Psychological Bulletin

Towards a Hierarchical Model of Social Cognition: A Neuroimaging Meta-Analysis and Integrative Review on Empathy and Theory of Mind

--Manuscript Draft--

Manuscript Number:	BUL-2019-1649R2
Full Title:	Towards a Hierarchical Model of Social Cognition: A Neuroimaging Meta-Analysis and Integrative Review on Empathy and Theory of Mind
Abstract:	In parallel with the rise in interest and volume of social cognition research, there has been increasing awareness of a lack of agreement on the concepts and taxonomy used to study social processes. Two central concepts in the field, Empathy and Theory of Mind (ToM), have been applied as overlapping umbrella terms for a variety of different processes of limited convergence. Here, we review and integrate evidence of brain activation, brain organization, and behavior into a coherent model of social cognitive processes. We start with a meta-analytic clustering of neuroimaging data across different social cognitive tasks. Results show that understanding others' mental states can be described by a multilevel model of hierarchical structure, similar to models in intelligence and personality research. A higher level describes more broad and abstract classes of functioning, whereas a lower level explicates how functions are applied to concrete contexts, given by specific stimulus and task formats. Specifically, the higher level of our model suggests three groups of neurocognitive processes: (i) Predominantly cognitive processes that are engaged when mentalizing requires self-generated cognition decoupled from the physical world. (ii) More affective processes which are engaged when we witness emotions in others, based on shared emotional, motor, and somatosensory representations. (iii) Combined processes which engage cognitive and affective functions in parallel. We discuss how these processes are explained by an underlying principal gradient of structural brain organization. Finally, we validate the model by a review of Empathy and ToM task interrelations found in behavioral studies.
Article Type:	Article
Keywords:	social cognition; mentalizing; Perspective Taking; Emotion; Mirror Neurons
Corresponding Author:	Matthias Schurz, Ph.D., Mag., University of Oxford Oxford, UNITED KINGDOM
Corresponding Author E-Mail:	matthias.schurz@gmail.com
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	University of Oxford
Other Authors:	Joaquim Radua
	Matthias G. Tholen
	Lara Maliske
	Daniel S. Margulies
	Rogier B. Mars
	Jerome Sallet
	Philipp Kanske
Corresponding Author's Secondary Institution:	
First Author:	Matthias Schurz
Order of Authors Secondary Information:	
Manuscript Region of Origin:	UNITED KINGDOM

Suggested Reviewers:	Frank Van Overwalle, Ph.D. Prof., Free University of Brussels Frank.VanOverwalle@vub.ac.be Expertise in: Neuroimaging meta-analysis, Social cognition, Theory of Mind, Goal understanding, Mirror System, Neuroimaging research. Selected publications: Van Overwalle, F. (2009). Social cognition and the brain: a meta-analysis. Human brain mapping, 30(3), 829-858. Van Overwalle, F., & Baetens, K. (2009). Understanding others' actions and goals by mirror and mentalizing systems: a meta-analysis. Neuroimage, 48(3), 564-584.
	Danilo Bzdok, Ph.D., Ph.D. Prof., Rheinisch-Westfalische Technische Hochschule Aachen Medizinische Fakultat dbzdok@ukaachen.de Expertise in: Neuroimaging meta-analysis, Large-scale databases, Theory of Mind, Empathy, Morality, Advanced data-analysis methods. Selected publications: Alcalá-López, D., Smallwood, J., Jefferies, E., Van Overwalle, F., Vogeley, K., Mars, R. B., ... & Bzdok, D. (2018). Computing the social brain connectome across systems and states. Cerebral cortex, 28(7), 2207-2232. Bzdok, D., Schilbach, L., Vogeley, K., Schneider, K., Laird, A. R., Langner, R., & Eickhoff, S. B. (2012). Parsing the neural correlates of moral cognition: ALE meta-analysis on morality, theory of mind, and empathy. Brain Structure and Function, 217(4), 783-796.
	Stephanie Preston, PhD Prof., University of Michigan prestos@umich.edu Expertise in: Empathy, Altruism, Theory of Mind (Cognitive Empathy), Behavioral and neuroimaging research. Selected publications: Preston, S. D. (2013). Examining the origins of altruism in offspring care. Psychological Bulletin. 139(6), 1305-1341. Preston, S. D., & De Waal, F. B. (2002). Empathy: Its ultimate and proximate bases. Behavioral and brain sciences, 25(1), 1-20. Preston, S. D.; Bechara, A.; Grabowski, T. J.; Damasio, H.; Damasio A. R. (2007). The neural substrates of cognitive empathy. Social Neuroscience, 2 (3-4), 254-275.
	Francesca Happe, PhD Prof., King's College London francesca.happe@kcl.ac.uk Expertise in: Measurement of social cognition, Theory of Mind, Social Cognition, Action Understanding, Autism, Behavioral and neuroimaging research. Selected publications: Happé, F., Cook, J. L., & Bird, G. (2017). The structure of social cognition: In(ter)dependence of sociocognitive processes. Annual review of psychology, 68, 243-267. Happé, F. G. (1994). An advanced test of theory of mind: Understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. Journal of autism and Developmental disorders, 24(2), 129-154.
	Jennifer Cook, PhD Birmingham Fellow, University of Birmingham j.l.cook@bham.ac.uk Expertise in: Social cognition, Action observation, Imitation, Theory of Mind, Autism, Behavioral and neuroimaging research. Selected publications:

	Happé, F., Cook, J. L., & Bird, G. (2017). The structure of social cognition: In (ter) dependence of sociocognitive processes. Annual review of psychology, 68, 243-267. Eddy, C. M., & Cook, J. L. (2018). Emotions in action: the relationship between motor function and social cognition across multiple clinical populations. Prog Neuropsychopharmacol Biol Psychiatry, 86, 229-244.
	Marcel Brass, PhD Prof., Ghent University Marcel.Brass@UGent.be Expertise in: Mirror neurons, Imitation, Theory of Mind, Implicit mentalizing, Behavioral and neuroimaging research. Selected publications: Cracco, E., Bardi, L., Desmet, C., Genschow, O., Rigoni, D., De Coster, L., ... & Brass, M. (2018). Automatic imitation: A meta-analysis. Psychological Bulletin, 144(5), 453. Brass, M., Ruby, P., & Spengler, S. (2009). Inhibition of imitative behaviour and social cognition. Philosophical Transactions of the Royal Society B: Biological Sciences, 364(1528), 2359-2367.
	Caroline Catmur, PhD Senior Lecturer, King's College London caroline.catmur@kcl.ac.uk Expertise in: Social cognition, Action Understanding, Mirror Neurons, Imitation, Empathy, Theory of Mind, Autism, Behavioral and neuroimaging research. Selected publications: Catmur, C., Thompson, E. L., Bairaktari, O., Lind, F., & Bird, G. (2018). Sensorimotor training alters action understanding. Cognition, 171, 10-14. Conway, J. R., Catmur, C., & Bird, G. (2019). Understanding individual differences in theory of mind via representation of minds, not mental states. Psychonomic bulletin & review, 1-15. Oakley, B. F., Brewer, R., Bird, G., & Catmur, C. (2016). Theory of mind is not theory of emotion: A cautionary note on the Reading the Mind in the Eyes Test. Journal of abnormal psychology, 125(6), 818.
Opposed Reviewers:	
Order of Authors:	Matthias Schurz
	Joaquim Radua
	Matthias G. Tholen
	Lara Maliske
	Daniel S. Margulies
	Rogier B. Mars
	Jerome Sallet
	Philipp Kanske
Manuscript Classifications:	emotion theory; empathy; face recognition; fMRI; social cognition; social neuroscience; social neuroscience; theory of mind; theory of mind
Additional Information:	
Question	**Response**
I have read the Instructions to Authors on the <a href="http://www.apa.org/pubs/journals/bul/index.aspx?tab=4" target="_blank">Psychological Bulletin<a/> website.	Yes

I have included a cover letter that contains the names and contact information of all authors. Inclusion of such information indicates that these individuals have agreed to be an author.	Yes
I have read and agree to follow the APA Journals Internet Posting Guidelines. Manuscripts with dissemination pre-acceptance will likely not be considered for publication inPsychological Bulletin. Authors must disclose any news and online activity at any point of the editorial process.	Yes
I attest that this manuscript is not being considered by another journal nor has it been published elsewhere. If applicable, I have provided information in my cover letter about any closely related manuscripts that have been submitted for consideration to the same or another journal. If the manuscript has been previously published in conference proceedings I have explained this in my cover letter.	Yes
This manuscript has been prepared consistent with *APA Publication Manual* (6th ed.) guidelines.	Yes
I have obtained permission to reproduce or adapt any copyrighted material from other sources and am able to provide documentation for this permission.	Yes
I agree to comply with APA Ethics Code Standard 8.14a, Sharing Research Data for Verification, allowing other qualified professionals to confirm the analyses and results should my manuscript be accepted for publication. Per APA guidelines I will retain raw data for a minimum of 5 years after publication of the research.	Yes
If my manuscript is accepted for publication, I agree to transfer copyright to APA per the guidelines on the APA Publication Rights Form.	Yes
I have followed instructions to adequately conceal the identity of all authors to ensure masked review.	Yes
If my manuscript submission includes supplemental materials, I have explained	Yes

the relevance of these materials in the cover letter, in the article itself, and within the supplemental materials.	
Was this manuscript or a manuscript that covered any overlapping issues submitted to Psychological Bulletin in the past?	No
Has this manuscript or a portion of it been submitted to another outlet, including journals and edited books?	No
Has this manuscript or a related one been previously submitted to Psychological Bulletin and rejected without an invitation to resubmit?	No
If you paper is a commentary, are you submitting a commentary of a paper currently in press? (Psychological Bulletin does not accept commentaries of manuscripts published earlier, as commentaries are sought at the time of review of a manuscript. Other commentaries may be submitted to other outlets.)	No
Does your manuscript report a narrative review?	No
Please provide an explanation. as follow-up to "Does your manuscript report a narrative review?"	Reports a series of meta-analyses.
Does your manuscript report a meta-analysis (or meta-meta-analysis)?	Yes
Is the PRISMA flow diagram (http://prisma-statement.org/PRISMAStatement/FlowDiagram.aspx) included? as follow-up to "Does your manuscript report a meta-analysis (or meta-meta-analysis)?"	No
Please provide an explanation. as follow-up to "Is the PRISMA flow diagram (http://prisma-statement.org/PRISMAStatement/FlowDiagram.aspx) included? "	Our review covers 11 individual meta-analyses, gathered by a number of strategies described in detail in our methods section. We feel that the PRISMA diagram would be slightly too rigid for our review. The main aim of our meta-analysis was to capture interrelations among a range of social cognition tasks by hierarchical clustering, rather than to provide a single, most up-to-date meta-analysis on a specific task / topic. However, if requested, we can still provide PRISMA diagrams for our review.
Were foreign language reports included? as follow-up to "Does your manuscript report a meta-analysis (or meta-meta-analysis)?"	No
Please provide an explanation.	We found no foreign language reports.

as follow-up to "Were foreign language reports included? "	
Were publication bias analyses performed? as follow-up to "Does your manuscript report a meta-analysis (or meta-meta-analysis)?"	No
Please provide an explanation. as follow-up to "Were publication bias analyses performed? "	Analysis of publication bias was outside the scope of our meta-analytic clustering. We aimed to characterize interrelations among social cognition tasks, to produce an integrative model of processes. For this aim, publication bias is of little relevance.
Are appropriate intercoder reliability indexes reported? as follow-up to "Does your manuscript report a meta-analysis (or meta-meta-analysis)?"	No
Please provide an explanation. as follow-up to "Are appropriate intercoder reliability indexes reported? "	In our meta-analysis, studies were assigned to tasks groups by two expert raters. All decisions are transparently described and discussed in the methods and results section. The potential effect of including/excluding individual studies was estimated by jack-knife sensitivity analysis applied to our meta-analytic clustering. Reliability indexes were not created, as this is not applicable.
Did you include a table giving descriptive information for each included study, including effect size and sample size? as follow-up to "Does your manuscript report a meta-analysis (or meta-meta-analysis)?"	Yes
Is there proper consideration of methodological decisions in the synthesized literature in the form of quality ratings and/or detailed coding of methodological features of the study? as follow-up to "Does your manuscript report a meta-analysis (or meta-meta-analysis)?"	No
Please provide an explanation. as follow-up to "Is there proper consideration of methodological decisions in the synthesized literature in the form of quality ratings and/or detailed coding of methodological features of the study? "	We apply methodological selection to all studies entering our meta-analysis, in keeping with recommended guidelines for the method we use.
Were the meta-analysis (or meta-meta-analysis) and report conducted in accordance with all MARS guidelines http://www.apastyle.org/manual/related/JARS-MARS.pdf? as follow-up to "Does your manuscript report a meta-analysis (or meta-meta-analysis)?"	No

Please provide an explanation. as follow-up to "Were the meta-analysis (or meta-meta-analysis) and report conducted in accordance with all MARS guidelines http://www.apastyle.org/manual/related/JARS-MARS.pdf? "	Although MARS guidelines are tailored to meta-analyses of behavioral data, our neuroimaging meta-analysis shows almost perfect compliance. Method-specific differences in presentation of results.
Original color figures can be printed in color at the editor's and publisher's discretion provided the author agrees to pay a portion of the cost. A list of the color figure prices charged to the author is available here. Please indicate which statement applies to your manuscript:	Print figures in grayscale, display as color online only. There is no cost to the author for this option.
If your paper is accepted, do you have any disclosures that would need to be listed on the Full Disclosure of Interests form? This includes any interests or activities that might be seen as influencing the research (e.g., financial interests in a test or procedure, funding by pharmaceutical companies for research).	No, I have no disclosures to report.

Dr. Matthias Schurz
Donders Institute for Brain, Cognition, & Behaviour, Radboud University Nijmegen, Netherlands /
Wellcome Centre for Integrative Neuroimaging, Dept. of Experimental Psychology, University of Oxford, UK
www.ru.nl/english/people/schurz-m/
www.psy.ox.ac.uk/team/matthias-schurz
matthias.schurz@gmail.com

UNIVERSITY OF OXFORD

Dear Prof. Verhaeghen, dear Editors,

We would like to re-submit the second (minor) revised version of our manuscript "Towards a Hierarchical Model of Social Cognition: A Neuroimaging Meta-Analysis and Integrative Review on Empathy and Theory of Mind" (BUL-2019-1649R1).

We are grateful for the positive evaluation and further remark made by one reviewer, and have inserted the additional reference as requested. We have documented this change in the response letter attached.

Based on the positive feedback, we have also carefully re-read the manuscript ourselves, and have polished all figures to maximize clarity, image quality and readability.

Sincerely,

Matthias Schurz (for the authors)

Contact addresses (ordered after author-list in the manuscript)

Matthias Schurz, Erwin-Schrödinger Post-Doctoral Fellow,
Wellcome Centre for Integrative Neuroimaging, Department of Experimental Psychology, University of Oxford, Tinsley Building, 13 Mansfield Road, OX13SR, Oxford, United Kingdom
https://www.psy.ox.ac.uk/team/matthias-schurz
matthias.schurz@psy.ox.ac.uk

Joaquim Radua, Miguel Servet Post-Doctoral Fellow,
Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Rosselló 149-153, 08036 Barcelona, Spain
https://www.radua.net
jradua@fidmag.com

Matthias G. Tholen, fourth year PhD student, PhD programme 'Imaging The Mind',
Centre for Cognitive Neuroscience, University of Salzburg, Hellbrunnerstr. 34, 5020 Salzburg, Austria
https://phdim.ccns.sbg.ac.at/doctoral-college-im/students/
Matthias.Tholen@sbg.ac.at

Lara Maliske, second year PhD student, Clinical Psychology and Behavioral Neuroscience Lab,
Faculty of Psychology, Technische Universität Dresden, Chemnitzer Str. 46, 01187 Dresden, Germany
https://tu-dresden.de/mn/psychologie/ikpp/klipsy/die-professur/beschaeftigte/lara-maliske
lara.maliske@tu-dresden.de

Daniel S. Margulies, Tenured researcher (CR1),
Centre National de la Recherche Scientifique (CNRS) UMR 7225, Frontlab, Institut du Cerveau et de la Moelle épinière, 75013 Paris, France
https://www.neuroconnlab.org
daniel.margulies@icm-institute.org

Rogier B. Mars, Associate Professor, BBSRC David Phillips Fellow,
Wellcome Centre for Integrative Neuroimaging, Centre for Functional MRI of the Brain (FMRIB), Nuffield Department of Clinical Neurosciences, John Radcliffe Hospital, OX3 9DU, Oxford, United Kingdom
https://www.ndcn.ox.ac.uk/team/rogier-mars
rogier.mars@ndcn.ox.ac.uk

Jerome Sallet, University Research Lecturer, Sir Henry Dale Fellow,
Wellcome Centre for Integrative Neuroimaging, Department of Experimental Psychology, University of Oxford, Tinsley Building, 13 Mansfield Road, OX13SR, Oxford, United Kingdom
https://www.psy.ox.ac.uk/team/jerome-sallet
jerome.sallet@psy.ox.ac.uk

Philipp Kanske, Professor for Clinical Psychology and Behavioral Neuroscience,
Faculty of Psychology, Technische Universität Dresden, Chemnitzer Str. 46, 01187 Dresden, Germany
https://tu-dresden.de/mn/psychologie/ikpp/klipsy
philipp.kanske@tu-dresden.de

Response Letter "Towards a Hierarchical Model of Social Cognition: A Neuroimaging Meta-Analysis and Integrative Review on Empathy and Theory of Mind"

Response Letter

Reviewer #4:
I reviewed this manuscript in a prior submission, and am very pleased to see that in my opinion, the authors have done an excellent job responding to the reviewers' feedback and have a much improved and more comprehensive manuscript that more accurately represents recent research on social cognition. I only have one very minor comment. In addition to the existing relevant citations, I might recommend the authors cite the neurosynth DMN meta-analysis of Andrews-Hanna, Smallwood & Spreng, 2014 in their section on overlap with language. I suggest doing so because that manuscript used a similar neurosynth decoding approach and also noted strong overlap with language-related processes, which was discussed in relation to social cognition.

Response: Thanks for pointing out this convergent finding. We have cited the meta-analysis you are referring to now in the section "3.6. Possible roles of language processes in the three cluster solution", on p. 34:
"A notable finding from neurosynth decoding was that both the intermediate and the affective cluster showed high loadings on language related terms. This supports results from a previous meta-analysis (Andrews-Hanna, Smallwood & Spreng, 2014), which noted strong overlap between default mode and language-related processes, which was discussed in relation to social cognition."

Towards a Hierarchical Model of Social Cognition: A Neuroimaging Meta-Analysis and Integrative Review on Empathy and Theory of Mind

Matthias Schurz[1,2,3*], Joaquim Radua[4,5,6,7], Matthias G. Tholen[3], Lara Maliske[8], Daniel S. Margulies[9], Rogier B. Mars[1,2,10], Jerome Sallet[1], Philipp Kanske[8,11]

[1]Wellcome Centre for Integrative Neuroimaging, Department of Experimental Psychology, University of Oxford, Oxford, United Kingdom
[2]Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, Nijmegen, The Netherlands
[3]Centre for Cognitive Neuroscience, University of Salzburg, Salzburg, Austria
[4]Imaging of Mood- and Anxiety-Related Disorders (IMARD) Group, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Barcelona, Spain
[5]Mental Health Research Networking Center (CIBERSAM), Madrid, Spain
[6]Early Psychosis: Interventions and Clinical-detection (EPIC) Lab, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, United Kingdom
[7]Department of Clinical Neuroscience, Centre for Psychiatric Research and Education, Karolinska Institutet, Stockholm, Sweden
[8]Clinical Psychology and Behavioral Neuroscience, Faculty of Psychology, Technische Universität Dresden, Dresden, Germany
[9]Centre National de la Recherche Scientifique (CNRS) UMR 7225, Frontlab, Institut du Cerveau et de la Moelle épinière, Paris, France
[10]Wellcome Centre for Integrative Neuroimaging, Centre for Functional MRI of the Brain (FMRIB), Nuffield Department of Clinical Neurosciences, John Radcliffe Hospital, Oxford, United Kingdom
[11]Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

*Corresponding author. Permanent email address: matthias.schurz@gmail.com. Present postal address: Donders Institute for Brain, Cognition, & Behaviour, Radboud University Nijmegen, Montessorilaan 3, 6525 HR Nijmegen, The Netherlands.

Towards a Hierarchical Model of Social Cognition: A Neuroimaging Meta-Analysis and Integrative Review on Empathy and Theory of Mind

Abstract (247/250 words)

In parallel with the rise in interest and volume of social cognition research, there has been increasing awareness of a lack of agreement on the concepts and taxonomy used to study social processes. Two central concepts in the field, empathy and Theory of Mind (ToM), have been applied as overlapping umbrella terms for a variety of different processes of limited convergence. Here, we review and integrate evidence of brain activation, brain organization, and behavior into a coherent model of social cognitive processes. We start with a meta-analytic clustering of neuroimaging data across different social cognitive tasks. Results show that understanding others' mental states can be described by a multilevel model of hierarchical structure, similar to models in intelligence and personality research. A higher level describes more broad and abstract classes of functioning, whereas a lower level explicates how functions are applied to concrete contexts, given by specific stimulus and task formats. Specifically, the higher level of our model suggests three groups of neurocognitive processes: (i) Predominantly cognitive processes that are engaged when mentalizing requires self-generated cognition decoupled from the physical world. (ii) More affective processes which are engaged when we witness emotions in others, based on shared emotional, motor, and somatosensory representations. (iii) Combined processes which engage cognitive and affective functions in parallel. We discuss how these processes are explained by an underlying principal gradient of structural brain organization. Finally, we validate the model by a review of empathy and ToM task interrelations found in behavioral studies.

Key words: Social cognition, Mentalizing, Perspective Taking, Emotion, Mirror Neurons

Public significance statement: Empathy and Theory of Mind play a key role for understanding others. Here we present a meta-analysis that shows how both features can be deconstructed into several sub-processes. Moreover, we propose that understanding others' mental states is best described by a multilevel model of hierarchical structure, similar to models in intelligence and personality research.

1. Introduction

Successful social interaction requires not only representation of the overt behavior of people, but also of the covert underlying forces, such as thoughts and emotions. Over the last two decades, behavioral and brain imaging research has generated an abundance of evidence on how we manage to infer unobservable mental states of others. At the same time, there has been increasing awareness of a lack of agreement regarding the concepts and taxonomy used to study these social processes. It is clear that similar terms are being used to describe different processes, and at times different terms for describing similar processes (see Happé, Cook, & Bird, 2017). In particular, two terms have been of central importance - empathy, generally referring to an affective route for understanding others (Titchener, 1909; Gallese, 2003), and Theory of Mind (ToM) or mentalizing, referring to cognitive representations of others' mental states (Premack & Woodruff, 1978; Adolphs, 2009; Keysers & Gazzola, 2009; Kanske, 2018). On a conceptual level, it has been argued that both are used as umbrella terms for studying a variety of different processes of limited convergence (see Schaafsma, Pfaff, Spunt, & Adolphs, 2015 for ToM; see Bloom, 2017; Zaki, 2017 for empathy). This observation has also been reflected in recent literature reviews. Rather than finding broad and homogeneous networks, meta-analyses which explicitly categorized studies based on stimuli and instructions found multiple functional sub-divisions among the neural correlates for ToM (Molenberghs, Johnson, Henry, & Mattingley, 2016; Schurz, Radua, Aichhorn, Richlan, & Perner, 2014; see also Mar, 2011; Van Overwalle, 2009; Van Overwalle & Baetens, 2009) and empathy (Fan, Duncan, de Greck, & Northoff, 2011; Gu, Hof, Friston, & Fan, 2013; Lamm, Decety, & Singer, 2011; Timmers, Park, Fischer, Kronman, Heathcote, Hernandez et al., 2018).

The issues illustrated above are compounded by the fact that empathy and ToM are both thought to comprise affective and cognitive subforms, making a distinction between the terms difficult. For instance, definitions of ToM not only include the ability to make inferences about others’ cognitive mental states, such as beliefs and thoughts, but also about their desires and emotions (e.g. Frith & Frith, 2006; Premack & Woodruff, 1978). The latter feature has sometimes been referred to as 'affective ToM' (e.g. Kalbe et al., 2010; Schlaffke et al., 2015; Sebastian et al., 2011; Shamay-Tsoory & Aharon-Peretz, 2007). Relatedly, it was proposed that processing of other's mental states engages Theory of Mind, irrespective of cognitive or affective content, whenever it requires metarepresentation (Leslie, Friedman & German (2004), that is, representing a propositional attitude (e.g. 'Sally is happy that ...', or 'Sally wants that ...'). On the other hand, definitions of empathy contain emotional processes such as the sharing of others’ feelings, as well as cognitive processes such as reasoning about the others’ affective states (e.g. Dziobek et al., 2011; Shamay-Tsoory, Aharon-Peretz, & Perry, 2009), referred to as 'cognitive empathy' (Hooker, Verosky, Germine, Knight, & D'Esposito, 2010, Walter, 2012). Even broader conceptualizations of empathy additionally contain features like empathic concern or compassion for another person (Davis, 1994; see also Zaki & Ochsner, 2012). These variable models of empathy and ToM challenge an integration of findings across different fields and labs.

An important step for clarifying current theories of social cognition is to determine how different concepts relate to one another (e.g. in a multi-layered manner; De Waal, 2012; see also De Waal, 2007; Preston & Hofelich, 2012: Singer, 2006), and to explore how they are best grouped according to common underlying processes (Happe et al., 2017). This approach also advances a 'deconstruction' of social processes (see Schaafsma et al., 2015), i.e. the mapping of broad terms like ToM and empathy to a set of underlying processes that are in turn linked to concrete experimental tasks. Note that such an endeavor also dovetails with a strategy promoted by the National Institute of Mental Health (NIMH) for studying mental disorders (Insel et al., 2010) - the Research Domain Criteria (RDoC) Projects.

We begin our article with a meta-analysis of neuroimaging studies, and then move on to reviewing corresponding behavioral findings. Neuroimaging data provide an useful starting point, as brain activity reported in standard space is readily comparable across studies, and can be used to generate a comprehensive picture of task-by-task interrelations (i.e. activation overlaps). This serves us as a framework for reviewing behavioral studies on task-by-task correlations. Building on previous meta-analytic work (Schurz et al., 2014), we sorted studies according to stimuli and task instructions. In the present study, we not only include task groups from the ToM literature (e.g. False Belief, Reading Mind in the Eyes tasks), but also from the field of empathy (e.g. Observing Pain, Evaluating Situated Emotions tasks). A central additional feature of the present study is a hierarchical clustering of meta-analytic results (see Laird et al., 2011, Riedel et al., 2015). Specifically, after obtaining meta-analytic result maps for each task group, we determined their overlaps (via image correlation) and clustered similar maps together. This produced a hierarchical tree (dendrogram), which characterizes relations between brain activation networks for different task groups.

The main goal of this clustering is to estimate latent factors, i.e. common neurocognitive components engaged by different empathy and ToM tasks, which is in line with previous perspectives on empathy and ToM (e.g., De Waal, 2007; Preston & Hofelich, 2012). In theory, components engaged by different tasks could be related in various forms. Multiple components could be largely overlapping, one component could be a sub-component of another, or components could have some elements in common but additionally contain distinct processes (see Happe et al., 2017, for discussion). The clustering tree we generated is flexible enough to capture any of these relations. Identifying common components is also supported by the argument that whatever is pivotal for social cognition as measured in empathy and ToM studies should be reflected in common activation across a number of tasks. In this way, activation driven by idiosyncratic elements of task-design is ruled out by the overlap (see Mar, 2011). In addition, our clustering across empathy and ToM tasks clarifies theories of sub-processes for these domains. To illustrate, cognitive empathy and affective ToM could represent two independent and self-contained social abilities, or could be overlapping and to some extent redundant concepts.

2. Methods

2.1. Literature search and study selection

We reviewed the neuroimaging literature on ToM and empathy to get an overview of the different experimental tasks that were used, and sorted studies with similar stimuli and instructions into task groups. All studies from ToM and empathy task groups were retrieved from database searches using pubmed and isi web of science ('core collection'). Our review includes literature published up to November 2019. Neuroimaging studies were identified by the keywords ‘neuroimaging’ or ‘fMRI’ or ‘PET’, and were assigned to ToM groups when additionally including ‘theory of mind’ or ‘mentalizing’ or ‘mindreading’, and to the empathy groups when including ‘empathy’, or ‘empathetic’, or ‘altruism’, or ‘sympathy’, or ‘emotional contagion’, or ‘compassion’. For ToM, we incorporated the literature from a previous meta-analysis (Schurz et al., 2014) into the present sample. After preliminary lists of studies were created for each task group, existing literature reviews were searched for further compatible studies. Reference lists for ToM were adopted from the reviews by Bzdok et al.

(2012), Denny, Kober, Wager, & Ochsner (2012), Murray, Schaer, & Debbané (2012), Perner and Leekam (2008), Spreng, Mar, & Kim (2009), Van Overwalle (2009), as well as Van Overwalle and Baetens (2009). Likewise, reference lists for empathy covered the reviews by Bzdok et al. (2012), Fan, Duncan, de Greck, & Northoff (2011), Gu, Hof, Friston, & Fan (2013), Lamm, Decety, & Singer (2011), and Morelli, Sacchet, & Zaki (2015). In addition, the literature reviewed in a meta-analysis on emotion processing (Fusar-Poli et al., 2009) was considered. Papers were retained if they matched to one of the empathy and ToM task groups defined previously, irrespective of the terminology used by the authors. All studies included in task groups had to furthermore fulfil the following selection criteria (see Radua et al., 2012): Reported coordinates had to correspond to standard space (MNI or TAL) and stem from whole brain analysis, and use a consistent threshold throughout the whole brain. Data from clinical samples were removed. If a study reported more than one contrast, the one best corresponding to the other studies from the task group was selected. Altogether we included 103 studies from the ToM literature, divided into six task groups: False Belief (n=25), Trait Judgments (n=19), Strategic Games (n=13), Rational Actions (n=11), Social Animations (n=20), and Reading Mind in the Eyes (n=15). We identified 85 studies from the empathy literature that we divided into five task groups: Observing Pain (n=21), Observing Emotions (n=25), Sharing Emotions or Pain (n=12), Evaluating Situated Emotions (n=15), and Reasoning about Emotions (n=12). Thus, our meta-analyses covers in total 188 studies.

2.2. Meta-analysis methods

For each of the eleven task groups identified, we carried out an effect-size based meta-analysis (AES-SDM 4.31, Radua et al., 2012, www.sdmproject.com). The SDM method is based on the positive features from existing peak probability methods for meta-analysis, such as Activation Likelihood Estimation (ALE, Eickhoff et al., 2009) or Multilevel Kernel

Density Analysis (MKDA, Wager et al., 2007). In addition, it incorporates the effect-sizes of reported activations, thereby extracting more detailed information from the published literature. Based on t-values and sample sizes reported in studies, AES-SDM creates effect-size (Hedge's g values) and variance maps (the latter derived from the distribution of effect-sizes)[1]. Statistical significance of meta-analytic maps was assessed by permutation tests that randomize the location of the voxels within the standard gray matter template. One hundred random maps were generated by permutation for each meta-analysis. We report all results in MNI space and at a statistical threshold of $p<0.005$ uncorrected (voxel-level) and a cluster threshold of 10 voxels. This threshold was found to optimally balance sensitivity and specificity, and to be an approximate equivalent to a corrected threshold of $p< 0.05$ in original neuroimaging studies (Radua et al., 2012). For contrasts between meta-analyses, we used the SDM linear model function, calculating the difference between effect-size estimates from two meta-analyses while accounting for differences in sample size and within and between study variability. For determining common activation in multiple contrasts, we applied conjunction minimum analysis (e.g. Nichols, Brett, Andersson, Wager, & Poline, 2005) via the image calculator utility of SPM12 (www.fil.ion.ucl.ac.uk).

2.3. Hierarchical clustering analysis

After having obtained meta-analytic result maps for all eleven task groups, we applied agglomerative hierarchical clustering to them (see e.g. Laird et al., 2015; Riedel et al., 2015, 2018). Seeking for a hierarchical structure is in keeping with a number of previous conceptualizations of social cognition as a multi-layered or multi-level phenomenon (De

[1] Hedge's g values are derived from t-statistics, or equivalently from p-values or z-scores. Effect sizes around reported peak coordinates (<20 mm) are estimated based on an an anisotropic un-normalized Gaussian Kernel. Mean brain activity for each task group is determined by a random-effects model, with each study being weighted by the inverse of the sum of its variance plus an estimate of between-study heterogeneity (DerSimonian & Laird, 1986). This enables studies with larger sample size or lower variability to contribute more strongly.

Waal, 2012; Preston & De Waal, 2002; Singer, 2006; Schaafsma et al., 2015). To our knowledge, this is the first clustering of SDM effect-size meta-analysis maps. Therefore, we compared several settings for discriminative performance (see Supplementary Materials) and found unthresholded effect size maps (Hedges' g) and Pearson correlation coefficients to best capture image dissimilarity among our meta-analyses. Clustering consisted of three steps. In step 1, we transformed the unthresholded meta-analytic effect size maps into feature vectors containing voxel values, and concatenated them horizontally, forming a matrix of *n* task groups (i.e., 11) and *p* voxels (i.e., 902629). Based on that we calculated pairwise Pearson correlation coefficients between all feature vectors, from which we derived an *n* by *n* dissimilarity matrix (1-*r* values) reflecting the whole-brain multi-voxel dissimilarity between maps. In step 2, we grouped meta-analyses into clusters by applying agglomerative hierarchical clustering (in MATLAB 2013a). For linkage, we selected the linkage 'average' method, which represents a compromise between the clustering's sensitivity to outliers, and its tendency to form 'long chains' of elements per cluster. As a distance measure, we selected 'euclidean', which takes into account both the profile and magnitudes of task-to-task similarities. 'Euclidean' is among the most widely used distance measures, and the default setting for MATLAB's hierarchical clustering. The resulting dendrogram is displayed in Figure 4B. In step 3 we evaluated different solutions based on the dendrogram from our clustering. Based on previous works (e.g. Laird et al., 2015; Riedel et al., 2015, 2018) we relied on two metrics for this step: (i) cophenetic distance and (ii) density of task separation. The cophenetic distance between clusters at a given model order (i.e. number of clusters) reflects dissimilarity between sub-clusters[2]. A clustering optimum is indicated by high difference in cophenetic distances when moving from lower to higher model orders, as this

[2] The relative difference in cophenetic distances d_c when transitioning from one model order (x) to the next higher ($x + 1$) can be derived from the cophenetic distances c_x and c_{x+1} in the form of $d_c = (c_{x+1} - c_x) / (c_{x+1})$.

indicates that introducing new sub-clusters produces substantially different (i.e. *distant*) brain activity patterns. The second metric we used, density of task separation, indicates whether separating clusters into sub-clusters maintains a balanced distribution of task groups across sub-clusters (as opposed to producing disproportionately large/small sub-clusters, such as a cluster consisting of only one task group)[3]. Decreases in the density of task separation indicate good solutions, and reflect a split into sub-clusters with balanced numbers of task groups. Taking our two clustering metrics together, an optimal solution is indicated by a model order with a *relatively high* difference in cophenetic distance and a *relatively low* density of task separation.

To check the stability of our clustering, we repeated the procedure with leave-one-out jackknife sensitivity analysis[4]. Figure 3A shows the range of clustering metrics found across our jackknife repeats (n=5000), which was taken as a guide for selecting the best clustering solutions. Furthermore, for the three cluster solution (which we present as a main result later) we show the consistency with which task groups were assigned to clusters across iterations of our jackknife analysis (see bar plots in Figure 3B). In general, good consistency (agreement for over 90% of iterations) was found for most task groups (except for Rational Actions).

2.4. Analyses of overlap with other maps

For a broader perspective on our meta-analysis, we compared our results with an extensive set of automatically generated meta-analyses across a wide range of topics. This step allows us to discuss our findings in relation to brain activation for a wide range of other social

---

[3] If cluster i_0 consisted of n_0 task groups at model order x, and at model order x+1 was separated into sub-clusters i_1 and i_2 with n_1 and n_2 task groups, then the density of task separation can be calculated as $d_s = n_1/n_0$, given that $n_1 > n_2$.
[4] That is, we removed one random study from each task group (simultaneously), calculated the meta-analyses with the remaining studies, clustered the new results, replaced the removed studies, and repeated the process. We carried out 5000 clusterings based on leave-one-out samples, and summarize the metrics across all of them in Figure 3B.

processes, and also non-social processes. We used the 'decoding' tool of the neurosynth database (Yarkoni, Poldrack, Nichols, Van Essen, & Wager, 2011) for this analysis, which we accessed via the larger neuroimaging repository neurovault.org (Gorgolewski et al., 2015). In keeping with the recommended input specifications, we created unthresholded versions of mean and contrast maps for decoding[5]. In Figures 6 and 7 we show neurosynth decoding results as Pearson correlation coefficients, describing image similarity between our maps and the automatically generated meta-analyses. For clarity, we show only most strongly correlated social and non-social meta-analyses found with neurosynth decoding[6]. In addition, we have discarded meta-analyses of little interest for the current article, such as e.g. for neuroanatomical labels (e.g. 'mpfc', 'psts')[7].

To further characterize the neurocognitive processes linked to our meta-analysis maps, we calculated the overlap of our meta-analysis with a whole-brain map of brain connectivity organization (Margulies et al., 2016), and a resting-state connectivity parcellation atlas (Yeo et al., 2011). We determined overlaps with these maps by conjunction analysis[8], and summarized them as a variant of the dice score: For each meta-analysis map, we calculated

---

[5] As result maps from meta-analytic contrasts originally contain positive (e.g. map1>map2) and negative (e.g. map2>map1) values, we next removed (masked) all negative values from these images to obtain clearly interpretable decoding results. To illustrate, after masking, results for decoding of the contrast (map1 > map2) gives results that can be specifically linked to activations in map 1.

