
Matthias Schurz
University of Oxford; Radboud University Nijmegen; and University of Salzburg

Joaquim Radua
Institut d’Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Barcelona, Spain; King’s College London; and Karolinska Institutet

Matthias G. Tholen
University of Salzburg

Lara Maliske
Technische Universität Dresden

Daniel S. Margulies
Centre National de la Recherche Scientifique (CNRS), Paris, France

Rogier B. Mars
University of Oxford and Radboud University Nijmegen

Jerome Sallet
University of Oxford and University of Lyon

Philipp Kanske
Technische Universität Dresden and Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

Along with the increased interest in and volume of social cognition research, there has been higher 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 identified as overlapping umbrella terms for 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.

This article was published Online First November 5, 2020. Matthias Schurz, Wellcome Centre for Integrative Neuroimaging, Department of Experimental Psychology, University of Oxford; Donders Institute for Brain, Cognition, and Behaviour, Radboud University Nijmegen; and Centre for Cognitive Neuroscience, University of Salzburg. Joaquim Radua, Imaging of Mood- and Anxiety-Related Disorders (IMARD) Group, Institut d’Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Barcelona, Spain; Institute of Psychiatry, Psychology and Neuroscience, King’s College London; and Department of Clinical Neuroscience, Centre for Psychiatric Research and Education, Karolinska Institutet. Matthias G. Tholen, Centre for Cognitive Neuroscience, University of Salzburg. Lara Maliske, Clinical Psychology and Behavioral Neuroscience, Faculty of Psychology, Technische Universität Dresden. Daniel S. Margulies, Frontlab, Institut du Cerveau et de la Moelle Epinière, Centre National de la Recherche Scientifique (CNRS) UMR 7225, Paris, France. Rogier B. Mars, Wellcome Centre for Integrative Neuroimaging, Centre for Functional MRI of the Brain (FMRIB), Nuffield Department of Clinical Neurosciences, John Radcliffe Hospital, University of Oxford; and Donders Institute for Brain, Cognition, and Behaviour, Radboud University Nijmegen. Jerome Sallet, Wellcome Centre for Integrative Neuroimaging, Department of Experimental Psychology, University of Oxford; and Stem Cell and Brain Research Institute U1208, Institut national de la santé et de la recherche médicale (INSERM), Université Lyon 1, University of Lyon. Philipp Kanske, Clinical Psychology and Behavioral Neuroscience, Faculty of Psychology, Technische Universität Dresden; and Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany.

Matthias Schurz was supported by an Erwin Schrödinger Fellowship (FWF-P26099-B27), a Marie Skłodowska-Curie Individual Fellowship (MSCA-IF 844734), and the German Research Foundation Network ‘Understanding Others’ (SCHN 1481/2-1). The Wellcome Centre for Integrative Neuroimaging is supported by core funding from the Wellcome Trust (203159/Z/16/Z).

Joaquim Radua has received support from the Instituto de Salud Carlos III, Ministry of Economy and Competitiveness of Spain (PI19/00394, CPII19/0009), integrated into the Plan Nacional de I+D+i, and cofounded by the ISCIII-Subdirección General de Evaluación and Fondo Europeo de Desarrollo Regional (FEDER). Matthias G. Tholen was supported by the Austrian Science Fund’s Doctoral College ‘Imaging the Mind’ (W1233). Lara Maliske was supported by a scholarship of the Friedrich Ebert Foundation. Rogier B. Mars was supported by the Biotechnology and Biological Sciences Research Council (BBSRC) UK (BB/N019814/1) and the Netherlands Organization for Scientific Research NWO (452-13-015). Jerome Sallet was supported by a Wellcome Trust Henry Dale Fellowship (105651/Z/14/Z) and an IDEXLYON IMPULSION 2020 Grant (IDEX/JMP/2020/14). Philipp Kanske was supported by the German Federal Ministry of Education and Research within the ASD-Net (BMBF FKZ 01AE1409A), the German Research Foundation (KA 4412/2-1, KA 4412/4-1, KA 4412/5-1, CRC940/C07), Die Junge Akademie at the Berlin-Brandenburg Academy of Sciences and Humanities. The German National Academy of Sciences Leopoldina, and the German Research Foundation Network ‘Understanding Others’ (SCHN 1481/2-1). Figure 8 shows brain images created by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH90640) and the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University.

Maps and files from the present analyses are available at: https://osf.io/pavz7/

Correspondence concerning this article should be addressed to Matthias Schurz, Donders Institute for Brain, Cognition, and Behaviour, Radboud University Nijmegen, Montessorilaan 3, 6525 HR Nijmegen, the Netherlands. E-mail: matthias.schurz@gmail.com
Successful social interaction requires representing not only the overt behavior of people but also its 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 the 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. Clearly, different terms are being used to describe similar processes and, at times, similar terms for describing different 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’ mental states (Adolphs, 2009; Kanske, 2018; Keysers & Gazzola, 2009; Premack & Woodruff, 1978). 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 that explicitly categorized studies based on stimuli and instructions found multiple functional subdivisions 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 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, complicating the distinction between the terms. For instance, definitions of ToM include the ability to make inferences not only 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., 2012; Shamay-Tsoory & Aharon-Peretz, 2007). Relatedly, it was proposed that the processing of others’ mental states engages ToM, 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 others’ affective states (e.g., Dziobek et al., 2011; Shamay-Tsoory, Aharon-Peretz, & Perry, 2009), referred to as “cognitive empathy” (Hooker, Verosky, Germain, Knight, & D’Esposito, 2010; Walter, 2012). Even broader conceptualizations of empathy contain additional features such as 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 to clarify current theories of social cognition is determining how different concepts relate to one another (e.g., in a multilayered manner; De Waal, 2012; see also De Waal, 2007; Preston & Horie, 2012; Singer, 2006) and exploring how they are best grouped according to common underlying processes (Happé et al., 2017). This approach also advances a “deconstruction” of social processes (see Schaafsma et al., 2015), that is, the mapping of broad terms such as ToM and empathy to a set of underlying processes that are in turn linked to concrete experimental tasks. Such an endeavor also dovetails with the strategy promoted by the National Institute of Mental Health (NIMH) for
TOWARD A HIERARCHICAL MODEL OF SOCIAL COGNITION

studying mental disorders (Insel et al., 2010)—the Research Domain Criteria (RDoC) Projects.

