0.9	1.0	0.02	0.9	1.0
Misattribution bias	-0.01	0.99	[0.94–1.05]	0.8	1.0	0.15	0.5	1.0	0.22	0.3	1.0
Detection bias	-0.002	1.00	[0.92–1.08]	1.0	1.0	-0.09	0.7	1.0	0.21	0.4	1.0
Affective memory	0.003	1.00	[0.94–1.07]	0.9	1.0	0.20	0.5	1.0	0.19	0.5	1.0
Social cognition											
Guilt ratings	-0.68	0.50	[0.13–2.04]	0.3	1.0	7.18	0.2	1.0	-3.25	0.6	1.0
Shame ratings	-0.07	0.93	[0.28–3.12]	0.9	1.0	4.55	0.3	1.0	0.66	0.9	1.0
Information sampling	0.004	1.00	[0.96–1.05]	0.8	1.0	0.11	0.6	1.0	0.02	0.9	1.0
Social interpretation bias	0.02	1.02	[0.97–1.07]	0.4	1.0	-0.17	0.4	1.0	0.00	1.0	1.0
Cold cognition											
Verbal memory	0.09	1.09	[0.87–1.38]	0.5	1.0	-0.70	0.4	1.0	-0.39	0.7	1.0
Working memory	0.02	1.02	[0.75–1.37]	0.9	1.0	0.50	0.7	1.0	1.00	0.5	1.0
Reaction time	0.0001	1.00	[0.98–1.02]	1.0	1.0	0.08	0.2	1.0	0.05	0.5	1.0

Table 1. Logistic regression showing association between cognitive scores at baseline and treatment status (remitter vs non-responder) at week 8. Also reported are linear regressions showing association between cognitive scores at baseline and relative change in $HDRS_6$ scores from baseline to week 8 and week 12. Age and sex are included in all models and *p*-values have been corrected for 11 tests in accordance with the Bonferroni-Holm method.

Table 2. Antidepressant treatment response for cognitive profiles

	Cluster A			Cluster B			Cluster C		
	Mean ± SD	Range	N	Mean ± SD	Range	N	Mean ± SD	Range	N
Age in years	26.6 ± 9.1	18–57	38	28.6 ± 8.8	20–56	28	26.8 ± 5.8	18–40	26
$HDRS_6$ score	12.2 ± 1.6	8–17	38	11.9 ± 1.3	7–14	28	13.1 ± 1.9	9–17	26
Week 8 $\Delta HDRS_6$	58.8 ± 29.4	-8.3–100	34	53.0 ± 29.8	-18.2–92.3	23	36.8 ± 29.0	14.3–100	22
Week 12 $\Delta HDRS_6$	62.3 ± 32.5	-33.3–100	33	59.9 ± 31.0	-8.3–100	22	56.7 ± 31.6	0–100	21

Table 3. Age and depression symptom severity indexed with the Hamilton Depression Rating subscale 6 ($HDRS_6$) at baseline are reported. Relative change in $HDRS_6$ scores ($\Delta HDRS_6$) between baseline and week 8 and baseline and week 12 are also shown. Note, $\Delta HDRS_6$ scores represent improvement (i.e. decrease in symptom severity in %).

Table 3. Correlation between change in cognitive score and week 12 $\Delta HDRS_6$

	Week 12 $\Delta HDRS_6$		
	rho	*p*	$p_{corrected}$
Emotion processing			
ΔRecognition bias	-0.09	0.5	1.0
ΔMisattribution bias	-0.21	0.1	1.0
ΔDetection bias	-0.05	0.7	1.0
ΔAffective memory	-0.13	0.3	1.0
Social cognition			
ΔGuilt ratings	0.16	0.2	1.0
ΔShame ratings	0.09	0.5	1.0
ΔInformation sampling	-0.01	1.0	1.0
ΔSocial interpretation bias	0.14	0.3	1.0
Cold cognition			
ΔVerbal memory	0.05	0.7	1.0
ΔWorking memory	-0.03	0.8	1.0
ΔReaction time	0.09	0.4	1.0

Table 2. Correlation between absolute changes in cognitive scores from baseline to Week 12 and changes in depressive symptom severity indexed as relative change in $HDRS_6$ scores from baseline to week 12. *p*-values have been corrected for 11 tests in accordance with the Bonferroni-Holm method.