[6] A large number of correlated terms is found by neurosynth decoding. For clarity, we decided to select a small number (five) of top-ranked associations, in order to highlight major themes linked to our results. Note that terms found formed topically coherent clusters (see Figures 6 and 7 - terms cluster around the topics language, default mode network function, and sensory-affective processing). Therefore, in our discussion we interpret the relation of our maps to major themes identified by neurosynth, rather than individual terms.

[7] In addition, we discarded technical/unspecific terms (e.g. 'independent component', 'task', 'state') and conceptually redundant terms (e.g. 'dmn' and 'mode' are redundant with 'default') from our selection.

8 For overlaps with resting-state networks, we used the 7-networks parcellation by Yeo et al. (2011), more specifically a MNI transformed version ("liberal mask", see https://surfer.nmr.mgh.harvard.edu/fswiki/CorticalParcellation_Yeo2011). For the connectivity gradient, a partitioned version was used, consisting of 20 maps, each corresponding to a 5-percentile step along its progression. All resting-state / gradient-percentile maps were in turn overlaid with our meta-analysis result maps, by determining conjunction images after binarizing all inputs (in SPM12). Since all inputs conformed to MNI space, only adjustment of images in terms of size and resolution was required (which we implemented via the reslice function in SPM12, with the resting-state / gradient-percentile maps being the image defining space).

the percentage of voxels it comprised that fell within different parts of each connectivity map[9].

3. Results and Discussion

In the next sections, we describe the task groups that we found in the ToM and empathy literature, and briefly review some of the rationale behind each type of task. These task groups will then be input for our key analysis, the meta-analytic clustering, which follows in section 3.3. Concrete task examples are given in Table 1 for ToM, and Table 2 for empathy task groups. We also illustrate the meta-analytic result maps for each task group in Figures 1 and 2.

3.1. Task Groups ToM

3.1.1. False Belief

False Belief tasks have been widely used in developmental psychology and were quickly adapted for neuroimaging using the story format (e.g., Gallagher et al., 2000). Since the logical structure of False Belief stories differs from simpler control conditions, we only included more recent studies presenting false photograph control stories (e.g. Saxe & Kanwisher, 2003; Dodell-Feder, Koster-Hale, Bedny, & Saxe, 2011) in our meta-analysis. As illustrated in Table 1, such stories may describe an old photo of an object, showing it the way it was earlier on. By now, the object has changed in some way, and therefore the photo corresponds to an outdated representation of the object (therefore sometimes referred to as *false* photo). Note that for False Belief tasks, we found many more eligible studies in the

[9] That is, if *i1* is a meta-analysis map and *i2* a gradient-percentile map, we calculated *n voxels in (i1&i2)* / *n voxels in i1.*

literature than for other task groups (in total 38 studies). To avoid large differences in size between task groups, we randomly selected 25 studies from this large sample of False Belief tasks.

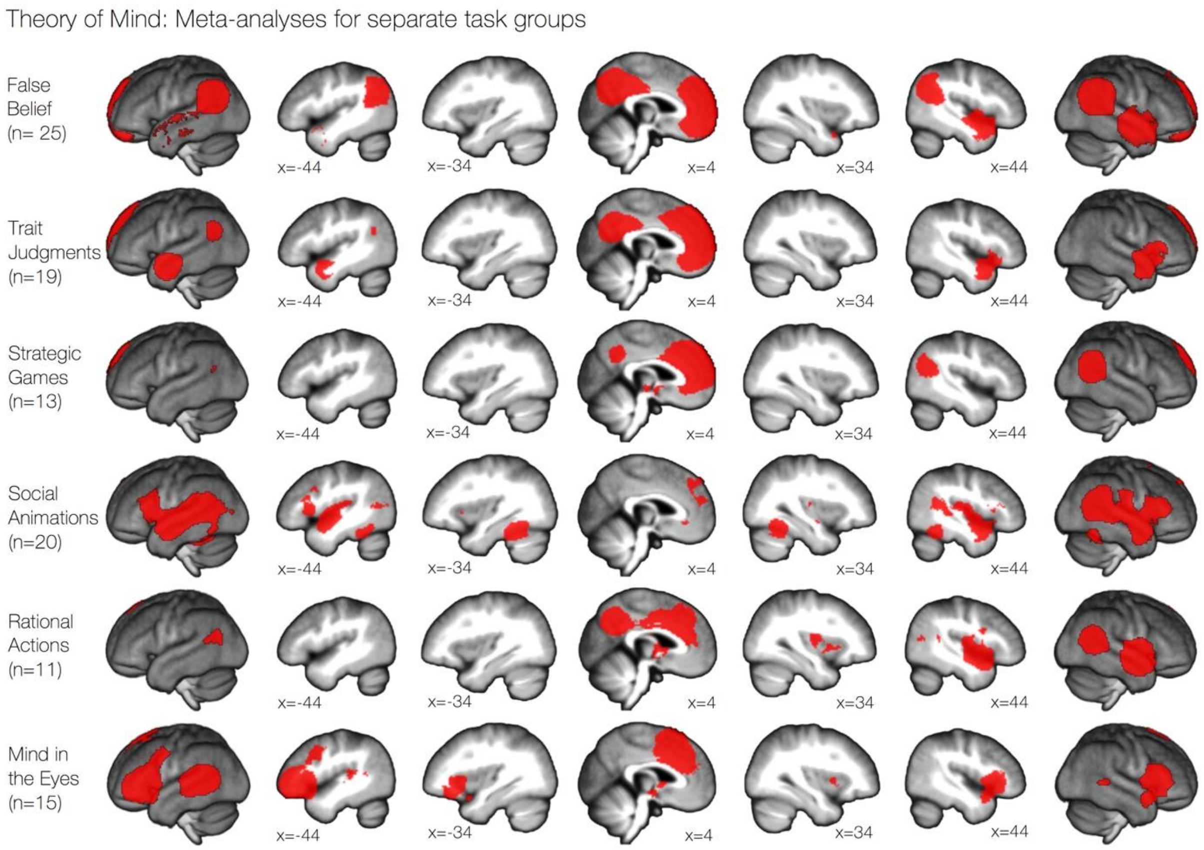


Figure 1. Results of separate meta-analyses for task groups from Theory of Mind. Maps were thresholded at voxel-wise threshold of $p < .005$ uncorrected and a cluster extent threshold of 10 voxels.

3.1.2. Trait Judgments

Inspired by the discovery of brain areas specialized for conceptual knowledge about different classes of inanimate stimuli (e.g, tools, houses), Trait Judgment tasks were introduced. These tasks, aimed at finding brain areas with a comparable level of specialization for conceptual knowledge about persons (e.g. Mitchell, Heatherton, & Macrae, 2002), quickly gained popularity. Trait Judgment tasks in our meta-analysis presented participants with written material concerning another person’s traits (e.g. adjectives, opinions, or personal episodes).

Usually, the person was only described verbally to participants. However, a few studies presented photographs of the person characterized (e.g. faces, body parts, or person as a whole). Control conditions for Trait Judgments diverted attention away from these mental states by asking for linguistic judgments on trait words (e.g. 'is this word written in upper- or lower-case?') or presented words/statements that did not contain mental states.

3.1.3. Strategic Games

Early studies used Strategic Games for studying ToM, based on the idea that feedback from a social partner – indicated by his/her moves in the game – may trigger spontaneous mentalizing. Even when not explicitly asked to 'mindread', participants would spontaneously try to guess (i.e. infer) the intentions of the other player (e.g. Rilling, Sanfey, Aronson, Nystrom, & Cohen, 2004). This relates to the notion of an 'intentional stance' (e.g. Dennett, 1971), i.e., a disposition to reason about the beliefs, desires, and intentions of others (Gallagher, Jack, Roeperstorff, & Frith, 2002). Our meta-analysis included studies where participants were asked to play a game with another player with whom they could compete or cooperate (for example in the prisoner's dilemma game). Players received feedback about the decision of the other player, but they could not see each other. The contrast of interest was typically brain activation for playing strategic games with a human partner compared to playing with a computer (which follows a simple algorithm).

3.1.4. Social Animations

Social Animations were introduced (Castelli, Happé, Frith, & Frith, 2000; Castelli, Frith, Happé, & Frith, 2002) as a low-level alternative to verbal or cartoon-based materials used in the field (e.g. Baron-Cohen et al., 1994; Fletcher et al., 1995; Happé et al., 1996). The idea was to trigger ToM processes with minimal input in order to distinguish central mechanisms for mental-state attribution from other, potentially stimulus-related, processes. Studies in this task group presented video animations of simple geometrical shapes or objects which performed movements resembling intentional or social interactions. This type of stimulus is based on the classical triangles task by Heider and Simmel (1994). In control conditions, the animations showed random or purely mechanical movements (e.g. resembling the movement of billiard balls on the table). For each movie, participants were asked to explain / decide if an interaction between two shapes was portrayed. One study in this task group showed similar movies in the experimental and control conditions but varied instructions (e.g. by asking participants to focus on physical properties in the control condition).

Table 1.
Examples for studies in Theory of Mind task groups. For a complete list, see Supplementary Materials.

Author	Experimental Condition	Control Condition
False Belief (25 studies)		
Oliver et al., 2018	Read a short vignette involving a person with a false belief. Predict the behavior of this person based on her belief. Stimuli adapted from Dodell-Feder et al. (2011). e.g. *'The morning of high school dance Sarah placed her high heel shoes under her dress and then went shopping. That afternoon, her sister borrowed the shoes and later put them under Sarah's bed. Sarah gets ready assuming her shoes are under the dress.' (<u>Yes</u> or No).*	Read a short vignette involving a photograph/physical representation of the past, and a description how things depicted have changed by now. Answer a question about the outdated scene shown. e.g. *'Old maps of the islands near Titan are displayed in the Maritime museum. Erosion has since taken its toll, leaving only the three largest islands. Near Titan today there are many islands.' (Yes or <u>No</u>).*
Saxe & K., 2003	Read a short vignette involving a person with a false belief. Answer a question about her belief. e.g. *'John told Emily that he had a Porsche. Actually, his car is a Ford. Emily doesn't know anything about cars so she believed John. When Emily sees John's car, she thinks it is a …?' (<u>Porsche</u> or Ford).*	Read a false-photograph vignette. Answer a question concerning the outdated content in the photo. e.g. *'A photograph was taken of an apple hanging on a tree branch. The film took half an hour to develop. In the meantime, a strong wind blew the apple to the ground. The developed photograph shows the apple on the …?' (<u>tree</u> or ground).*
Trait Judgements (19 studies)		
Ma et al., 2011	Read written statements conveying trait diagnostic information about persons (describing behavior). Then read a single trait-adjective and indicate whether it is consistent with the behavior of that person. e.g. *'Tolvan gave her sister a hug ... consistent with "friendly"?'*	Read written statement about a person doing something. This behavior is neutral and does not convey trait diagnostic information about the person. Indicate the gender of the person in the sentence. e.g. *'Tolvan gave her mother a bottle … is Tolvan male or female?'*
Zhu et al., 2007	Read a personality trait adjective (e.g., *'brave', 'childish')* and indicate if it correctly describes a former American president (*Bill Clinton*, press *'yes'* or *'no'*).	Read a personality trait adjective (e.g., *'brave', 'childish')* and indicate if it is written in lower- or upper-case (press *'yes'* or *'no'*).
Strategic Games (13 studies)		
Takahashi et al., 2015	Play the matching pennies game against a human. Both players are asked to choose one out of two options at the same time. For one player, the goal is to choose the same options as the other. For the other player, goal is to choose the different option.	Play the matching pennies game against a computer. Both players are asked to choose one out of two options at the same time. For one player, the goal is to choose the same options as the other. For the other player, goal is to choose the different option.
Kircher et al., 2009	Play the prisoner's dilemma game (iterated version). You play with a human player for game points. Both players choose a cooperative or defective strategy on each trial. If both players choose defective, they gain almost no game points at all. If both choose cooperative, both gain some game points. If players choose differently, the defective player gains more points.	Play the prisoner's dilemma game (iterated version). You play with a computer for game points.
Social Animations (20 studies)		
Moessnang et al., 2016	Watch a video animation of two interacting triangles, which involve influence on each other's mental states (e.g. *coaxing*). Indicate if a social/goal-directed/random movement was shown, and indicate the feeling of both triangles (*positive/negative*). Respond via button press to both questions.	Watch video animation of two triangles, which interact in a goal-directed manner (e.g. *one chasing the other*). Indicate if a social/goal-directed/random movement was shown, and indicate the feeling of both triangles (*positive/negative*). Respond via button press to both questions.
Castelli et al., 2000	Watch video animation of two interacting triangles (e.g. '*mother and child are playing*'). Explain verbally what was happing (after fMRI).	Watch video animation of two randomly moving triangles. Explain verbally what was happing (after fMRI).
Reading Mind in the Eyes (15 studies)		
Baron-Cohen et al., 1999	View a photograph showing the eye-region of a face. Indicate which of two words (e.g., *concerned* versus *unconcerned*) describes the mental state of that person (button press).	View a photograph showing the eye-region of a face. Indicate if the person is male or female (button press).
Bos et al., 2016	View a photograph showing the eye-region of a face. Indicate if a mental state word presented before (e.g., *shy, hostile, playful*) matches the photo.	View a photograph showing the eye-region of a face. Indicate if a non-mental state word presented before (e.g., *woman, curly-hair, heavy eye-brows*) matches the photo.
Rational Actions (11 studies)		

Brunet et al., 2000	View a cartoon story and predict what will happen based on intentions of a character (no false belief). Choose a logical story ending from several options shown in pictures. e.g. '*A prisoner is in his cell. First, he breaks the bars of his prison window. Then he walks to his bed.*' Participants must indicate what will happen next … '*the prisoner ties a rope from the sheets on his bed /the prisoner shouts out loud.*'	View cartoon stories and predict what will happen (press button) based on physical causality. e.g. *A cartoon story shows a person standing in front of a slide. A large ball is coming down the slide, heading towards the person standing there. Participants must indicate what will happen next; pictures show the ball knocking over the person or the ball resting on the ground and the person standing next to it.*
Heleven et al., 2019	View cartoon stories showing a person over a sequence of events. Order of pictures (i.e. events) is jumbled. Indicate the correct order of pictures based on the intentions of the character (button press).	View cartoon stories showing a sequence of events. Order of pictures (i.e. events) is jumbled. Indicate the correct order of pictures based on physical causality (button press).

3.1.5. Reading Mind in the Eyes

The Reading Mind in the Eyes Test was introduced in neuroimaging research based on its capacity to dissociate social from more general abilities or intelligence, and was linked to Theory of Mind and mindreading ability in earlier work (Baron-Cohen et al., 1999b). To illustrate, it was found that adults with high-functioning autism spectrum disorder (Baron-Cohen, Jolliffe, Mortimore, & Robertson, 1997), as well as parents of children with autism spectrum disorder (Baron-Cohen & Hammer, 1997), show deficits on this task, but not children with William’s syndrome (Tager-Flusberg, Boshart, & Baron-Cohen, 1998). Neuroimaging studies using the Reading Mind in the Eyes tasks presented a photograph of the eye-region of a face, and asked participants to think about the person’s mental state, or to indicate which adjective (among several options) best described the person’s mental state. Control conditions showed similar photographs but asked for physical judgments of the persons depicted (e.g. gender or age), or in one exceptional case simply asked for passive viewing of a fixation cross. Note that, for sake of sample coherence (and in the light of the empathy tasks we compare here), we did not include two studies from the Reading Mind in the Eyes sample in Schurz et al. (2014), as these studies asked participants for more basic emotion judgments.

3.1.6. Rational Actions

Early studies presenting cartoons were introduced as a non-verbal alternative to story-based mentalizing (see e.g. Brunet, Sarfati, Hardy-Baylé, & Decety, 2000). This was in part to circumvent difficulties in studying social cognition in schizophrenia accompanied by speech disorganization (Brunet, Sarfati, & Hardy-Baylé, 2003; Sarfati, Hardy-Baylé, Besche, & Widlöcher, 1997). All tasks in the Rational Actions group presented short cartoons and asked participants to predict a likely ending, based on rational intentions of the protagonist (implied in his/her current actions). To keep the task group conceptually homogeneous and distinct from others, we did not include studies which featured false-belief related cartoons (e.g. deception), emotional scenes, or centrally featured communicative acts (communicative intentions). In control conditions, questions about non-mental aspects of the scenes were asked, as for example physical causality.

3.2. Task Groups empathy

3.2.1. Observing Pain

Pain has been a central theme for empathy research, as it was argued that it represents one of the most salient forms of suffering in others (Ochsner et al., 2008; Zaki, Ochsner, Henlin, Wager, & Mackey, 2007). Empathizing with another's suffering is an essential feature of human social behavior, seen as a critical precursor for more sophisticated forms of empathy and central for moral reasoning (e.g. Morrison, Lloyd, Di Pellegrino, & Roberts, 2004). Furthermore, because of its high saliency, pain is an effective stimulus for engaging participants in a task and measuring their brain activity. Studies in this task group presented pictures or videos showing a person’s face or body parts in painful situations. Tasks did not ask for an explicit judgment related to the painful stimuli, but rather for passive viewing or simple tasks demonstrating attentional engagement (e.g. asking which trial type was shown or to detect visual changes in a fixation cross between trials). Whereas all experimental conditions presented painful stimulation of body parts or faces, the control conditions presented neutral physical stimulation (e.g. being touched by a Q-tip) or no stimulation.

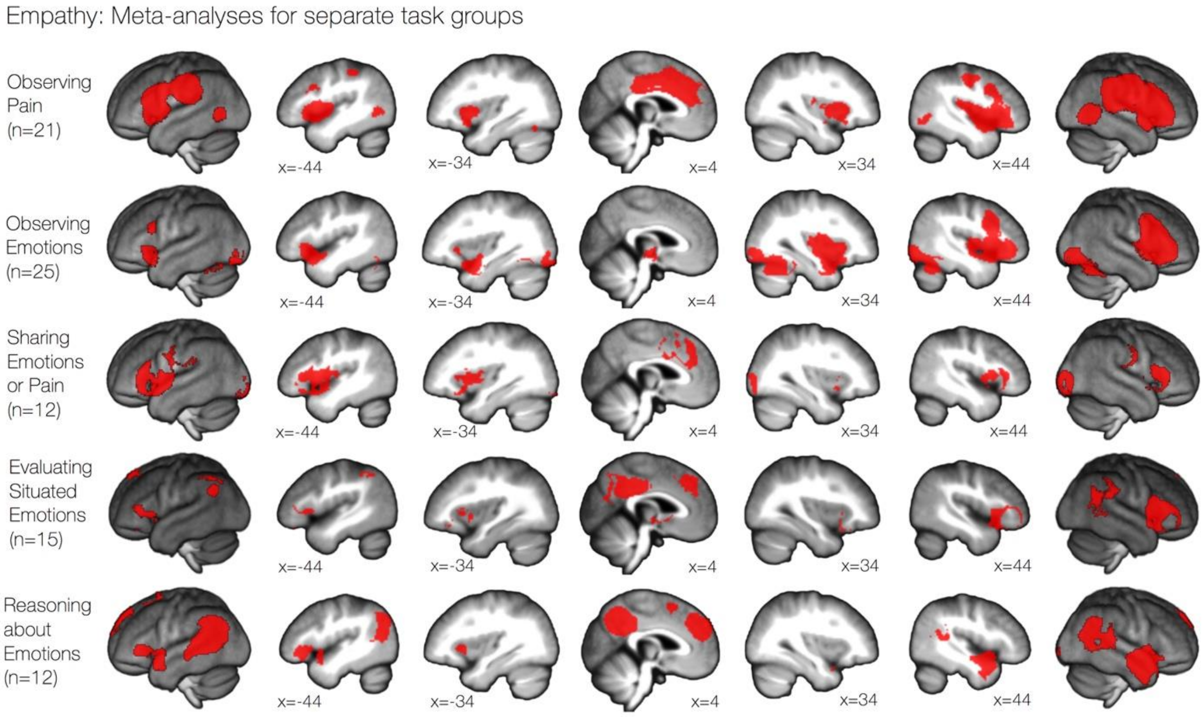


Figure 2. Results of separate meta-analyses for task groups for empathy. Maps were thresholded at voxel-wise threshold of $p < .005$ uncorrected and a cluster extent threshold of 10 voxels. For Reasoning about Emotions, we selected a common type of task that we found not only in the empathy literature, but also in studies on ToM.

3.2.2. Observing Emotions

Recognizing others' emotional states constitutes a process supporting empathy (e.g. Baron-Cohen, 2002; Schulte-Rüther, Markowitsch, Fink, & Piefke, 2007). However, an even stronger relationship between the observation of (facial) emotions and empathy has been suggested. Based on 'Mirror Neuron' theories (e.g. Rizzolatti & Craighero, 2004), it was hypothesized that (covert) mirroring of observed facial expressions triggers activity in emotional brain areas, and thus, an empathic response (Carr, Iacoboni, Dubeau, Mazziotta, & Lenzi, 2003; Wicker et al., 2003). Researchers explored these relationships in the context of simple passive viewing tasks presenting facial expressions. As the same task-type has also been widely used in other fields of neuroimaging research, we obtained additional studies for this task group from a large sample of suitable research (Fusar-Poli et al., 2009), to achieve n=25 as the sample size[10]. All studies in our task group presented pictures or videos of faces displaying basic emotional expressions (such as anger, fear, happiness, or disgust). Participants were not asked for explicit emotion judgments but passively viewed the stimuli or made judgments regarding non-emotional stimulus characteristics (e.g. gender or physical properties). In control conditions, faces showed neutral expressions or no stimuli were presented at all.

3.2.3. Sharing Emotions or Pain

While the previous task groups have been related to more automatically occurring processes, the present task group probes more voluntary and explicit forms of empathy (see e.g. de Greck et al., 2012a,b) that possibly link to top-down regulatory processes (see van der Heiden, Scherpiet, Konicar, Birbaumer, & Veit, 2013). The task group contains comparable stimuli as the previous categories (faces with basic emotions, body parts in painful situations), but here participants were explicitly instructed to share the emotional state of the target (e.g. 'feeling into' her). In some instances participants were additionally asked to rate the experienced or expressed emotion. In control conditions, participants made similar judgments but based on faces with neutral emotional expressions or body parts under non-painful stimulation.

3.2.4. Evaluating Situated Emotions

Another group of studies added situational context to empathy tasks. Here, researchers argued that empathy involves more than just focusing on another person, but also taking into account how he or she is embedded in a situation and contextual background (e.g. Ruby & Decety,

[10] After identifying the facial emotion viewing studies in the Empathy literature, we topped up this sample to achieve a task group of 25 studies. We randomly selected additional studies from a meta-analysis on implicit (uninstructed) facial emotion recognition (Fusar-Poli et al., 2009).

2004; Regenbogen, 2012). Tasks presenting such contextual information were taken to measure a more naturalistic form of empathy (see Mathur, Harada, Lipke, & Chiao, 2010; Zaki, Weber, Bolger, & Ochsner, 2009), different from previous task groups presenting (over-) simplified stimuli. Studies asked participants to judge a person's emotion based on the situation he or she was experiencing (e.g. by selecting among alternatives). The situational context was either verbally described by the target person, or given as a written narrative. The target person was additionally shown to participants in all tasks, either in a picture or video (telling the story of how the event happened). Control conditions either asked for similar judgments in the context of emotionally neutral content, or diverted attention away from emotional material by asking for physical judgments (e.g. judging whether the person was shown on the left or right side of the screen).

Table 2.
Examples for studies in empathy task groups. For a complete list, see Supplementary Materials.

Author	Experimental Condition	Control Condition
Observing Pain (21 studies)		
Olsson et al., 2007	Watch a video of a model undergoing fear conditioning situations (receiving the shock). No response.	Watch a video of a model undergoing fear conditioning situations (not receiving the shock). No response.
Bos et al., 2015	Watch videos of hands under painful (needle) stimulation. Watch attentively and detect changes in fixation cross appearing between videos.	Watch videos of hands under non-painful (cotton swab) stimulation. Watch attentively and detect changes in fixation cross appearing between videos.
Observing Emotions (25 studies)		
Critchley et al., 2000	View a picture of a face with an emotional expression (anger, fear). Judge the gender of the person (button press).	View a picture of a face with a neutral expressions. Judge the gender of the person (button press).
Toller et al., 2015	View video of face showing fearful expression. Relax and focus on the actors eyes.	View video of landscapes. Relax and do nothing.
Sharing Emotions or Pain (12 studies)		
Preis et al., 2013	View a picture of a body part (hand) under painful stimulation. Imagine how the person in the picture feels and rate her pain (VAS).	View a picture of a body part (hand) under non-painful stimulation. Imagine how the person in the picture feels and rate her pain (VAS).
Reniers et al., 2014	View a picture of a face with a sad expression. Imagine what the person in the picture is feeling. No response in the scanner.	View a picture of a face with a neutral expression. Imagine what the person in the picture is feeling. No response in the scanner.
Evaluating Situated Emotions (15 studies)		
Kanske et al., 2015	Watch a video of a person telling about a negative autobiographical event. Rate how you feel and how much compassion you feel with the person in the video (button press.	Watch a video of a person telling about a neutral autobiographical event. Rate how you feel and how much compassion you feel with the person in the video (button press).
Reyes-A. et al., 2017	Read a short vignette about an emotionally negative or positive event, then view a picture of the person to whom this happened. Think about what this person is feeling (no overt response).	Read a short vignette about an emotionally neutral event, then view a picture of the person to whom this happened. Think about what this person is feeling (no overt response).
Reasoning about Emotions (12 studies)		
Hooker et al., 2010	View a series of pictures with two persons. One person realizes that she mistakenly had a false belief regarding an emotion-triggering state of affairs. That leads to a change of her emotions. Indicate if the pictures showed a social (i.e. emotional), physical, or no change (button press).	View a series of pictures with two persons. One person has a false belief regarding an emotion-triggering state of affairs. Nothing changes over the pictures (i.e. so this person does not realize that she had a false belief). Indicate if pictures showed a social, physical, or no change (button press).
Voellm et al., 2006	View a cartoon story showing two persons. One is in an emotion-triggering situation. Predict what the other person will do to make her feel better (button press).	View a cartoon story showing two persons in neutral everyday situation. Predict what will happen next based on physical causality (button press).

VAS... Visual Analog Scale (allows ratings along a continuous scale)

3.2.5. Reasoning about Emotions

The intersection between empathy and ToM is a recurrent topic in research, bearing concepts such as cognitive empathy (Hooker, Verosky, Germine, Knight, & D'Esposito, 2010; Preston & de Waal, 2002; Schnell, Bluschke, Konradt, & Walter, 2011; Shamay-Tsoory et al., 2009), affective ToM (Mier et al., 2010; Schlaffke et al., 2015; Sebastian et al., 2012), mentalizing about emotion (Hooker Verosky, Germine, Knight, & D'Esposito, 2008), and emotional perspective taking (Derntl et al., 2010; 2012). Across the diversity of labels, we identified a number of tasks with coherent stimuli and instructions, which were characterized by combining mental state reasoning with emotion judgments. The interrelation between those elements could go in either direction. One set of studies asked participants to infer a future rational action (and therefore, rational intention) triggered by an emotion. A second set of studies asked for inferences about an emotion or emotional change triggered by beliefs or a belief-revision (e.g. a person becomes aware of an emotionally relevant object or event). The typical stimulus format in this task group were pictures or cartoons (but we accepted one additional task with a verbal story format). Control conditions asked for inferences about future events based on physical causality or other forms of less complex inference. Although it can be strongly linked to both the empathy and ToM literature, we cover this task group in the section on empathy tasks. This is because a (weak) majority of its studies were found by empathy-related keywords. Note, however, that labelling this task as an empathy (or ToM) task has no effect on our clustering, which is purely driven by the features of our meta-analytic brain activation maps, irrespective of terminology.

3.3. Clustering

Based on the meta-analytic result maps obtained for all eleven task groups, we carried out clustering. This is the central step of our analysis, which allows us estimating an appropriate number of sub-components of ToM and empathy. For an overview of correspondences between task groups, we show the image dissimilarity of the eleven meta-analyses in Figure 4A. In Figure 4B we present a dendrogram that illustrates task-by-task and cluster-by-cluster relations in a dendrogram. Based on the information shown in the dendrogram, we selected an optimal clustering based on two features (see Figure 3A). First, our desired model ensures a good separation of brain activation patterns between clusters. Second, components of a good model should be sufficiently abstract to generalize across concrete instances, i.e. multiple tasks groups, rather than picking up variance related to one outlier task group.

Metrics show that among all clusterings, the three cluster solution shows best performance (see Figure 3A). This result is of central relevance for our aim to find common neurocognitive components across ToM and empathy tasks. We will argue throughout the next sections and again in our Conclusion section (5.1.) that the three cluster solution reflects central processes for social cognition.

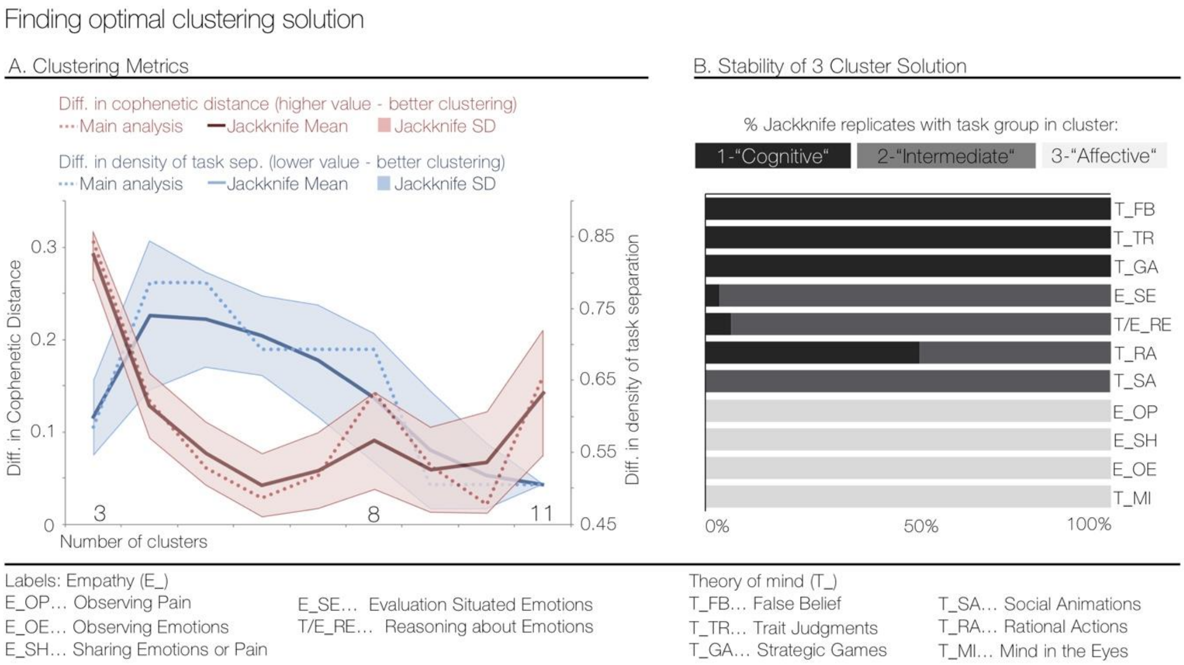


Figure 3. (A) Evaluation and comparison of different clustering solutions. Plot shows changes in two metrics when moving from 2 to 3, 3 to 4, …, and 10 to 11 cluster solutions. The first metric, shown in red and on the left Y-axis, gives the relative difference in cophenetic

distances when moving from one model to the next. A relatively high difference in cophenetic distances indexes that introducing new sub-clusters results in better separation of brain activity patterns, and thus, good clustering. The second metric, shown in blue, gives the density of task separation, reflecting whether separating clusters into sub-clusters maintains a balanced number of task groups in each sub-cluster. Relatively low density of task separation indexes good clustering. Preferred clusterings in terms of both metrics are indicated on the X-axis: 3, 8, and 11. Metric changes are shown for clustering of complete meta-analyses (main analysis), and clusterings based on leave-one-out jackknife sensitivity analysis with 5000 repeats (jackknife mean, standard deviation). (A) Stability of the assignment of task groups (i.e., associated meta-analytic maps) to clusters for the three cluster solution, based on leave-one-out jackknife sensitivity analysis. Bars indicate percentage of jackknife-repeats for which task groups were assigned to the same clusters as in the main analysis.

While we also observed that the simpler two cluster solution already explains a part of the variance in brain activation, the three cluster solution explains substantial additional variance on top of that[11]. Therefore, we will discuss both the two- and three cluster solution as high-level clustering solutions (i.e. those that divide the data only in a small number of clusters). Across all clustering solutions, we found clusterings with further explanatory value at model orders 8 and 11, with the 11 cluster solution performing particularly well. As we will lay out in the next sections, the hierarchical structure of our clustering allows us to adopt a multilevel perspective on our results, consistent with previous conceptualizations of social cognition as a multi-layered or multi-level phenomenon (De Waal, 2012; Preston & De Waal, 2002; Singer, 2006; Schaafsma et al., 2015). The "Russian-doll" model, for instance, proposes that the core functions of motor mimicry and emotional contagion are embedded in several layers of more complex processes, ranging up to perspective taking (De Waal, 2007; De Waal & Preston, 2017). In the next sections, we will discuss the 2, 3, and 11 cluster solution. Finally, we will integrate them by proposing a multilevel model of social cognitive processes in our Conclusion section (5.1).

[11] While the two-fold clustering fell out of range for our metric calculation, its explanatory power is shown in the dendrogram in Figure 4B. Due to the hierarchical nature of our clustering, good performance for both the two- and three cluster solution are compatible observations: The three cluster solution contains the division made in the two cluster solution (i.e. it does not lose this information). On top of that, it adds another division (by splitting the cluster containing mostly ToM tasks in two halves), which explains substantial additional variance.

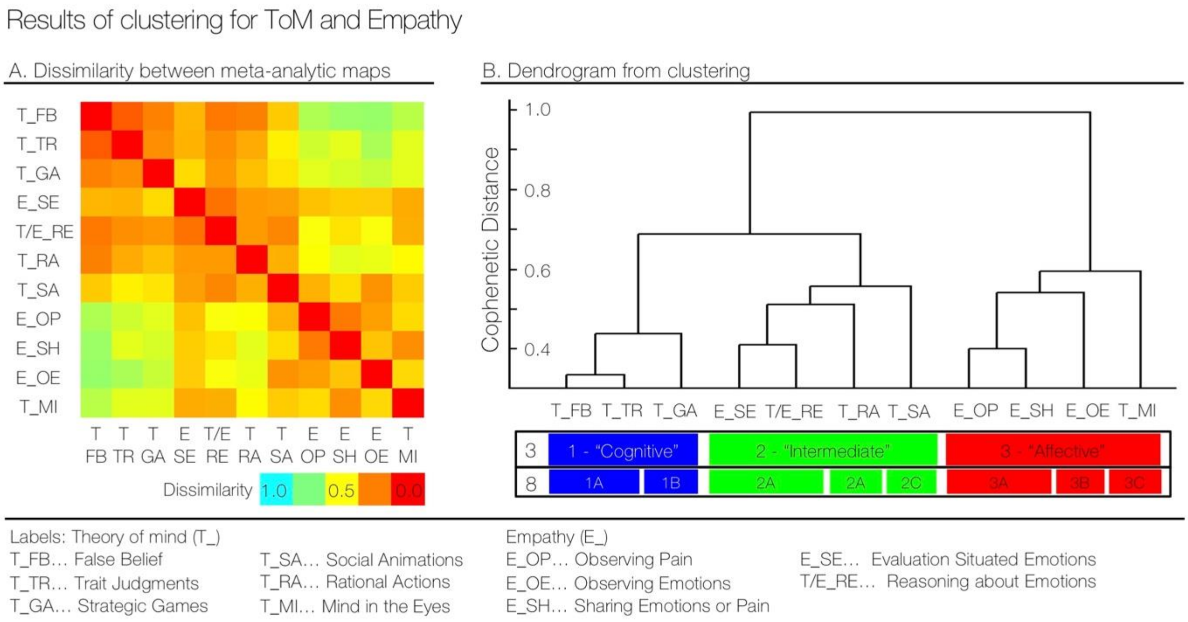


Figure 4. (A) Degree of dissimilarity between meta-analytic result maps (given as 1-Pearson's *r*). High dissimilarity reflects little correspondence or overlap between brain maps. (B) Dendrogram from hierarchical clustering. Height of the branches indexes cophenetic distances, and thus the dissimilarity between sub-clusters at a model order. Color bars indicate positions of 3- and 8-cluster solutions with respect to the dendrogram.

3.4. High-level clusterings

3.4.1. Two cluster solution

The two cluster solution shows that our approach could retrace large parts of the classical ToM vs. empathy distinction made in the literature (see e.g., Bzdok et al. 2012; Kanske, Böckler, & Singer, 2017; Preston & Hofelich, 2012; Walter, 2012). The networks that we found for the two cluster solution (Figure 5) broadly converged with typical ToM (e.g. Koster-Hale & Saxe, 2013; Mitchell, 2009; Molenberghs et al. 2016; Saxe & Kanwisher, 2003; Van Overwalle, 2009) and empathy areas, respectively (e.g. Bzdok et al. 2012; Singer & Lamm, 2009; Timmers et al., 2018). Out of 6 task groups retrieved from the ToM literature, 5 ended up in a common cluster. On the other hand, 3 out of 5 task groups from the empathy literature were grouped together. Figure 5 illustrates how these three task groups end up at somewhat unexpected positions in higher levels of the clustering.