We begin our analysis with a meta-analysis of neuroimaging studies and then move on to review the corresponding behavioral findings. Neuroimaging data provide a useful starting point because 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 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 included task groups not only from the ToM literature (e.g., False Belief, Reading the 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 among brain activation networks for different task groups.

The main goal of this clustering is to estimate latent factors, that is, 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 largely overlap, one component could be a subcomponent of another, or components could have some elements in common but additionally contain distinct processes (see Happé 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 several tasks. 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 subprocesses 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.

Method

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 included 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 they included theory of mind or mentalizing or mindreading and to empathy groups when they included 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. Articles were retained if they matched one of the empathy and ToM task groups defined previously, irrespective of the terminology used by the authors. All studies included in the task groups had to further fulfill 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 using a consistent threshold throughout the whole brain. Data from clinical samples were removed. If a study reported more than one contrast, the one that best corresponded 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 the 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 cover a total of 188 studies.

Meta-Analysis Methods

For each of the 11 task groups, we carried out an effect-size-based meta-analysis using the Signed Differential Mapping method (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., 2016) or multilevel kernel density analysis (MKDA, Wager, Lindquist, & Kaplan, 2007). In addition, it incorporates the effect sizes of reported activations, thus extracting more detailed information from the published literature. Based on t values and sample sizes reported in studies, SDM creates effect-size (Hedge’s g values) and variance maps (derived from the distribution of effect sizes). 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 for the MNI space and at a statistical threshold of p < .005 uncorrected (voxel-level) and a cluster threshold of 10 voxels (see Figures 1 and 2). This threshold was found to optimally balance sensitivity and specificity and to be an approximate equivalent to a corrected threshold of p < .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

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 anisotropic un-normalized Gaussian kernel. The 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 sizes or lower variabilities to contribute more strongly.
variability. To determine 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). Note that, as the ultimate goal of our meta-analyses was to cluster result-maps based on similarities in whole-brain activation patterns, we did not carry out analyses of publication bias.

Hierarchical Clustering Analysis

After obtaining meta-analytic result maps for all 11 task groups, we applied agglomerative hierarchical clustering to them (see, e.g., Laird et al., 2015; Riedel et al., 2015, 2018). Searching for a hierarchical structure is consistent with several previous conceptualizations of social cognition as a multilayered or multilevel phenomenon (De Waal, 2012; Preston & De Waal, 2002; Schaalma et al., 2015; Singer, 2006). To our knowledge, this is the first clustering of SDM effect-size meta-analysis maps. Therefore, we compared several settings for discriminative performance (see the online supplemental 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 these, we calculated pairwise Pearson correlation coefficients between all feature vectors, from which we derived an n-by-n dissimilarity matrix (1 – r values) reflecting whole-brain multivoxel dissimilarity between maps. In Step 2, we grouped meta-analyses into clusters by applying agglomerative hierarchical clustering in MATLAB 8.1 (The Math Works, Inc., 2013). 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

2 Our meta-analytic clustering is based on patterns (i.e., data-vectors) which capture activation across thousands of brain locations, that have been computed separately for 11 task types. As publication bias in neuroimaging studies can be region- and task-specific, an estimation thereof would produce a very high number of possible sub-analyses. Although such an extensive analysis goes beyond the scope of our meta-analytic clustering project, further work is needed to systematically address publication bias in the social neurosciences.
distance measure, we selected Euclidean, which considers 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. 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: (a) cophenetic distance and (b) density of task separation. Figure 3 shows both measures for different clustering solutions; Figure 4 shows the underlying dendrogram. The cophenetic distance between clusters at a given model order (i.e., number of clusters) reflects dissimilarity between subclusters. A clustering optimum is indicated by high difference in cophenetic distances when moving from lower to higher model orders as this indicates that introducing new subclusters produces substantially different (i.e., distant) brain activity patterns. The second metric we used, density of task separation, indicates whether separating clusters into subclusters maintains a balanced distribution of task groups across subclusters. A clustering optimum is indicated by high difference in cophenetic distances when moving from lower to higher model orders as this indicates that introducing new subclusters produces substantially different (i.e., distant) brain activity patterns. The second metric we used, density of task separation, indicates whether separating clusters into subclusters maintains a balanced distribution of task groups across subclusters (as opposed to producing disproportionately large/small subclusters, such as a cluster consisting of only one task group). Decreases in the density of task separation indicate good solutions and reflect a split into subclusters 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. 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 more than 90% of iterations) was found for most task groups (except for Rational Actions).

---

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

4 If cluster $i_x$ consisted of $n_x$ task groups at model order $x$, and at model order $x + 1$ was separated into subclusters $i_{x1}$ and $i_{x2}$ with $n_{x1}$ and $n_{x2}$ task groups, then the density of task separation can be calculated as $d_x = n_x / n_x$, with $n_{x1} > n_{x2}$.