Figure 1. Antidepressant treatment effects on cognition

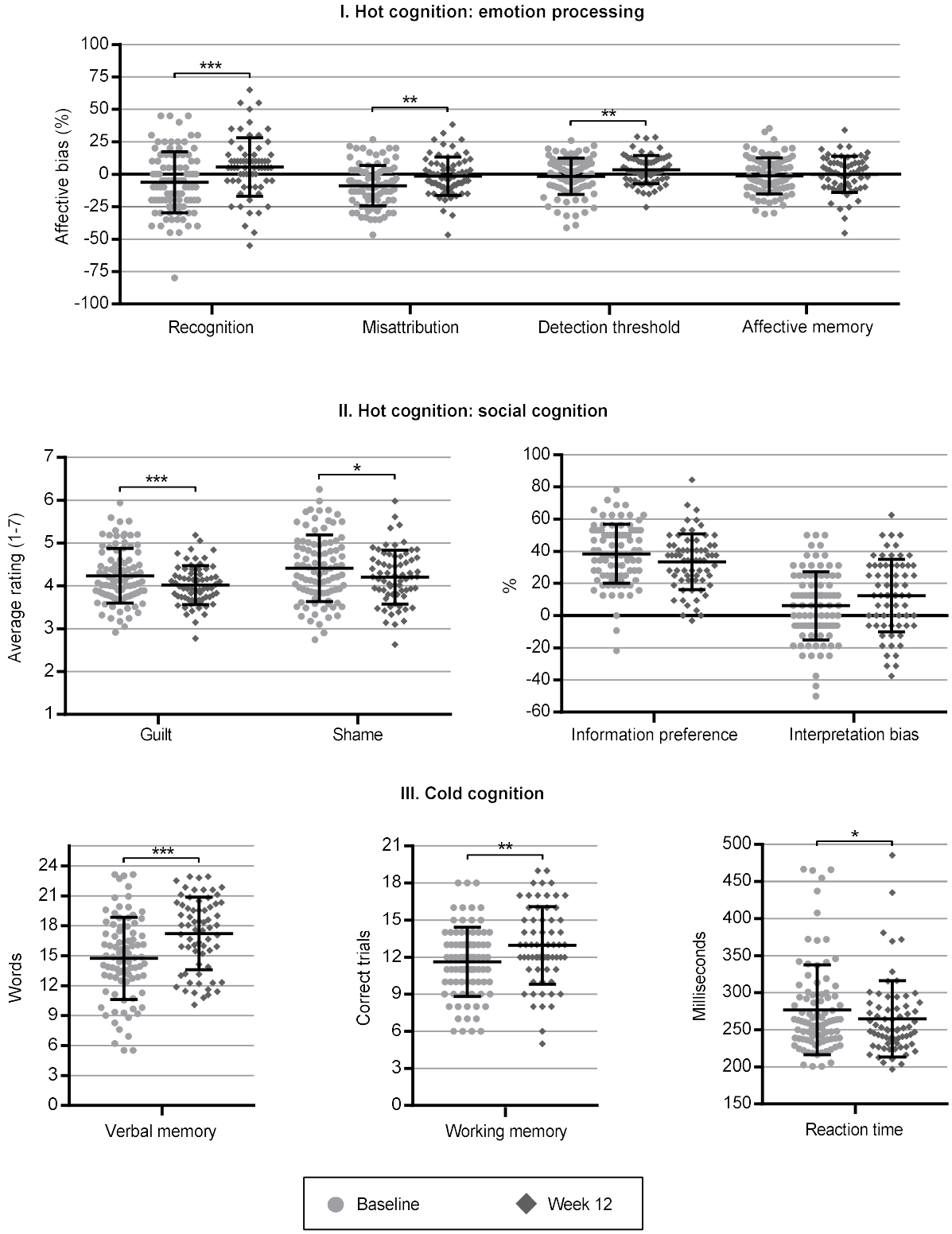


Figure 2. Changes in cognitive performance from baseline to week 12 across cognitive profile clusters