The first group was the Reading Mind in the Eyes task, which researchers originally described as 'an advanced test of theory of mind' (e.g. Baron-Cohen et al., 1999b). This task clustered together with other task groups drawn from the empathy literature (see Figure 5). Brain activation in the left inferior frontal gyrus, anterior insula, and anterior cingulate cortex (see also Molenberghs et al., 2016) was largely overlapping with areas found for the other empathy tasks in the cluster. Despite asking participants to select high-level mental-state-related words such as 'interested', 'affectionate' or 'contented', it has been suggested that the Reading Mind in the Eyes test might engage processing of emotional states and detection of subtle facial cues. Oakley, Brewer, Bird, & Catmur (2016) found that performance on the Reading Mind in the Eyes task is more strongly related to individual differences in alexithymia than in autism spectrum disorder, suggesting that it measures emotion recognition ability in addition to, or even rather than, ToM abilities. Furthermore, as we review in more detail in section 3.7, task-performance correlations in non-impaired participants link the Reading Mind in the Eyes test to face-based emotion categorization (Olderbak et al., 2015), but not to processing of beliefs (e.g. Strange Stories task: Rice, Anderson, Velnoskey, Thompson, & Redcay, 2016; Dziobek et al., 2006). In addition, a link to more intermediate tasks combining cognitive and affective elements has been reported (Ferguson & Austin, 2010, see section 3.7 for further explanation).

The other task groups that clustered unexpectedly were Evaluating Situated Emotions and Reasoning about Emotions. We will discuss potential processes underlying these tasks later in section 3.5.2.3. While Evaluating Situated Emotions tasks have been labelled 'empathy' tasks in previous research, studies from the Reasoning about Emotions group have been described more heterogeneously. Similar paradigms have been linked to 'cognitive empathy' (Hooker et al., 2010), 'affective ToM' (Schlaffke et al., 2015), or 'mentalizing about emotion' (Hooker et al., 2008). We assigned the task group to empathy as it contained more studies with empathy- than ToM-related keywords. However, such an a priori assignment is debatable. Note however, that our data driven clustering would have assigned this task group at the same position irrespective of our labelling as 'empathy' or 'Theory of Mind' task.

3.5.2. Three cluster solution

The three way clustering reflects a central result of our analysis, and will be discussed in depth throughout the next sections. Finally, in the Conclusion section (5.1), we propose that it reflects the higher, central level of a hierarchical multilevel model of social cognitive processes. Figure 5 shows the activated networks for the three clusters. For simplicity, we will refer to these clusters as 'cognitive' (1), 'affective' (3), and 'intermediate' (2). The task-by-task dissimilarity matrix in Figure 4A suggests that the cognitive and affective clusters form opposite poles with largely distinct brain activity profiles while the intermediate cluster bears similarities to both poles. This observation is further supported by a weaker image correlation between the (untresholded) cognitive and affective cluster maps (r = .46) compared to correlations between the intermediate and other two maps (intermediate-cognitive r = .80, intermediate-affective r = .76).

For the functional description of the three clusters that follows, we applied neurosynth functional decoding (Yarkoni et al., 2011): a tool that allows a brain activation map to be compared with an extensive set of automatically generated meta-analyses across a wide range of topics ('terms'). Highest convergence with social and non-social terms is shown in Figures

6 and 7. On the basis of this broad decoding, we can discuss our findings not only in the context of classical theories of social processes, but also alternative theories that suggest more general purpose processes to underlie ToM (e.g. Buckner & Carroll, 2007; Corbetta, Patel, & Shulman, 2008; Heyes & Frith, 2014; Heyes, 2018) and empathy (e.g. Barrett, Lindquist, & Gendron, 2007; Lindquist, Sapute, & Gendron, 2015; Wager et al., 2016, but see Lieberman & Eisenberger, 2015, 2016).

3.5.2.1. Cognitive cluster (Cluster 1)

Brain activation for the cognitive cluster (see Figure 5) was mainly found in cortical midline and temporo parietal areas. Strongest activation was found in the anterior cingulate cortex and medial Prefrontal Cortex. This cluster of activation extended along the cortical midline to precuneus and parts of mid cingulate cortex. Bilateral temporo parietal areas included the right posterior superior temporal gyrus, right supramarginal gyrus, left posterior middle temporal gyrus and inferior parietal lobule. Additional areas were found in bilateral anterior temporal cortices, and a smaller subcortical cluster (caudate). Compared to a resting-state network atlas of the brain (Yeo et al., 2011), activations for cluster 1 showed most prominent overlap with the Default Mode Network (DMN). More specifically, 56 % of voxel of cluster 1 fell within the DMN, followed by smaller overlaps with the Frontoparietal Network (9 %), and the Ventral Attention Network (also known as Salience Network, 9 %).

Neurosynth decoding (Figure 6) for social terms found the strongest associations with ‘theory of mind’, ‘mentalizing’ and related terms, characterizing the cognitive cluster as most prototypical of 'ToM'. For non-social terms, the strongest associations were found for ‘default’, ‘self-referential’, and ‘autobiographical’. These later decoding results mirror neurocognitive accounts of understanding others that emphasize a role of the Default Mode Network (DMN; see Mars et al., 2012; Meyer, Davachi, Ochsner, & Lieberman, 2018;

Spreng et al., 2010; Spunt, Meyer, & Lieberman, 2015; Buckner & Carroll, 2007; Bzdok et al., 2013).

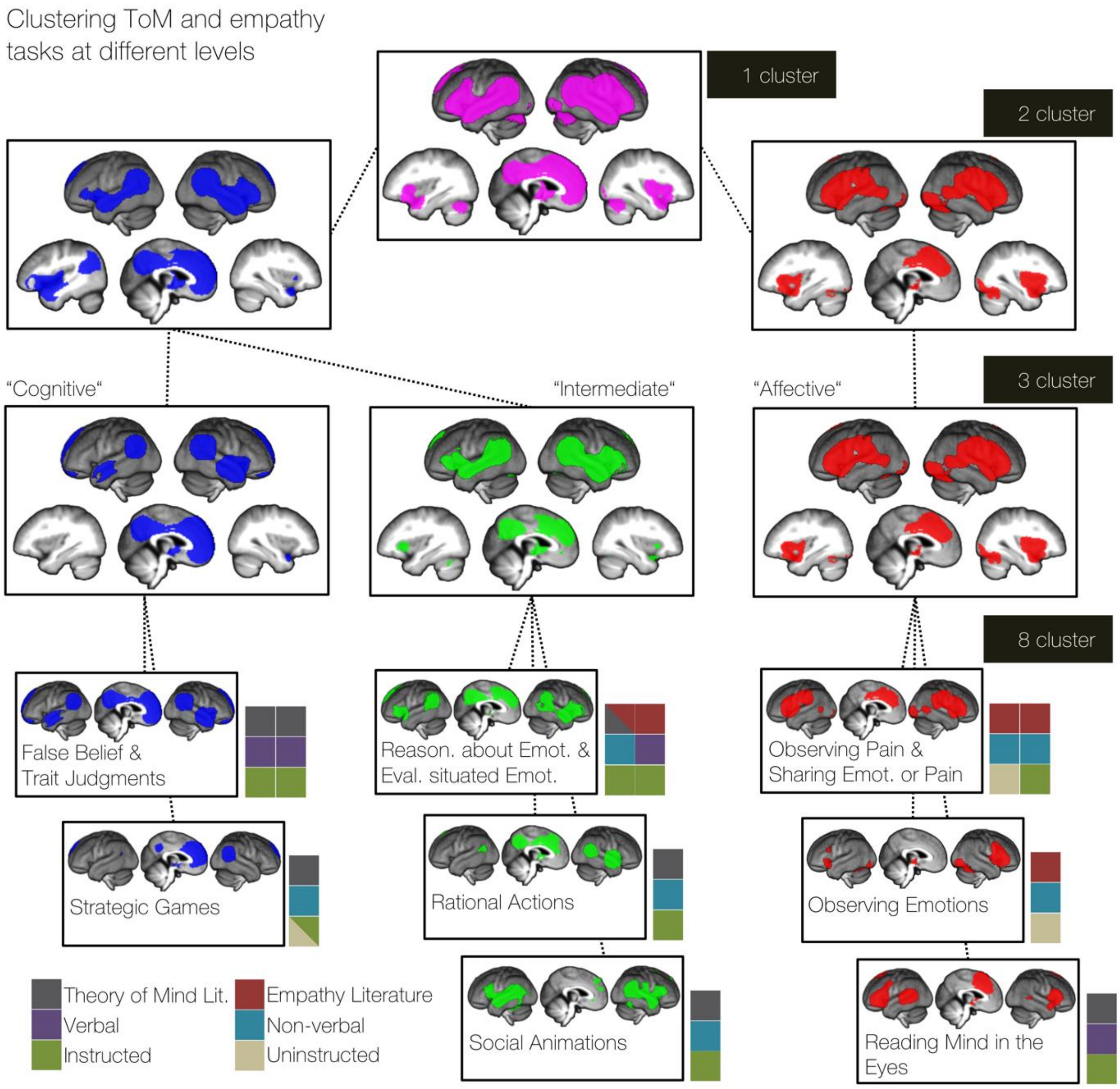


Figure 5. Mean brain activation for clusters at different model orders. We carried out pooled meta-analyses, i.e. one separate meta-analysis per cluster, where all its task groups are joined together. Analyses were thresholded at voxel-wise threshold of $p < .005$ uncorrected and a cluster extent threshold of 10 voxels. The 1-cluster solution is shown for illustrative purposes only, and was not evaluated or compared against other clusterings. Colors indicate how the 3 cluster solution relates to both higher- and lower-level clusterings (blue - cognitive cluster, green - intermediate cluster, red - affective cluster). At the lowest-level of the dendrogram, we indicate for each cluster some exemplary stimulus- and task-categorizations. The 8 cluster solution was selected as a representative low-level clustering. However, note that the 11 cluster solution shows favorable clustering metrics, but corresponds to what has been shown in Figures 1 and 2 (i.e. complete separation into individual task groups).

It has been argued (e.g. Andrews-Hannah, 2014) that self-generated cognition decoupled from the physical world is mediated by the DMN. This becomes relevant for ToM as we do not have immediate perceptual access to others' mental states (see Frith & Frith, 2003; Lieberman, 2007). For example, the 'self-projection' hypothesis (Buckner & Carroll, 2007) states that the DMN uses past experiences in an adaptive fashion to imagine perspectives and events beyond those that emerge from the immediate environment. In line with that, overlapping parts of the DMN have been found implicated in other- as well as self-related mental state reasoning (e.g., Mitchell et al., 2005; Murray et al., 2012).

Related theories were formulated in the ToM field (independent of the DMN), linking ToM to a decoupling mechanism that allows the separation of beliefs from reality (Frith & Frith, 2003, 2012; Gallagher & Frith, 2003, metarepresentation of mental states in the form of propositional attitudes (Leslie, Friedman, & German, 2004), the processing of covert (i.e. unobservable) mental states (Gobbini, Koralek, Bryan, Montgomery, & Haxby, 2007) or perspective differences (Perner & Leekam, 2008, Perner & Rössler, 2012).

An interesting feature of the cognitive cluster is that it contained a specific subset of tasks from the ToM literature: False Belief tasks, Trait Judgments, and Strategic Games. While general-purpose DMN theories usually make no predictions regarding the type of ToM task that should engage this network, theories from the ToM field usually associate False Belief tasks with decoupling (e.g. Frith & Frith, 2003) and processing of covert mental states (Gobbini et al., 2007). Relatedly, Strategic Games require tracking of potential deception (and thus, beliefs) and therefore, can also be linked to decoupling.

Less frequently mentioned by DMN/decoupling theories is the processing of personality traits (i.e. Trait Judgments). Arguably, personality traits could be seen as mental states that are abstracted (i.e. generalized) across concrete instances, and are sometimes also decoupled from observable behavior (e.g. a person might perform the same dangerous action either out of courage or recklessness). Moreover, it was hypothesized that transient (e.g. beliefs) and stable (e.g. traits) mental states of others are jointly processed by a multilevel representation (Tamir & Thornton, 2018), where knowledge about a person's traits is used to guide expectations about transient mental states (Conway, Catmur, & Bird, 2019; Tamir & Thornton, 2018; see also Thornton, Weaverdyck, & Tamir, 2019). To illustrate, a particularly distrustful person might be harder to deceive, and therefore, less likely to have a false belief (Conway et al., 2019).

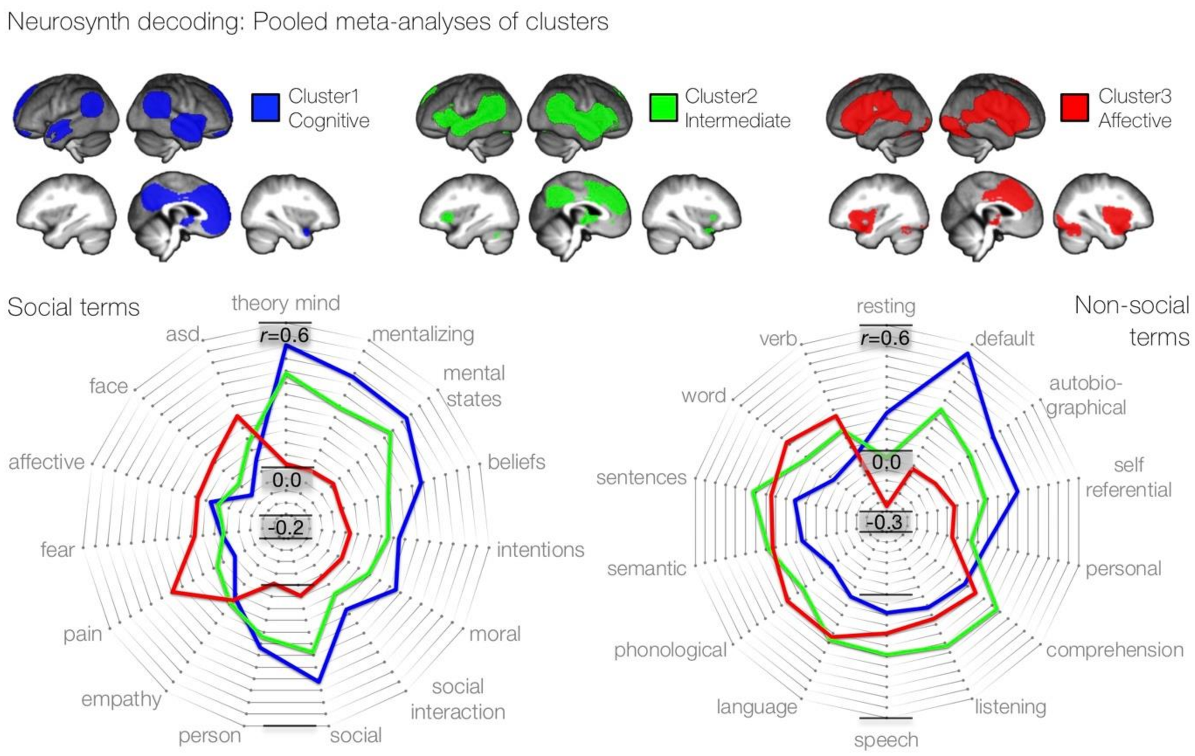


Figure 6. Mean brain activation from meta-analyses for the three cluster solution (pooled meta-analyses), at a voxel-wise threshold of $p <$ .005 uncorrected and cluster extent threshold of 10 voxels. Polar plots show results from large-scale meta-analytic decoding (neurosynth.org; Yarkoni et al., 2011). Meta-analysis maps were compared to an extensive set of automatically generated meta-analyses across a wide range of topics ('terms'), and findings of highest convergence (image correlation, i.e. Pearson's r) are shown. For each of our three clusters, we show the 5 most strongly correlated terms found for social and non-social topics (for further details, see methods section 2.4).

3.5.2.2. Affective cluster (Cluster 3)

For the affective cluster, we found brain activation (see red in Figure 5) across right frontal cortex, peaking in inferior frontal gyrus, and extending to right insula and temporal pole, pre- and postcentral gyri, as well as the supramarginal gyrus. Further areas were activated in left inferior frontal gyrus, insula, temporal areas and supramarginal gyrus. Another large area was activated in supplementary motor area and adjacent medial frontal gyrus and mid cingulate cortex. Two smaller areas were also found in left inferior occipital gyrus and left cerebellum. Overlaps with resting-state networks (Yeo et al., 2011) for cluster 3 were mainly found in the Ventral Attention Network (26 %), the Somatosensory Network (16 %), and the Default Mode Network (14 %). The most prominent social terms found with neurosynth decoding where ‘pain’, ‘fear’, ‘affective' and 'face' (as well as the clinical term 'asd'). For non-social terms, we found high loadings on a number of language-related terms, such as ‘word’, ‘phonological’, 'language' or ‘semantic’ (for a discussion of possible roles of language processes, see section 3.6). In line with the terms ‘pain’ and ’fear’ found by neurosynth decoding, strong activation for the affective cluster was found in the left insula. This structure has been described as part of a 'core network' that activates whenever we witness the suffering of others (e.g. Bzdok et al., 2012; Lamm et al., 2011; Preckel, Kanske, & Singer, 2018). To illustrate, studies on empathy for pain and other negative emotions consistently found activation in anterior insula and anterior cingulate cortex (Singer et al., 2004; Jackson, Meltzoff, & Decety, 2005; Jabbi, Swart, & Keysers, 2007; Kanske et al., 2015; Tholen, Trautwein, Böckler, Singer, & Kankse, 2020). Activation patterns in these areas were found to predict the affective and emotional states of an observed other. Such a relationship could be found across different modalities such as pain, disgust, or unfairness (Corradi-Dell’Acqua, Tusche, Vuilleumier, & Singer, 2016; but see Krishnan et al., 2016), which suggests that parts of these brain areas encode affective rather than sensory features of stimuli. Moreover, the same areas were found not only active during emotion observation, but also when participants themselves experienced an emotion first-hand (e.g. Lamm et al., 2011; Rütgen et al., 2015). This has been taken as evidence for 'shared networks' in empathy (Carrillo et al., 2019; Corradi-Dell'Acqua et al., 2016; Preston & de Waal, 2002; Gallese & Goldman, 1998; see also Alcalá-Lopez, Vogeley, Binkofski, & Bzdok, 2019). In addition to left insula, the affective cluster showed activation in the left inferior frontal/precentral gyrus, postcentral (somatosensory) and supramarginal gyri, which again have been linked to 'shared networks'. In particular, premotor cortex has been linked to 'mirroring' of emotional expressions, that is, covert (or overt) imitation of observed emotional facial expressions (c.f. Adolphs, 2009; Carr et al., 2003; Decety & Jackson, 2004; Gazzola, Aziz-Zadeh, & Keysers, 2006; Pfeifer & Dapretto, 2011; Shamay-Tsoory et al., 2009). The shared networks hypothesis was embedded in the general framework of 'Mirror Neurons' (Gallese & Goldman, 1998; Gallese, 2003, Rizzolatti & Craighero, 2004), and is compatible with theories of common coding for action and perception (Keysers, Kaas, & Gazzola, 2010; Keysers & Gazzola, 2009) and perception-action models (e.g. de Waal & Preston, 2017; Preston, 2007; Preston & Hofelich, 2012; Preston & Waal, 2002). These models assume that seeing an emotional expression automatically activates the corresponding motor- and somatosensory representations in the observer (producing an embodied representation), which facilitates decoding/understanding of these emotional states. Such an interpretation is also supported by our neurosynth decoding results, identifying the terms ‘action’ and ‘motor’ for the affective cluster.

Tasks in the affective cluster broadly align with the notion of an empathy core network for witnessing pain or emotions of others. The cluster contained those 3 task groups from the empathy literature which present relatively simple stimuli concerning others' pain and emotions without any context or additional information (Observing Pain, Observing Emotions, Sharing Emotions or Pain). In addition, it included the Reading the Mind in the

Eyes task from the ToM literature, which also fits to this description. Interestingly, shared network (e.g. Preston & de Waal, 2002) and mirroring (e.g. Gallese & Goldman, 1998) accounts often contain the notion that the implicated processes are taking place spontaneously (or automatically), i.e. in the absence of explicit task instructions (see also Cracco et al., 2018; Heyes, 2011). In fact, for half of the task groups in the affective cluster, participants were not explicitly instructed to empathize with others. This suggests that at least part of the processes linked to affect sharing do not need to be volitionally initiated, in keeping with the aforementioned accounts.

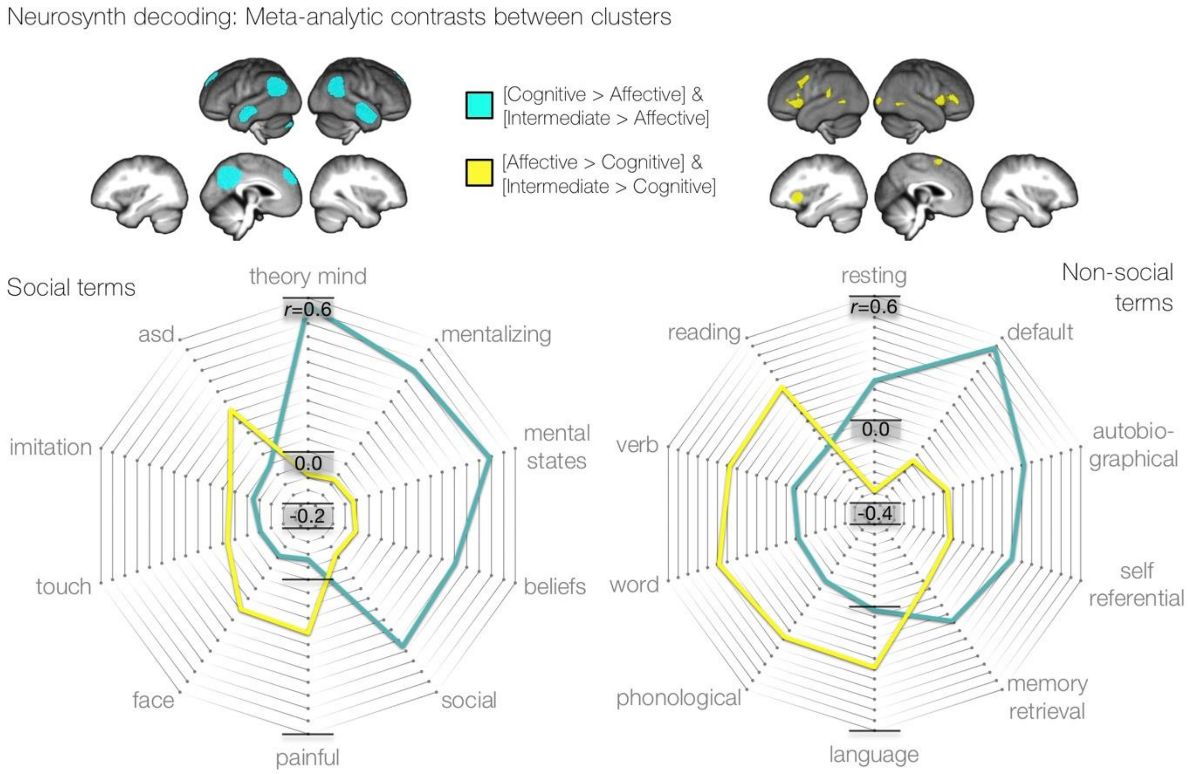


Figure 7. Meta-analytic contrasts for the three clusters, tailored to find specific commonalities between the intermediate and each of the other two clusters. Maps are shown at a voxel-wise threshold of $p < .005$ uncorrected and cluster extent threshold of 10 voxels. Meta-analytic decoding (neurosynth.org; Yarkoni et al., 2011) results are given in polar plots, expressed as image correlation, i.e. Pearson's *r*. For each contrast, we show the 5 most strongly correlated terms found for social and non-social topics. As we did not find any specific commonalities among the cognitive and affective cluster (at the statistical threshold), no decoding was done for this contrast.

3.5.2.3. Intermediate cluster (Cluster 2)

While the cognitive and affective clusters reflect two largely independent processes, the intermediate cluster takes an interesting position between them. In this section, we illustrate how it combines cognitive and affective elements in terms of activated brain areas and included task groups. We found brain activation for the intermediate cluster in large parts of the bilateral temporal lobes, spanning from the posterior superior temporal gyri to the anterior temporal lobes. We observed activation in areas overlapping with parts of the cognitive cluster (cluster 1), including bilateral temporo-parietal cortex and precuneus. Interestingly, we also found some convergence with activations found for the affective cluster, e.g. in left insula and inferior frontal gyrus. In terms of overlaps with resting-state networks (Yeo et al., 2011), activations for cluster 2 fell in the DMN (43 %), the Ventral Attention Network (18 %), and the Frontoparietal Network (10 %). To further investigate activation for this cluster, we carried out a set of meta-analytic contrasts tailored to identify commonalities between the intermediate cluster and the other two clusters[12]. Analyses confirmed significant activation specific for the cognitive and intermediate clusters in the precuneus, dorsal-posterior medial prefrontal cortex, bilateral temporo-parietal, and anterior temporal areas (see turquoise in Figure 7). Overall, these activations were more pronounced in right hemispheric areas. For the intermediate and affective clusters (see yellow in Figure 7), strongest common activations were found in left insula, and furthermore in bilateral inferior frontal gyri, left precentral gyrus, left superior temporal gyrus, and supplementary motor area. Decoding of the intermediate cluster showed largely a combination of terms already found for the cognitive and affective clusters. Notably, however, some language related terms showed higher loadings on the intermediate cluster compared to both other clusters: ‘sentences’, ‘speech’, 'listening', and ‘comprehension’. When decoding activations common to the intermediate and

[12] To find specific commonalities for the intermediate and the cognitive cluster (i.e. that go beyond what is also found for the affective cluster), we computed the conjunction of contrasts cognitive>affective & intermediate>affective. Likewise, for specific commonalities between the intermediate and the affective cluster, we computed intermediate>cognitive & affective>cognitive. For completeness, we also computed the contrast cognitive>intermediate & affective>intermediate, which found no significant activations.

the cognitive cluster (see Figure 7), we found social terms like ‘theory of mind’, ‘mentalizing’, and ‘mental states’, and non-social terms such as ‘default’, ‘self-referential’, and ‘autobiographical’. Decoding commonalities between the intermediate and the affective cluster, we found social terms such as ‘touch, ‘painful’, ‘face’, and ‘imitation' (as well as the clinical term 'asd'), and non-social terms like ‘phonological, ‘reading’, ‘word’ and ‘language.

Together, meta-analytic contrast maps and decoding results (Figure 7) highlight that the intermediate cluster combines cognitive and affective processes. Although theories proposed several variants of this combination, the fact that we found one intermediate cluster suggests also one common process. This common process could account for the concepts affective ToM (Mier et al., 2010; Schlaffke et al., 2015; Sebastian et al., 2012), cognitive empathy (Hooker et al., 2010; Preston, Bechara, Damasio, Grabowski, Stansfield, Mehta et al., 2007; Preston & de Waal, 2002; Schnell et al., 2011; Shamay-Tsoory et al., 2009, Walter, 2012), and emotional perspective taking (Derntl et al., 2010; 2012).

The notion that the intermediate cluster combines cognitive and affective processes is also reflected in elements of the task groups found in this cluster. As we will argue in our Conclusion section (5.2), these features could make the intermediate cluster particularly relevant with respect to measuring ecologically valid social cognition. With respect to cognitive processes, the task groups Reasoning about Emotions, Social Animations and Rational Actions all involve mental-state reasoning (i.e. about goals, intentions, or beliefs) by definition. Also the event narratives presented in the task group Evaluating Situated Emotions provide rich information for inferring not only the feelings of a person, but also his or her thoughts and desires. In line with this notion, Zaki et al. (2009) found that accurate empathic judgments based on event narratives are accompanied by co-recruitment of more affective areas (premotor cortex, inferior parietal lobule), and more cognitive areas that have been linked to mentalizing (e.g. medial Prefrontal Cortex, superior temporal sulcus). Besides cognitive features, most tasks from the intermediate cluster also call for affective processing. While the two task Evaluating Situated Emotions and Reasoning about Emotions do so per definition, there is evidence that also Social Animations trigger affective processes. Tasks from this group mostly depict affective scenarios (e.g. mocking, tricking, coaxing), and have been reported to elicit subjective emotional responses (Rimé, Boulanger, Laubin, Richir, & Stroobants, 1985). The involvement of affective processes is less clear for Rational Actions, which did not include explicitly affective stimuli. Possibly in relation to this, our clustering stability analysis (Figure 3B) found that this task group frequently changed membership in favour of the cognitive cluster. Taken together, our task groups suggest that the degree to which affect-laden contents and mental-state inference occur together determines whether cognitive and affective networks work in concert.

3.6. Possible roles of language processes in the three cluster solution

A notable finding from neurosynth decoding was that both the intermediate and the affective cluster showed high loadings on language related terms. This supports results from a previous meta-analysis (Andrews-Hanna, Smallwood & Spreng, 2014), which noted strong overlap between default mode and language-related processes, which was discussed in relation to social cognition. Possible contributions of language processes were discussed both in the empathy and ToM literature. In the empathy field, at least two alternative views have been proposed on the topic. Based on mirroring accounts, a first line of theorizing followed the observation that areas with ‘mirror properties' (e.g. Broca’s area, BA 44/45) overlap with areas traditionally linked to speech production in humans (e.g. Rizzolatti & Craighero, 2004). It was hypothesized this overlap might reflect that human language, including phonology and syntax, is embedded in the organizational properties of the motor system. Accordingly, the motor system represents an evolutionary precursor of language functions (e.g. Gallese, 2008; but see Toni, De Lange, Noordzij, & Hagoort, 2008). Our decoding might therefore not necessarily reflect language processing per se, but rather motor/mirror processes taking place in the same regions. A second line of theorizing suggests a more direct role of language for emotion processing and categorization. It has been repeatedly found that brain networks for emotion processing overlap with large parts of networks for semantic cognition (Lindquist, Wager, Kober, Bliss-Moreau, & Barrett, 2012; Binder, Desai, Graves, & Conant, 2009; see also Brooks et al., 2017). Based on that, psychological constructionist accounts have suggested that concepts available as words (e.g. 'anger,' 'disgust,' 'fear') shape how people understand their experiences as specific emotions (e.g. Barrett et al., 2007; Lindquist et al., 2015; 2017; Wilson-Mendenhall, Barrett, Simmons, & Barsalou, 2011). Emotions have been taken to arise from a combination of conceptual processing supported by language, and more basic experiences of affect such as the feeling of pleasure or displeasure. Furthermore, a role of semantic knowledge has been discussed in ToM-related accounts. For example, it was hypothesized that ToM involves mental-state concepts that we learn (as children) by communication with expert 'mindreaders' (e.g. Heyes & Frith, 2014). Moreover, some forms of mentalizing have been linked to retrieval of social semantic scripts (e.g. Frith & Frith, 2003; Gallagher & Frith, 2003), which contain knowledge about which activities take place in different contexts. Taken together, these theories suggest that both ToM- and empathy-related processing involves language capabilities. Interestingly, while decoding showed weaker loadings for the cognitive cluster (containing some classical verbal ToM tasks) on language related terms (e.g. 'sentences', 'language), stronger relations were found for the affective and intermediate clusters.

3.7. Behavioral separability of tasks from the three cluster solution

Exploring the broader implications of the three cluster model, we next evaluated whether our neural task clustering is mirrored by a behavioral task 'clustering'. Therefore, we reviewed behavioral literature on healthy adults (i.e. different studies than those in our meta-analysis[13]) for reports of performance correlations among social cognition tasks. If the three clusters we found on the neural level reflect different neurocognitive processes of social cognition, then behavioral studies presenting tasks similar to those in our neuroimaging meta-analysis should find a corresponding pattern of task-by-task intercorrelations. In particular, we followed up two observations from our neural clustering. First, the cognitive and affective cluster showed largely distinct neural networks. Therefore, behavior linked to tasks from cognitive versus affective clusters may be uncorrelated (or weakly correlated). Second, the intermediate cluster showed neural overlap with both the cognitive and the affective cluster. We therefore expect that tasks linked to this cluster show more widespread behavioral intercorrelations. Specifically, studies should report intercorrelations both between tasks from the cognitive and intermediate cluster, as well as the affective and intermediate cluster.

With respect to the first prediction, behavioral studies are clearly in support of the independence between processes associated with the cognitive and affective clusters. Kanske, Böckler, Trautwein, Parianen-Lesemann, & Singer (2016) tested both belief reasoning (cognitive cluster) and evaluation of emotions (affective cluster) in a large sample of non-impaired adults using a combined task set-up (the EmpaToM task). The authors found that behavioral performance on these two measures was uncorrelated (as well as neural activation strength). Also when additional data driven composites of several ToM and empathy tasks

[13] We reviewed independent literature to find behavioral correlations. All studies we report here are different from the neuroimaging studies in our meta-analysis, with the exception of Kanske et al. (2015). Due to its large sample size and thus robust results, we also discuss the behavioral data of this neuroimaging study.

were used, no association was found. Similarly, no correlation was found between performance on Reading Mind in the Eyes (affective) and Strange Stories (cognitive) tasks in 7-12 year old children (Rice et al., 2016) and adults (Dziobek et al., 2006). In addition, double-dissociations in cognitive versus affective task performance have been found for patients with lesions in different brain areas (e.g. Shamay-Tsoory & Aharon-Peretz, 2007; Shamay-Tsoory et al., 2009).

With respect to our second, and maybe less obvious prediction, a number of studies report relevant results. With respect to correlations between tasks from the affective and intermediate clusters, Lockwood et al. (2013) found a positive correlation between accuracy for ratings of social animations (in our intermediate cluster) and judgments of one's own affective reaction to emotional faces (affective cluster). Interestingly, correlations were also found between biological motion perception (affective/intermediate cluster) and the Reading Mind in the Eyes task (affective) both in adults (Miller & Saygin, 2013) and 7-12 year old children (Rice et al., 2016)[14]. Biological motion perception can be linked to the affective cluster due to its relation to action perception (and thus, perception-action cycles for affect sharing, see e.g. de Waal & Preston, 2017; Keysers, Kaas, & Gazzola, 2010). However, stimuli are also related to a task falling in the intermediate cluster (Social Animations), and as we illustrate in Figure 9, brain activation for biological motion overlaps with that for our intermediate cluster.

Another behavioral task that we tentatively link to the intermediate cluster is the faux pas test (e.g. Baron-Cohen, O'Riordan, Jones, Stone, & Plaisted, 1999a). Understanding a faux pas

[14] Also note that our assignment of the Reading Mind in the Eyes task to the affective cluster is reflected by behavioral findings. For example, Olderbak et al. (2015) reported a positive correlation between the Reading Mind in the Eyes task and a face-based emotion categorization task.

(e.g. observing someone who is hurting another person's feelings out of ignorance) requires not only processing of belief-like states (e.g. someone's ignorance), but also their affective consequences (e.g. subsequent regret/embarrassment due to accidentally hurting someone's feelings). Correspondingly, Ferguson & Austin (2010) found an association in performance on the Faux Pas test (intermediate) and the Reading Mind in the Eyes task (affective) in a large sample of non-impaired adults.

With respect to correlations between tasks from cognitive and intermediate clusters, results are more mixed. In a sample of non-impaired adults, Brewer et al. (2017) found no correlation between behavioral performance on social animations (intermediate cluster) and the strange stories task (cognitive cluster, since strange stories present various belief-related contents, see e.g. Happe, 1994; White, Hill, Happe, & Frith, 2009). The authors noted, however, that this could be linked to range restrictions in scores (as non-impaired adults often show ceiling effects on ToM tests linked to our cognitive cluster). Interestingly, when re-running the correlation analysis for a pooled sample of non-impaired adults and adults with autism spectrum disorder, Brewer et al. (2017) did find an association between performance on social animations and strange stories. Similarly, for 7-12 year old children (Rice et al., 2016), a correlation between behavioral performance for biological motion perception (affective/intermediate) and the strange stories (cognitive) was found.

3.8. Positioning the three clusters along a principal gradient of macroscale cortical organization

To further characterize functional relations between the intermediate and the cognitive and affective clusters, we projected our clusters' activation maps along a principal gradient of macroscale cortical organization, which describes a functional spectrum along the cortical surface based on patterns of functional connectivity (Margulies et al., 2016). The gradient characterizes continuous changes in patterns of functional connectivity, which reflect changes in functional roles: increased distance (in terms of connectivity) from primary sensory and motor areas reflects increasingly abstract and multimodal processing (Margulies et al., 2016). As shown in Figure 8A, unimodal sensory and motor representations lie at one end, while abstract multimodal representations lie at the other end of the connectivity gradient. Grounding this map also in brain structure, Margulies et al. (2016) showed that each region's position on the connectivity gradient strongly predicts the area's spatial (i.e. geodesic) distance along the cortical surface to higher-order association areas (at the top of the gradient).

We reasoned that if the intermediate cluster represents the parallel involvement of cognitive and affective processes, its position along the gradient should coextend with the positions of both other clusters. Conversely, if the intermediate cluster represents a unique and largely independent functional process, its position along the gradient should be distinct. As shown in Figure 8B, the intermediate cluster tended to overlap with the locations of both the cognitive and affective clusters, rather than any other positions. The cognitive cluster was located close to the transmodal-end of the gradient (85-95th percentiles), supporting its functional interpretation in terms of stimulus independent thought. The affective cluster was located towards the middle of the of the gradient (35th-55th percentiles). Note however that the very low-level aspects the tasks in the affective cluster, such as visual characteristics, may have been removed with the respective control conditions (at least partially).