5 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 5,000 clusterings based on leave-one-out samples and summarized the metrics across all of them in Figure 3.
Analysis of Overlap With Other Maps

For a broader perspective on our meta-analysis, we compared our main results (see Figure 5) 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 and also nonsocial processes. We used the decoding tool of the neurosynth database (Yarkoni, Polcrock, 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.6 In Figures 6 and 7, we showed the neurosynth decoding results as Pearson correlation coefficients, describing image similarity between our maps and automatically generated meta-analyses for research topics (“terms”) identified by text mining of literature databases. For clarity, we show only the most strongly correlated social and nonsocial terms found with neurosynth decoding.7 In addition, we discarded terms of little interest for the current article, such as for example neuroanatomical labels (e.g., mpfc, psts).

Figure 3. (A) Evaluation and comparison of different clustering solutions. The 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 (dark grey) 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 indicates that introducing new subclusters results in a better separation of brain activity patterns and thus good clustering. The second metric, shown in blue (light grey), shows the density of task separation, reflecting whether separating clusters into subclusters maintains a balanced number of task groups in each subcluster. The relatively low density of task separation indicates good clustering. Preferred clusterings in terms of both metrics are indicated on the x axis: 3, 8, and 11. Metric changes are shown for the clustering of complete meta-analyses (main analysis) and clusterings based on leave-one-out jackknife sensitivity analysis with 5,000 repeats (jackknife mean, standard deviation [SD]). (B) 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. The bars indicate the percentage of jackknife repeats for which task groups were assigned to the same clusters as in the main analysis. See the online article for the color version of this figure.
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 and summarized them as a variant of the dice score: For each meta-analysis map, we calculated the percentage of voxels it comprised that fell within different parts of each connectivity map.

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 the Clustering section. 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.

Task Groups: ToM

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). These False Belief stories had a particular logical structure. As illustrated in Table 1, stories typically first introduce a person and his/her true belief about a state of affairs. Unknown to that person, this state changes thereafter. As a result, the person’s belief becomes false. Participants are then asked a question related to that belief. To account for this structure of False Belief stories, our meta-analysis only included more recent studies presenting false photograph control conditions of similar form (e.g., Dodell-Feder, Koster-Hale, Bedny, & Saxe, 2011; Saxe & Kanwisher, 2003), see Table 1 for examples. Note that for False Belief tasks, we found many more eligible studies in the literature than for other task groups (38 studies in total). To avoid large differences in size between task groups, we randomly selected 25 studies from this large sample of False Belief tasks.

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, which aimed to find brain areas with a specific relationship to the type of stimulus, were designed to test for the ability to differentiate between different classes of stimuli. As illustrated in Table 2, examples include evaluating situational emotions (e.g., observation of pain), understanding emotions (e.g., sharing emotions or pain), and reasoning about emotions.

For overlaps with resting-state networks, we used the seven-network parcellation by Yeo et al. (2011), more specifically a MNI transformed version (liberal mask, see https://surfer.nmr.mgh.harvard.edu/fswiki/CorticalParcellation; Yeo et al., 2011).

The connectivity gradient, a partitioned version consisting of 20 maps was used, each corresponding to a five-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). Because all inputs conformed to the 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). That is, if $i_1$ is a meta-analysis map and $i_2$ a connectivity map, we calculated $n$ voxels in $(i_1 \& i_2) / n$ voxels in $i_2$. 
We carried out pooled meta-analyses, that is, one separate meta-analysis per cluster, where all its task groups are joined together. Analyses were thresholded at a voxel-wise threshold of \( p < .005 \) uncorrected and a cluster extent threshold of 10 voxels. The one-cluster solution is shown for illustrative purposes only and was not evaluated or compared against other clusterings. Colors indicate how the three-cluster solution relates to both higher- and lower-level clusterings: blue—cognitive cluster (left column), green—intermediate cluster (middle column), red—affective cluster (right column). At the lowest level of the dendrogram, we indicate for each cluster some exemplary stimulus and task categorizations. The eight-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). See the online article for the color version of this figure.

Figure 5. Mean brain activation for clusters at different model orders. We carried out pooled meta-analyses, that is, one separate meta-analysis per cluster, where all its task groups are joined together. Analyses were thresholded at a voxel-wise threshold of \( p < .005 \) uncorrected and a cluster extent threshold of 10 voxels. The one-cluster solution is shown for illustrative purposes only and was not evaluated or compared against other clusterings. Colors indicate how the three-cluster solution relates to both higher- and lower-level clusterings: blue—cognitive cluster (left column), green—intermediate cluster (middle column), red—affective cluster (right column). At the lowest level of the dendrogram, we indicate for each cluster some exemplary stimulus and task categorizations. The eight-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). See the online article for the color version of this figure.
comparable level of specialization for conceptual knowledge about persons (e.g., Mitchell, Heatherton, & Macrae, 2002), quickly became widely used. 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 the 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 lowercase?”) or presented words/statements that did not contain mental states.

**Strategic games.** Early studies used Strategic Games for studying ToM based on the idea that feedback from a social partner—indicated by their moves in the game—may trigger spontaneous mentalizing. Even when not explicitly asked to “mind-read,” 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), that is, a disposition to reason about the beliefs, desires, and intentions of others (Gallagher, Jack, Roepstorff, & 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 (e.g., 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 with playing with a computer (which follows a simple algorithm).