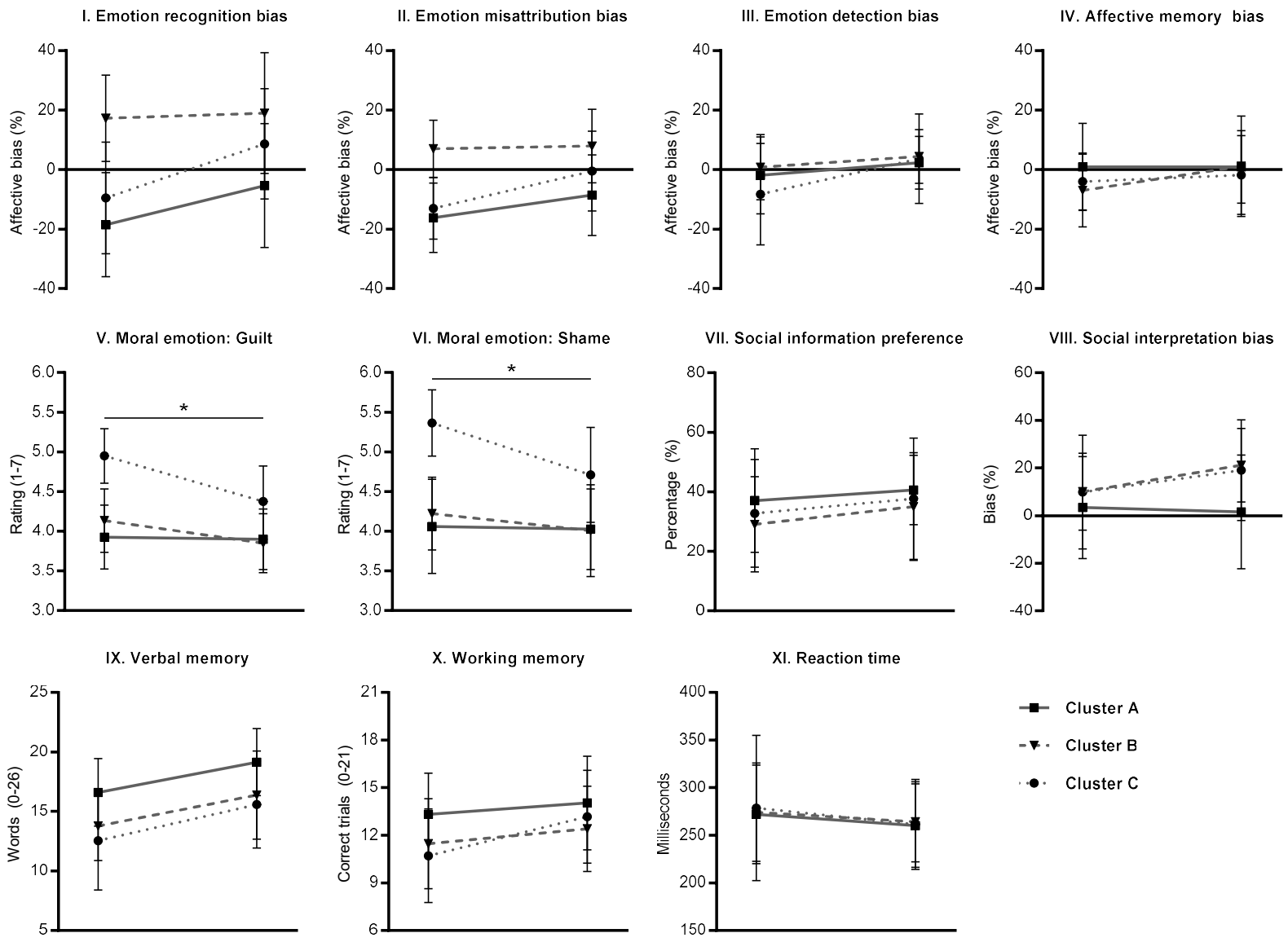


Figure 1. Group differences on affective, social, and cold cognitive outcomes between depressed patients and healthy controls. I. Affective cognition: *Recognition = affective bias for hit rate in the Emotional Recognition Task; Misattribution = affective bias for false alarm rate in the Emotional Recognition Task; Detection threshold = affective bias for the Intensity Morphing Task; Affective memory = affective bias for the Verbal Affective Memory Task 26.* II. Social cognition: *Guilt = average ratings of guilt in the Moral Emotions task; Shame = average ratings of shame in the Moral Emotions task; Information preference = choice of theory of mind-related information relative to facts in the Social Information Preference task; Interpretation bias = affective bias in choice of outcome in the Social Information Preference task.* III. Cold cognition: *Verbal memory = Total recall score for the Verbal Affective Memory Task; Working memory = Letter-Number Sequence task; Reaction time = Simple Reaction Time.* All models were corrected for age and sex. * $p < .05$, ** $p < .01$, *** $p < .001$.

Figure 2. Difference in magnitude of changes in cognitive performance from baseline to week 12 for the three cognitive clusters. The graphs represent mean raw scores at baseline and follow-up at week 12 and the error bars denote standard deviations. Significant main effect of changes in cognitive performance was observed at group level for guilt and shame ratings in the *Moral Emotions task* (graph V and VI). * $p < .05$

CONSORT

TRANSPARENT REPORTING of TRIALS