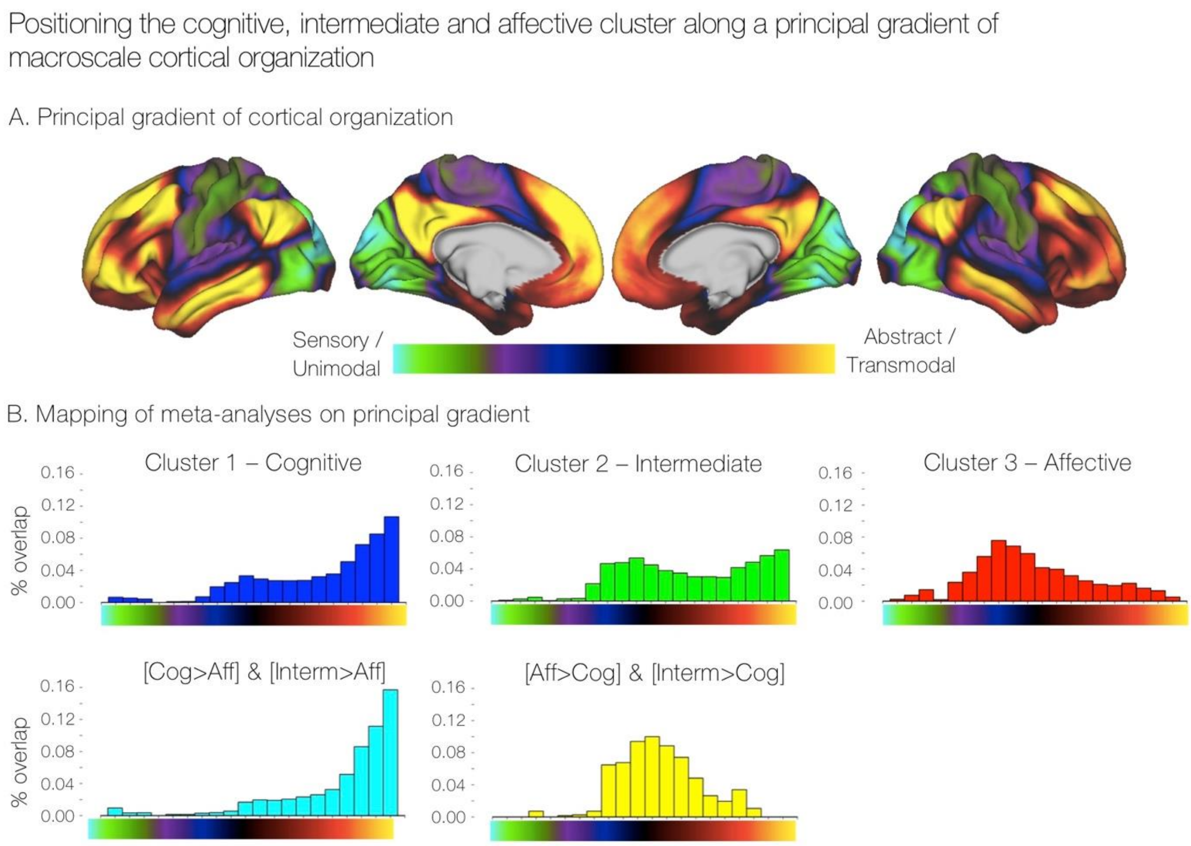


Figure 8. (A) Whole brain map of connectivity organization, capturing continuous changes (i.e. a principal gradient) of brain connectivity patterns across the cortex. Gradient ranges from sensory and unimodal areas in sensorimotor, auditory and visual cortex (green-blue), to increasingly abstract and transmodal areas (red-yellow) which largely correspond to the brain's default mode network. This gradient was determined by analysis of resting state connectivity data from a large sample of healthy adults (Human Connectome Project, n=820). (B) Percentiles along the gradient were split into 20 distinct maps. For each map, we calculated the overlap with our meta-analyses. Bar-charts give percent overlap between percentile maps and meta-analyses, i.e. how many percent of voxels of a meta-analysis map overlapped with each percentile map (normalizing results across differently sized meta-analysis maps).

These findings support our interpretation that the intermediate cluster combines cognitive and affective processes, which is notable for several reasons. During passive rest, spontaneous fluctuations of brain activity in corresponding areas and networks (e.g. affective: ventral attention network, cognitive: default mode network) has been found to be unrelated (e.g. Alcalá-Lopez et al., 2018) and sometimes even anti-correlated (Bzdok et al., 2013; Chai, Castañón, Öngür, & Whitfield-Gabrieli, 2012; Fox et al., 2005; Zhou et al., 2017). During tasks requiring externally focused attention, an inhibitory relation was found between the ventral attention network and the default mode network, considered to reflect the former down-regulating the latter to reduce interference from task-unrelated processes (Goulden et al., 2014; Wen, Liu, Yao, & Ding, 2013; Trautwein, Singer, & Kanske, 2016, see also Anticevic et al., 2012). The same inhibitory relationship has also been observed in a social setting (Kanske et al., 2016), where participants were independently asked to empathize with another person or to judge her beliefs. Note that contrary to the tasks in the intermediate cluster, Kanske et al. (2016) did not link cognitive and affective processing. While these findings suggest that areas linked to the cognitive and affective clusters are functionally segregated in many task contexts, our meta-analysis identified a cluster of tasks (intermediate) where cognitive and affective processes operate conjointly.

3.9. Low-level clustering

As we take a multilevel perspective on our results, we also consider clusterings beyond our central, high-level three cluster solution. We will focus on the relation between high- and low-level clusterings again in our Conclusion section (5.1). Across the entire spectrum of solutions, our metrics also pointed at two more low-level solutions, having 8 and 11 clusters. This suggests that brain activation contains additional variability that goes beyond what is captured at our central high level (three clusters). In particular the 11 cluster solution showed good performance, implying that task-to-task variability is large enough that each gets assigned to a cluster of its own by the algorithm. One possible explanation for this additional heterogeneity are differences in stimuli and task instructions, which might be unrelated to central processes linked to empathy and ToM. We illustrate the role of stimuli and task-instructions in our additional color-coding of Figure 5. For the sake of brevity, we only map out two popular distinctions: Verbal versus non-verbal stimuli, and instructed versus uninstructed tasks[15]. Note that several other (and partially related) categorizations have been

[15] We define "verbal" stimuli as items of written or spoken language, which contain task-relevant information and go beyond trivial task-cues (e.g. re-occurring instruction cues or response category reminders). The category "instructed" tasks contains all tasks that explicitly instruct participants to infer, judge or think about the mental states (cognitive or affective) of others.

discussed elsewhere, but go beyond the scope of our exemplary illustration (see e.g. Fan et al., 2011; Mar, 2010; Molenberghs et al., 2016; Timmers et al., 2018; Van Overwalle, 2009). Consistent with the idea of finding more abstract classes of functioning at higher levels, none of the clusters at model order three contains tasks from only one category (e.g. only verbal tasks)[16]. At lower levels, part of sub-clusterings are accompanied by separations in terms of task or stimulus-formats. For example, the three task groups False Belief, Trait Judgments, and Strategic Games cluster together at a higher-level (3 cluster solution), but Strategic Games are separated at a lower-level (8 cluster solution). These tasks differ in that the two former are verbal (sentences or words), whereas the latter is non-verbal (strategic decision making). While this pattern demonstrates that stimulus-format is implicated in parts of lower-level cluster separations, other observations demonstrate that this does not provide a perfect explanation. For example, as also illustrated in Figure 5, the Mind in the Eyes task falls into the same categories as Trait Judgment tasks: Verbal and instructed. In addition, both tasks are drawn from the Theory of Mind literature. Nevertheless, the two tasks end up at opposite positions in the clustering (see Figure 4B), and their activation maps show relatively low similarity (see Figure 4A). Therefore, we conclude that while stimulus- and task format drives part of the variability in meta-analytic activations, it does not provide a complete account of them. As we will argue in detail in our Conclusion section (5.1), this pattern supports a multilevel model for social cognitive processes, similar to what has been suggested in other fields such as intelligence and personality research. The three cluster solution (higher level) may reflect more abstract and broad classes of functioning. The 8 and 11 cluster solutions (lower level) capture additional variability, which possibly describes how central

[16] Figure 5 shows that at model order 3, the cognitive cluster contains only one non-verbal task, which may give rise to the impression that it features more verbal tasks than the other two. Note however, that our jack-knife clustering analysis found that the non-verbal task group Rational Actions frequently changes membership between the cognitive and the intermediate cluster, which weakens the coherence of a 'predominantly verbal' interpretation of the cognitive cluster.

processes are applied in concrete contexts of particular tasks. For such "task contexts", both the stimulus-format and the task-instructions play important roles.

4. Relation to other meta-analyses on social cognition

The present meta-analysis and review focused on processes involved in inferring other's unobservable mental states, and thus on the literature on empathy and Theory of Mind. For the sake of coherence and practical limitations, we did not carry out meta-analyses for other topics in social cognition. For example, we did not cover related processes, such as action observation ('mirroring'). Furthermore, we did not include tasks which involved mental state inference as a sub-component, employed in addition to diverse other processes. This was the case, for example, for the topics of moral cognition, social exclusion, or social decision making. However, we illustrate recently published meta-analyses on these topics below, and characterize the overlap with our meta-analytic clusters.

Figure 9 shows overlaps between thresholded maps from our three cluster solution and maps from other meta-analyses on social topics. For a meta-analysis on biological motion perception (whole body movement, e.g. waking, dancing), we found highest overlap with the intermediate cluster. Moreover, the map also showed overlaps to both the cognitive and affective clusters. However, note that the intermediate cluster spatially overlaps with both other clusters (see top row in Figure 9). For a meta-analysis on action observation or 'mirroring' (Hardwick et al., 2018), we illustrate overlaps specifically for meta-analysis on (i) all types of action observation, (ii) observation of actions performed with arms, and (iii) observation of actions performed with faces. For all three maps, we found largest overlap with the affective cluster, and to lesser extent also with the intermediate cluster. For a meta-analysis on inhibition of imitation (Darda & Ramsey, 2019; see also Brass, Ruby, &

Spengler, 2009; Hogeveen et al., 2014), we again found preferential overlaps for the affective and the intermediate cluster.

Next, we overlaid our three cluster solution to maps from related meta-analyses on more 'affective' topics. For a meta-analysis on Emotion Matching (Dricu & Frühholz, 2020), that is, matching expressions of a target face and a number of other faces, overlap was clearly highest for our affective cluster. For an additional meta-analysis on Emotion Labeling (Dricu & Frühholz, 2020), we found considerable overlaps with both the affective and the intermediate cluster. In this meta-analysis, tasks required participants to match facial, vocal or bodily expressed emotions with one label out of several alternatives. For a meta-analysis on vicarious reward processing (Morelli, Sacchet, & Zaki, 2015; see also Apps, Rushworth, & Chang, 2016; Lockwood, Apps, Roiser, & Viding, 2015) we found limited overlaps of equal size for the cognitive and affective clusters. In this group of tasks, participants would, for example, witness how another person wins money, gets praised by another person, or receives pleasant touch.

We further included a meta-analysis on Social Exclusion tasks (Vijayakumar et al., 2017) in our overlap analysis. Here, participants experienced being rejected mostly in the context of a virtual ball-game ('Cyberball'), and some additional social judgment / chatroom contexts. Overlaps for this meta-analysis were mainly found with our cognitive cluster, and to lesser extent also with our affective cluster. For a multi-study analysis (n=68) on Spontaneous Theory of Mind (Boccadoro et al., 2019), we found equal amounts of overlap with our three clusters, and thus no preference for the affective, intermediate or cognitive clusters became clear. Tasks in this multi-study analysis presented an 'uninstructed' false belief task, in which participants watched videos of an agent witnessing some but not all events happening in a room, and thus develops a false belief about where a certain object has been placed (e.g., inside versus outside a box). Somewhat similarly, for a meta-analysis on visual perspective taking (Schurz, Aichhorn, Martin, & Perner, 2013) we found no prominent overlap to any one of our three clusters (but see Dumontheil, Küster, Apperly, & Blakemore, 2010; Santiesteban, Banissy, Catmur, & Bird, 2012). Visual perspective taking tasks asked participants to judge what another person can see, or how another person sees an object. For a summary map of studies on Computational Modeling of ToM (Cheong et al., 2017), that is, studies using game theoretic approaches to model how individuals engage strategic reasoning in competitive and cooperative contexts, we observed a slight preference in overlap with our cognitive cluster, and further overlap with the intermediate cluster. Finally, for a meta-analysis on moral cognition (Bzdok et al., 2012; see also Eres, Louis, & Molenberghs, 2018) large overlaps were found for the cognitive and the intermediate clusters, and comparatively little for the affective cluster. Moral cognition tasks featured scenarios with moral violations or dilemmas, and required participants to make appropriateness judgments on actions of one individual towards others.

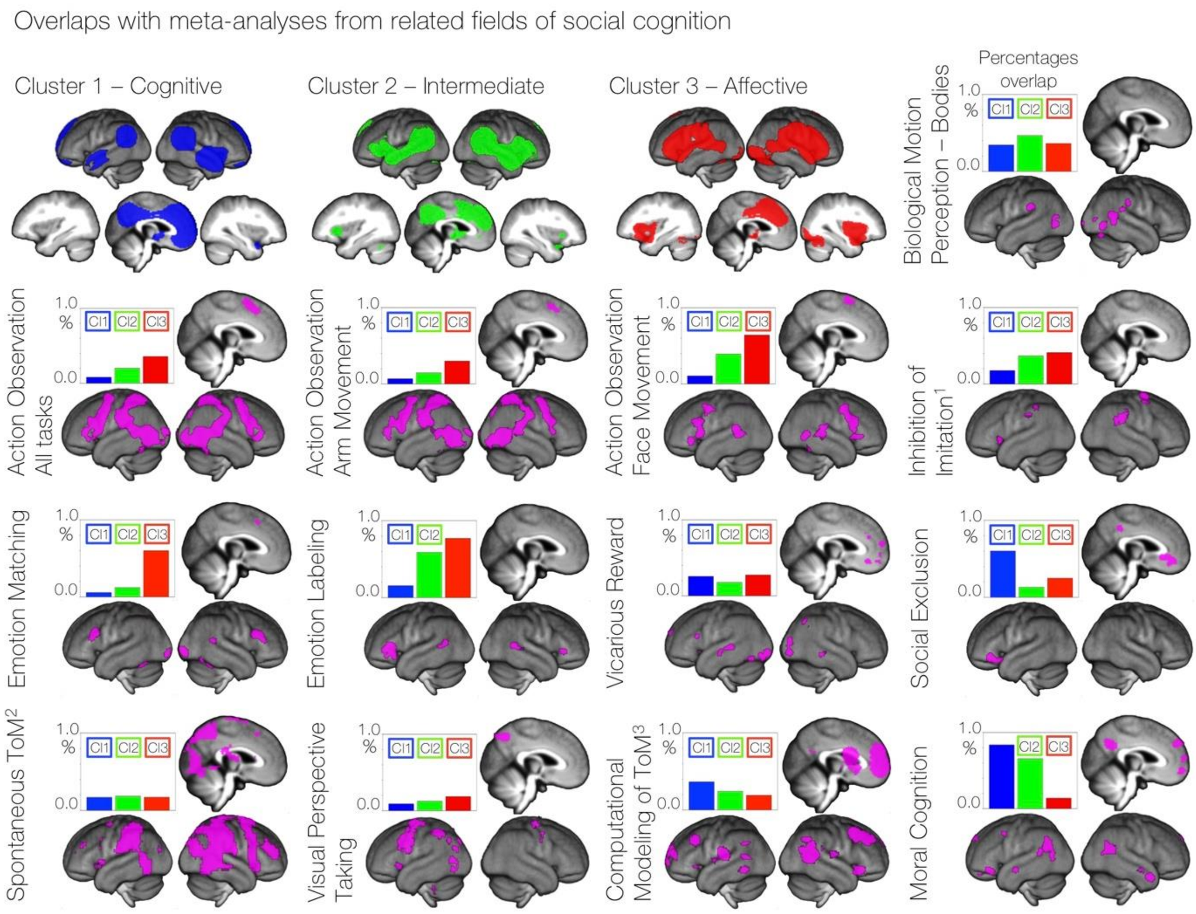


Figure 9. Overlaps between our three cluster solution and meta-analyses on related topics of social cognition. For each topic map, we indicate the percentage of voxels falling into each of the three meta-analytic clusters. Note that percentage values are relative to the size of each topic-related map (e.g., maps containing extensive areas of activation are more likely to feature regions outside of our three clusters). Therefore, the relative patterns of overlaps within each topic-map are of main interest. See section "4. Relation to other meta-analyses on social cognition" for more details. Footnotes: [1] For exploratory purposes we show an uncorrected map of the meta-analysis results by Darda & Ramsey (2018). [2] Map from Boccadoro et al. (2019) shows results from a multi-study analysis on spontaneous ToM (not a meta-analysis, based on original data), 3 studies, n=68. [3] Map from Cheong et al. (2017) shows most consistent overlaps across 5 published studies, summarized based on expanding reported coordinates into 15 mm radius spheres.

5. Summary and Conclusion

5.1. A hierarchical perspective

Over two decades of neuroimaging and behavioral research produced an impressive amount of data and a rich variety of theories and perspectives on the neurocognitive processes underlying the human ability to understand other minds. However, there has been increasing awareness of disagreement concerning the concepts and taxonomy underlying social processes (Happe et al., 2017; Schaafsma et al., 2015; Spunt & Adolphs, 2017). This meta-analysis was aimed at supporting the development of a coherent and balanced theoretical model of major cognitive factors underlying the ability to understand other minds. We have argued that neuroimaging data provides a good starting point for such a model, as other sources - such as behavioral data - do not (yet) provide a complete picture of task-by-task interrelations. As we have reviewed in section 3.7, behavioral intercorrelations were only studied for a limited number of tasks, among which some suffer from range restrictions in scores (e.g. non-impaired adults showing ceiling effects in some traditional ToM tasks).

The present meta-analysis used hierarchical clustering to sort and group neural patterns elicited across a range of empathy and ToM tasks. Such a hierarchical approach shares characteristics with previous work conceptualizing social cognition as a multi-layered or multi-level phenomenon (De Waal, 2012; Preston & De Waal, 2002; Singer, 2006; Schaafsma et al., 2015). An advantage of the clustering method applied here is that it provides a data-driven answer to the question of how many factors or latent variables are sufficient and appropriate for modeling social cognition. Results suggest the answer is twofold: On the one hand, the best overall clustering performance was reached when dividing the data into a three cluster solution. However, further local peaks in performance were found when splitting the data into eight and eleven clusters. The higher-level, three cluster solution provides a solid foundation of evidence for the assumption that empathy and ToM share certain processes, and therefore brain activity, across different tasks and stimuli. However, the concurrent existence of an additional, lower level of clustering (essentially by task) highlights the question of the appropriate level of concreteness and detail in neurocognitive accounts of social cognition. Will it be possible to formulate a highly specific mechanism that can be applied to all tasks and contexts? We rather suggest modelling social cognition by multilevel theories, akin to models in other fields, as for example intelligence research. There, some accounts feature a central construct which reflects a latent variable that indexes a broader, more abstract neurocognitive function (e.g. dynamic control via a multiple-demand network; Duncan et al., 2013). Turning to individual tasks specifies how this function manifests in a concrete context and for a particular problem. Task-contexts can have different degrees of relevance for ecological social cognition. While some tasks may contain spurious processes related to idiosyncratic elements of a paradigm (see e.g. Mar, 2011 for discussion), other tasks may resemble to some extent a real-life problem (such as for example, trying to guess a person's mental state based on his or her facial expression). Modeling our data as a multilevel construct accommodates for the diversity in how social abilities manifest, as well as for the specific demands placed by its particular instances (e.g. in experimental tasks). To illustrate, a number of accounts have described mechanisms targeted at explaining false belief understanding, which is assumed to be a hallmark of human social cognition (e.g. Wimmer & Perner, 1983; Premack & Woodruff, 1978; Saxe & Kanwisher, 2003). It turns out to be difficult to predict whether and how these mechanisms are recruited by other social cognition tasks. For example, how would a mechanism for decoupling beliefs from knowledge about reality (e.g. Frith & Frith, 2003) work in the context of a Trait Judgment task? This question is relevant, as meta-analytic maps for False Belief and Trait Judgment tasks showed high similarity (see e.g. Figure 4A). Our approach addresses this issue by placing the most situation-specific mechanisms, such as belief decoupling, at the lower level of our hierarchical model. Decoupling beliefs represents a specific implementation of the broader function of self-generated and decoupled thought in social cognition (cognitive cluster). The multilevel perspective accommodates the fact that (i) a mechanism for decoupling false beliefs is not perfectly applicable to other task contexts but (ii) a broader functional class of self-generated thought in social cognition forms the basis for belief decoupling (but does not completely specify it).

5.2. Separated versus combined social processes

The central novel finding of our three cluster solution is the intermediate cluster. Our results show that rather than being a distinct (sub-)form of processing, this cluster combines processes from the cognitive and affective clusters. While previous labels such as 'affective ToM' and 'cognitive empathy' claim this process for either the domain of ToM or empathy, we suggest that referring to this function as 'conjoint ToM and empathy' provides a more unbiased positioning in the terminological landscape. Perhaps more important than its terminology, however, is the functional relevance of this cluster. Co-occurrence of cognitive and affective processes has been linked to more naturalistic forms of social cognition. In an elegant study, Zaki et al. (2009) demonstrated the relevance of combining cognitive and affective processes for understanding others. In an fMRI experiment, participants viewed videos of persons talking about emotional autobiographical events. After filming, the person shown in the video had been asked to rate his/her own emotional states during the clip, on a moment-to-moment basis. In addition, participants observing in the scanner were asked to rate what the person in the video likely felt at each point in time. By comparing the person's own with participants' ratings, an index of empathic accuracy could be generated. At time points where empathic accuracy of the participants was high, activation levels were increased both in areas linked to cognitive (e.g. mPFC, superior temporal sulcus) and affective (e.g. inferior parietal lobule, premotor cortex) processes. The task by Zaki et al. (2009) is part of the task group Evaluating Situated Emotions, which ended up in the intermediate cluster of our analysis.

Another central case of co-activation of cognitive and affective processes are everyday social interactions, which are quintessential for understanding other minds (see Schilbach et al., 2013). For example, Schilbach, Eickhoff, Mojzisch, & Vogeley (2008) recorded facial mimicry-related brain activity during online social interaction and found conjoint activation in cortical midline (e.g. precuneus) and motor areas (e.g. precentral gyrus). This pattern of co-activated areas resembles our intermediate cluster, i.e. the combination of cognitive and affective processes. Also other studies presenting scenes or videos of social interactions reported co-activation of cognitive and affective processes (e.g. Deuse et al., 2016; Wolf, Dziobek, & Heekeren, 2010). Note, however, that both naturalistic social cognition (see Zaki & Ochsner, 2009) and social interaction (Schilbach et al., 2013) were argued to engage additional processes not covered by the cognitive and affective clusters (for example, reward-related areas during social interactions). Another group of tasks where co-activation of cognitive and affective clusters was observed concerns altruistic decisions (Hare, Camerer, Knoepfle, O'Doherty, & Rangel, 2010; Tusche, Böckler, Kanske, Trautwein, & Singer, 2016; see also Hein, Silani, Preuschoff, Batson, & Singer, 2010). Tusche et al. (2016) found that the level of generosity in donations for charitable organizations was predicted by brain activity patterns in both right TPJ and right anterior insula during that task, again pointing out co-occurrence of cognitive and affective processes.

Taken together, the reviewed studies suggest that co-activation of cognitive and affective processes might have particular relevance for ecologically valid social cognition. Our meta-analytic clustering demonstrates that a portion of the tasks typically assumed to measure either ToM or empathy (the intermediate cluster) engage cognitive and affective processes concurrently. This has implications for selecting tasks when measuring different aspects of social cognition. Furthermore, our review of behavioral studies showed that in terms of performance, tasks linked to the intermediate cluster (e.g. social animations) are correlated to both cognitive cluster (e.g. strange stories) and affective cluster (e.g. facial emotion recognition) tasks. In addition, tasks linked to the intermediate cluster often show good clinical discrimination, such as the faux pas test (e.g. for autism spectrum: Baron-Cohen et al., 1999a; frontotemporal dementia: Gregory et al., 2002; schizophrenia: Konstantakopoulos et al., 2014) or social animations (e.g. for autism spectrum: White, Coniston, Rogers, & Frith, 2011; schizophrenia: Bliksted et al., 2018). While these findings turn the spotlight on task from the intermediate cluster with respect to ecological and clinical relevance, they also trigger a novel question regarding their interpretation. Since such tasks contain both cognitive and affective processes, inter-individual differences found for them could reflect different sources - a difference in cognitive processes only, affective processes only, or both of them (combined). Therefore, we suggest that studies selecting tasks according to the three cluster solution identified in this meta-analysis, and ideally presenting a combination of them (cognitive, affective and intermediate), will increase the precision in measuring discriminable as well as conjoint social processes.

5.3. Outlook

The capacity to understand what other people think and feel remains one of the most elusive mental faculties. However, the extensive research of the last decades that we have brought together here, meta-analytically, has enabled a major step forward. We propose to model social cognition as a hierarchical multilevel construct, thereby seeking to preserve and accommodate the diversity of proposed processes and features of social cognition (on a lower level), while at the same time making explicit a level of coherence among them (i.e. the higher level). Based on our research focus and the evidence that is available to date, this meta-analysis produced a first sample of processes, levels, and mechanisms that such a model might have. At the highest level of our model, we suggest that social cognitive processes can be captured in terms of two overarching networks which are flexibly combined, and relate to more sensory-affective versus more abstract and decoupled representations of others' states.

Acknowledgements

To ensure masked review, we will add acknowledgements at a later point. It will include the following statement: "We are grateful for the very helpful comments of four reviewers and the editor on a previous version of this manuscript."

6. References

Adolphs, R. (2009). The social brain: neural basis of social knowledge. *Annual Review of Psychology, 60*, 693-716. https://doi.org/10.1146/annurev.psych.60.110707.163514

Aichhorn, M., Perner, J., Weiss, B., Kronbichler, M., Staffen, W., & Ladurner, G. 2009. Temporo-parietal junction activity in theory-of-mind tasks: falseness, beliefs, or attention. *Journal of Cognitive Neuroscience, 21*(6), 1179-1192. https://doi.org/10.1162/jocn.2009.21082.

Alcalá-López, D., Smallwood, J., Jefferies, E., Van Overwalle, F., Vogeley, K., Mars, R. B., ... & Bzdok, D. (2018). Computing the social brain connectome across systems and states. *Cerebral Cortex, 28*(7), 2207-2232. https://doi.org/10.1093/cercor/bhx121

Alcalá-López, D., Vogeley, K., Binkofski, F., & Bzdok, D. (2019). Building blocks of social cognition: Mirror, mentalize, share? *Cortex, 118*, 4-18. https://doi.org/10.1016/j.cortex.2018.05.006

Andersson, F., Glaser, B., Spiridon, M., Debbané, M., Vuilleumier, P., & Eliez, S. (2008). Impaired activation of face processing networks revealed by functional magnetic resonance imaging in 22q11.2 deletion syndrome. *Biological Psychiatry, 63*(1), 49-57. https://doi.org/10.1016/j.biopsych.2007.02.022

Andrews- Hanna, J. R., Smallwood, J., & Spreng, R. N. (2014). The default network and self- generated thought: component processes, dynamic control, and clinical relevance. *Annals of the New York Academy of Sciences, 1316*(1), 29-52. https://doi.org/10.1111/nyas.12360

Anticevic, A., Cole, M. W., Murray, J. D., Corlett, P. R., Wang, X. J., & Krystal, J. H. (2012). The role of default network deactivation in cognition and disease. *Trends in Cognitive Sciences, 16*(12), 584-592. https://doi.org/10.1016/j.tics.2012.10.008

Apps, M. A., Rushworth, M. F., & Chang, S. W. (2016). The anterior cingulate gyrus and social cognition: tracking the motivation of others. *Neuron, 90*(4), 692-707. https://doi.org/10.1016/j.neuron.2016.04.018

Baron-Cohen, S. (2002). The extreme male brain theory of autism. *Trends in Cognitive Sciences, 6*(6), 248-254.https://doi.org/10.1016/S1364-6613(02)01904-6

Baron-Cohen, S., & Hammer, J. (1997). Parents of children with Asperger syndrome: what is the cognitive phenotype? *Journal of Cognitive Neuroscience, 9*(4), 548-554. https://doi.org/10.1162/jocn.1997.9.4.548

Baron- Cohen, S., Jolliffe, T., Mortimore, C., & Robertson, M. (1997). Another advanced test of theory of mind: Evidence from very high functioning adults with autism or Asperger syndrome. *Journal of Child Psychology and Psychiatry, 38*(7), 813-822. https://doi.org/10.1111/j.1469-7610.1997.tb01599.x

Baron-Cohen, S., O'Riordan, M., Jones, R., Stone, V., & Plaisted, K. (1999a). A new test of social sensitivity: Detection of faux pas in normal children and children with Asperger syndrome. *Journal of Autism and Developmental Disorders, 29*(5), 407-418. https://doi.org/10.1023/A:1023035012436

Baron-Cohen, S., Ring, H., Moriarty, J., Schmitz, B., Costa, D., & Ell, P. (1994). Recognition of mental state terms: Clinical findings in children with autism and a functional neuroimaging study of normal adults. *The British Journal of Psychiatry, 165*(5), 640-649. https://doi.org/10.1192/bjp.165.5.640

Baron-Cohen, S., Ring, H.A., Wheelwright, S., Bullmore, E.T., Brammer, M.J., Simmons, A., & Williams, S.C. (1999b). Social intelligence in the normal and autistic brain: An fMRI study. *European Journal of Neuroscience, 11*(6), 1891-1898. https://doi.org/10.1046/j.1460-9568.1999.00621.x

Barrett, L. F., Lindquist, K. A., & Gendron, M. (2007). Language as context for the perception of emotion. *Trends in Cognitive Sciences, 11*(8), 327-332. https://doi.org/10.1016/j.tics.2007.06.003

Bliksted, V., Frith, C., Videbech, P., Fagerlund, B., Emborg, C., Simonsen, A., ... & Campbell-Meiklejohn, D. (2018). Hyper-and Hypomentalizing in Patients with First-Episode Schizophrenia: fMRI and Behavioral Studies. *Schizophrenia Bulletin, 2,* 377-385. https://doi.org/10.1093/schbul/sby027

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex, 19*(12), 2767-2796. https://doi.org/10.1093/cercor/bhp055

Blakemore, S.-J., Boyer, P., Pachot-Clouard, M., Meltzoff, A., Segebarth, C., & Decety, J. (2003). The Detection of Contingency and Animacy from Simple Animations in the Human Brain. *Cerebral Cortex, 13*(8), 837-844. https://doi.org/10.1093/cercor/13.8.837

Bloom, P. (2017). Empathy, schmempathy: response to Zaki. Trends in cognitive sciences, 21(2), 60-61. https://doi.org/10.1016/j.tics.2016.12.003

Boccadoro, S., Cracco, E., Hudson, A. R., Bardi, L., Nijhof, A. D., Wiersema, J. R., ... & Mueller, S. C. (2019). Defining the neural correlates of spontaneous theory of mind

(ToM): An fMRI multi-study investigation. *NeuroImage, 203*, 116193. https://doi.org/10.1016/j.neuroimage.2019.116193

Bos, P. A., Hofman, D., Hermans, E. J., Montoya, E. R., Baron-Cohen, S., & van Honk, J. (2016). Testosterone reduces functional connectivity during the 'Reading the Mind in the Eyes' Test. *Psychoneuroendocrinology, 68*, 194-201. https://doi.org/10.1016/j.psyneuen.2016.03.006.

Brass, M., Ruby, P., & Spengler, S. (2009). Inhibition of imitative behaviour and social cognition. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364*(1528), 2359-2367. https://doi.org/10.1098/rstb.2009.0066

Brewer, N., Young, R. L., & Barnett, E. (2017). Measuring theory of mind in adults with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 47*(7), 1927-1941. https://doi.org/10.1007/s10803-017-3080-x

Brooks, J. A., Shablack, H., Gendron, M., Satpute, A. B., Parrish, M. H., & Lindquist, K. A. (2017). The role of language in the experience and perception of emotion: a neuroimaging meta-analysis. *Social Cognitive and Affective Neuroscience, 12*(2), 169-183. https://doi.org/10.1093/scan/nsw121

Brunet, E., Sarfati, Y., & Hardy-Baylé, M. C. (2003). Reasoning about physical causality and other's intentions in schizophrenia. *Cognitive Neuropsychiatry, 8*(2), 129-139. https://doi.org/10.1080/13546800244000256

Brunet, E., Sarfati, Y., Hardy-Baylé, M. C., & Decety, J. (2000). A PET investigation of the attribution of intentions with a nonverbal task. *NeuroImage*, *11*(2), 157-166. https://doi.org/10.1006/nimg.1999.0525

Buckner, R. L., & Carroll, D. C. (2007). Self-projection and the brain. *Trends in Cognitive Sciences, 11*(2), 49-57. https://doi.org/10.1016/j.tics.2006.11.004

Bzdok, D., Schilbach, L., Vogeley, K., Schneider, K., Laird, A. R., Langner, R., & Eickhoff, S. B. (2012). Parsing the neural correlates of moral cognition: ALE meta-analysis on morality, theory of mind, and empathy. *Brain Structure and Function, 217*(4), 783-796. https://doi.org/10.1007/s00429-012-0380-y

Bzdok, D., Langner, R., Schilbach, L., Jakobs, O., Roski, C., Caspers, S., ... & Eickhoff, S. B. (2013). Characterization of the temporo-parietal junction by combining data-driven parcellation, complementary connectivity analyses, and functional decoding. *NeuroImage, 81,* 381-392.

Carr, L., Iacoboni, M., Dubeau, M. C., Mazziotta, J. C., & Lenzi, G. L. (2003). Neural mechanisms of empathy in humans: a relay from neural systems for imitation to limbic areas. *Proceedings of the National Academy of Sciences, 100*(9), 5497-5502. https://doi.org/10.1073/pnas.0935845100

Carrillo, M., Han, Y., Migliorati, F., Liu, M., Gazzola, V., & Keysers, C. (2019). Emotional Mirror Neurons in the Rat's Anterior Cingulate Cortex. *Current Biology, 29*(8), 1301-1312. https://doi.org/10.1016/j.cub.2019.03.024

Castelli, F., Happé, F., Frith, U., & Frith, C. (2000). Movement and mind: a functional imaging study of perception and interpretation of complex intentional movement patterns. *NeuroImage, 12*(3), 314-325. https://doi.org/10.1006/nimg.2000.0612

Castelli, F., Frith, C., Happé, F., & Frith, U. (2002). Autism, Asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes. *Brain, 125*(8), 1839-1849. https://doi.org/10.1093/brain/awf189

Chai, X. J., Castañón, A. N., Öngür, D., & Whitfield-Gabrieli, S. (2012). Anticorrelations in resting state networks without global signal regression. *NeuroImage, 59*(2), 1420-1428. https://doi.org/10.1016/j.neuroimage.2011.08.048

Cheong, J. H., Jolly, E., Sul, S., & Chang, L. J. (2017). Computational models in social neuroscience. In Moustafa, A. A. (ed.), *Computational Models of Brain and Behavior*, John Wiley & Sons, Ltd., pp. 229–44. https://doi.org/10.1002/9781119159193

Conway, J. R., Catmur, C., & Bird, G. (2019). Understanding individual differences in theory of mind via representation of minds, not mental states. *Psychonomic Bulletin & Review, 26*, 798–812. https://doi.org/10.3758/s13423-018-1559-x

Corbetta, M., Patel, G., & Shulman, G. L. (2008). The reorienting system of the human brain: from environment to theory of mind. *Neuron, 58*(3), 306-324. https://doi.org/10.1016/j.neuron.2008.04.017

Corradi-Dell'Acqua, C., Tusche, A., Vuilleumier, P., & Singer, T. (2016). Cross-modal representations of first-hand and vicarious pain, disgust and fairness in insular and cingulate cortex. *Nature Communications, 7*, 10904. https://doi.org/10.1038/ncomms10904

Cracco, E., Bardi, L., Desmet, C., Genschow, O., Rigoni, D., De Coster, L., ... & Brass, M. (2018). Automatic imitation: A meta-analysis. *Psychological Bulletin, 144*(5), 453. http://dx.doi.org/10.1037/bul0000143

Critchley, H., Daly, E., Phillips, M., Brammer, M., Bullmore, E., Williams, S., ... & Murphy, D. (2000). Explicit and implicit neural mechanisms for processing of social information from facial expressions: a functional magnetic resonance imaging study. *Human Brain Mapping, 9(*2), 93-105. https://doi.org/10.1002/(SICI)1097-0193(200002)9:2<93::AID-HBM4>3.0.CO;2-Z

Darda, K. M., & Ramsey, R. (2019). Darda, K. M., & Ramsey, R. (2019). The inhibition of automatic imitation: a meta-analysis and synthesis of fMRI studies. *NeuroImage, 197*, 320-329. doi: https://doi.org/10.1101/33493

Davis, M. H. (1994). *Empathy: A social psychological approach*. Westview Press. https://doi.org/10.4324/9780429493898

Decety, J., & Jackson, P. L. (2004). The functional architecture of human empathy. *Behavioral and Cognitive Neuroscience Reviews, 3*(2), 71-100. https://doi.org/10.1177/1534582304267187

de Greck, M., Shi, Z., Wang, G., Zuo, X., Yang, X., Wang, X., ... & Han, S. (2012a). Culture modulates brain activity during empathy with anger. *NeuroImage, 59*(3), 2871-2882. https://doi.org/10.1016/j.neuroimage.2011.09.052

de Greck, M., Wang, G., Yang, X., Wang, X., Northoff, G., & Han, S. (2012b). Neural substrates underlying intentional empathy. *Social Cognitive and Affective Neuroscience, 7*(2), 135-144. https://doi.org/10.1093/scan/nsq093

Dennett, D. C. (1971). Intentional systems. *The Journal of Philosophy, 68*(4), 87-106. https://doi.org/10.2307/2025382

Denny, B. T., Kober, H., Wager, T. D., & Ochsner, K. N. (2012). A meta-analysis of functional neuroimaging studies of self-and other judgments reveals a spatial gradient for mentalizing in medial prefrontal cortex. *Journal of Cognitive Neuroscience, 24*(8), 1742-1752. https://doi.org/10.1162/jocn_a_00233

Derntl, B., Finkelmeyer, A., Eickhoff, S., Kellermann, T., Falkenberg, D. I., Schneider, F., & Habel, U. (2010). Multidimensional assessment of empathic abilities: neural correlates and gender differences. *Psychoneuroendocrinology, 35*(1), 67-82. https://doi.org/10.1016/j.psyneuen.2009.10.006

Derntl, B., Finkelmeyer, A., Voss, B., Eickhoff, S. B., Kellermann, T., Schneider, F., & Habel, U. (2012). Neural correlates of the core facets of empathy in schizophrenia. *Schizophrenia Research, 136*(1-3), 70-81. https://doi.org/10.1016/j.schres.2011.12.018

DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials, 7*(3), 177-188. https://doi.org/10.1016/0197-2456(86)90046-2

Deuse, L., Rademacher, L. M., Winkler, L., Schultz, R. T., Gründer, G., & Lammertz, S. E. (2016). Neural correlates of naturalistic social cognition: brain-behavior relationships in healthy adults. *Social Cognitive and Affective Neuroscience, 11*(11), 1741-1751. https://doi.org/10.1093/scan/nsw094

de Waal, F. B. M. (2007). *The 'Russian doll' model of empathy and imitation*. In S. Bråten (Ed.), *Advances in consciousness research: Vol. 68. On being moved: From mirror neurons to empathy* (p. 49–69). John Benjamins Publishing Company. https://doi.org/10.1075/aicr.68.06waa

de Waal, F. B. (2012). The antiquity of empathy. *Science, 336*(6083), 874-876.

de Waal, F. B., & Preston, S. D. (2017). Mammalian empathy: behavioural manifestations and neural basis. *Nature Reviews Neuroscience*, *18*(8), 498. https://doi.org/10.1038/nrn.2017.72

Dodell-Feder, D., Koster-Hale, J., Bedny, M., & Saxe, R. (2011). fMRI item analysis in a theory of mind task. *NeuroImage, 55*(2), 705-712. https://doi.org/10.1016/j.neuroimage.2010.12.040

Dricu, M., & Frühholz, S. (2020). A neurocognitive model of perceptual decision-making on emotional signals. *Human Brain Mapping, 41*(6), 1532-1556. https://doi.org/10.1002/hbm.24893

Dumontheil, I., Küster, O., Apperly, I. A., & Blakemore, S. J. (2010). Taking perspective into account in a communicative task. *NeuroImage, 52*(4), 1574-1583. https://doi.org/10.1016/j.neuroimage.2010.05.056

Duncan, J. (2013). The structure of cognition: attentional episodes in mind and brain. *Neuron, 80*(1), 35-50. https://doi.org/10.1016/j.neuron.2013.09.015

Dziobek, I., Fleck, S., Kalbe, E., Rogers, K., Hassenstab, J., Brand, M., ... & Convit, A. (2006). Introducing MASC: a movie for the assessment of social cognition. *Journal of Autism and Developmental Disorders*, *36*(5), 623-636. https://doi.org/10.1007/s10803-006-0107-0

Dziobek, I., Preißler, S., Grozdanovic, Z., Heuser, I., Heekeren, H. R., & Roepke, S. (2011). Neuronal correlates of altered empathy and social cognition in borderline personality disorder. *NeuroImage*, *57*(2), 539-548. https://doi.org/10.1016/j.neuroimage.2011.05.005

Eickhoff, S. B., Nichols, T. E., Laird, A. R., Hoffstaedter, F., Amunts, K., Fox, P. T., ... & Eickhoff, C. R. (2016). Behavior, sensitivity, and power of activation likelihood estimation characterized by massive empirical simulation. *NeuroImage*, *137*, 70-85.