**Social animations.** Social Animations were introduced (Castelli, Frith, Happé, & Frith, 2002; Castelli, Happé, Frith, & Frith, 2000) 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 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 (1944). 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 took place. 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).

Reading the mind in the eyes. The Reading the 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 ToM and mind-reading abilities in earlier work (Baron-Cohen, Ring, et al., 1999). To illustrate, it was found that adults with high-functioning autism spectrum disorder (Baron-Cohen, Joliffe, 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 the Mind in the Eyes task presented a photograph of the eye region of a face and asked participants to think about the person’s mental state or 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 the sake of sample coherence (and in light of the empathy tasks we compare here), we did not include two studies from the Reading the Mind in the Eyes sample in Schurz et al. (2014) because these studies asked participants for more basic emotion judgments.

Rational actions. Early studies presenting cartoons were introduced as a nonverbal 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
### Examples of Studies in Theory of Mind Task Groups

<table>
<thead>
<tr>
<th>Author</th>
<th>Experimental condition</th>
<th>Control condition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>False belief (25 studies)</strong></td>
<td>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, Koster-Hale, Bedny, and Saxe (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])</td>
<td>Read a short vignette involving a photograph/physical representation of the past, and a description of 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])</td>
</tr>
<tr>
<td>Oliver, Vieira, Neufeld, Dziobek, &amp; Mitchell, 2018</td>
<td>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])</td>
<td>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])</td>
</tr>
<tr>
<td>Saxe &amp; Kanwisher, 2003</td>
<td>Read a personality trait adjective (e.g., brave, childish) and indicate whether it correctly describes a former American president (Bill Clinton, Yes or No).</td>
<td>Read a personality trait adjective (e.g., brave, childish) and indicate whether it is written in lower- or uppercase (Yes or No).</td>
</tr>
<tr>
<td>Ma et al., 2011</td>
<td>Read a 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?”)</td>
<td>Read a 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]?”)</td>
</tr>
<tr>
<td></td>
<td>Read a single statement about a person doing something. Indicate the gender of the person in the sentence (e.g., “Tolvan gave her mother a bottle . . . is Tolvan male or female?”)</td>
<td>Read a single statement about a person doing something. Indicate the gender of the person in the sentence (e.g., “Tolvan gave her mother a bottle . . . is Tolvan male or female?”)</td>
</tr>
<tr>
<td>Zhu, Zhang, Fan, &amp; Han, 2007</td>
<td>Read a personality trait adjective (e.g., brave, childish) and indicate whether it correctly describes a former American president (Bill Clinton, Yes or No).</td>
<td>Read a personality trait adjective (e.g., brave, childish) and indicate whether it is written in lower- or uppercase (Yes or No).</td>
</tr>
<tr>
<td><strong>Strategic games (13 studies)</strong></td>
<td>Play the matching-pennies game against a computer. Both players are asked to choose one 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 a different option.</td>
<td>Play the matching-pennies game against a computer. Both players are asked to choose one 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 a different option.</td>
</tr>
<tr>
<td>Takahashi, Izuma, Matsumoto, Matsumoto, &amp; Omori, 2015</td>
<td>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.</td>
<td>Play the prisoner’s dilemma game (iterated version). You play with a computer for game points.</td>
</tr>
<tr>
<td>Kircher et al., 2009</td>
<td>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.</td>
<td>Play the prisoner’s dilemma game (iterated version). You play with a computer for game points.</td>
</tr>
<tr>
<td><strong>Social animations (20 studies)</strong></td>
<td>Watch a video animation of two interacting triangles, which involve influence on each other’s mental states (e.g., coaxing). Indicate whether a social/goal-directed/ random movement was shown, and indicate the feeling of both triangles (Positive/Negative). Respond via button press to both questions.</td>
<td>Watch a video animation of two triangles which interact in a goal-directed manner (e.g., one chasing the other). Indicate whether a social/goal-directed/random movement was shown, and indicate the feeling of both triangles (Positive/Negative). Respond via button press to both questions.</td>
</tr>
<tr>
<td>Moessnang et al., 2016</td>
<td>Watch video animation of two interacting triangles (e.g., portraying a scene where mother and child are playing). Explain verbally what was happening (after fMRI).</td>
<td>Watch video animation of two randomly moving triangles. Explain verbally what was happening (after fMRI).</td>
</tr>
</tbody>
</table>

*(table continues)*
Table 1 (continued)

<table>
<thead>
<tr>
<th>Author</th>
<th>Experimental condition</th>
<th>Control condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baron-Cohen et al., 1999</td>
<td>View a photograph showing the eye region of a face. Indicate which of two words (e.g., concerned versus uninterested) describes the mental state of that person (button press).</td>
<td>View a photograph showing the eye region of a face. Indicate whether the person is male or female (button press).</td>
</tr>
<tr>
<td>Bos et al., 2016</td>
<td>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.</td>
<td>View a photograph showing the eye region of a face. Indicate whether a nonmental state word presented before (e.g., woman, curly hair, heavy eyebrows) matches the photo.</td>
</tr>
<tr>
<td>Brunet, Sarfati, Hardy-Baylé, &amp; Decety, 2000</td>
<td>View a cartoon story and predict what will happen based on the 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.” Indicate what will happen next . . . “The prisoner ties a rope from the sheets on his bed/the prisoner shouts out loud to get some attention.”)</td>
<td>View cartoon stories and predict what will happen (press button) based on physical causality (e.g., “A person is standing in front of a slide. A large ball is coming down the slide, heading towards that person.” Indicate what will happen next . . . “The ball knocks the person/the ball rests on the ground and the person stands beside it.”)</td>
</tr>
<tr>
<td>Heleven, van Dun, &amp; Van Overwalle, 2019</td>
<td>View cartoon stories showing a person over a sequence of events. The order of pictures (i.e., events) is jumbled. Indicate the correct order of pictures based on the intentions of the character (button press).</td>
<td>View cartoon stories showing a sequence of events. The order of pictures (i.e., events) is jumbled. Indicate the correct order of pictures based on physical causality (button press).</td>
</tr>
</tbody>
</table>

Note. For a complete list, see the online supplemental materials.