Eres, R., Louis, W. R., & Molenberghs, P. (2018). Common and distinct neural networks involved in fMRI studies investigating morality: an ALE meta-analysis. *Social Neuroscience*, *13*(4), 384-398. https://doi.org/10.1080/17470919.2017.1357657

Fan, Y., Duncan, N. W., de Greck, M., & Northoff, G. (2011). Is there a core neural network in empathy? An fMRI based quantitative meta-analysis. *Neuroscience & Biobehavioral Reviews*, *35*(3), 903-911. https://doi.org/10.1016/j.neubiorev.2010.10.009

Ferguson, F. J., & Austin, E. J. (2010). Associations of trait and ability emotional intelligence with performance on Theory of Mind tasks in an adult sample. *Personality and Individual Differences*, *49*(5), 414-418. https://doi.org/10.1016/j.paid.2010.04.009

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., & Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(27), 9673-9678. https://doi.org/10.1073/pnas.0504136102

Fletcher, P. C., Happe, F., Frith, U., Baker, S. C., Dolan, R. J., Frackowiak, R. S., & Frith, C. D. (1995). Other minds in the brain: a functional imaging study of “theory of mind” in story comprehension. *Cognition, 57*(2), 109-128. https://doi.org/10.1016/0010-0277(95)00692-R

Friston, K. J., Price, C. J., Fletcher, P., Moore, C., Frackowiak, R. S. J., & Dolan, R. J. (1996). The trouble with cognitive subtraction. *NeuroImage, 4*(2), 97-104. https://doi.org/10.1006/nimg.1996.0033

Frith, U., & Frith, C. D. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society B: Biological Sciences, 358*(1431), 459-473. https://doi.org/10.1098/rstb.2002.1218

Frith, C. D., & Frith, U. (2006). The neural basis of mentalizing. *Neuron, 50*(4), 531-534. https://doi.org/10.1016/j.neuron.2006.05.001

Frith, C. D., & Frith, U. (2012). Mechanisms of social cognition. *Annual Review of Psychology, 63*, 287-313. https://doi.org/10.1146/annurev-psych-120710-100449

Fusar-Poli, P., Placentino, A., Carletti, F., Landi, P., Allen, P., Surguladze, S., ... & Perez, J. (2009). Functional atlas of emotional faces processing: a voxel-based meta-analysis of 105 functional magnetic resonance imaging studies. *Journal of Psychiatry & Neuroscience, 34*(6), 418–432. https://www.ncbi.nlm.nih.gov/pubmed/19949718

Gallagher, H. L., Happé, F., Brunswick, N., Fletcher, P. C., Frith, U., & Frith, C. D. (2000). Reading the mind in cartoons and stories: an fMRI study of 'theory of mind'in verbal and nonverbal tasks. *Neuropsychologia, 38*(1), 11-21. https://doi.org/10.1016/S0028-3932(99)00053-6

Gallagher, H. L., & Frith, C. D. (2003). Functional imaging of 'theory of mind'. *Trends in Cognitive Sciences, 7*(2), 77-83. https://doi.org/10.1016/S1364-6613(02)00025-6

Gallagher, H. L., Jack, A. I., Roepstorff, A., & Frith, C. D. (2002). Imaging the intentional stance in a competitive game. *NeuroImage, 16*(3), 814-821. https://doi.org/10.1006/nimg.2002.1117

Gallese, V. (2003). The roots of empathy: the shared manifold hypothesis and the neural basis of intersubjectivity. *Psychopathology, 36*(4), 171-180. https://doi.org/10.1159/000072786

Gallese, V. (2008). Mirror neurons and the social nature of language: The neural exploitation hypothesis. *Social Neuroscience*, *3*(3-4), 317-333. https://doi.org/10.1080/17470910701563608

Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences, 2*(12), 493-501. https://doi.org/10.1016/S1364-6613(98)01262-5

Gazzola, V., Aziz-Zadeh, L., & Keysers, C. (2006). Empathy and the somatotopic auditory mirror system in humans. *Current Biology, 16*(18), 1824-1829. https://doi.org/10.1016/j.cub.2006.07.072

Gobbini, M. I., Koralek, A. C., Bryan, R. E., Montgomery, K. J., & Haxby, J. V. (2007). Two takes on the social brain: a comparison of theory of mind tasks. *Journal of Cognitive Neuroscience*, *19*(11), 1803-1814. https://doi.org/10.1162/jocn.2007.19.11.1803

Gorgolewski, K. J., Varoquaux, G., Rivera, G., Schwarz, Y., Ghosh, S. S., Maumet, C., ... & Yarkoni, T. (2015). NeuroVault. org: a web-based repository for collecting and sharing unthresholded statistical maps of the human brain. *Frontiers in Neuroinformatics, 9*, 8. https://doi.org/10.3389/fninf.2015.00008

Goulden, N., Khusnulina, A., Davis, N. J., Bracewell, R. M., Bokde, A. L., McNulty, J. P., & Mullins, P. G. (2014). The salience network is responsible for switching between the default mode network and the central executive network: replication from DCM. *Neuroimage, 99,* 180-190. https://doi.org/10.1016/j.neuroimage.2014.05.052

Gregory, C., Lough, S., Stone, V., Erzinclioglu, S., Martin, L., Baron- Cohen, S., & Hodges, J. R. (2002). Theory of mind in patients with frontal variant frontotemporal dementia and

Alzheimer’s disease: theoretical and practical implications. *Brain, 125*(4), 752-764. https://doi.org/10.1093/brain/awf079

Grosbras, M. H., Beaton, S., & Eickhoff, S. B. (2012). Brain regions involved in human movement perception: A quantitative voxel-based meta-analysis. *Human Brain Mapping, 33*(2), 431-454. https://doi.org/10.1002/hbm.21222Gu, X., Hof, P. R., Friston, K. J., & Fan, J. (2013). Anterior insular cortex and emotional awareness. *Journal of Comparative Neurology, 521*(15), 3371-3388. https://doi.org/10.1002/cne.23368

Happé, F. G. (1994). An advanced test of theory of mind: Understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *Journal of Autism and Developmental Disorders*, *24*(2), 129-154. https://doi.org/10.1007/BF02172093

Happé, F., Cook, J. L., & Bird, G. (2017). The structure of social cognition: In (ter) dependence of sociocognitive processes. *Annual Review of Psychology, 68,* 243-267. https://doi.org/10.1146/annurev-psych-010416-044046

Happe, F., Ehlers, S., Fletcher, P., Frith, U., Johansson, M., Gillberg, C., ... & Frith, C. (1996). ‘Theory of mind’ in the brain. Evidence from a PET scan study of Asperger syndrome. *NeuroReport, 8*(1), 197-201. https://www.ncbi.nlm.nih.gov/pubmed/9051780

Hare, T. A., Camerer, C. F., Knoepfle, D. T., O'Doherty, J. P., & Rangel, A. (2010). Value computations in ventral medial prefrontal cortex during charitable decision making incorporate input from regions involved in social cognition. *Journal of Neuroscience, 30*(2), 583-590. https://doi.org/10.1523/JNEUROSCI.4089-09.2010

Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *The American Journal of Psychology, 57*(2), 243-259. http://dx.doi.org/10.2307/1416950

Hein, G., Silani, G., Preuschoff, K., Batson, C. D., & Singer, T. (2010). Neural responses to ingroup and outgroup members' suffering predict individual differences in costly helping. *Neuron, 68*(1), 149-160. https://doi.org/10.1016/j.neuron.2010.09.003

Heyes, C. (2011). Automatic imitation. *Psychological Bulletin, 137*(3), 463. http://dx.doi.org/10.1037/a0022288

Heyes, C. (2018). *Cognitive gadgets: the cultural evolution of thinking*. Harvard University Press. https://doi.org/10.4159/9780674985155

Heyes, C. M., & Frith, C. D. (2014). The cultural evolution of mind reading. *Science, 344*(6190), 1243091. https://doi.org/10.1126/science.1243091

Hogeveen, J., Obhi, S. S., Banissy, M. J., Santiesteban, I., Press, C., Catmur, C., & Bird, G. (2014). Task-dependent and distinct roles of the temporoparietal junction and inferior frontal cortex in the control of imitation. *Social Cognitive and Affective Neuroscience, 10*(7), 1003-1009. https://doi.org/10.1093/scan/nsu148

Hooker, C. I., Verosky, S. C., Germine, L. T., Knight, R. T., & D'Esposito, M. (2008). Mentalizing about emotion and its relationship to empathy. *Social Cognitive and Affective Neuroscience, 3*(3), 204-217. https://doi.org/10.1093/scan/nsn019

Hooker, C. I., Verosky, S. C., Germine, L. T., Knight, R. T., & D'Esposito, M. (2010). Neural activity during social signal perception correlates with self-reported empathy. *Brain Research, 1308*, 100-113. https://doi.org/10.1016/j.brainres.2009.10.006

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., ... & Wang, P. (2010). Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. *American Journal of Psychiatry, 167*(7), 748-51. https://doi.org/10.1176/appi.ajp.2010.09091379

Jabbi, M., Swart, M., & Keysers, C. (2007). Empathy for positive and negative emotions in the gustatory cortex. *NeuroImage*, *34*(4), 1744-1753. https://doi.org/10.1016/j.neuroimage.2006.10.032

Jackson, P. L., Meltzoff, A. N., & Decety, J. (2005). How do we perceive the pain of others? A window into the neural processes involved in empathy. *NeuroImage, 24*(3), 771-779. https://doi.org/10.1016/j.neuroimage.2004.09.006

Kalbe, E., Schlegel, M., Sack, A. T., Nowak, D. A., Dafotakis, M., Bangard, C., ... & Kessler, J. (2010). Dissociating cognitive from affective theory of mind: a TMS study. *Cortex*, *46*(6), 769-780. https://doi.org/10.1016/j.cortex.2009.07.010

Kanske, P. (2018). The social mind: disentangling affective and cognitive routes to understanding others. *Interdisciplinary Science Reviews*, *43*(2), 115-124. https://doi.org/10.1080/03080188.2018.1453243

Kanske, P., Böckler, A., Trautwein, F. M., Parianen Lesemann, F. H., & Singer, T. (2016). Are strong empathizers better mentalizers? Evidence for independence and interaction between the routes of social cognition. *Social Cognitive and Affective Neuroscience, 11*(9), 1383-1392. https://doi.org/10.1093/scan/nsw052

Kanske, P., Böckler, A., & Singer, T. (2017). Models, Mechanisms and Moderators Dissociating Empathy and Theory of Mind. *Current Topics in Behavioral Neurosciences 30*,193-206. https://doi.org/10.1007/7854_2015_412

Kanske, P., Böckler, A., Trautwein, F. M., & Singer, T. (2015). Dissecting the social brain: Introducing the EmpaToM to reveal distinct neural networks and brain–behavior relations for empathy and Theory of Mind. *NeuroImage, 122,* 6-19. https://doi.org/10.1016/j.neuroimage.2015.07.082

Keysers, C., & Gazzola, V. (2009). Expanding the mirror: vicarious activity for actions, emotions, and sensations. *Current Opinion in Neurobiology, 19*(6), 666-671. https://doi.org/10.1016/j.conb.2009.10.006

Keysers, C., Kaas, J. H., & Gazzola, V. (2010). Somatosensation in social perception. *Nature Reviews Neuroscience, 11*(6), 417. https://doi.org/10.1038/nrn2833

Kim, Y. T., Lee, J. J., Song, H. J., Kim, J. H., Kwon, D. H., Kim, M. N., ... & Chang, Y. (2010). Alterations in cortical activity of male methamphetamine abusers performing an empathy task: fMRI study. *Human Psychopharmacology: Clinical and Experimental, 25*(1), 63-70. https://doi.org/10.1002/hup.1083.

Kircher, T., Bluemel, I., Marjoram, D., Lataster, T., Krabbendam, L., Weber, J., van Os, J., & Krach, S. (2009). Online mentalising investigated with functional MRI. *Neuroscience Letters, 454*, 176-181. https://doi.org/10.1016/j.neulet.2009.03.026.

Konstantakopoulos, G., Oulis, P., Ploumpidis, D., Patrikelis, P., Nikitopoulou, S., Papadimitriou, G. N., & David, A. S. (2014). Self-rated and performance-based empathy in schizophrenia: the impact of cognitive deficits. *Social Neuroscience, 9*(6), 590-600. https://doi.org/10.1080/17470919.2014.934395

Koster-Hale, J., & Saxe, R. (2013). Theory of mind: a neural prediction problem. *Neuron*, *79*(5), 836-848. https://doi.org/10.1016/j.neuron.2013.08.020

Krishnan, A., Woo, C. W., Chang, L. J., Ruzic, L., Gu, X., Lopez-Sola, M., ... & Wager, T. D. (2016). Somatic and vicarious pain are represented by dissociable multivariate brain patterns. *Elife, 5*, e15166. https://doi.org/10.1016/10.7554/eLife.15166

Laird, A. R., Fox, P. M., Eickhoff, S. B., Turner, J. A., Ray, K. L., McKay, D. R., ... & Fox, P. T. (2011). Behavioral interpretations of intrinsic connectivity networks. *Journal of Cognitive Neuroscience, 23*(12), 4022-4037. https://doi.org/10.1162/jocn_a_00077

Laird, A. R., Riedel, M. C., Sutherland, M. T., Eickhoff, S. B., Ray, K. L., Uecker, A. M., ... & Fox, P. T. (2015). Neural architecture underlying classification of face perception paradigms. *NeuroImage, 119*, 70-80. https://doi.org/10.1016/j.neuroimage.2015.06.044

Lamm, C., Decety, J., & Singer, T. (2011). Meta-analytic evidence for common and distinct neural networks associated with directly experienced pain and empathy for pain. *NeuroImage, 54*(3), 2492-2502. https://doi.org/10.1016/j.neuroimage.2010.10.014

Leslie, A. M., Friedman, O., & German, T. P. (2004). Core mechanisms in 'theory of mind'. *Trends in Cognitive Sciences, 8*(12), 528-533. https://doi.org/10.1016/j.tics.2004.10.001

Lieberman, M. D. (2007). Social cognitive neuroscience: a review of core processes. *Annual Reviews of Psychology*, *58*, 259-289. https://doi.org/10.1146/annurev.psych.58.110405.085654

Lieberman, M. D., Burns, S. M., Torre, J. B., & Eisenberger, N. I. (2016). Reply to Wager et al.: Pain and the dACC: The importance of hit rate-adjusted effects and posterior probabilities with fair priors. *Proceedings of the National Academy of Sciences, 113*(18), E2476-E2479. https://doi.org/10.1073/pnas.1603186113

Lieberman, M. D., & Eisenberger, N. I. (2015). The dorsal anterior cingulate cortex is selective for pain: Results from large-scale reverse inference. *Proceedings of the National Academy of Sciences, 112*(49), 15250-15255. https://doi.org/10.1073/pnas.1515083112

Lindquist, K. A. (2017). The role of language in emotion: existing evidence and future directions. *Current Opinion in Psychology, 17*, 135-139. https://doi.org/10.1016/j.copsyc.2017.07.006

Lindquist, K. A., Satpute, A. B., & Gendron, M. (2015). Does language do more than communicate emotion? *Current Directions in Psychological Science*, *24*(2), 99-108. https://doi.org/10.1177/0963721414553440

Lindquist, K. A., Wager, T. D., Kober, H., Bliss-Moreau, E., & Barrett, L. F. (2012). The brain basis of emotion: a meta-analytic review. *Behavioral and Brain Sciences, 35*(3), 121-143.https://doi.org/10.1017/S0140525X11000446

Lockwood, P. L., Apps, M. A., Roiser, J. P., & Viding, E. (2015). Encoding of vicarious reward prediction in anterior cingulate cortex and relationship with trait empathy. *Journal of Neuroscience, 35*(40), 13720-13727. https://doi.org/10.1523/JNEUROSCI.1703-15.2015

Lockwood, P. L., Bird, G., Bridge, M., & Viding, E. (2013). Dissecting empathy: high levels of psychopathic and autistic traits are characterized by difficulties in different social information processing domains. *Frontiers in Human Neuroscience, 7,* 760. https://doi.org/10.3389/fnhum.2013.00760

Ma, N., Vanderkerckhove, M., van Overwalle, F., Seurinck, R., & Fias, W. (2011). Spontaneous and intentional trait inferences recruit a common mentalizing network to a different degree: spontaneous inferences activate only its core areas. *Social Neuroscience, 6*(2), 123-138. https://doi.org/10.1080/17470919.2010.485884.

Mar, R. A. (2011). The neural bases of social cognition and story comprehension. *Annual Review of Psychology, 62,* 103-134. https://doi.org/10.1146/annurev-psych-120709-145406

Margulies, D. S., Ghosh, S. S., Goulas, A., Falkiewicz, M., Huntenburg, J. M., Langs, G., ... & Jefferies, E. (2016). Situating the default-mode network along a principal gradient of macroscale cortical organization. *Proceedings of the National Academy of Sciences, 113*(44), 12574-12579. https://doi.org/10.1073/pnas.1608282113

Mars, R. B., Neubert, F. X., Noonan, M. P., Sallet, J., Toni, I., & Rushworth, M. F. (2012). On the relationship between the “default mode network” and the “social brain”. *Frontiers in Human Neuroscience, 6,* 189. https://doi.org/10.3389/fnhum.2012.00189

Mars, R. B., Sallet, J., Schüffelgen, U., Jbabdi, S., Toni, I., & Rushworth, M. F. (2011). Connectivity-based subdivisions of the human right “temporoparietal junction area”: evidence for different areas participating in different cortical networks. *Cerebral Cortex, 22*(8), 1894-1903. https://doi.org/10.1093/cercor/bhr268

Mathur, V. A., Harada, T., Lipke, T., & Chiao, J. Y. (2010). Neural basis of extraordinary empathy and altruistic motivation. *NeuroImage, 51*(4), 1468-1475. https://doi.org/10.1016/j.neuroimage.2010.03.025

Meyer, M. L., Davachi, L., Ochsner, K. N., & Lieberman, M. D. (2018). Evidence that default network connectivity during rest consolidates social information. *Cerebral Cortex, 1*, 11. doi: 10.1093/cercor/bhy071

Mier, D., Lis, S., Neuthe, K., Sauer, C., Esslinger, C., Gallhofer, B., & Kirsch, P. (2010). The involvement of emotion recognition in affective theory of mind. *Psychophysiology, 47*(6), 1028-1039. https://doi.org/10.1111/j.1469-8986.2010.01031.x

Miller, L. E., & Saygin, A. P. (2013). Individual differences in the perception of biological motion: links to social cognition and motor imagery. *Cognition, 128*(2), 140-148. https://doi.org/10.1016/j.cognition.2013.03.013

Mitchell, J. P. (2009). Inferences about mental states. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364*(1521), 1309-1316. https://doi.org/10.1098/rstb.2008.0318

Mitchell, J.P., Banaji, M.R., MacRae, C.N. (2005). The link between social cognition and self-referential thought in the medial prefrontal cortex. *Journal of Cognitive Neuroscience, 17*, 1306-1315.

Mitchell, J. P., Heatherton, T. F., & Macrae, C. N. (2002). Distinct neural systems subserve person and object knowledge. *Proceedings of the National Academy of Sciences, 99*(23), 15238-15243. https://doi.org/10.1073/pnas.232395699

Morelli, S. A., Sacchet, M. D., & Zaki, J. (2015). Common and distinct neural correlates of personal and vicarious reward: A quantitative meta-analysis. *NeuroImage, 112*, 244-253. https://doi.org/10.1016/j.neuroimage.2014.12.056

Morrison, I., Lloyd, D., Di Pellegrino, G., & Roberts, N. (2004). Vicarious responses to pain in anterior cingulate cortex: is empathy a multisensory issue? *Cognitive, Affective, & Behavioral Neuroscience, 4*(2), 270-278. https://doi.org/10.3758/CABN.4.2.270

Molenberghs, P., Johnson, H., Henry, J. D., & Mattingley, J. B. (2016). Understanding the minds of others: A neuroimaging meta-analysis. *Neuroscience & Biobehavioral Reviews, 65*, 276-291. https://doi.org/10.1016/j.neubiorev.2016.03.020

Müller, V. I., Cieslik, E. C., Laird, A. R., Fox, P. T., Radua, J., Mataix-Cols, D., ... & Wager, T. D. (2018). Ten simple rules for neuroimaging meta-analysis. *Neuroscience & Biobehavioral Reviews, 84*, 151-161.

Murray, R. J., Schaer, M., & Debbané, M. (2012). Degrees of separation: A quantitative neuroimaging meta-analysis investigating self-specificity and shared neural activation between self-and other-reflection. *Neuroscience & Biobehavioral Reviews, 36*(3), 1043-1059. https://doi.org/10.1016/j.neubiorev.2011.12.013

Nichols, T., Brett, M., Andersson, J., Wager, T., & Poline, J. B. (2005). Valid conjunction inference with the minimum statistic. *NeuroImage, 25*(3), 653-660. https://doi.org/10.1016/j.neuroimage.2004.12.005

Oakley, B. F., Brewer, R., Bird, G., & Catmur, C. (2016). Theory of mind is not theory of emotion: A cautionary note on the Reading the Mind in the Eyes Test. *Journal of Abnormal Psychology, 125*(6), 818. http://dx.doi.org/10.1037/abn0000182

Ochsner, K. N., Zaki, J., Hanelin, J., Ludlow, D. H., Knierim, K., Ramachandran, T., ... & Mackey, S. C. (2008). Your pain or mine? Common and distinct neural systems supporting the perception of pain in self and other. *Social Cognitive and Affective Neuroscience, 3*(2), 144-160. https://doi.org/10.1093/scan/nsn006

Olderbak, S., Wilhelm, O., Olaru, G., Geiger, M., Brenneman, M. W., & Roberts, R. D. (2015). A psychometric analysis of the reading the mind in the eyes test: toward a brief form for research and applied settings. *Frontiers in Psychology, 6*, 1503. https://doi.org/10.3389/fpsyg.2015.01503

Olsson, A., Nearing, K. I., & Phelps, E. A. (2007). Learning fears by observing others: the neural systems of social fear transmission. *Social Cognitive and Affective Neuroscience, 2*(1), 3-11. https://doi.org/10.1093/scan/nsm005

Perner, J., & Leekam, S. (2008). The curious incident of the photo that was accused of being false: Issues of domain specificity in development, autism, and brain imaging. *Quarterly Journal of Experimental Psychology, 61*(1), 76-89. https://doi.org/10.1080/17470210701508756

Perner, J., & Roessler, J. (2012). From infants' to children's appreciation of belief. *Trends in Cognitive Sciences, 16*(10), 519-525. https://doi.org/10.1016/j.tics.2012.08.004

Pfeifer, J. H., & Dapretto, M. (2009). *"Mirror, mirror, in my mind": Empathy, interpersonal competence, and the mirror neuron system*. In J. Decety & W. Ickes (Eds.), *Social neuroscience. The social neuroscience of empathy* (p. 183–197). MIT Press. https://doi.org/10.7551/mitpress/9780262012973.003.0015

Preckel, K., Kanske, P., & Singer, T. (2018). On the interaction of social affect and cognition: empathy, compassion and theory of mind. *Current Opinion in Behavioral Sciences, 19,* 1-6. https://doi.org/10.1016/j.cobeha.2017.07.010

Preis, M. A., Schmidt-Samoa, C., Dechent, P., & Kroener-Herwig, B. (2013). The effects of prior pain experience on neural correlates of empathy for pain: An fMRI study. *Pain, 154*(3), 411-418. https://doi.org/10.1016/j.pain.2012.11.014.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind?. *Behavioral and Brain Sciences, 1*(4), 515-526. https://doi.org/10.1017/S0140525X00076512

Preston, S. D. (2007). *A perception-action model for empathy*. In T. Farrow & P. Woodruff (Eds.), *Empathy in mental illness* (p. 428–447). Cambridge University Press. https://doi.org/10.1017/CBO9780511543753.024

Preston, S. D., Bechara, A., Damasio, H., Grabowski, T. J., Stansfield, R. B., Mehta, S., & Damasio, A. R. (2007). The neural substrates of cognitive empathy. *Social Neuroscience, 2*(3-4), 254-275.

Preston, S. D., & de Waal, F. B. (2002). Empathy: Its ultimate and proximate bases. *Behavioral and Brain Sciences, 25*(1), 1-20. https://doi.org/10.1017/S0140525X02000018

Preston, S. D., & Hofelich, A. J. (2012). The many faces of empathy: Parsing empathic phenomena through a proximate, dynamic-systems view of representing the other in the self. *Emotion Review, 4*(1), 24-33.

Price, C. J., Devlin, J. T., Moore, C. J., Morton, C., & Laird, A. R. (2005). Meta- analyses of object naming: Effect of baseline. *Human Brain Mapping, 25*(1), 70-82. https://doi.org/10.1002/hbm.20132

Radua, J., Mataix-Cols, D., Phillips, M. L., El-Hage, W., Kronhaus, D. M., Cardoner, N., & Surguladze, S. (2012). A new meta-analytic method for neuroimaging studies that combines reported peak coordinates and statistical parametric maps. *European Psychiatry, 27*(8), 605-611. https://doi.org/10.1016/j.eurpsy.2011.04.001

Regenbogen, C., Schneider, D. A., Gur, R. E., Schneider, F., Habel, U., & Kellermann, T. (2012). Multimodal human communication—targeting facial expressions, speech content and prosody. *NeuroImage, 60*(4), 2346-2356. https://doi.org/10.1016/j.neuroimage.2012.02.043

Reniers, R. L., Völlm, B. A., Elliott, R., & Corcoran, R. (2014). Empathy, ToM, and self–other differentiation: An fMRI study of internal states. *Social Neuroscience, 9*(1), 50-62. https://doi.org/10.1080/17470919.2013.86136

Rice, K., Anderson, L. C., Velnoskey, K., Thompson, J. C., & Redcay, E. (2016). Biological motion perception links diverse facets of theory of mind during middle childhood. *Journal of Experimental Child Psychology, 146*, 238-246. https://doi.org/10.1016/j.jecp.2015.09.003

Riedel, M. C., Ray, K. L., Dick, A. S., Sutherland, M. T., Hernandez, Z., Fox, P. M., ... & Laird, A. R. (2015). Meta-analytic connectivity and behavioral parcellation of the human cerebellum. *NeuroImage, 117,* 327-342. https://doi.org/10.1016/j.neuroimage.2015.05.008

Riedel, M. C., Yanes, J. A., Ray, K. L., Eickhoff, S. B., Fox, P. T., Sutherland, M. T., & Laird, A. R. (2018). Dissociable meta- analytic brain networks contribute to coordinated emotional processing. *Human Brain Mapping, 39*(6), 2514-2531. https://doi.org/10.1002/hbm.24018

Rilling, J. K., Sanfey, A. G., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2004). The neural correlates of theory of mind within interpersonal interactions. *NeuroImage, 22*(4), 1694-1703. https://doi.org/10.1016/j.neuroimage.2004.04.015

Rimé, B., Boulanger, B., Laubin, P., Richir, M., & Stroobants, K. (1985). The perception of interpersonal emotions originated by patterns of movement. *Motivation and Emotion, 9*(3), 241-260. https://doi.org/10.1007/BF00991830

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience, 27*, 169-192. https://doi.org/10.1146/annurev.neuro.27.070203.144230

Ruby, P., & Decety, J. (2004). How would you feel versus how do you think she would feel? A neuroimaging study of perspective-taking with social emotions. *Journal of Cognitive Neuroscience, 16*(6), 988-999. https://doi.org/10.1162/0898929041502661

Rütgen, M., Seidel, E. M., Silani, G., Riečanský, I., Hummer, A., Windischberger, C., ... & Lamm, C. (2015). Placebo analgesia and its opioidergic regulation suggest that empathy for pain is grounded in self pain. *Proceedings of the National Academy of Sciences, 112*(41), E5638-E5646. https://doi.org/10.1073/pnas.1511269112

Santiesteban, I., Banissy, M. J., Catmur, C., & Bird, G. (2012). Enhancing social ability by stimulating right temporoparietal junction. *Current Biology, 22*(23), 2274-2277. https://doi.org/10.1016/j.cub.2012.10.018

Sarfati, Y., Hardy-Baylé, M. C., Besche, C., & Widlöcher, D. (1997). Attribution of intentions to others in people with schizophrenia: a non-verbal exploration with comic strips. *Schizophrenia Research, 25*(3), 199-209. https://doi.org/10.1016/S0920-9964(97)00025-X

Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people: the role of the temporo-parietal junction in "theory of mind". *NeuroImage*, *19*(4), 1835-1842. https://doi.org/10.1016/S1053-8119(03)00230-1

Schilbach, L., Eickhoff, S. B., Mojzisch, A., & Vogeley, K. (2008). What's in a smile? Neural correlates of facial embodiment during social interaction. *Social Neuroscience, 3*(1), 37-50. https://doi.org/10.1080/17470910701563228

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). Toward a second-person neuroscience. *Behavioral and Brain Sciences, 36*(4), 393-414. https://doi.org/10.1017/S0140525X12000660

Schlaffke, L., Lissek, S., Lenz, M., Juckel, G., Schultz, T., Tegenthoff, M., ... & Brüne, M. (2015). Shared and nonshared neural networks of cognitive and affective theory- of-mind: A neuroimaging study using cartoon picture stories. *Human Brain Mapping, 36*(1), 29-39. https://doi.org/10.1002/hbm.22610

Schurz, M., Aichhorn, M., Martin, A., & Perner, J. (2013). Common brain areas engaged in false belief reasoning and visual perspective taking: a meta-analysis of functional brain imaging studies. *Frontiers in Human Neuroscience, 7*, 712. https://doi.org/10.3389/fnhum.2013.00712

Schurz, M., Radua, J., Aichhorn, M., Richlan, F., & Perner, J. (2014). Fractionating theory of mind: a meta-analysis of functional brain imaging studies. *Neuroscience & Biobehavioral Reviews, 42*, 9-34. https://doi.org/10.1016/j.neubiorev.2014.01.009

Schaafsma, S. M., Pfaff, D. W., Spunt, R. P., & Adolphs, R. (2015). Deconstructing and reconstructing theory of mind. *Trends in Cognitive Sciences*, *19*(2), 65-72. https://doi.org/10.1016/j.tics.2014.11.007

Schnell, K., Bluschke, S., Konradt, B., & Walter, H. (2011). Functional relations of empathy and mentalizing: an fMRI study on the neural basis of cognitive empathy. *NeuroImage, 54*(2), 1743-1754. https://doi.org/10.1016/j.neuroimage.2010.08.024

Schulte-Rüther, M., Markowitsch, H. J., Fink, G. R., & Piefke, M. (2007). Mirror neuron and theory of mind mechanisms involved in face-to-face interactions: a functional magnetic resonance imaging approach to empathy. *Journal of Cognitive Neuroscience, 19*(8), 1354-1372. https://doi.org/10.1162/jocn.2007.19.8.1354

Schurz, M., & Perner, J. (2015). An evaluation of neurocognitive models of theory of mind. *Frontiers in Psychology, 6*, 1610.

Sebastian, C. L., Fontaine, N. M., Bird, G., Blakemore, S. J., De Brito, S. A., McCrory, E. J., & Viding, E. (2012). Neural processing associated with cognitive and affective Theory of Mind in adolescents and adults. *Social Cognitive and Affective Neuroscience, 7*(1), 53-63. https://doi.org/10.1093/scan/nsr023

Shamay-Tsoory, S. G., & Aharon-Peretz, J. (2007). Dissociable prefrontal networks for cognitive and affective theory of mind: a lesion study. *Neuropsychologia, 45*(13), 3054-3067. https://doi.org/10.1016/j.neuropsychologia.2007.05.021

Shamay-Tsoory, S. G., Aharon-Peretz, J., & Perry, D. (2009). Two systems for empathy: a double dissociation between emotional and cognitive empathy in inferior frontal gyrus versus ventromedial prefrontal lesions. *Brain, 132*(3), 617-627. https://doi.org/10.1093/brain/awn279

Singer, T. (2006). The neuronal basis and ontogeny of empathy and mind reading: review of literature and implications for future research. *Neuroscience & Biobehavioral Reviews, 30*(6), 855-863.

Singer, T., & Lamm, C. (2009). The social neuroscience of empathy. *Annals of the New York Academy of Sciences, 1156*(1), 81-96. https://doi.org/10.1111/j.1749-6632.2009.04418.x

Singer, T., Seymour, B., O'Doherty, J., Kaube, H., Dolan, R. J., & Frith, C. D. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science, 303*(5661), 1157-1162. https://doi.org/10.1126/science.1093535

Spreng, R. N., Mar, R. A., & Kim, A. S. (2009). The common neural basis of autobiographical memory, prospection, navigation, theory of mind, and the default mode: a quantitative meta-analysis. *Journal of Cognitive Neuroscience, 21*(3), 489-510. https://doi.org/10.1162/jocn.2008.21029

Spreng, R. N., & Grady, C. L. (2010). Patterns of brain activity supporting autobiographical memory, prospection, and theory of mind, and their relationship to the default mode network. *Journal of Cognitive Neuroscience, 22*(6), 1112-1123. https://doi.org/10.1162/jocn.2009.21282

Spunt, R. P., Meyer, M. L., & Lieberman, M. D. (2015). The default mode of human brain function primes the intentional stance. *Journal of Cognitive Neuroscience, 27*(6), 1116-1124. https://doi.org/10.1162/jocn_a_00785

Spunt, R. P., & Adolphs, R. (2017). A new look at domain specificity: insights from social neuroscience. *Nature Reviews Neuroscience*, *18*(9), 559. https://doi.org/doi: 10.1038/nrn.2017.76

Tager-Flusberg, H., Boshart, J., & Baron-Cohen, S. (1998). Reading the windows to the soul: Evidence of domain-specific sparing in Williams syndrome. *Journal of Cognitive Neuroscience, 10*(5), 631-639. https://doi.org/10.1162/089892998563031

Titchener, E.B. 1909. Experimental Psychology of the Thought-Processes. New York: MacMillan.

Timmers, I., Park, A. L., Fischer, M. D., Kronman, C. A., Heathcote, L. C., Hernandez, J. M., & Simons, L. E. (2018). Is empathy for pain unique in its neural correlates? A meta-analysis of neuroimaging studies of empathy. *Frontiers in Behavioral Neuroscience, 12*. doi.org/10.3389/fnbeh.2018.00289

Tamir, D. I., & Thornton, M. A. (2018). Modeling the predictive social mind. *Trends in cognitive sciences, 22*(3), 201-212. https://doi.org/10.1016/j.tics.2017.12.005

Tholen, M. G., Trautwein, F. M., Böckler, A., Singer, T., & Kanske, P. (2020). Functional magnetic resonance imaging (fMRI) item analysis of empathy and theory of mind. *Human Brain Mapping, 41*(10), 2611-2628. https://doi.org/10.1002/hbm.24966

Thornton, M. A., Weaverdyck, M. E., & Tamir, D. I. (2019). The social brain automatically predicts others' future mental states. *Journal of Neuroscience, 39*(1), 140-148. https://doi.org/10.1523/JNEUROSCI.1431-18.2018

Toni, I., De Lange, F. P., Noordzij, M. L., & Hagoort, P. (2008). Language beyond action. *Journal of Physiology-Paris, 102*(1-3), 71-79. https://doi.org/10.1016/j.jphysparis.2008.03.005

Trautwein, F. M., Singer, T., & Kanske, P. (2016). Stimulus-driven reorienting impairs executive control of attention: Evidence for a common bottleneck in anterior insula. *Cerebral Cortex, 26*(11), 4136-4147. https://doi.org/10.1093/cercor/bhw225

Tusche, A., Böckler, A., Kanske, P., Trautwein, F. M., & Singer, T. (2016). Decoding the charitable brain: empathy, perspective taking, and attention shifts differentially predict altruistic giving. *Journal of Neuroscience, 36*(17), 4719-4732. https://doi.org/10.1523/JNEUROSCI.3392-15.201

Ushida, T., Ikemoto, T., Tanaka, S., Shinozaki, J., Taniguchi, S., Murata, Y., ... & Tamura, Y. (2008). Virtual needle pain stimuli activates cortical representation of emotions in normal volunteers. *Neuroscience Letters, 439*(1), 7-12. https://doi.org/10.1016/j.neulet.2008.04.085

van der Heiden, L., Scherpiet, S., Konicar, L., Birbaumer, N., & Veit, R. (2013). Inter-individual differences in successful perspective taking during pain perception mediates emotional responsiveness in self and others: an fMRI study. *NeuroImage, 65*, 387-394. https://doi.org/10.1016/j.neuroimage.2012.10.003

Van Overwalle, F. (2009). Social cognition and the brain: a meta- analysis. *Human Brain Mapping, 30*(3), 829-858. https://doi.org/10.1002/hbm.20547

Van Overwalle, F., & Baetens, K. (2009). Understanding others' actions and goals by mirror and mentalizing systems: a meta-analysis. *NeuroImage, 48*(3), 564-584. https://doi.org/10.1016/j.neuroimage.2009.06.009

Vijayakumar, N., Cheng, T. W., & Pfeifer, J. H. (2017). Neural correlates of social exclusion across ages: A coordinate-based meta-analysis of functional MRI studies. *NeuroImage, 153*, 359-368. https://doi.org/10.1016/j.neuroimage.2017.02.050

Wager, T. D., Atlas, L. Y., Botvinick, M. M., Chang, L. J., Coghill, R. C., Davis, K. D., ... & Yarkoni, T. (2016). Pain in the ACC?. *Proceedings of the National Academy of Sciences, 113*(18), E2474-E2475. https://doi.org/10.1073/pnas.1600282113

Walter, H. (2012). Social cognitive neuroscience of empathy: concepts, circuits, and genes. *Emotion Review*, *4*(1), 9-17. https://doi.org/10.1177/1754073911421379

Walter, H.W., Adenzato, M., Ciaramidaro, A., Enrici, I., Lorenzo, P., & Bara, B.G. (2004). Understanding intentions in social interaction: the role of the anterior paracingulate cortex. *Journal of Cognitive Neuroscience, 16*(10), 1854-1863. https://doi.org/10.1162/0898929042947838

Wen, X., Liu, Y., Yao, L., & Ding, M. (2013). Top-down regulation of default mode activity in spatial visual attention. *Journal of Neuroscience, 33*(15), 6444-6453. https://doi.org/10.1523/JNEUROSCI.

White, S. J., Coniston, D., Rogers, R., & Frith, U. (2011). Developing the Frith- Happé animations: A quick and objective test of Theory of Mind for adults with autism. *Autism Research, 4*(2), 149-154. https://doi.org/10.1002/aur.174

White, S., Hill, E., Happe, F. & Frith, U. Revisiting the Strange Stories: revealing mentalizing impairments in autism. *Child Development, 80*, 1097-1117 (2009). https://doi.org/10.1111/j.1467-8624.2009.01319.x

Wicker, B., Keysers, C., Plailly, J., Royet, J. P., Gallese, V., & Rizzolatti, G. (2003). Both of us disgusted in My insula: the common neural basis of seeing and feeling disgust. *Neuron, 40*(3), 655-664. https://doi.org/10.1016/S0896-6273(03)00679-2

Wilson-Mendenhall, C. D., Barrett, L. F., Simmons, W. K., & Barsalou, L. W. (2011). Grounding emotion in situated conceptualization. *Neuropsychologia, 49*(5), 1105-1127. https://doi.org/10.1016/j.neuropsychologia.2010.12.032

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13*(1), 103-128. https://doi.org/10.1016/0010-0277(83)90004-5

Wolf, I., Dziobek, I., & Heekeren, H. R. (2010). Neural correlates of social cognition in naturalistic settings: a model-free analysis approach. *NeuroImage, 49*(1), 894-904. https://doi.org/10.1016/j.neuroimage.2009.08.060

Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., & Wager, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nature Methods, 8*(8), 665. https://doi.org/ 10.1038/nmeth.1635

Yeo, B. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., ... & Fischl, B. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology, 106*(3), 1125-1165. https://doi.org/10.1152/jn.00338.2011

Zaki, J. (2017). Moving beyond stereotypes of empathy. *Trends in Cognitive Sciences, 21*(2), 59. https://doi.org/10.1016/j.tics.2016.12.004

Zaki, J., Ochsner, K. N., Hanelin, J., Wager, T. D., & Mackey, S. C. (2007). Different circuits for different pain: patterns of functional connectivity reveal distinct networks for processing pain in self and others. *Social Neuroscience, 2*(3-4), 276-291. https://doi.org/10.1080/17470910701401973

Zaki, J., & Ochsner, K. N. (2012). The neuroscience of empathy: progress, pitfalls and promise. *Nature Neuroscience, 15*(5), 675. https://doi.org/10.1038/nn.3085

Zaki, J., Weber, J., Bolger, N., & Ochsner, K. (2009). The neural bases of empathic accuracy. *Proceedings of the National Academy of Sciences, 106*(27), 11382-11387. https://doi.org/10.1073/pnas.0902666106

Zhou, Y., Friston, K. J., Zeidman, P., Chen, J., Li, S., & Razi, A. (2017). The hierarchical organization of the default, dorsal attention and salience networks in adolescents and young adults. *Cerebral Cortex, 28*(2), 726-737. https://doi.org/10.1093/cercor/bhx307

Zhu, Y., Zhang, L., Fan, J., & Han, S. (2007). Neural basis of cultural influence on self-representation. *NeuroImage, 34*(3), 1310-1316.

References in the meta-analyses

Adams, R.B. Jr., Rule, N.O., Franklin, R.G. Jr., Wang, E., Stevenson, M.T., Yoshikawa, S., Nomura, M., Sato, W., Kveraga, K., Ambady, N. (2010). Cross-cultural reading the mind in the eyes: an fMRI investigation. *Journal of Cognitive Neuroscience. 22*(1), 97-108. https://doi.org/10.1162/jocn.2009.21187

Aichhorn, M., Perner, J., Weiss, B., Kronbichler, M., Staffen, W., & Ladurner, G. (2009). Temporo-parietal junction activity in theory-of-mind tasks: falseness, beliefs, or attention. Journal of *Cognitive Neuroscience 21*(6), 1179-1192. https://doi.org/10.1162/jocn.2009.21082

Akitsuki, Y., & Decety, J. (2009). Social context and perceived agency affects empathy for pain: an event-related fMRI investigation. *NeuroImage, 47*(2), 722-734. https://doi.org/10.1016/j.neuroimage.2009.04.091

Ammons, C., Doss, C., Bala, D., & Kana, R. (2018). Brain responses underlying anthropomorphism, agency, and social attribution in Autism Spectrum Disorder. *The Open Neuroimaging Journal, 12,* 16-29. https://doi.org/10.2174/1874440001812010016

Andersson, F., Glaser, B., Spiridon, M., Debbané, M., Vuilleumier, P., & Eliez, S. (2008). Impaired activation of face processing networks revealed by functional magnetic resonance imaging in 22q11. 2 deletion syndrome. *Biological Psychiatry, 63*(1), 49-57. https://doi.org/10.1016/j.biopsych.2007.02.022

Andrews-Hanna, J., Saxe, R., & Yarkoni, T. (2014). Contributions of episodic retrieval and mentalizing to autobiographical thought: Evidence from functional neuroimaging, resting-state connectivity, and fMRI meta-analyses. *NeuroImage, 91,* 324-335. https://doi.org/10.1016/j.neuroimage.2014.01.032

Assaf, M., Kahn, I., Pearlson, G.D., Johnson, M.R., Yeshurun, Y., Calhoun, V.D., & Hendler, T. (2009). Brain Activity Dissociates Mentalization from Motivation During an Interpersonal Competitive Game. *Brain Imaging and Behavior, 3*(1), 24-37. https://doi.org/10.1007/s11682-008-9047-y

Azevedo, R. T., Macaluso, E., Avenanti, A., Santangelo, V., Cazzato, V., & Aglioti, S. M. (2013). Their pain is not our pain: brain and autonomic correlates of empathic resonance with the pain of same and different race individuals. *Human Brain Mapping, 34*(12), 3168-3181. https://doi.org/10.1002/hbm.22133

Azevedo, R. T., Macaluso, E., Viola, V., Sani, G., & Aglioti, S. M. (2014). Weighing the stigma of weight: An fMRI study of neural reactivity to the pain of obese individuals. *NeuroImage, 91*, 109-119. https://doi.org/10.1016/j.neuroimage.2013.11.041

Bahnemann, M., Dziobek, I., Prehn, K., Wolf, I., & Heekeren, H.R. (2010). Sociotopy in the temporoparietal cortex: common versus distinct processes. *Social Cognititve and Affective Neuroscience, 5*(1),48-58. https://doi.org/10.1093/scan/nsp045

Ballotta, D., Lui, F., Porro, C., Nichello, P., & Benuzzi, F. (2018). Modulation of neural circuits underlying temporal production by facial expressions of pain. *PLoS ONE, 13*(2), e0193100. https://doi.org/10.1371/journal.pone.0193100

Baron-Cohen, S., Ring, H.A., Wheelwright, S., Bullmore, E.T., Brammer, M.J., Simmons, A., & Williams, S.C. (1999). Social intelligence in the normal and autistic brain: an fMRI study. *European Journal of Neuroscience, 11*(6), 1891-1898. https://doi.org/10.1046/j.1460-9568.1999.00621.x

Batut, A.-C., Gounot, D., Namer, I., Hirsch, E., Kehrli, P., & Metz-Lutz, M.-N. (2006). *Epilepsy & Behavior, 9,* 415-423. https://doi.org/10.1016/j.yebeh.2006.07.013

Blakemore, S.-J., Boyer, P., Pachot-Clouard, M., Meltzoff, A., Segebarth, C., & Decety, J. (2003). The Detection of Contingency and Animacy from Simple Animations in the Human Brain. *Cerebral Cortex, 13*(8), 837-844. https://doi.org/10.1093/cercor/13.8.837

Bliksted, V., Frith, C., Videbech, P., Fagerlund, B., Emborg, C., Simonsen, A., … Campbell-Meiklejohn, D. (2019). Hyper- and Hypomentalizing in Patients with First-Episode Schizophrenia: fMRI and Behavioral Studies. *Schizophrenia Bulletin, 45*(2), 377-385. https://doi.org/10.1093/schbul/sby027

Bos, P. A., Hofman, D., Hermans, E. J., Montoya, E. R., Baron-Cohen, S., & van Honk, J. (2016). Testosterone reduces functional connectivity during the 'Reading the Mind in the Eyes' Test. *Psychoneuroendocrinology, 68*, 194-201. https://doi.org/10.1016/j.psyneuen.2016.03.006

Bos, P. A., Montoya, E. R., Hermans, E. J., Keysers, C., & van Honk, J. (2015). Oxytocin reduces neural activity in the pain circuitry when seeing pain in others. *NeuroImage, 113*, 217-224. https://doi.org/10.1016/j.neuroimage.2015.03.049

Bruene, M., Lissek, S., Fuchs, N., Witthaus, H., Peters, S., Nicolas, V., Juckel, G., & Tegenthoff, M. (2008). An fMRI study of theory of mind in schizophrenic patients with 'passivity' symptoms. *Neuropsychologia, 46*(7), 1992-2001. https://doi.org/10.1016/j.neuropsychologia.2008.01.023

Brunet, E., Sarfati, Y., Hardy-Baylé, M.C., & Decety, J. (2000). A PET investigation of the attribution of intentions with a nonverbal task. *NeuroImage, 11*(2), 157-166. https://doi.org/10.1006/nimg.1999.0525

Carr, L., Iacoboni, M., Dubeau, M. C., Mazziotta, J. C., & Lenzi, G. L. (2003). Neural mechanisms of empathy in humans: a relay from neural systems for imitation to limbic areas. *Proceedings of the national Academy of Sciences, 100*(9), 5497-5502. https://doi.org/10.1073/pnas.0935845100

Castelli, F., Happé, F., Frith, U., & Frith, C. (2000). Movement and mind: a functional imaging study of perception and interpretation of complex intentional movement patterns. *NeuroImage, 12*(3), 314-315. https://doi.org/10.1006/nimg.2000.0612

Castelli, I., Baglio, F., Blasi, V., Alberoni, M., Falini, A., Liverta-Sempio, O., Nemni, R., & Marchetti, A. (2010). Effects of aging on mindreading ability through the eyes: an fMRI study. *Neuropsychologia. 48*(9), 2586-2594. https://doi.org/10.1016/j.neuropsychologia.2010.05.005

Chakrabarti, B., Bullmore, E., & Baron-Cohen, S. (2006). Empathizing with basic emotions: common and discrete neural substrates. *Social Neuroscience, 1*(3-4), 364-384. https://doi.org/10.1080/17470910601041317

Chaminade, T., Rosset, D., Da Fonseca, D., Nazarian, B., Lutcher, E., Cheng, G., & Deruelle, C. (2012). How do we think machines think? An fMRI study of alleged competition with an artificial intelligence. *Frontiers in Human Neuroscience, 6*:103. https://doi.org/10.3389/fnhum.2012.00103

Christov-Moore, L., Conway, P., & Iacoboni, M. (2017). Deontological dilemma response tendencies and sensorimotor representations of harm to others. *Frontiers in Integrative Neuroscience, 11*, 34. https://doi.org/10.3389/fnint.2017.00034

Contreras, J. M., Schirmer, J., Banaji, M., & Mitchell, J.(2013a,b). Common brain regions with distinct patterns of neural responses during mentalizing about groups and individuals. *Journal of Cognitive Neuroscience*, *25*(9). 1406-1417. https://doi.org/10.1162/jocn_a_00403

Cools, R., Calder, A., Lawrence, A., Clark, L., Bullmore, E., & Robbins, T. (2005). Individual differences in threat sensitivity predict serotonergic modulation of amygdala response to fearful faces. *Psychopharmacology*, *180*, 670-679. https://doi.org/10.1007/s00213-005-2215-5

Corradi-Dell'Acqua, C., Hofstetter, C., & Vuilleumier, P. (2014). Cognitive and affective theory of mind share the same local patterns of activity in posterior temporal but not medial prefrontal cortex. *Social Cognitive and Affective Neuroscience, 9*, 1175-1184. https://doi.org/10.1093/scan/nst097

Craik, F.I.M., Moroz, T.M., Moscovitch, M., Stuss, D.T., Winocur, G., Tulving, E., & Kapur, S. (1999). In Search of the Self: A Positron Emission Tomography Study. *Psychological Science, 10*(1), 26-34. https://doi.org/10.1111/1467-9280.00102

Critchley, H., Daly, E., Phillips, M., Brammer, M., Bullmore, E., Williams, S., ... & Murphy, D. (2000). Explicit and implicit neural mechanisms for processing of social information from facial expressions: a functional magnetic resonance imaging study. *Human Brain Mapping, 9*(2), 93-105. https://doi.org/10.1002/(SICI)1097-0193(200002)9:2<93::AID-HBM4>3.0.CO;2-Z

Das, P., Lagopoulus, J., Coulston, C.M., Henderson, A.F., & Malhi, G.S. (2012). Mentalizing impairment in schizophrenia: a functional MRI study. *Schizophrenia Research, 134*(2-3), 158-164. https://doi.org/10.1016/j.schres.2011.08.019

De Achával, D., Villareal, M.F., Costanzo, E.Y., Douer, J., Castro, M.N., Mora, M.C., Nemeroff, C.B., Chu, E., Bär, K.J., & Guinjoan, S.M. (2012). Decreased activity in right-hemisphere structures involved in social cognition in siblings discordant for schizophrenia. *Schizophrenia Research, 134*(2-3), 171-179. https://doi.org/10.1016/j.schres.2011.11.010

Decety, J., Echols, S., & Correll, J. (2010). The blame game: the effect of responsibility and social stigma on empathy for pain. *Journal of Cognitive Neuroscience, 22*(5), 985-997. https://doi.org/10.1162/jocn.2009.21266

Dodell-Feder, D., Koster-Hale, J., Bedny, M., & Saxe, R. (2011). fMRI item analysis in a theory of mind task. *NeuroImage, 55*(2), 705-712. https://doi.org/10.1016/j.neuroimage.2010.12.040

Eddy, C. M., Cavanna, A. E., & Hansen, P. C. (2017). Empathy and aversion: the neural signature of mentalizing in Tourette syndrome. *Psychological Medicine, 47*(3), 507-517. https://doi.org/10.1017/S0033291716002725

Ernst, J., Northoff, G., Böker, H., Seifritz, E., & Grimm, S. (2013). Interoceptive awareness enhances neural activity during empathy. *Human Brain Mapping, 34*(7), 1615-1624. https://doi.org/10.1002/hbm.22014

Fan, Y. T., Chen, C., Chen, S. C., Decety, J., & Cheng, Y. (2013). Empathic arousal and social understanding in individuals with autism: evidence from fMRI and ERP measurements. *Social Cognitive and Affective Neuroscience, 9*(8), 1203-1213. https://doi.org/10.1093/scan/nst101

Focquaert, F., Steven-Wheeler, M.S., Vanneste, S., Doron, K.W., & Platek, S.M. (2010). Mindreading in individuals with an empathizing versus systemizing cognitive style: An fMRI study. *Brain Research Bulletin, 83*(5), 214-222. https://doi.org/10.1016/j.brainresbull.2010.08.008

Fourie, M., Stein, D., Solms, M., Gobodo-Madikizela, P., & Decety, J. (2017). Empathy and moral emotions in post-apartheid South Africa: an fMRI investigation. *Social Cognitive and Affective Neuroscience, 12*(6), 881-892. https://doi.org/10.1093/scan/nsx019

Fourie, M., Stein, D., Solms, M., Gobodo-Madikizela, P., & Decety, J. (2019). Effects of early adversity and social discrimination on empathy for complex mental states: An fMRI investigation. *Scientific Reports, 9*,12959. https://doi.org/10.1038/s41598-019-49298-4

Fujino, J., Yamasaki, N., Miyata, J., Kawada, R., Sasaki, H., Matsukawa, N., ... & Aso, T. (2014). Altered brain response to others′ pain in major depressive disorder. *Journal of Affective Disorders, 165,* 170-175. https://doi.org/10.1016/j.jad.2014.04.058

Fukui, H., Murai, T., Shinozaki, J., Aso, T., Fukuyama, H., Hayashi, T., Hanakawa, T. (2006). The neural basis of social tactics: An fMRI study. *NeuroImage, 32*(2), 913-920. https://doi.org/10.1016/j.neuroimage.2006.03.039

Gallagher, H.L., & Frith, C.D. (2004). Dissociable neural pathways for the perception and recognition of expressive and instrumental gestures. *Neuropsychologia, 42*(13), 1725-1736. https://doi.org/10.1016/j.neuropsychologia.2004.05.006

Gallagher, H.L., Jack, A.I., Roepstorff, A., & Frith, C.D. (2002). Imaging the intentional stance in a competitive game. *NeuroImage, 16*(3), 814-821. https://doi.org/10.1006/nimg.2002.1117

Gao, X., Pan, W., Li, C., Wng, L., Yao, M., & Chen, A. (2017). Long-time exposure to violent video games does not show desensitization on empathy for pain: An fMRI study. *Frontiers in Psychology, 8*, 650. https://doi.org/10.3389/ fpsyg.2017.00650

Gobbini, M. I., Koralek, A., Bryan, R., Montgomery, K., & Haxby, J. (2007). Two takes on the social brain: A comparison of theory of mind tasks. *Journal of Cognitive Neuroscience, 19*(11), 1803-1814. https://doi.org/10.1162/jocn.2007.19.11.1803

Gorno-Tempini, M. L., Pradelli, S., Serafini, M., Pagnoni, G., Baraldi, P., Porro, C., ... & Nichelli, P. (2001). Explicit and incidental facial expression processing: an fMRI study. *NeuroImage, 14*(2), 465-473. https://doi.org/10.1006/nimg.2001.0811

Greven, I., Downing, P., & Ramsey, R. (2016). Linking person perception and person knowledge in the human brain. *Social Cognitive and Affective Neuroscience, 11*(4), 641-651. https://doi.org/10.1093/scan/nsv148 de Greck, M., Shi, Z., Wang, G., Zuo, X., Yang, X., Wang, X., ... & Han, S. (2012a). Culture modulates brain activity during empathy with anger. *NeuroImage, 59*(3), 2871-2882. https://doi.org/10.1016/j.neuroimage.2011.09.052

de Greck, M., Wang, G., Yang, X., Wang, X., Northoff, G., & Han, S. (2012b). Neural substrates underlying intentional empathy. *Social Cognitive and Affective Neuroscience, 7*(2), 135-144. https://doi.org/10.1093/scan/nsq093

Guo, X., Zheng, L., Wang, H., Zhu, L., Li, J., Wang, Q., ... & Yang, Z. (2013). Exposure to violence reduces empathetic responses to other's pain. *Brain and Cognition, 82*(2), 187-191. https://doi.org/10.1016/j.bandc.2013.04.005

Habel, U., Windischberger, C., Derntl, B., Robinson, S., Kryspin-Exner, I., Gur, R., & Moser, E. (2007). Amygdala activation and facial expressions: explicit emotion discrimination versus implicit emotion processing. *Neuropsychologia, 45*(10), 2369-2377. https://doi.org/10.1016/j.neuropsychologia.2007.01.023

Hartwright, C., Apperly, I., & Hansen, P. (2015). The special case of self-perspective inhibition in mental, but not non-mental, representation. *Neuropsychologia, 67*, 183-192. https://doi.org/10.1016/j.neuropsychologia.2014.12.015

Harvey, P. O., Zaki, J., Lee, J., Ochsner, K., & Green, M. F. (2013). Neural substrates of empathic accuracy in people with schizophrenia. *Schizophrenia Bulletin, 39*(3), 617-628. https://doi.org/10.1093/schbul/sbs042

Heatherton, T.F., Wyland, C.L., Macrae, C.N., Demos, K.E., Denny, B.T. & Kelley, W.M. (2006). Medial prefrontal activity differentiates self from close others. *Social Cognitive and Affective Neuroscience, 1*(1), 18-25. https://doi.org/10.1093/scan/nsl001

Heil, L., Colizoli, O., Hartstra, E., Kwisthout, J., van Pelt, S., van Rooij, I., Bekkering, H. (2019). Processing of prediction errors in mentalizing areas. *Journal of Cognitive Neuroscience, 31*(6), 900-912. https://doi.org/10.1162/jocn_a_01381

Heleven, E., van Dun, K., & van Overwalle, F. (2019). The posterior Cerebellum is involved in constructing Social Action Sequences: An fMRI study. *Scientific Reports, 9*, 11110. https://doi.org/10.1038/s41598-019-46962-7

Hennenlotter, A., Schroeder, U., Erhard, P., Castrop, F., Haslinger, B., Stoecker, D., ... & Ceballos-Baumann, A. O. (2005). A common neural basis for receptive and expressive communication of pleasant facial affect. *NeuroImage, 26*(2), 581-591. https://doi.org/10.1016/j.neuroimage.2005.01.057

Hennion, S., Delbeuck, X., Koelkebeck, K., Brion, M., Tyvart, L., Plomhause, L., … Szurhaj, W. (2016). A functional magnetic resonance imaging investigation of theory of mind impairments in patients with temporal lobe epilepsy. Neuropsychologia, 93, 271-279. https://doi.org/10.1016/j.neuropsychologia.2016.11.007

Hooker, C. I., Verosky, S. C., Germine, L. T., Knight, R. T., & D'Esposito, M. (2008). Mentalizing about emotion and its relationship to empathy. *Social Cognitive and Affective Neuroscience, 3*(3), 204-217. https://doi.org/10.1093/scan/nsn019

Hooker, C. I., Verosky, S. C., Germine, L. T., Knight, R. T., & D'Esposito, M. (2010). Neural activity during social signal perception correlates with self-reported empathy. *Brain Research, 1308*, 100-113. https://doi.org/10.1016/j.brainres.2009.10.006

Horan, W., Iacoboni, M., Cross, K., Korb, A., Lee, J., Nori, P., … Green, M. (2014). Self-reported empathy and neural activity during action imitation and observation in schizophrenia. *NeuroImage: Clinical, 5*, 100-108. https://doi.org/10.1016/j.nicl.2014.06.006

Hughes, C., Cassidy, B., Faskowitz, J., Avena-Koenigsberger, A., Sporns, O., & Krendl, A. (2019). Age differences in specific neural connections within the Default Mode Network underlie theory of mind. *NeuroImage, 191*, 269-277. https://doi.org/10.1016/j.neuroimage.2019.02.024

Iidaka, T., Omori, M., Murata, T., Kosaka, H., Yonekura, Y., Okada, T., & Sadato, N. (2001). Neural interaction of the amygdala with the prefrontal and temporal cortices in the processing of facial expressions as revealed by fMRI. *Journal of Cognitive Neuroscience, 13*(8), 1035-1047. https://doi.org/10.1162/089892901753294338

Immordino-Yang, M. H., McColl, A., Damasio, H., & Damasio, A. (2009). Neural correlates of admiration and compassion. *Proceedings of the National Academy of Sciences, 106*(19), 8021-8026. https://doi.org/10.1073/pnas.0810363106

Jack, A. & Pelphrey, K. (2016). Neural correlates of animacy attribution include neocerebellum in healthy adults. *Cerebral Cortex, 25,* 4240-4247. https://doi.org/10.1093/cercor/bhu146

Kana, R.K., Keller, T.A., Cherkassky, V.L., Minshew, N.J., & Just, M.A. (2009). Atypical frontal-posterior synchronization of Theory of Mind regions in autism during mental state attribution. *Social Neuroscience, 4*(2), 135-152. https://doi.org/10.1080/17470910802198510

Kana, R., Libero, L., Hu, C., Deshpande, H., & Colburn, J. (2014). Functional brain networks and white matter underlying theory-of-mind in autism. *Social Cognitive and Affective Neuroscience, 9,* 98-105. https://doi.org/10.1093/scan/nss106

Kanske, P., Böckler, A., Trautwein, F. M., & Singer, T. (2015a,b). Dissecting the social brain: Introducing the EmpaToM to reveal distinct neural networks and brain–behavior relations for empathy and Theory of Mind. *NeuroImage, 122,* 6-19. https://doi.org/10.1016/j.neuroimage.2015.07.082

Kim, Y. T., Lee, J. J., Song, H. J., Kim, J. H., Kwon, D. H., Kim, M. N., ... & Chang, Y. (2010). Alterations in cortical activity of male methamphetamine abusers performing an empathy task: fMRI study. *Human Psychopharmacology: Clinical and Experimental, 25*(1), 63-70. https://doi.org/10.1002/hup.1083

Kircher, T., Bluemel, I., Marjoram, D., Lataster, T., Krabbendam, L., Weber, J., van Os, J., & Krach, S. (2009). Online mentalising investigated with functional MRI. *Neuroscience Letters, 454*, 176-181. https://doi.org/10.1016/j.neulet.2009.03.026

Kliemann, D., Young, L., Scholz, J., & Saxe, R. (2008). The influence of prior record on moral judgment. *Neuropsychologia. 46*(12), 2949-2957. https://doi.org/10.1016/j.neuropsychologia.2008.06.010

Koelkebeck, K., Hirao, K., Kawada, R., Miyata, J., Saze, T., Ubukata, S., … Murai, T. (2011). Transcultural differences in brain activation patterns during theory of mind (ToM) task performance in Japanese and Caucasian participants. *Social Neuroscience 6*(5-6), 615-626. https://doi.org/10.1080/17470919.2011.620763

Krach, S., Bluemel, I, Marjoram, D., Lataster, T., Krabbendam, L., Weber, J., van Os, J., & Kircher, T. (2009). Are women better mindreaders? Sex differences in neural correlates of mentalizing detected with functional MRI. *BMC Neuroscience, 4*, 10:9. https://doi.org/10.1186/1471-2202-10-9

Krach, S., Kamp-Becker, I., Einhäuser, W., Sommer, J., Frässle, S., Jansen, A., … Paulus, F. (2015). Evidence from pupillometry and fMRI indicates reduced neural response during vicarious social pain but not physical pain in autism. *Human Brain Mapping, 36*, 4730-4744. https://doi.org/10.1002/hbm.22949

Kühn, S., Kugler, D., Schmalen, K., Weichenberger, M., Witt, C., & Gallinat, J. (2018). The myth of blunted gamers: No evidence for desensitization in empathy for pain after a violent video game intervention in a longitudinal fMRI study on non-gamers. *Neurosignals, 26*, 22-30. https://doi.org/10.1159/000487217

Kullmann, J., Grigoleit, J.-S., Wolf, O., Engler, H., Oberbeck, R., Elsenbruch, S., … Gizewski, E. (2014). Experimental human endotoxemia enhances brain activity during social cognition. *Social Cognitive and Affective Neuroscience, 9*, 786-793. https://doi.org/10.1093/scan/nst049

Lamm, C., Meltzoff, A. N., & Decety, J. (2010). How do we empathize with someone who is not like us? A functional magnetic resonance imaging study. *Journal of Cognitive Neuroscience*, *22*(2), 362-376. https://doi.org/10.1162/jocn.2009.21186

Lange, K., Williams, L. M., Young, A. W., Bullmore, E. T., Brammer, M. J., Williams, S. C., ... & Phillips, M. L. (2003). Task instructions modulate neural responses to fearful facial expressions. *Biological Psychiatry, 53*(3), 226-232. https://doi.org/10.1016/s0006-3223(02)01455-5

Lee, J., Quintana, J., Nori, P., & Green, M.F. (2011). Theory of mind in schizophrenia: exploring neural mechanisms of belief attribution. *Social Neuroscience, 6*(5-6), 569-581. https://doi.org/10.1080/17470919.2011.620774

Lee, S. & McCarthy, G. (2016). Functional heterogeneity and convergence in the right temporoparietal junction. *Cerebral Cortex, 26,* 1108-1116. https://doi.org/10.1093/cercor/bhu292

Lou, H.C., Luber, B., Crupain, M., Keenan, J.P., Nowak, M., Kjaer, T.W., … Lisanby, S.H. (2004). Parietal cortex and representation of the mental Self. *Proc. Natl. Acad. Sci. U. S. A. 101*(17), 6827-6832. https://doi.org/10.1073/pnas.0400049101

Luo, S., Shi, Z., Yang, X., Wang, X., & Han, S. (2013). Reminders of mortality decrease midcingulate activity in response to others' suffering. *Social Cognitive and Affective Neuroscience, 9*(4), 477-486. https://doi.org/10.1093/scan/nst010

Ma, Y., Wang, C., & Han, S. (2011b). Neural responses to perceived pain in others predict real-life monetary donations in different socioeconomic contexts. *NeuroImage, 57*(3), 1273-1280. https://doi.org/10.1016/j.neuroimage.2011.05.003

Ma, N., Vanderkerckhove, M., van Overwalle, F., Seurinck, R., & Fias, W. (2011a). Spontaneous and intentional trait inferences recruit a common mentalizing network to a different degree: spontaneous inferences activate only its core areas. *Social Neuroscience, 6*(2), 123-138. https://doi.org/10.1080/17470919.2010.485884

Mackes, N., Golm, D., O'Daly, O., Sarkar, S., Sonuga-Barke, E., Fairchield, G., & Mehta, M. (2018). Tracking emotions in the brain – Revisiting the Empathic Accuracy Test. *NeuroImage, 178*, 677-686. https://doi.org/10.1016/j.neuroimage.2018.05.080

Malhi, G.S., Lagopoulos, J., Das, P., Moss, K., Berk, M., & Coulston, C.M. (2008). A functional MRI study of Theory of Mind in euthymic bipolar disorder patients. *Bipolar Disorder*, *10*(8), 943-956. https://doi.org/10.1111/j.1399-5618.2008.00643.x

Mano, Y., Harada, T., Sugiura, M., Saito, D. N., & Sadato, N. (2009). Perspective-taking as part of narrative comprehension: a functional MRI study. *Neuropsychologia*, *47*(3), 813-824. https://doi.org/10.1016/j.neuropsychologia.2008.12.011

Martin, A., & Weisberg, J. (2003). Neural foundations for understanding social and mechanical concepts. *Cognitive Neuropsychology*, *20*(3-6), 575-587. https://doi.org/10.1080/02643290342000005

Mascaro, J., Rilling, J., Tenzin Negi, L., & Raison, C. (2013). Compassion meditation enhances empathic accuracy and related neural activity. *Social Cognitive and Affective Neuroscience, 8*, 48-55. https://doi.org/10.1093/scan/nss095

McAdams, C.J., & Krawczyk, D.C. (2011). Impaired neural processing of social attribution in anorexia nervosa. *Psychiatry Research*, *194*(1), 54-63. https://doi.org/10.1016/j.pscychresns.2011.06.016

Mitchell, J.P. (2008). Activity in right temporo-parietal junction is not selective for theory-of-mind. *Cerebral Cortex. 18*(2), 262-271. https://doi.org/10.1093/cercor/bhm051

Mitchell, J. P., Heatherton, T. F., & Macrae, C. N. (2002). Distinct neural systems subserve person and object knowledge. *Proc. Natl. Acad. Sci. U S A. 99,* 15238-15243. https://doi.org/10.1073/pnas.232395699

Mitchell, J.P., Macrae, C.N., & Banaji, M.R. 2004. Encoding-specific effects of social cognition on the neural correlates of subsequent memory. *Journal of Neuroscience, 24*(21), 4912-4917. https://doi.org/10.1523/JNEUROSCI.0481-04.2004

Mitchell, J.P., Banaji, M.R., & Macrae, C.N. (2005a). General and specific contributions of the medial prefrontal cortex to knowledge about mental states. *NeuroImage, 28*(4), 757-762. https://doi.org/10.1016/j.neuroimage.2005.03.011

Mitchell, J.P., Banaji, M.R., & Macrae, C.N. (2005b). The link between social cognition and self-referential thought in the medial prefrontal cortex. *Journal of Cognitive Neuroscience, 17*(8), 1306-1315. https://doi.org/10.1162/0898929055002418

Mitchell, J.P., Cloutier, J., Banaji, M.R., & Macrae, C.N. (2006). Medial prefrontal dissociations during processing of trait diagnostic and nondiagnostic person information. *Social Cognitive and Affective Neuroscience, 1*(1), 49-55. https://doi.org/10.1093/scan/nsl007

Mitchell, J.P., Macrae, C.N., & Banaji, M.R. (2005c). Forming impressions of people versus inanimate objects: social-cognitive processing in the medial prefrontal cortex. *NeuroImage, 26*(1), 251-257. https://doi.org/10.1016/j.neuroimage.2005.01.031

Mochizuki, H., Baumgärtner, U., Kamping, S., Ruttorf, M., Schad, L., Flor, H., … Treede, R.-D. (2013). Cortico-subcortical activation patterns for itch and pain imagery. *Pain, 154,* 1989-1998. https://doi.org/10.1016/j.pain.2013.06.007

Modinos, G., Renken, R., Ormel, J., & Aleman, A. (2011). Self-reflection and the psychosis-prone brain: an fMRI study. *Neuropsychology, 25*(3), 295-305. https://doi.org/10.1037/a0021747

Moessnang, C., Schäfer, A., Bilek, E., Roux, P., Otto, K., Baumeister, S., … Tost, H. (2016). Specificity, reliablity and sensitivity of social brain responses during spontaneous mentalizing. *Social Cognitive and Affective Neuroscience, 11*(11), 1687-1697. https://doi.org/10.1093/scan/nsw098

Moor, B.G., Macks, Z.A., Gueroglu, B., Rombouts, S.A., Molen, M.W., & Crone, E.A. (2012). Neurodevelopmental changes of reading the mind in the eyes. *Social Cognitive and Affective Neuroscience, 7(1),* 44-52. https://doi.org/10.1093/scan/nsr020

Moriguchi, Y., Ohnishi, T., Kawachi, T., Mori, T., Hirakata, M., Yamada, M., ... & Komaki, G. (2005). Specific brain activation in Japanese and Caucasian people to fearful faces. *Neuroreport*, *16*(2), 133-136. https://doi.org/10.1097/00001756-200502080-00012

Moriguchi, Y., Ohnishi, T., Lane, R.D., Maeda, M., Mori, T., Nemoto, K., Matsuda, H., & Komaki, G. (2006). Impaired self-awareness and theory of mind: an fMRI study of mentalizing in alexithymia. *NeuroImage, 32*(3), 1472-1482. https://doi.org/10.1016/j.neuroimage.2006.04.186

Morrison, I., Lloyd, D., Di Pellegrino, G., & Roberts, N. (2004). Vicarious responses to pain in anterior cingulate cortex: is empathy a multisensory issue?. *Cognitive, Affective, & Behavioral Neuroscience, 4*(2), 270-278. https://doi.org/10.3758/CABN.4.2.270

Murphy, E.R., Brent, B.K., Benton, M., Pruitt, P., Diwadkar, V., Rajarethinam, R.P., Keshavan, M.S. (2010). Differential processing of metacognitive evaluation and the neural circuitry of the self and others in schizophrenia: a pilot study. *Schizophrenia Research*, *116*(2-3), 252-258. https://doi.org/10.1016/j.schres.2009.11.009

Naughtin, C., Horne, K., Schneider, D., Venini, D., York, A., & Dux, P. (2017). Do implicit and explicit belief processing share neural substrates? *Human Brain Mapping, 38,* 4760-4772. https://doi.org/10.1002/hbm.23700

Nolte, T., Bolling, D., Hudac, C., Fonagy, P., Mayes, L., & Pelphrey, K. (2013). Brain mechanisms underlying the impact of attachment-related stress on social cognition. *Frontiers in Human Neuroscience, 7,* 816. https://doi.org/10.3389/fnhum.2013.00816

Nomi, J. S., Scherfeld, D., Friederichs, S., Schäfer, R., Franz, M., Wittsack, H. J., ... & Seitz, R. J. (2008). On the neural networks of empathy: A principal component analysis of an fMRI study. *Behavioral and Brain Functions, 4*(1), 41. https://doi.org/10.1186/1744-9081-4-41

Ochsner, K.N., Beer, J.S., Robertson, E.R., Cooper, J.C., Gabrieli, J.D., Kihlstrom, J.F., & D'Esposito, M. (2005). The neural correlates of direct and reflected self-knowledge. *NeuroImage*, *28*(4), 797-814. https://doi.org/10.1016/j.neuroimage.2005.06.069

Ochsner, K. N., Zaki, J., Hanelin, J., Ludlow, D. H., Knierim, K., Ramachandran, T., ... & Mackey, S. C. (2008). Your pain or mine? Common and distinct neural systems supporting the perception of pain in self and other. *Social Cognitive and Affective Neuroscience, 3*(2), 144-160. https://doi.org/10.1093/scan/nsn006

Özdem, C., Brass, M., van der Cruyssen, L., & van Overwalle, F. (2017). The overlap between false belief and spatial reorientation in the temporo-parietal junction: The role of input modality and task. *Social Neuroscience, 12*(2). 207-217. https://doi.org/10.1080/17470919.2016.1143027

Oliver, L., Vieira, J., Neufeld, R., Dziobek, I., & Mitchell, D. (2018). Greater involvement of action simulation mechanisms in emotional vs cognitive empathy. *Social Cognitive and Affective Neuroscience*, *13(*4), 367-380. https://doi.org/10.1093/scan/nsy013

Olsson, A., Nearing, K. I., & Phelps, E. A. (2007). Learning fears by observing others: the neural systems of social fear transmission. *Social Cognitive and Affective Neuroscience, 2*(1), 3-11. https://doi.org/10.1093/scan/nsm005

Otti, A., Wohlschlaeger, A., & Noll-Hussong, M. (2015). Is the medial prefrontal cortex necessary for theory of mind*? PLoS ONE, 10*(8), e0135912. https://doi.org/10.1371/journal.pone.0135912

Pehrs, C., Zaki, J., Schlochtermeier, L., Jacobs, A., Kuchinke, L., & Koelsch, S. (2017). The temporal pole top-down modulates the ventral visual stream during social cognition. *Cerebral Cortex, 27*, 777-792. https://doi.org/10.1093/cercor/bhv226

Perner, J., Aichhorn, M., Kronbichler, M., Staffen, W., & Ladurner, G. (2006). Thinking of mental and other representations: The roles of left and right temporo-parietal junction. *Social Neuroscience, 1*(3-4), 245-258. https://doi.org/10.1080/17470910600989896

Phillips, M. L., Young, A. W., Senior, C., Brammer, M., Andrew, C., Calder, A. J., ... & Gray, J. A. (1997). A specific neural substrate for perceiving facial expressions of disgust. N*ature, 389*(6650), 495. /https://doi.org/10.1038/39051

Platek, S.M., Keenan, J.P., Gallup, G.G. Jr., & Mohamed, F.B. (2004). Where am I? The neurological correlates of self and other. *Brain Research. Cognitive Brain Research, 19*(2), 114-122. https://doi.org/10.1016/j.cogbrainres.2003.11.014

Powell, J., Grossi, D., Corcoran, R., Gobet, F., & García-Fiñana, M. (2017). The neural correlates of theory of mind and their role during empathy and game of chess: A functional magnetic resonance imaging study. *Neuroscience, 355,* 149-160. https://doi.org/10.1016/j.neuroscience.2017.04.042

Preis, M. A., Schmidt-Samoa, C., Dechent, P., & Kroener-Herwig, B. (2013). The effects of prior pain experience on neural correlates of empathy for pain: An fMRI study. *Pain, 154*(3), 411-418. https://doi.org/10.1016/j.pain.2012.11.014

Quaio-Tasserit, E., Corradi-Dell'Acqua, C., & Vuilleumier, P. (2018). The good, the bad, and the suffering. Transient emotional episodes modulate the neural circuits of pain and empathy. *Neuropsychologia, 116*(A), 99-116. https://doi.org/10.1016/j.neuropsychologia.2017.12.027.