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 the (covert) mirroring of observed facial expressions triggers activity in emotional brain areas and thus an empathetic 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. Because 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.\(^{11}\) All studies in our task group presented pictures or videos of faces displaying basic emotional expressions (e.g., anger, fear, happiness, or disgust). Participants were not asked for explicit emotion judgments but passively viewed the stimuli or made judgments regarding nonemotional stimulus characteristics (e.g., gender or physical properties). In control conditions, faces showed neutral expressions, or no stimuli were presented at all.

Sharing emotions or pain. Whereas the previous task groups have been related to more automatically occurring processes, the ending based on the rational intentions of the protagonist (implied in their 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 the nonmental aspects of the scenes were asked, for example, physical causality.

Task Groups: Empathy

Observing pain. Pain has been a central theme in 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, Hanelin, 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 to 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 stimulations of body parts or faces, the control conditions presented neutral physical stimulations (e.g., being touched by a Q-tip) or no stimulation.

\(^{11}\) After identifying facial emotion viewing studies in the empathy literature, we added studies to 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).
The task group contains comparable stimuli as the previous instances, participants were additionally asked to rate the emotional state of the target (e.g., “feeling into” her). In some categories (faces with basic emotions, body parts in painful situations; see van der Heiden, Scherpiet, Konicar, Birbaumer, & Veit, 2012) that are possibly linked to top-down regulatory processes, empathy (see, e.g., de Greck, Shi, et al., 2012; de Greck, Wang, et al., 2012) added situational context to empathy tasks. Here, researchers argued that empathy involves more than just focusing on another person but also considers how they are embedded in a situation and contextual background (e.g., Regenbogen et al., 2012; Ruby & Decety, 2004). Tasks presenting such contextual information were

### Table 2

<table>
<thead>
<tr>
<th>Author</th>
<th>Experimental condition</th>
<th>Control condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Olsson, Nearing, &amp; Phelps, 2007</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Bos, Montoya, Hermans, Keyser, &amp; van Honk, 2015</strong></td>
<td>Watch videos of hands under painful (needle) stimulation. Watch attentively and detect changes in the fixation cross appearing between videos.</td>
<td>Watch videos of hands under nonpainful (cotton swab) stimulation. Watch attentively and detect changes in the fixation cross appearing between videos.</td>
</tr>
<tr>
<td><strong>Observing emotions (25 studies)</strong></td>
<td>View a picture of a face with an emotional expression (anger, fear). Judge the gender of the person (button press).</td>
<td>View a picture of a face with a neutral expressions. Judge the gender of the person (button press).</td>
</tr>
<tr>
<td>Critchley et al., 2000</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Toller et al., 2015</strong></td>
<td>View a video of a face showing fearful expression. Relax and focus on the actor’s eyes.</td>
<td>View a video of landscapes. Relax and do nothing.</td>
</tr>
<tr>
<td><strong>Sharing emotions or pain (12 studies)</strong></td>
<td>View a picture of a body part (hand) under painful stimulation. Imagine how the person in the picture feels and rate their pain (VAS).</td>
<td>View a picture of a body part (hand) under nonpainful stimulation. Imagine how the person in the picture feels and rate their pain (VAS).</td>
</tr>
<tr>
<td>Preis, Schmidt-Samoa, Dechent, &amp; Kroener-Herwig, 2013</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reniers, Völlm, Elliott, &amp; Corcoran, 2014</td>
<td>View a picture of a face with a sad expression. Imagine what the person in the picture is feeling. No response in the scanner.</td>
<td>View a picture of a face with a neutral expression. Imagine what the person in the picture is feeling. No response in the scanner.</td>
</tr>
<tr>
<td><strong>Evaluating situated emotions (15 studies)</strong></td>
<td>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).</td>
<td>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).</td>
</tr>
<tr>
<td>Kanske et al., 2015</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reyes-Aguilar et al., 2017</td>
<td>Read a short vignette about an emotionally negative or positive event, then view a picture of the person to whom it happens. Think about what this person is feeling (no overt response).</td>
<td>Read a short vignette about an emotionally neutral event, then view a picture of the person to whom it happens. Think about what this person is feeling (no overt response).</td>
</tr>
<tr>
<td><strong>Reasoning about emotions (12 studies)</strong></td>
<td>View a series of pictures with two persons. One person realizes that he/she mistakenly had a false belief regarding an emotion-triggering state of affairs. That leads to a change in his/her emotions. Indicate whether the pictures show a social (i.e., emotional) change, a physical change, or no change (button press).</td>
<td>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 next pictures (so the person does not realize that he/she had a false belief). Indicate whether the pictures show a social (i.e., emotional) change, a physical change, or no change (button press).</td>
</tr>
<tr>
<td>Hooker, Verosky, Germine, Knight, &amp; D’Esposito, 2010</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Völlm et al., 2006</td>
<td>View a cartoon story showing two persons. One is in an emotion-triggering situation. Predict what the other person will do to make this person feel better (button press).</td>
<td>View a cartoon story showing two persons in a neutral everyday situation. Predict what will happen next based on physical causality (button press).</td>
</tr>
</tbody>
</table>

**Note.** For a complete list, see the online supplemental materials. VAS = Visual Analog Scale (allows ratings along a continuous scale).
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 they were 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).