Rameson, L. T., Morelli, S. A., & Lieberman, M. D. (2012). The neural correlates of empathy: experience, automaticity, and prosocial behavior. *Journal of Cognitive Neuroscience*, *24*(1), 235-245. https://doi.org/10.1162/jocn_a_00130

Regenbogen, C., Schneider, D. A., Gur, R. E., Schneider, F., Habel, U., & Kellermann, T. (2012). Multimodal human communication—targeting facial expressions, speech content and prosody. *NeuroImage, 60*(4), 2346-2356. https://doi.org/10.1016/j.neuroimage.2012.02.043

Reniers, R. L., Völlm, B. A., Elliott, R., & Corcoran, R. (2014). Empathy, ToM, and self–other differentiation: An fMRI study of internal states. *Social Neuroscience, 9*(1), 50-62. https://doi.org/100.1080/17470919.2013.861360

Reyes-Aguilar, A., Fernandez-Ruiz, J., Pasaye, E., & Barrios, F. (2017). Executive mechanisms for thinking about negative situations in both cooperative and non-cooperative contexts. *Frontiers in Human Neuroscience*, *11*, 275. https://doi.org/10.3389/fnhum.2017.00275

Riedl, R., Mohr, P., Kenning, P., Davis, F., & Heekeren, H. (2014). Trusting humans and avatars: A brain imaging study based on evolution theory. *Journal of Management Information Systems, 30*(4), 83-114. https://doi.org/10.2753/MIS0742-1222300404

Rilling, J.K., Sanfey, A.G., Aronson, J.A., Nystrom, L.E., & Cohen, J.D. (2004). The neural correlates of theory of mind within interpersonal interactions. *NeuroImage, 22*(4), 1694-1703. https://doi.org/10.1016/j.neuroimage.2004.04.015

Roser, P., Lissek, S., Tegenthoff, M., Nicolas, V., Juckel, G., & Brüne, M. (2012). Alterations of theory of mind network activation in chronic cannabis users. *Schizophrenia Research, 139*(1-3), 19-26. https://doi.org/10.1016/j.schres.2012.05.020

Ross, L.A., Olson, I.R., 2010. Social cognition and the anterior temporal lobes. *NeuroImage, 49*, 3452-3462. https://doi.org/10.1016/j.neuroimage.2009.11.012

Russell, T.A., Rubia, K., Bullmore, E.T., Soni, W., Suckling, J., Brammer, M.J., Simmons, A., Williams, S.C., & Sharma, T. (2000). Exploring the social brain in schizophrenia: left prefrontal underactivation during mental state attribution. *American Journal of Psychiatry, 157*(12), 2040-2042. https://doi.org/10.1176/appi.ajp.157.12.2040

Santos, N.S., Kuzmanovic, B., David, N., Rotarska-Jagiela, A., Eickhoff, S.B., Shah, J.N., Fink, G.R., Bente, G., & Vogeley, K. (2010). Animated brain: a functional neuroimaging study on animacy experience. *NeuroImage, 53*(1), 291-302. https://doi.org/10.1016/j.neuroimage.2010.05.080

Saxe, R., Kanwisher, N., (2003). People thinking about thinking people. The role of the temporo-parietal junction in 'theory of mind'. *NeuroImage, 19*(4), 1835-42. https://doi.org/10.1016/s1053-8119(03)00230-1

Saxe, R., & Powell, L., 2006. It's the thought that counts: specific brain regions for one component of theory of mind. *Psychological Science, 17*(8), 692-9. https://doi.org/10.1111/j.1467-9280.2006.01768.x

Saxe, R., Schulz, L.E., Jiang, Y.V. (2006). Reading minds versus following rules: dissociating theory of mind and executive control in the brain. *Social Neuroscience, 1*(3-4), 284-298. https://doi.org/10.1080/17470910601000446

Saxe, R., & Wexler, A. (2005). Making sense of another mind: the role of the right temporo-parietal junction. *Neuropsychologia, 43*(10), 1391-1399. https://doi.org/10.1016/j.neuropsychologia.2005.02.013

Schiffer, B., Pawliczek, C., Müller, B., Gizewski, E., & Walter, H. (2013). Why don't men understand women? Altered neural networks for reading the language of male and female eyes. *PLoS ONE, 8*(4), e60278. https://doi.org/10.1371/journal.pone.0060278

Schlaffke, L., Lissek, S., Lenz, M., Juckel, G., Schultz, T., Tegenthoff, M., ... & Brüne, M. (2015). Shared and nonshared neural networks of cognitive and affective theory-of-mind: A neuroimaging study using cartoon picture stories. *Human Brain Mapping, 36*(1), 29-39. https://doi.org/10.1002/hbm.22610

Schmitz, T.W., Kawahara-Baccus, T.N., & Johnson, S.C. (2004). Metacognitive evaluation, self-relevance, and the right prefrontal cortex. *NeuroImage, 22*(2), 941-947. https://doi.org/10.1016/j.neuroimage.2004.02.018

Schmitgen, M., Walter, H., Drost, S., Rückl, S., & Schnell, K. (2016). Stimulus-dependent amygdala involvement in affective theory of mind generation. *NeuroImage, 129,* 450-459. https://doi.org/10.1016/j.neuroimage.2016.01.029

Schneider, K., Regenbogen, C., Pauly, K. D., Gossen, A., Schneider, D. A., Mevissen, L., ... & Schneider, F. (2013). Evidence for gender-specific endophenotypes in high-functioning autism spectrum disorder during empathy. *Autism Research, 6*(6), 506-521. https://doi.org/10.1002/aur.1310

Schnell, K., Bluschke, S., Konradt, B., & Walter, H. (2011). Functional relations of empathy and mentalizing: an fMRI study on the neural basis of cognitive empathy. *NeuroImage, 54*(2), 1743-1754. https://doi.org/10.1016/j.neuroimage.2010.08.024

Schlaffke, L., Lissek, S., Lenz, M., Juckel, G., Schultz, T., Tegenthoff, M., ... & Brüne, M. (2015). Shared and nonshared neural networks of cognitive and affective theory-of-mind: A neuroimaging study using cartoon picture stories. *Human Brain Mapping, 36*(1), 29-39. https://doi.org/10.1002/hbm.22610

Schroeder, U., Hennenlotter, A., Erhard, P., Haslinger, B., Stahl, R., Lange, K., & Ceballos-Baumann, A. (2004). Functional neuroanatomy of perceiving surprised faces. H*uman Brain Mapping, 23*(4), 181-187. https://doi.org/10.1002/hbm.20057

Schulte-Rüther, M., Greimel, E., Markowitsch, H. J., Kamp-Becker, I., Remschmidt, H., Fink, G. R., & Piefke, M. (2011). Dysfunctions in brain networks supporting empathy: an fMRI study in adults with autism spectrum disorders. *Social Neuroscience*, *6*(1), 1-21. https://doi.org/10.1080/17470911003708032

Schulte-Rüther, M., Greimel, E., Piefke, M., Kamp-Becker, I., Remschmidt, H., Fink, G. R., ... & Konrad, K. (2014). Age-dependent changes in the neural substrates of empathy in autism spectrum disorder. *Social Cognitive and Affective Neuroscience*, *9*(8), 1118-1126. https://doi.org/10.1093/scan/nst088

Schultz, J., Imamizu, H., Kawato, M., & Frith, C.D. (2004). Activation of the human superior temporal gyrus during observation of goal attribution by intentional objects. *Journal of Cognitive Neuroscience*, *16*(10), 1695-1705. https://doi.org/10.1162/0898929042947874

Seara-Cardoso, A., Viding, E., Lickley, R., & Sebastian, C. (2015). Neural responses to others' pain vary with psychopathic traits in healthy adult males. *Cognitive, Affective, and Behavioral Neuroscience, 15*, 578-588. https://doi.org/10.3758/s13415-015-0346-7

Sebastian, C. L., Fontaine, N. M., Bird, G., Blakemore, S. J., De Brito, S. A., McCrory, E. J., & Viding, E. (2012a,b). Neural processing associated with cognitive and affective Theory of Mind in adolescents and adults. *Social Cognitive and Affective Neuroscience, 7*(1), 53-63. https://doi.org/10.1093/scan/nsr023

Sprengelmeyer, R., Rausch, M., Eysel, U., & Przuntek, H. (1998). Neural structures associated with recognition of facial expressions of basic emotions. *Proceedings of the Royal Society B: Biological Sciences, 265,* 1927-1931. https://doi.org/10.1098/rspb.1998.0522

Sripada, C.S., Angstadt, M., Banks, S., Nathan, P.J., Liberzon, I., & Phan, K.L. (2009). Functional neuroimaging of mentalizing during the trust game in social anxiety disorder. *Neuroreport. 20*(11), 984-989. https://doi.org/10.1097/WNR.0b013e32832d0a67

Surguladze, S. A., Brammer, M. J., Young, A. W., Andrew, C., Travis, M. J., Williams, S. C., & Phillips, M. L. (2003). A preferential increase in the extrastriate response to signals of danger. *NeuroImage*, *19*(4), 1317-1328. https://doi.org/10.1016/s1053-8119(03)00085-5

Surguladze, S., Elkin, A., Ecker, C., Kalidindi, S., Corsico, A., Giampetro, V., … Philips, M. (2008). Genetic variation in the serotonin transporter modulates neural system-wide response to fearful faces. *Genes, Brain and Behavior, 7*, 543-551. https://doi.org/10.1111/j.1601-183X.2008.00390.x

Suzuki, S., Niki, K., Fujisaki, S., & Akiyama, E. (2011). Neural basis of conditional cooperation. *Social Cognitive and Affective Neuroscience, 6*(3), 338-347. https://doi.org/10.1093/scan/nsq042

Takahashi, H., Izuma, K., Matsumoto, M., Matsumoto, K., & Omori, T. (2015). The anterior insula tracks behavioral entropy during an interpersonal competitive game. *PLoS ONE, 10*(6), e0123329. https://doi.org/10.1371/journal.pone.0123329

Takahashi, H., Terada, K., Morita, T., Suzuki, S., Haji, T., Kozima, H., … Naito, E. (2014). Different impressions of other agents obtained through social interaction uniquely modulate dorsal and ventral pathway activities in the social human brain. *Cortex, 58,* 289-300. https://doi.org/10.1016/j.cortex.2014.03.011

Tavares, P., Lawrence, A.D., Barnard, P.J. 2008. Paying attention to social meaning: an FMRI study. *Cerebral Cortex, 18*(8), 1876-1885. https://doi.org/10.1093/cercor/bhm212

Tei, S., Becker, C., Kawada, R., Fujino, J., Jankowski, K. F., Sugihara, G., ... & Takahashi, H. (2014). Can we predict burnout severity from empathy-related brain activity?. *Translational Psychiatry, 4*(6), e393. https://doi.org/10.1038/tp.2014.34

Thye, M., Murdaugh, D., & Kana, R. (2018a,b). Brain mechanisms underlying reading the mind from eyes, voices, and actions. *Neuroscience, 374*, 172-186. https://doi.org/10.1016/j.neuroscience.2018.01.045

Todorov, A., Gobbini, M.I., Evans, K.K., & Haxby, J.V. (2007). Spontaneous retrieval of affective person knowledge in face perception. *Neuropsychologia, 45*(1), 163-173. https://doi.org/10.1016/j.neuropsychologia.2006.04.018

Toller, G., Adhimoolam, B., Grunwald, T., Huppert, H.-J., Kurthen, M., Rankin, K., & Jokeit, H. (2015). Right mesial temporal lobe epilepsy impairs empathy-related brain responses to dynamic fearful faces. *Journal of Neurology, 262,* 729-741. https://doi.org/10.1007/s00415-014-7622-2

Ushida, T., Ikemoto, T., Tanaka, S., Shinozaki, J., Taniguchi, S., Murata, Y., ... & Tamura, Y. (2008). Virtual needle pain stimuli activates cortical representation of emotions in normal volunteers. *Neuroscience Letters, 439*(1), 7-12. https://doi.org/10.1016/j.neulet.2008.04.085

Van der Cruyssen, L., Heleven, E., Ma, N., Vandekerckhove, M., & Van Overwalle, F. (2015). Distinct neural correlates of social categories and personality traits. *NeuroImage, 104*, 336-346. https://doi.org/10.1016/j.neuroimage.2014.09.022

Van der Gaag, C., Minderaa, R. B., & Keysers, C. (2007). Facial expressions: what the mirror neuron system can and cannot tell us. *Social Neuroscience, 2*(3-4), 179-222. https://doi.org/10.1080/17470910701376878

van der Heiden, L., Scherpiet, S., Konicar, L., Birbaumer, N., & Veit, R. (2013). Inter-individual differences in successful perspective taking during pain perception mediates emotional responsiveness in self and others: an fMRI study. *NeuroImage, 65*, 387-394. https://doi.org/10.1016/j.neuroimage.2012.10.003

Völlm, B. A., Taylor, A. N., Richardson, P., Corcoran, R., Stirling, J., McKie, S., ... & Elliott, R. (2006a,b). Neuronal correlates of theory of mind and empathy: a functional magnetic resonance imaging study in a nonverbal task. *NeuroImage*, *29*(1), 90-98. https://doi.org/10.1016/j.neuroimage.2005.07.022

Vuilleumier, P., Armony, J. L., Driver, J., & Dolan, R. J. (2001). Effects of attention and emotion on face processing in the human brain: an event-related fMRI study. *Neuron, 30*(3), 829-841. https://doi.org/10.1016/s0896-6273(01)00328-2

Vuilleumier, P., Richardson, M. P., Armony, J. L., Driver, J., & Dolan, R. J. (2004). Distant influences of amygdala lesion on visual cortical activation during emotional face processing. *Nature Neuroscience, 7*(11), 1271. https://doi.org/10.1038/nn1341

Walter, H.W., Adenzato, M., Ciaramidaro, A., Enrici, I., Lorenzo, P., & Bara, B.G. (2004a,b). Understanding intentions in social interaction: the role of the anterior paracingulate cortex. *Journal of Cognitive Neuroscience, 16*(10), 1854-1863. https://doi.org/10.1162/0898929042947838

Walter, H., Ciaramidaro, A., Adenzato, M., Vasic, N., Ardito, R.B., Erk, S., & Bara, B.G. (2009). Dysfunction of the social brain in schizophrenia is modulated by intention type: an fMRI study. *Social Cognitive and Affective Neuroscience, 4*(2), 166-176. https://doi.org/10.1093/scan/nsn047

Wang, Y., Liu, W.-H., Li, Z., Wei, X.-H., Jiang, X.-Q., Neumann, D., Chan, R. (2015). Dimensional schizotypy and social cognition: an fMRI imaging study. *Frontiers in Behavioral Neuroscience, 9,* 133. https://doi.org/10.3389/fnbeh.2015.00133

Wicker, B., Keysers, C., Plailly, J., Royet, J. P., Gallese, V., & Rizzolatti, G. (2003). Both of us disgusted in My insula: the common neural basis of seeing and feeling disgust. *Neuron, 40*(3), 655-664. https://doi.org/10.1016/S0896-6273(03)00679-2

Wicker, B., Perrett, D.I., Baron-Cohen, S., & Decety, J. (2003). Being the target of another's emotion: a PET study. *Neuropsychologia, 41*(2), 139-146. https://doi.org/10.1016/s0028-3932(02)00144-6

Willert, A., Mohnke, S., Erk, S., Schnell, K., Romanczuk-Seiferth, N., Quinlivan, E., … Walter, A. (2015). Alterations in neural theory of mind processing in euthymic patients with bipolar disorder and unaffected relatives. *Bipolar Disorders, 17*, 880-891. https://doi.org/10.1111/bdi.12352

Williams, L., Lidell, B., Kemp, A., Bryant, R., Meares, R., Peduto, A., & Gordon, E. (2006). Amygdala-prefrontal dissociation of subliminal and supraliminal fear. *Human Brain Mapping, 27*, 652-661. https://doi.org/10.1002/hbm.20208

Wolfensberger, S., Veltman, D., Hoogendijk, W., Boomsma, D., & de Geus, E. (2008). Amygdala responses to emotional faces in twins discordant or concordant for the risk for anxiety and depression. *NeuroImage, 41*, 544-552. https://doi.org/10.1016/j.neuroimage.2008.01.053

Young, L., Cushman, F., Hauser, M., & Saxe, R. (2007a,b). The neural basis of the interaction between theory of mind and moral judgment. *Proc. Natl. Acad. Sci. U S A. 104*(20), 8235-8240. https://doi.org/10.1073/pnas.0701408104

Young, L., & Saxe, R. 2009. An FMRI investigation of spontaneous mental state inference for moral judgment. *Journal of Cognitive Neuroscience, 21*(7), 1396-1405. https://doi.org/10.1162/jocn.2009.21137

Young, L., Dodell-Feder, D., Saxe, R. (2010). What gets the attention of the temporo-parietal junction? An fMRI investigation of attention and theory of mind. *Neuropsychologia. 48*(9), 2658-2664. https://doi.org/10.1016/j.neuropsychologia.2010.05.012

Young, L., Scholz, J., & Saxe, R. (2011). Neural evidence for 'intuitive prosecution': the use of mental state information for negative moral verdicts. *Social Neuroscience, 6*(3), 302-315. https://doi.org/10.1080/17470919.2010.529712

Zaki, J., Ochsner, K. N., Hanelin, J., Wager, T. D., & Mackey, S. C. (2007). Different circuits for different pain: patterns of functional connectivity reveal distinct networks for processing pain in self and others. *Social Neuroscience, 2*(3-4), 276-291. https://doi.org/10.1080/17470910701401973

Zaki, J., Weber, J., Bolger, N., & Ochsner, K. (2009). The neural bases of empathic accuracy. *Proceedings of the National Academy of Sciences, 106*(27), 11382-11387. https://doi.org/10.1073/pnas.0902666106

Zaki, J., Weber, J., & Ochsner, K. (2012). Task-dependent neural bases of perceiving emotionally expressive targets. *Frontiers in Human Neuroscience, 6*, 228. https://doi.org/10.3389/fnhum.2012.00228

Zhou, Y., Wang, Y., Rao, L.-L., Yang, L.-Q, & Li, S. (2014). Money talks: neural substrate of modulation of fairness by monetary incentives. *Frontiers in Behavioral Neuroscience, 8*,150. https://doi.org/10.3389/fnbeh.2014.00150

Zhu, Y., Zhang, L., Fan, J., & Han, S. (2007). Neural basis of cultural influence on self-representation. *NeuroImage, 34*(3), 1310-1316. https://doi.org/10.1016/j.neuroimage.2006.08.047

Figure Captions

Figure 1. Results of separate meta-analyses for task groups from Theory of Mind. Maps were thresholded at voxel-wise threshold of $p < .005$ uncorrected and a cluster extent threshold of 10 voxels.

Figure 2. Results of separate meta-analyses for task groups for empathy. Maps were thresholded at voxel-wise threshold of $p < .005$ uncorrected and a cluster extent threshold of 10 voxels. For Reasoning about Emotions, we selected a common type of task that we found not only in the empathy literature, but also in studies on ToM.

Figure 3. (A) Evaluation and comparison of different clustering solutions. Plot shows changes in two metrics when moving from 2 to 3, 3 to 4, …, and 10 to 11 cluster solutions. The first metric, shown in red and on the left Y-axis, gives the relative difference in cophenetic distances when moving from one model to the next. A relatively high difference in cophenetic distances indexes that introducing new sub-clusters results in better separation of brain activity patterns, and thus, good clustering. The second metric, shown in blue, gives the density of task separation, reflecting whether separating clusters into sub-clusters maintains a balanced number of task groups in each sub-cluster. Relatively low density of task separation indexes good clustering. Preferred clusterings in terms of both metrics are indicated on the X-axis: 3, 8, and 11. Metric changes are shown for clustering of complete meta-analyses (main analysis), and clusterings based on leave-one-out jackknife sensitivity analysis with 5000 repeats (jackknife mean, standard deviation). (A) Stability of the assignment of task groups (i.e., associated meta-analytic maps) to clusters for the three cluster solution, based on leave-one-out jackknife sensitivity analysis. Bars indicate percentage of jackknife-repeats for which task groups were assigned to the same clusters as in the main analysis.

Figure 4. (A) Degree of dissimilarity between meta-analytic result maps (given as 1-Pearson's r). High dissimilarity reflects little correspondence or overlap between brain maps. (B) Dendrogram from hierarchical clustering. Height of the branches indexes cophenetic distances, and thus the dissimilarity between sub-clusters at a model order. Color bars indicate positions of 3- and 8-cluster solutions with respect to the dendrogram.

Figure 5. Mean brain activation for clusters at different model orders. We carried out pooled meta-analyses, i.e. one separate meta-analysis per cluster, where all its task groups are joined together. Analyses were thresholded at voxel-wise threshold of $p < .005$ uncorrected and a cluster extent threshold of 10 voxels. The 1-cluster solution is shown for illustrative purposes only, and was not evaluated or compared against other clusterings. Colors indicate how the 3 cluster solution relates to both higher- and lower-level clusterings (blue - cognitive cluster, green - intermediate cluster, red - affective cluster). At the lowest-level of the dendrogram, we indicate for each cluster some exemplary stimulus- and task-categorizations. The 8 cluster solution was selected as a representative low-level clustering. However, note that the 11 cluster solution shows favorable clustering metrics, but corresponds to what has been shown in Figures 1 and 2 (i.e. complete separation into individual task groups).

Figure 6. Mean brain activation from meta-analyses for the three cluster solution (pooled meta-analyses), at a voxel-wise threshold of $p < .005$ uncorrected and cluster extent threshold of 10 voxels. Polar plots show results from large-scale meta-analytic decoding (neurosynth.org; Yarkoni et al., 2011). Meta-analysis maps were compared to an extensive set of automatically generated meta-analyses across a wide range of topics ('terms'), and findings of highest convergence (image correlation, i.e. Pearson's r) are shown. For each of

our three clusters, we show the 5 most strongly correlated terms found for social and non-social topics (for further details, see methods section 2.4).

Figure 7. Meta-analytic contrasts for the three clusters, tailored to find specific commonalities between the intermediate and each of the other two clusters. Maps are shown at a voxel-wise threshold of $p < .005$ uncorrected and cluster extent threshold of 10 voxels. Meta-analytic decoding (neurosynth.org; Yarkoni et al., 2011) results are given in polar plots, expressed as image correlation, i.e. Pearson's r. For each contrast, we show the 5 most strongly correlated terms found for social and non-social topics. As we did not find any specific commonalities among the cognitive and affective cluster (at the statistical threshold), no decoding was done for this contrast.

Figure 8. (A) Whole brain map of connectivity organization, capturing continuous changes (i.e. a principal gradient) of brain connectivity patterns across the cortex. Gradient ranges from sensory and unimodal areas in sensorimotor, auditory and visual cortex (green-blue), to increasingly abstract and transmodal areas (red-yellow) which largely correspond to the brain's default mode network. This gradient was determined by analysis of resting state connectivity data from a large sample of healthy adults (Human Connectome Project, n=820). (B) Percentiles along the gradient were split into 20 distinct maps. For each map, we calculated the overlap with our meta-analyses. Bar-charts give percent overlap between percentile maps and meta-analyses, i.e. how many percent of voxels of a meta-analysis map overlapped with each percentile map (normalizing results across differently sized meta-analysis maps).

Figure 9. Overlaps between our three cluster solution and meta-analyses on related topics of social cognition. For each topic map, we indicate the percentage of voxels falling into each of the three meta-analytic clusters. Note that percentage values are relative to the size of each topic-related map (e.g., maps containing extensive areas of activation are more likely to feature regions outside of our three clusters). Therefore, the relative patterns of overlaps within each topic-map are of main interest. See section "4. Relation to other meta-analyses on social cognition" for more details. Footnotes: [1] For exploratory purposes we show an uncorrected map of the meta-analysis results by Darda & Ramsey (2018). [2] Map from Boccadoro et al. (2019) shows results from a multi-study analysis on spontaneous ToM (not a meta-analysis, based on original data), 3 studies, n=68. [3] Map from Cheong et al. (2017) shows most consistent overlaps across 5 published studies, summarized based on expanding reported coordinates into 15 mm radius spheres.

Table 1.
Examples for studies in Theory of Mind task groups. For a complete list, see Supplementary Materials.

Author	Experimental Condition	Control Condition
False Belief (25 studies)		
Oliver et al., 2018	Read a short vignette involving a person with a false belief. Predict the behavior of this person based on her belief. Stimuli adapted from Dodell-Feder et al. (2011). e.g. *'The morning of high school dance Sarah placed her high heel shoes under her dress and then went shopping. That afternoon, her sister borrowed the shoes and later put them under Sarah's bed. Sarah gets ready assuming her shoes are under the dress.' (Yes or No).*	Read a short vignette involving a photograph/physical representation of the past, and a description how things depicted have changed by now. Answer a question about the outdated scene shown. e.g. *'Old maps of the islands near Titan are displayed in the Maritime museum. Erosion has since taken its toll, leaving only the three largest islands. Near Titan today there are many islands.' (Yes or No).*
Saxe & K., 2003	Read a short vignette involving a person with a false belief. Answer a question about her belief. e.g. *'John told Emily that he had a Porsche. Actually, his car is a Ford. Emily doesn't know anything about cars so she believed John. When Emily sees John's car, she thinks it is a …?' (Porsche or Ford).*	Read a false-photograph vignette. Answer a question concerning the outdated content in the photo. e.g. *'A photograph was taken of an apple hanging on a tree branch. The film took half an hour to develop. In the meantime, a strong wind blew the apple to the ground. The developed photograph shows the apple on the …?' (tree or ground).*
Trait Judgements (19 studies)		
Ma et al., 2011	Read written statements conveying trait diagnostic information about persons (describing behavior). Then read a single trait-adjective and indicate whether it is consistent with the behavior of that person. e.g. *'Tolvan gave her sister a hug ... consistent with "friendly"?'*	Read written statement about a person doing something. This behavior is neutral and does not convey trait diagnostic information about the person. Indicate the gender of the person in the sentence. e.g. *'Tolvan gave her mother a bottle … is Tolvan male or female?'*
Zhu et al., 2007	Read a personality trait adjective (e.g., *'brave', 'childish')* and indicate if it correctly describes a former American president (*Bill Clinton*, press *'yes'* or *'no'*).	Read a personality trait adjective (e.g., *'brave', 'childish')* and indicate if it is written in lower- or upper-case (press *'yes'* or *'no'*).
Strategic Games (13 studies)		
Takahashi et al., 2015	Play the matching pennies game against a human. Both players are asked to choose one out of two options at the same time. For one player, the goal is to choose the same options as the other. For the other player, goal is to choose the different option.	Play the matching pennies game against a computer. Both players are asked to choose one out of two options at the same time. For one player, the goal is to choose the same options as the other. For the other player, goal is to choose the different option.
Kircher et al., 2009	Play the prisoner's dilemma game (iterated version). You play with a human player for game points. Both players choose a cooperative or defective strategy on each trial. If both players choose defective, they gain almost no game points at all. If both choose cooperative, both gain some game points. If players choose differently, the defective player gains more points.	Play the prisoner's dilemma game (iterated version). You play with a computer for game points.
Social Animations (20 studies)		
Moessnang et al., 2016	Watch a video animation of two interacting triangles, which involve influence on each other's mental states (e.g. *coaxing*). Indicate if a social/goal-directed/random movement was shown, and indicate the feeling of both triangles (*positive/negative*). Respond via button press to both questions.	Watch video animation of two triangles, which interact in a goal-directed manner (e.g. *one chasing the other*). Indicate if a social/goal-directed/random movement was shown, and indicate the feeling of both triangles (*positive/negative*). Respond via button press to both questions.
Castelli et al., 2000	Watch video animation of two interacting triangles (e.g. '*mother and child are playing*'). Explain verbally what was happing (after fMRI).	Watch video animation of two randomly moving triangles. Explain verbally what was happing (after fMRI).
Reading Mind in the Eyes (15 studies)		
Baron-Cohen et al., 1999	View a photograph showing the eye-region of a face. Indicate which of two words (e.g., *concerned* versus *unconcerned*) describes the mental state of that person (button press).	View a photograph showing the eye-region of a face. Indicate if the person is male or female (button press).
Bos et al., 2016	View a photograph showing the eye-region of a face. Indicate if a mental state word presented before (e.g., *shy*, *hostile*, *playful*) matches the photo.	View a photograph showing the eye-region of a face. Indicate if a non-mental state word presented before (e.g., *woman*, *curly-hair*, *heavy eye-brows*) matches the photo.
Rational Actions (11 studies)		
Brunet et al., 2000	View a cartoon story and predict what will happen based on intentions of a character (no false belief). Choose a logical story ending from several options shown in pictures. e.g. '*A prisoner is in his cell. First, he breaks the bars of his prison window. Then he walks to his bed.'* Participants must indicate what will happen next … *'the prisoner ties a rope from the sheets on his bed /the prisoner shouts out loud.'*	View cartoon stories and predict what will happen (press button) based on physical causality. e.g. *A cartoon story shows a person standing in front of a slide. A large ball is coming down the slide, heading towards the person standing there. Participants must indicate what will happen next; pictures show the ball knocking over the person or the ball resting on the ground and the person standing next to it.*
Heleven et al., 2019	View cartoon stories showing a person over a sequence of events. Order of pictures (i.e. events) is jumbled. Indicate the correct order of pictures based on the intentions of the character (button press).	View cartoon stories showing a sequence of events. Order of pictures (i.e. events) is jumbled. Indicate the correct order of pictures based on physical causality (button press).

Table 2.
Examples for studies in empathy task groups. For a complete list, see Supplementary Materials.

Author	Experimental Condition	Control Condition
Observing Pain (21 studies)		
Olsson et al., 2007	Watch a video of a model undergoing fear conditioning situations (receiving the shock). No response.	Watch a video of a model undergoing fear conditioning situations (not receiving the shock). No response.
Bos et al., 2015	Watch videos of hands under painful (needle) stimulation. Watch attentively and detect changes in fixation cross appearing between videos.	Watch videos of hands under non-painful (cotton swab) stimulation. Watch attentively and detect changes in fixation cross appearing between videos.
Observing Emotions (25 studies)		
Critchley et al., 2000	View a picture of a face with an emotional expression (anger, fear). Judge the gender of the person (button press).	View a picture of a face with a neutral expressions. Judge the gender of the person (button press).
Toller et al., 2015	View video of face showing fearful expression. Relax and focus on the actors eyes.	View video of landscapes. Relax and do nothing.
Sharing Emotions or Pain (12 studies)		
Preis et al., 2013	View a picture of a body part (hand) under painful stimulation. Imagine how the person in the picture feels and rate her pain (VAS).	View a picture of a body part (hand) under non-painful stimulation. Imagine how the person in the picture feels and rate her pain (VAS).
Reniers et al., 2014	View a picture of a face with a sad expression. Imagine what the person in the picture is feeling. No response in the scanner.	View a picture of a face with a neutral expression. Imagine what the person in the picture is feeling. No response in the scanner.
Evaluating Situated Emotions (15 studies)		
Kanske et al., 2015	Watch a video of a person telling about a negative autobiographical event. Rate how you feel and how much compassion you feel with the person in the video (button press.	Watch a video of a person telling about a neutral autobiographical event. Rate how you feel and how much compassion you feel with the person in the video (button press).
Reyes-A. et al., 2017	Read a short vignette about an emotionally negative or positive event, then view a picture of the person to whom this happened. Think about what this person is feeling (no overt response).	Read a short vignette about an emotionally neutral event, then view a picture of the person to whom this happened. Think about what this person is feeling (no overt response).
Reasoning about Emotions (12 studies)		
Hooker et al., 2010	View a series of pictures with two persons. One person realizes that she mistakenly had a false belief regarding an emotion-triggering state of affairs. That leads to a change of her emotions. Indicate if the pictures showed a social (i.e. emotional), physical, or no change (button press).	View a series of pictures with two persons. One person has a false belief regarding an emotion-triggering state of affairs. Nothing changes over the pictures (i.e. so this person does not realize that she had a false belief). Indicate if pictures showed a social, physical, or no change (button press).
Voellm et al., 2006	View a cartoon story showing two persons. One is in an emotion-triggering situation. Predict what the other person will do to make her feel better (button press).	View a cartoon story showing two persons in neutral everyday situation. Predict what will happen next based on physical causality (button press).

VAS... Visual Analog Scale (allows ratings along a continuous scale)

Figure1

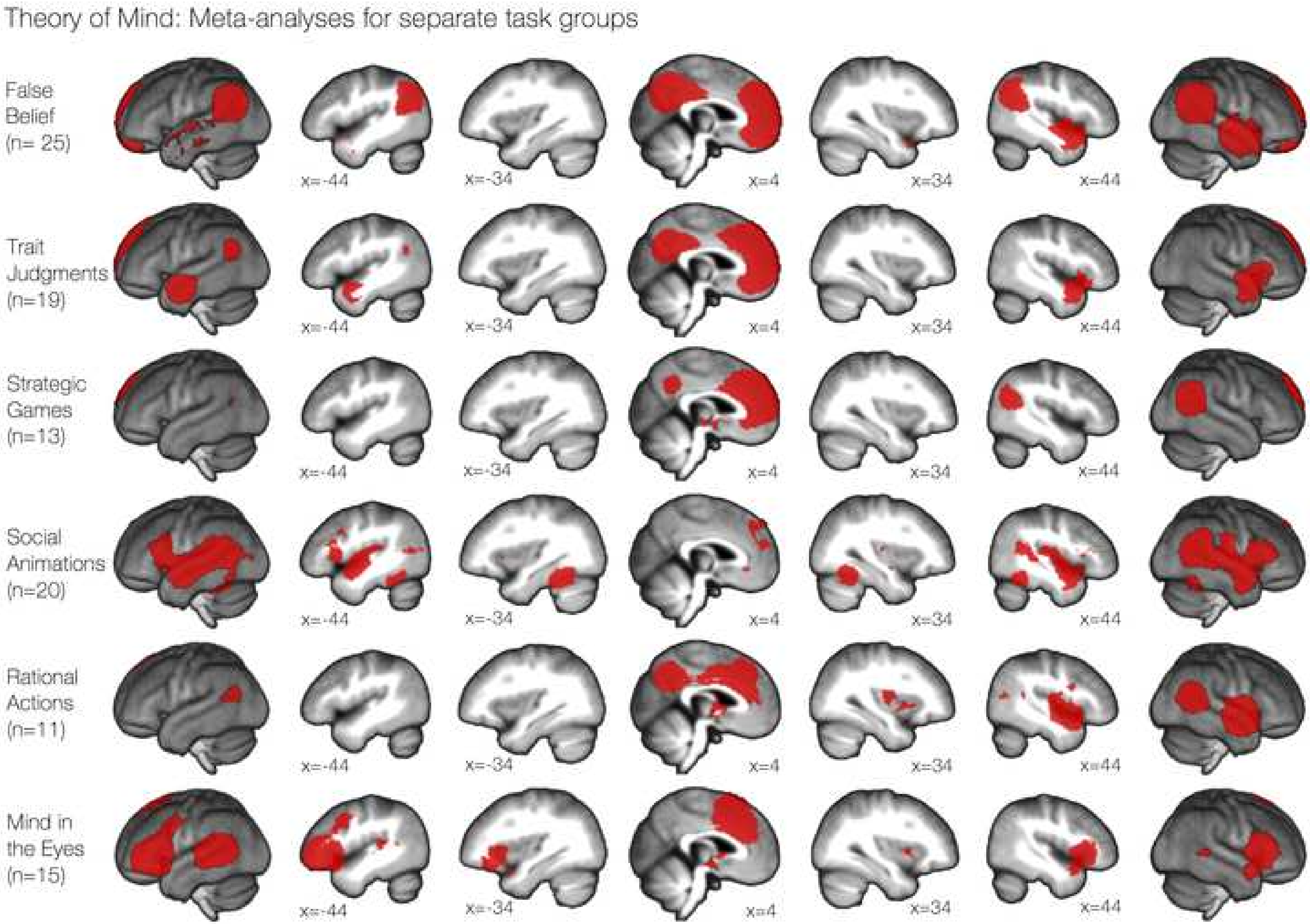

Theory of Mind: Meta-analyses for separate task groups
False Belief (n= 25)
Trait Judgments (n=19)
Strategic Games (n=13)
Social Animations (n=20)
Rational Actions (n=11)
Mind in the Eyes (n=15)
x=-44
x=-34
x=4
x=34
x=44

Figure2

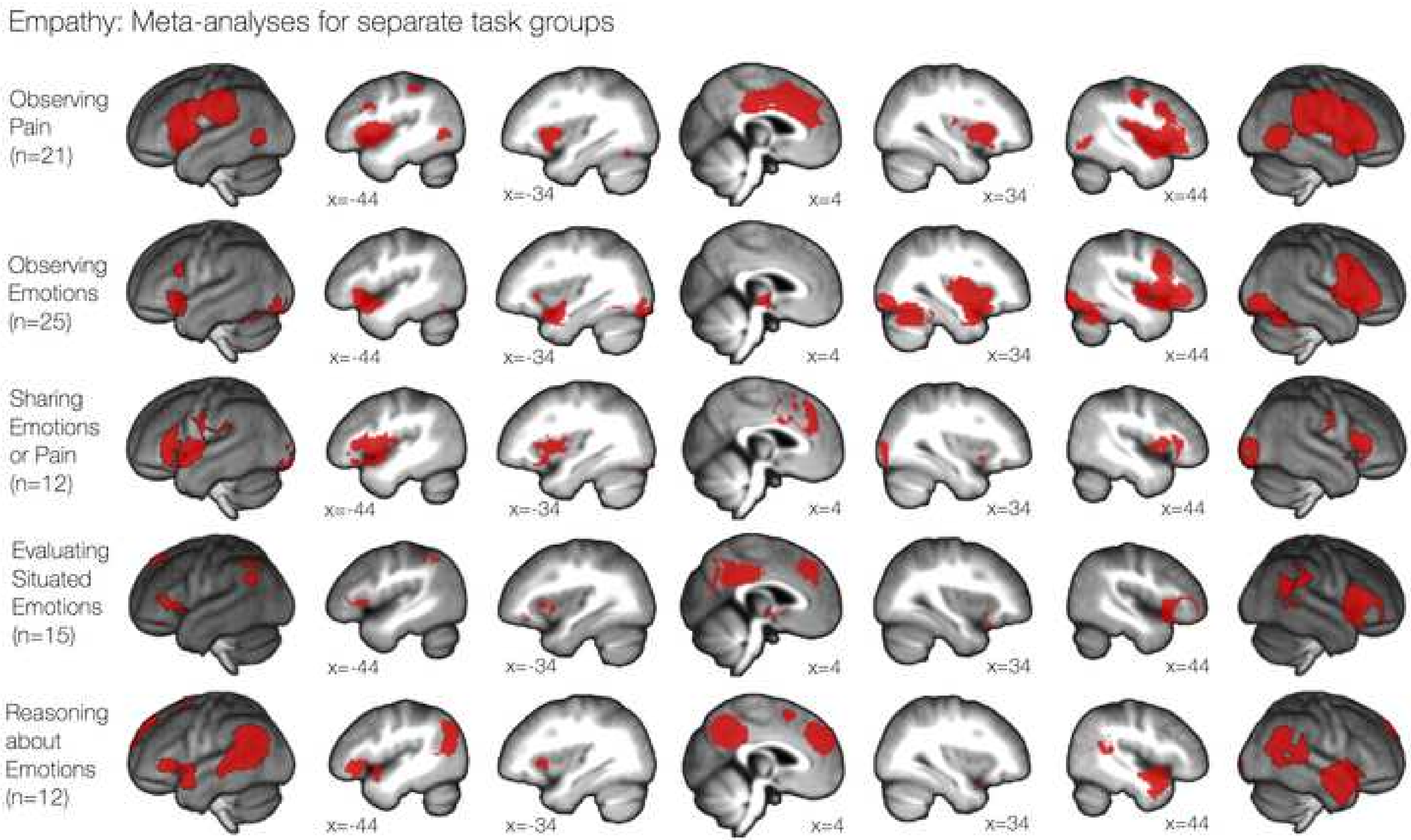
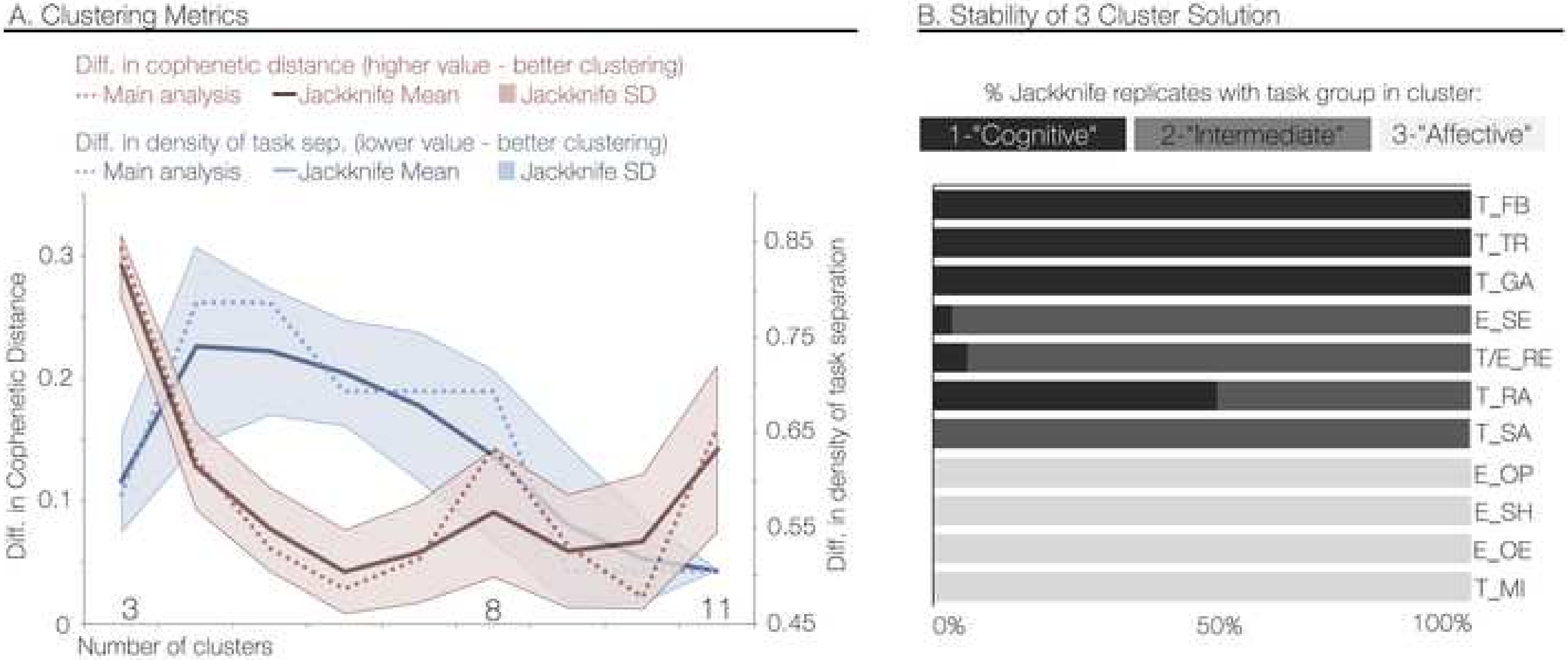

Empathy: Meta-analyses for separate task groups
Observing Pain (n=21)
Observing Emotions (n=25)
Sharing Emotions or Pain (n=12)
Evaluating Situated Emotions (n=15)
Reasoning about Emotions (n=12)
x=-44
x=-34
x=4
x=34
x=44

Figure3

Finding optimal clustering solution



Labels: Empathy (E_)
E_OP... Observing Pain
E_OE... Observing Emotions
E_SH... Sharing Emotions or Pain
E_SE... Evaluation Situated Emotions
T/E_RE... Reasoning about Emotions

Theory of mind (T_)
T_FB... False Belief
T_TR... Trait Judgments
T_GA... Strategic Games
T_SA... Social Animations
T_RA... Rational Actions
T_MI... Mind in the Eyes

Figure4

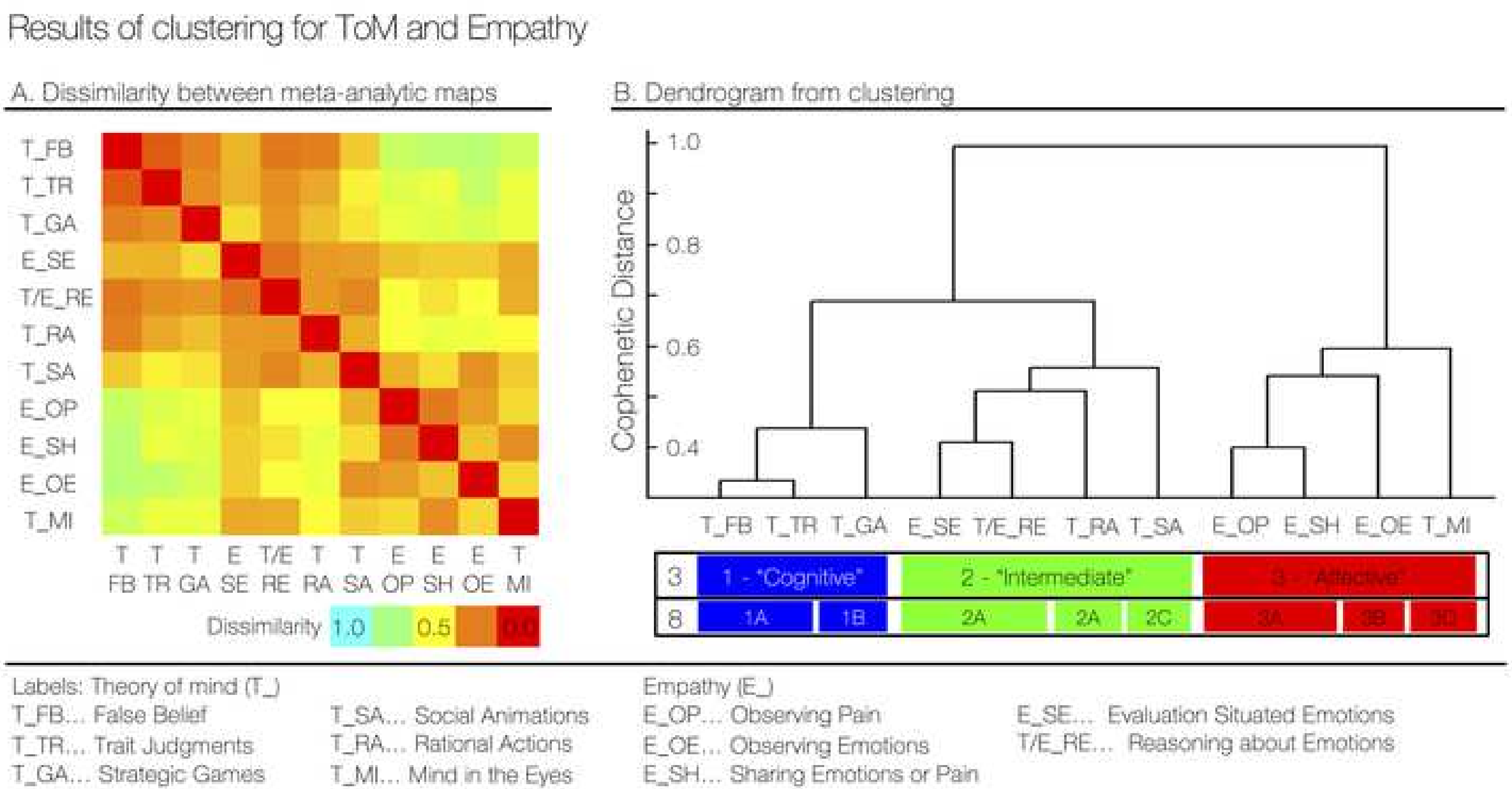
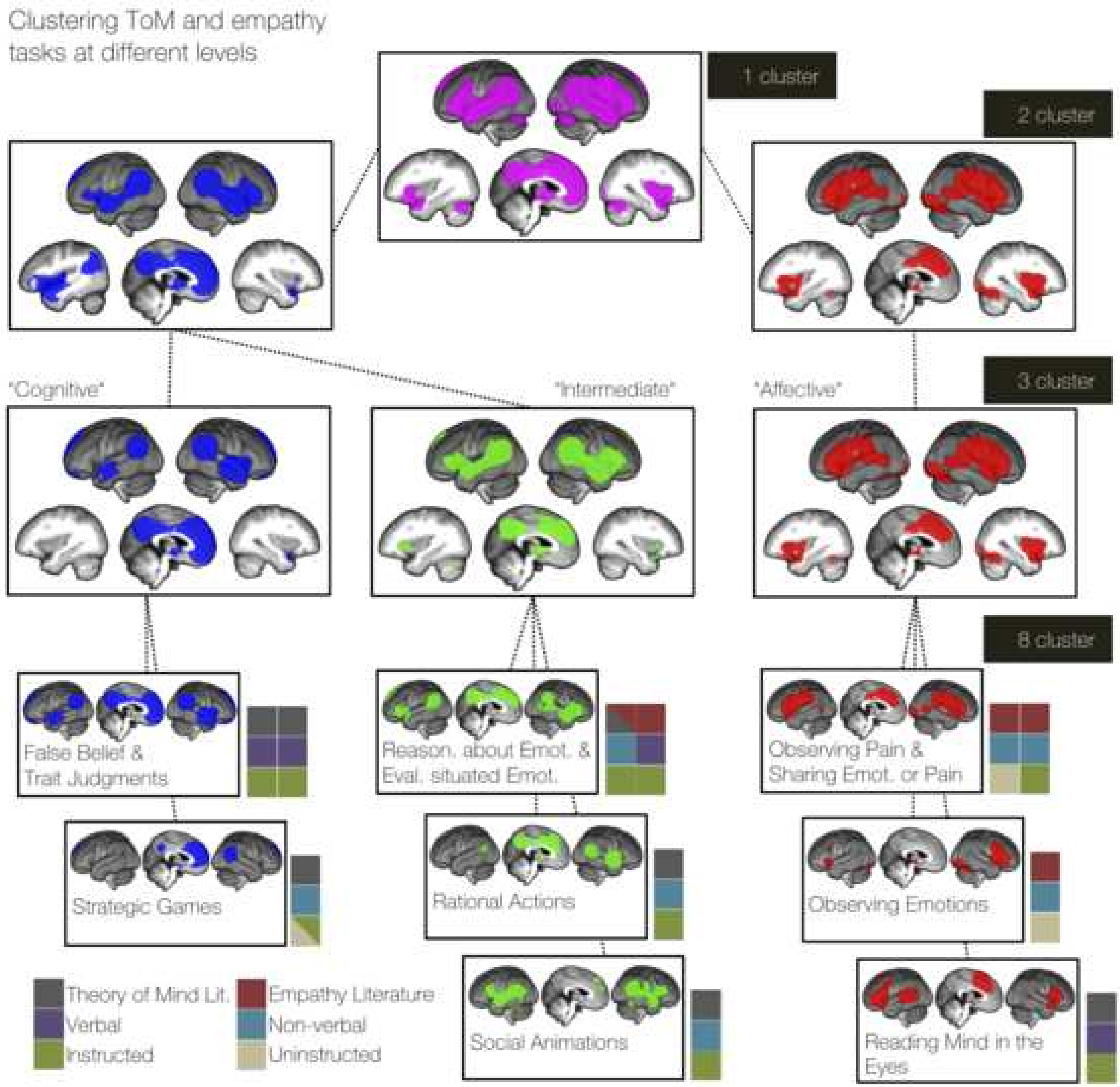

Results of clustering for ToM and Empathy
A. Dissimilarity between meta-analytic maps
T_FB
T_TR
T_GA
E_SE
T/E_RE
T_RA
T_SA
E_OP
E_SH
E_OE
T_MI
T FB T TR T GA E SE T/E RE T RA T SA E OP E SH E OE T MI
Dissimilarity 1.0 0.5 0.0
B. Dendrogram from clustering
Cophenetic Distance
1.0
0.8
0.6
0.4
T_FB T_TR T_GA E_SE T/E_RE T_RA T_SA E_OP E_SH E_OE T_MI
3
1 - "Cognitive"
2 - "Intermediate"
3 - "Affective"
8
1A
1B
2A
2A
2C
3A
3B
Labels: Theory of mind (T_)
T_FB... False Belief
T_TR... Trait Judgments
T_GA... Strategic Games
T_SA... Social Animations
T_RA... Rational Actions
T_MI... Mind in the Eyes
Empathy (E_)
E_OP... Observing Pain
E_OE... Observing Emotions
E_SH... Sharing Emotions or Pain
E_SE... Evaluation Situated Emotions
T/E_RE... Reasoning about Emotions

Figure5



Clustering ToM and empathy tasks at different levels
1 cluster
2 cluster
3 cluster
8 cluster
"Cognitive"
"Intermediate"
"Affective"
False Belief & Trait Judgments
Strategic Games
Reason. about Emot. & Eval. situated Emot.
Rational Actions
Social Animations
Observing Pain & Sharing Emot. or Pain
Observing Emotions
Reading Mind in the Eyes
Theory of Mind Lit.
Verbal
Instructed
Empathy Literature
Non-verbal
Uninstructed

Figure6

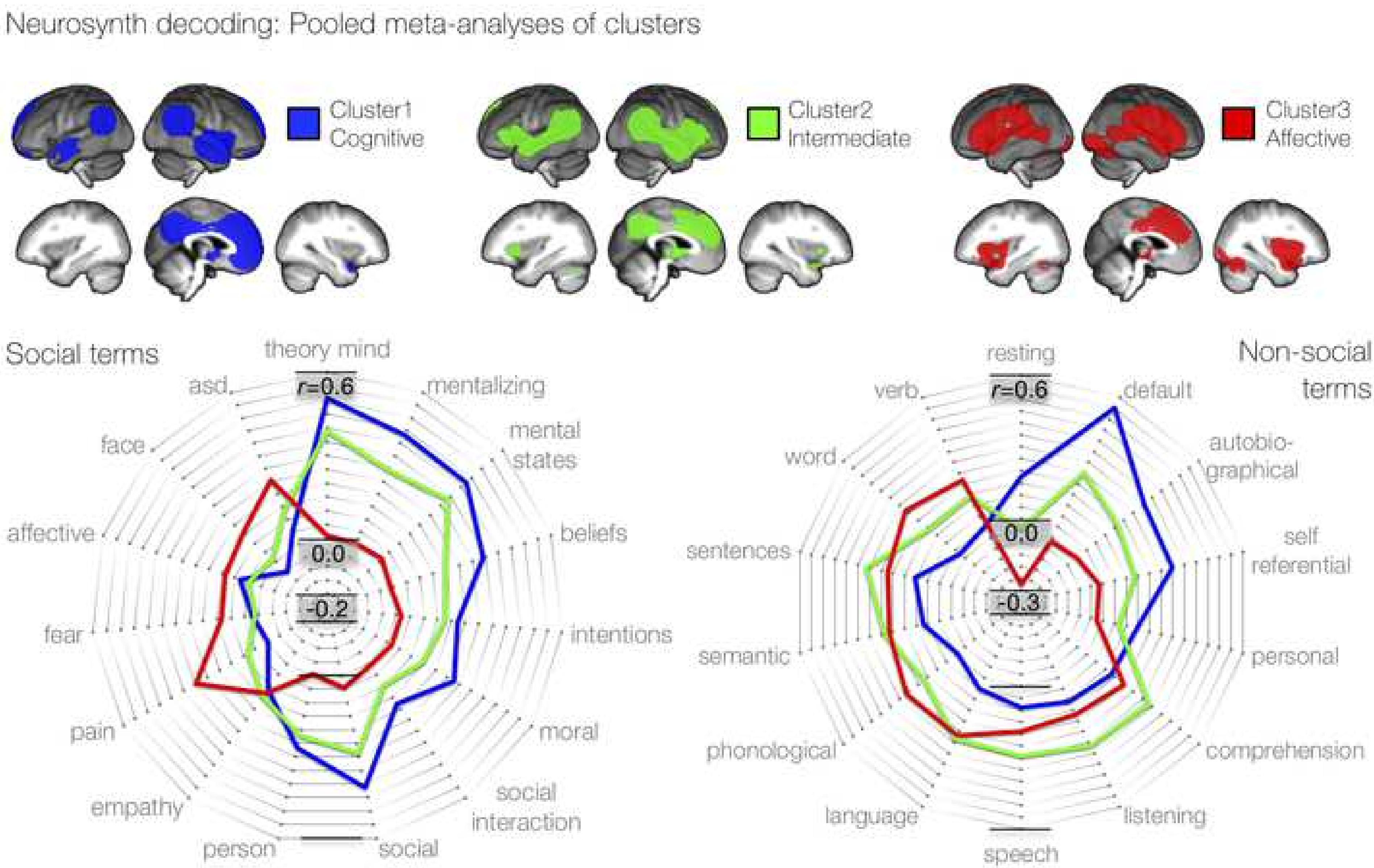

Neurosynth decoding: Pooled meta-analyses of clusters
Cluster1
Cognitive
Cluster2
Intermediate
Cluster3
Affective
Social terms
theory mind
r=0.6
mentalizing
mental
states
beliefs
intentions
moral
social
interaction
social
person
empathy
pain
fear
affective
face
asd
0.0
-0.2
Non-social
terms
resting
r=0.6
default
autobio-
graphical
self
referential
personal
comprehension
listening
speech
language
phonological
semantic
sentences
word
verb
0.0
-0.3

Figure7

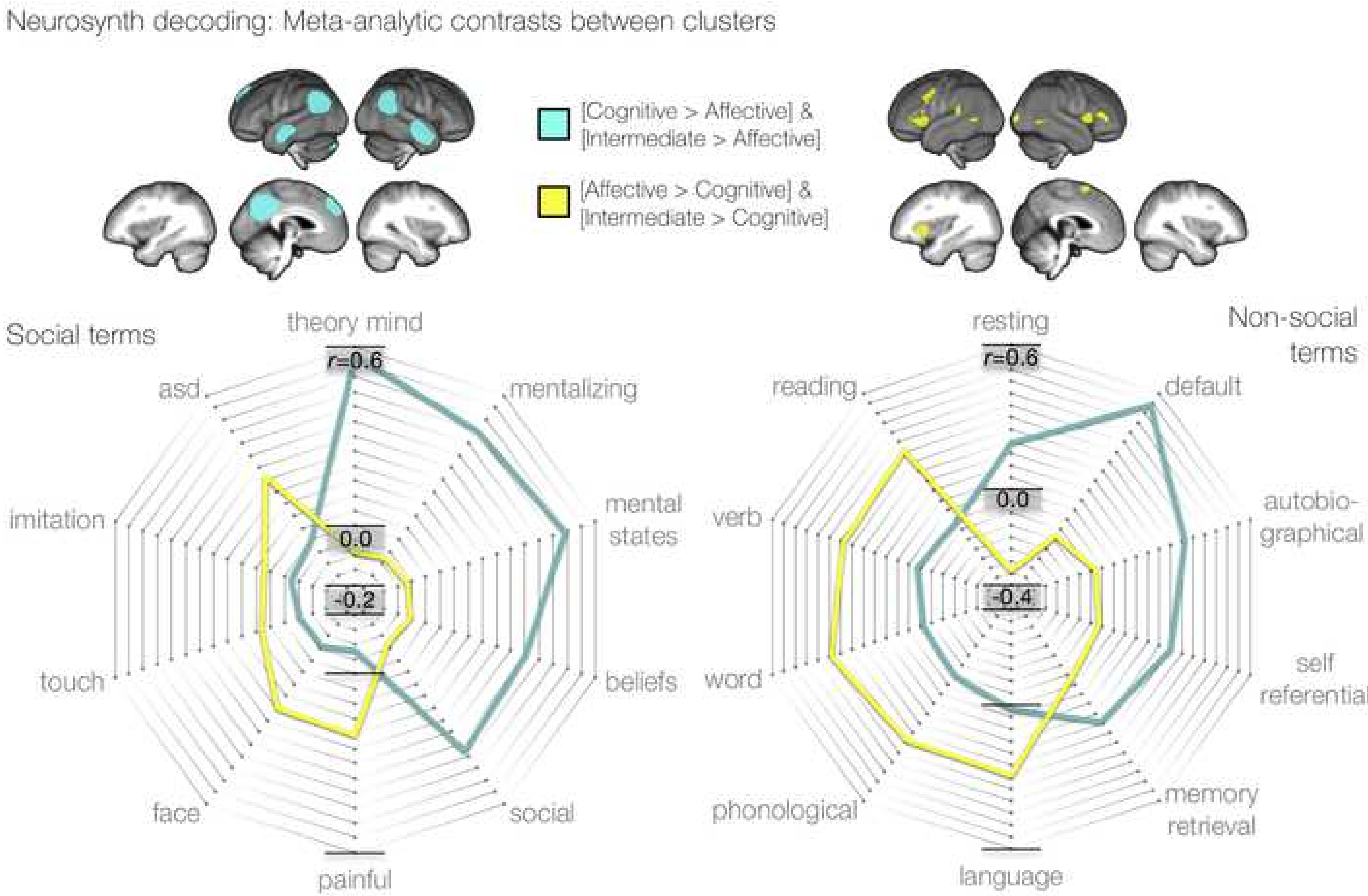

Neurosynth decoding: Meta-analytic contrasts between clusters
[Cognitive > Affective] &
[Intermediate > Affective]
[Affective > Cognitive] &
[Intermediate > Cognitive]
Social terms
theory mind
r=0.6
mentalizing
mental states
beliefs
social
painful
face
touch
imitation
asd
0.0
-0.2
Non-social terms
resting
r=0.6
default
autobio-graphical
self referential
memory retrieval
language
phonological
word
verb
reading
0.0
-0.4

Figure8

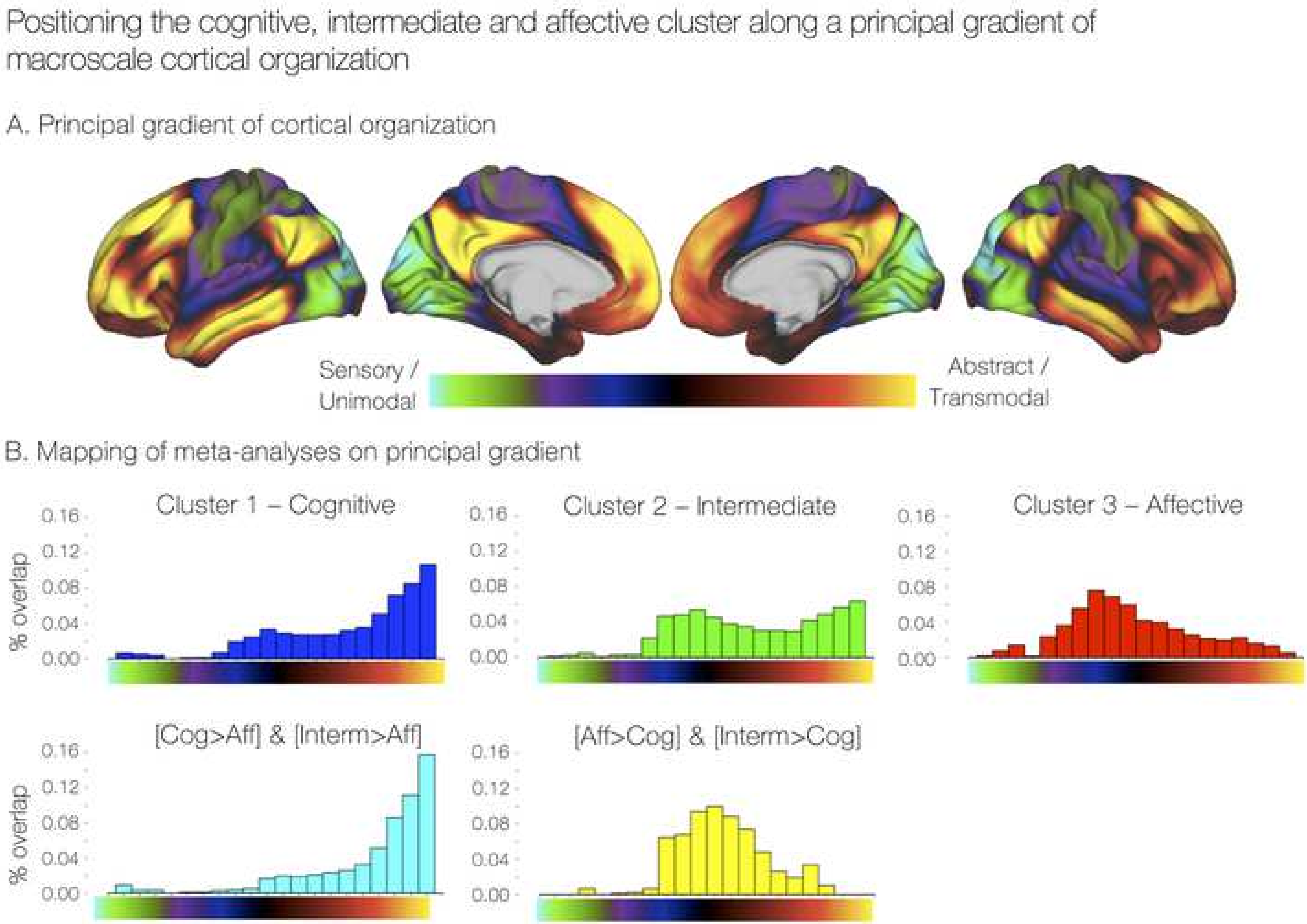
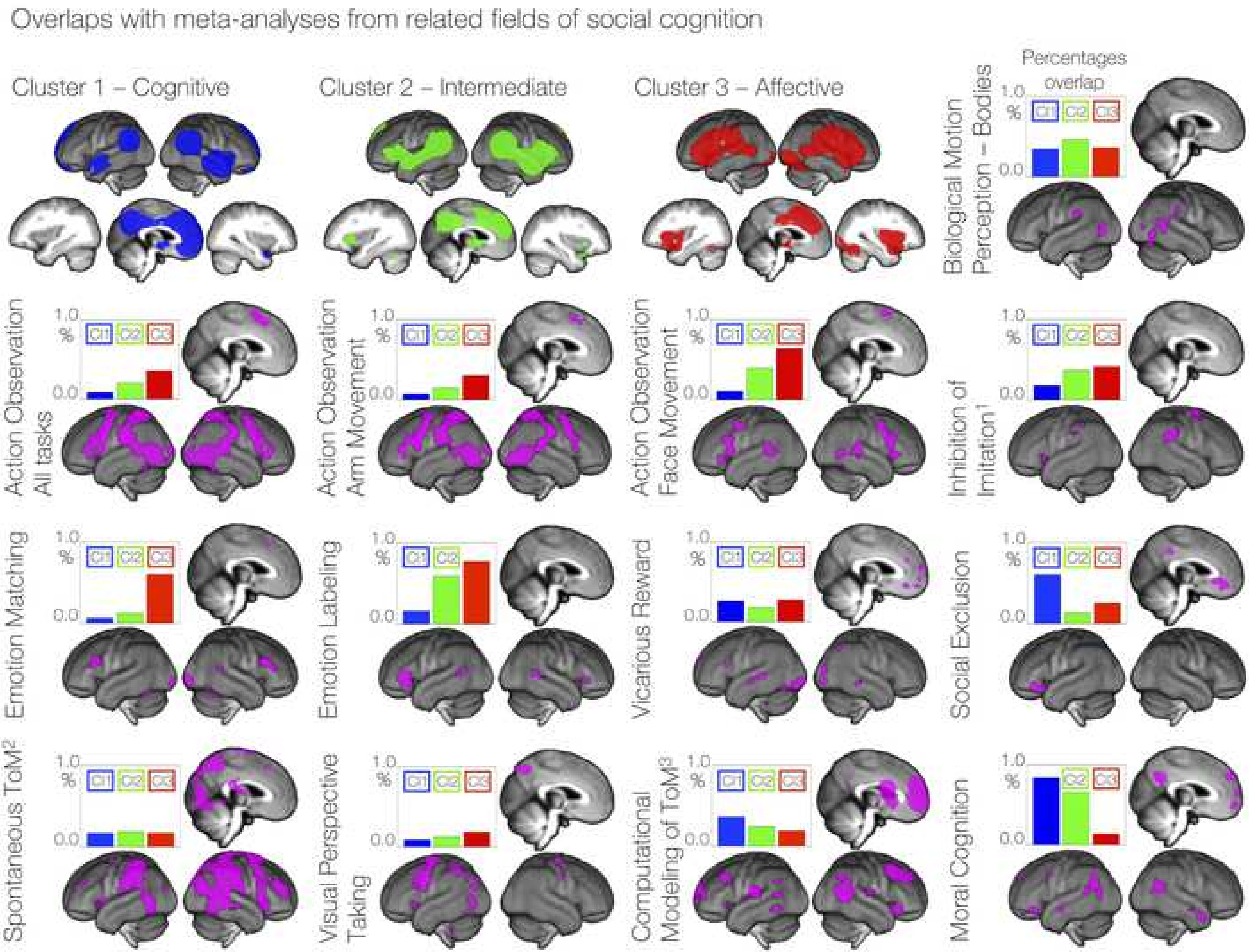

Positioning the cognitive, intermediate and affective cluster along a principal gradient of macroscale cortical organization
A. Principal gradient of cortical organization
Sensory / Unimodal
Abstract / Transmodal
B. Mapping of meta-analyses on principal gradient
Cluster 1 – Cognitive
Cluster 2 – Intermediate
Cluster 3 – Affective
[Cog>Aff] & [Interm>Aff]
[Aff>Cog] & [Interm>Cog]
% overlap
0.16
0.12
0.08
0.04
0.00

Figure9



Overlaps with meta-analyses from related fields of social cognition
Cluster 1 – Cognitive
Cluster 2 – Intermediate
Cluster 3 – Affective
Percentages overlap
Biological Motion Perception – Bodies
Action Observation All tasks
Action Observation Arm Movement
Action Observation Face Movement
Inhibition of Imitation1
Emotion Matching
Emotion Labeling
Vicarious Reward
Social Exclusion
Spontaneous ToM2
Visual Perspective Taking
Computational Modeling of ToM3
Moral Cognition
1.0
%
0.0
Cl1
Cl2
Cl3