**Reasoning about emotions.** The intersection between empathy and ToM is a recurrent topic in research, bearing concepts such as cognitive empathy (Hooker et al., 2010; Preston & de Waal, 2002; Schnell, Blaschke, Konradt, & Walter, 2011; Shamay-Tsoory et al., 2009), affective ToM (Mier et al., 2010; Schlafke et al., 2015; Sebastian et al., 2012), mentalizing about emotion (Hooker Verosky, Germine, Knight, & D’Esposito, 2008), and emotional perspective taking (Dennet al., 2010, 2012). Across the diversity of labels, we identified several 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. Another 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 was 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 empathy and ToM literature, we cover this task group in the section on empathy tasks. This is because a (weak) majority of these studies were found by empathy-related keywords. Note, however, that labeling 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.

**Clustering**

Based on the meta-analytic result maps obtained for all 11 task groups, we carried out clustering. This is the central step of our analysis, which allows us to estimate an appropriate number of subcomponents of ToM and empathy. For an overview of correspondences between task groups, we show the image dissimilarity of the 11 meta-analyses in Figure 4A. In Figure 4B, we present a dendrogram that illustrates task-by-task and cluster-by-cluster relations. 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, the components of a good model should be sufficiently abstract to generalize across concrete instances, that is, multiple task groups, rather than picking up variance related to one outlier task group. Metrics show that among all clusterings, the three-cluster solution shows the best performance (see Figure 3A). This result is of central relevance to our aim to find common neurocognitive components across ToM and empathy tasks. We will argue throughout the next sections and in our Conclusion section (A Hierarchical Perspective) that the three-cluster solution reflects central processes for social cognition.

Although we also observed that the simpler two-cluster solution already explains part of the variation in brain activation, the three-cluster solution explains substantial additional variance. 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 values at model orders 8 and 11, with the 11-cluster solution performing particularly well. As we will explain 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 multilayered or multilevel phenomenon (De Waal, 2012; Preston & De Waal, 2002; Schaafsm a et al., 2015; Singer, 2006). 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 two-, three-, and 11-cluster solution. Finally, we will integrate them by proposing a multilevel model of social–cognitive processes in our Conclusion section (A Hierarchical Perspective). Also note that the multilevel nature of our results evades potential concerns regarding publication bias. For example, a common form of publication bias is the tendency to preferentially report results that correspond to well-established standard findings in a field. Arguably, this may drive increased coherence within empathy and ToM studies, respectively, and thus inflate evidence for a two-cluster solution (empathy-vs.-ToM). However, such a publication bias cannot account for our three- and 11-cluster solutions, because they cover results across empathy and ToM fields.

**High-Level Clusterings**

**Two-cluster solution.** The two-cluster solution shows that our approach could retrace large parts of the classical ToM versus 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 (see 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). Of six task groups retrieved from the ToM literature, five ended up in a common cluster. On the other hand, the remaining one consisted of empathy tasks that were different from previous task groups presenting (over) simplified stimuli. This is because a (weak) majority of these studies were found by empathy-related keywords. Note, however, that labeling 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.

---

12 Although the twofold clustering fell out of range for our metric calculation, its explanatory power is shown in the dendrogram in Figure 4B. Because of 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). In addition, it adds another division (by splitting the cluster containing mostly ToM tasks in two halves), which explains substantial additional variance.
hand, three of five task groups from the empathy literature were grouped together. Figure 5 illustrates how the remaining three task groups end up at somewhat unexpected positions in higher levels of the clustering.

The first group was the Reading the Mind in the Eyes task, which researchers originally described as “an advanced test of theory of mind” (e.g., Baron-Cohen, Ring, et al., 1999). 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) largely overlapped 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 the Mind in the Eyes task performance 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 the Behavioral Separability of Tasks From the Three-Cluster Solution section, task performance correlations in nonimpaired participants link the Reading the 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: Dziobek et al., 2006; Rice, Anderson, Veloskey, Thompson, & Redcay, 2016). In addition, a link to more intermediate tasks combining cognitive and affective elements has been reported (Ferguson & Austin, 2010; see the Behavioral Separability of Tasks From the Three-Cluster Solution section 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 the Intermediate Cluster (Cluster 2) section. While Evaluating Situated Emotions tasks have been labeled 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 (Schlafke 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 labeling as empathy or Theory of Mind.

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 (A Hierarchical Perspective), 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 (unthresholded) cognitive and affective cluster maps ($r = .46$) compared with correlations between the intermediate and other two maps (intermediate-cognitive $r = .80$, intermediate-affective $r = .76$).

For the functional description of the three clusters, 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 nonsocial terms is shown in Figures 6 and 7. Based on this broad decoding, we can discuss our findings in the context of not only 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, 2018; Heyes & Frith, 2014) and empathy (e.g., Barrett, Lindquist, & Gendron, 2007; Lindquist, Satpute, & Gendron, 2015; Wager et al., 2016; but see Lieberman, Burns, Torre, & Eisenberger, 2016; Lieberman & Eisenberger, 2015).

Cognitive cluster (Cluster 1). Brain activation for the cognitive cluster (see Figure 5) was mainly found in the cortical midline and temporoparietal areas. Strongest activation was found in the anterior cingulate cortex and medial prefrontal cortex. This activation cluster extended along the cortical midline to the precuneus and parts of the midcingulate cortex. Bilateral temporoparietal 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 with a resting-state network atlas of the brain (Yeo et al., 2011), activations for cluster 1 showed the most prominent overlap with the default mode network (DMN). More specifically, 56% of voxels in cluster 1 fell within the DMN, followed by smaller overlaps with the frontoparietal network (9%) and the ventral attention network (also known as the salience network, 9%).

Neurosynth decoding (see 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 nonsocial terms, the strongest associations were found for default, self-referential, and autobiographical. These later decoding results mirror neurocognitive accounts of understanding others that emphasize the role of the DMN (see Buckner & Carroll, 2007; Bzdok et al., 2013; Mars, Neubert, et al., 2012; Mars, Sallet, et al., 2012; Meyer, Davachi, Ochsner, & Lieberman, 2019; Spreng, Mar, & Kim, 2009; Spreng et al., 2010; Spunt, Meyer, & Lieberman, 2015).

It has been argued (e.g., Andrews-Hanna, Smallwood, et al., 2014) that self-generated cognition decoupled from the physical world is mediated by the DMN. This becomes relevant for ToM because 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- and self-related mental state reasoning (e.g., Mitchell et al., 2005; Murray, Schaer, & Debbané, 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 et al., 2004), the processing of covert (i.e., unobservable) mental states (Gobbini, Koralek, Bryan, Montgomery, & Haxby, 2007), or perspective differences (Perner & Leekam, 2008; Perner & Roesler, 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. Whereas 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 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 the 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, Vawderdyc, & 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).

**Affective cluster (Cluster 3).** For the affective cluster, we found brain activation (see red in Figure 5) across the right frontal cortex, peaking in the inferior frontal gyrus, and extending to the right insula and temporal pole, pre- and postcentral gyrus, as well as the supramarginal gyrus. Further areas were activated in the left inferior frontal gyrus, insula, temporal areas, and supramarginal gyrus. Another large area was activated in the supplementary motor area and the adjacent medial frontal gyrus and midcingulate cortex. Two smaller areas were also found in the 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 DMN (14%). The most prominent social terms found with neurosynth decoding were pain, fear, affective, and face (as well as the clinical term asd – autism spectrum disorder).

For nonsocial terms, we found high loadings on several language-related terms, such as word, phonological, language, or semantic (for a discussion of the possible roles of language processes, see the Possible Roles of Language Processes in the Three-Cluster Solution section). 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 the anterior insula and anterior cingulate cortex (Jabbi, Swart, & Keysers, 2007; Jackson, Meltzoff, & Decety, 2005; Kanske, Böckler, Trautwein, & Singer, 2015; Singer et al., 2004; 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 to be active not only 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; Gallese & Goldman, 1998; Preston & de Waal, 2002; see also Alcalá-López, Vogeley, Binkofski, & Bzdok, 2019). In addition to the left insula, the affective cluster showed activation in the left inferior frontal/precentral gyrus and postcentral (somatosensory) and supramarginal gyri, which again have been linked to shared networks. In particular, the premotor cortex has been linked to the mirroring of emotional expressions, that is, the covert (or overt) imitation of observed emotional facial expressions (cf. Adolphs, 2009; Carr et al., 2003; Decety & Jackson, 2004; Gazzola, Aziz-Zadeh, & Keysers, 2006; Pfeifer & Dapretto, 2009; Shamay-Tsoory et al., 2009). The shared-networks hypothesis was embedded in the general framework of mirror neurons (Gallese, 2003; Gallese & Goldman, 1998; RizzolATTI & Craighero, 2004) and is compatible with theories of common coding for action and perception (Keysers & Gazzola, 2009; Keysers, Kaas, & Gazzola, 2010) and perception-action models (e.g., de Waal & Preston, 2017; Preston, 2007; Preston & de Waal, 2002; Preston & Hofefich, 2012). 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 the 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 the pain or emotions of others. The cluster contained those three 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 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), that is, in the absence of explicit task instructions (see also Cracco et al., 2018; Heyes, 2011). Our results show that 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.

**Intermediate cluster (Cluster 2).** Whereas 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 gyr to the anterior temporal lobes. We observed activation in areas overlapping with parts of the cognitive cluster (cluster 1), including the bilateral temporoparietal cortex and precuneus. Furthermore, we also found some convergence among activations found for the affective cluster, such as in the 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. Analyses confirmed significant activation specific for the cognitive and intermediate clusters in the precuneus, dorsal-posterior medial prefrontal cortex, bilateral temporoparietal, and anterior temporal areas (see Figure 7). Overall, these activations were more pronounced in the right hemispheric areas. For the intermediate and affective clusters, the strongest common activations were found in the left insula and furthermore in the bilateral inferior frontal gyrus, 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 with both other clusters: sentences, speech, listening, and comprehension. When decoding activations common to the intermediate and cognitive clusters (see Figure 7), we found social terms such as theory of mind, mentalizing, and mental states and nonsocial 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 nonsocial terms such as phonological, reading, word, and language.

Together, meta-analytic contrast maps and decoding results (see 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 also suggests a common process. This common process could account for this combination, the fact that we found one intermediate cluster and affective processes. Although theories proposed several variants (e.g., Rizzolatti & Craighero, 2004), it was hypothesized that 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 in emotion processing and categorization. It has been repeatedly found that brain networks for emotion processing overlap with large parts of networks for semantic cognition (Binder, Desai, Graves, & Conant, 2009; Lindquist, Wager, Kober, Bliss-Moreau, & Barrett, 2012; see also Brooks et al., 2017). Thus, 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, the role of semantic knowledge has been discussed in ToM related accounts, and previous meta-analyses (e.g.,

---

13 To find specific commonalities in the intermediate and cognitive clusters (i.e. that go beyond what is also found for the affective cluster), we computed the conjunction of contrasts cognitive > intermediate & affective > affective. Likewise, for specific commonalities between the intermediate and affective clusters, we computed intermediate > cognitive & affective > cognitive. For completeness, we also computed the contrast cognitive > intermediate & affective > intermediate, which found no significant activations.
Andrews-Hanna, Smallwood, et al., 2014; Mar, 2011) showed widespread overlap between default mode and language-related processes. For example, it was hypothesized that ToM involves mental-state concepts that we learn (as children) by communicating with expert mind readers (e.g., Heyes & Frith, 2014). Moreover, some forms of mentalizing have been linked to the 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.

**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-analysis14) for reports of performance correlations among social cognition tasks. If the three clusters we found on the neural

---

14 We reviewed independent literature to find behavioral correlations. All studies we report here are different from the neuroimaging studies in our meta-analysis except Kanske et al. (2015). Because of its large sample size and thus robust results, we also discuss the behavioral data of this neuroimaging study.
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 clusters 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 clusters. 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 clusters, as well as the affective and intermediate clusters.

With respect to the first prediction, behavioral studies clearly support the independence between processes associated with the cognitive and affective clusters. Kanske, Böckler, Trautwein, Parianen-Lesemann, and Singer (2016) tested both belief reasoning (cognitive cluster) and evaluation of emotions (affective cluster) in a large sample of nonimpaired adults using a combined-task setup (the EmPaToM task). The authors found that behavioral performance in these two measures was uncorrelated. Also, when additional data-driven composites of several ToM and empathy tasks were used, no association was found. Similarly, no correlation was found between performance in the Reading the Mind in the Eyes (affective) and Strange Stories (cognitive, because strange stories present various belief-related contents, see, e.g., Happé, 1994; White, Hill, Happe, & Frith, 2009) tasks. This observation was made both for 7- to 12-year-old children (Rice et al., 2016) and for 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, several studies report relevant results. Regarding associations between tasks from the affective and intermediate clusters, Lockwood, Bird, Bridge, and Viding (2013) found a positive correlation between accuracy for ratings of social animations in our intermediate cluster and judgments of one’s affective reaction to emotional faces (affective cluster). Interestingly, correlations were also found between behavioral motion perception (affective/intermediate cluster) and the Reading the Mind in the Eyes task (affective) both in adults (Miller & Saygin, 2013) and 7- to 12-year-old children (Rice et al., 2016).15 Biological motion perception can be linked to the affective cluster because of its relation to action perception (and thus perception-action cycles for affect sharing; see, e.g., de Waal & Preston, 2017; Keysers et al., 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 most prominently 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, 1999). Understanding a faux pas (e.g., observing someone who is hurting another person’s feelings out of ignorance) requires the processing of not only 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 and Austin (2010) found an association in performance between the Faux Pas test (intermediate) and the Reading the Mind in the Eyes task (affective) in a large sample of nonimpaired adults.

With respect to correlations between tasks from cognitive and intermediate clusters, results are more mixed. In a sample of nonimpaired adults, Brewer, Young, and Barnett (2017) found no correlation between behavioral performance on social animations (intermediate cluster) and the Strange Stories task (cognitive cluster). The authors noted, however, that this could be linked to range restrictions in scores (as nonimpaired adults often show ceiling effects on ToM tests linked to our cognitive cluster). Interestingly, when rerunning the correlation analysis for a pooled sample of nonimpaired adults and adults with autism spectrum disorder, Brewer et al. (2017) found an association between performance on Social Animations and Strange Stories. Similarly, for 7- to 12-year-old children (Rice et al., 2016), a correlation between behavioral performance in biological motion perception (affective/intermediate) and Strange Stories (cognitive) was found.

### 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 functional connectivity patterns (Margulies et al., 2016). The gradient characterizes continuous changes in these patterns, 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 (85th–95th percentiles), supporting its functional interpretation in terms of abstract, stimulus-independent thought. The affective cluster was located toward the middle of the gradient (35th–55th percentiles), which indicates more sensory-based and unimodal processing.

These findings support our interpretation that the intermediate cluster combines cognitive and affective processes, which is notable

---

15 Also note that our assignment of the Reading the 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.
for several reasons. During passive rest, spontaneous fluctuations of brain activity in corresponding areas and networks (e.g., affective: Ventral Attention Network; cognitive: DMN) have been found to be unrelated (e.g., Alcalá-López 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., 2018). During tasks requiring externally focused attention, an inhibitory relation was found between the ventral attention network and the DMN, which is considered to reflect the former down-regulating the latter to reduce interference from task-unrelated processes (Goulden et al., 2014; Trautwein, Singer, & Kanske, 2016; Wen, Liu, Yao, & Ding, 2013; see also Anticevic et al., 2012). Although 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.

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 (A Hierarchical Perspective). Across the entire spectrum of solutions, our metrics also pointed to two more low-level solutions, with 8 and 11 clusters. This suggests that brain activation contains additional

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 may feature more regions outside our three clusters). Therefore, the relative patterns of overlaps within each topic map are of main interest. See the Relation to Other Meta-Analyses on Social Cognition section for more details. Footnotes: 1 For exploratory purposes, we show an uncorrected map of the meta-analysis results by Darda and Ramsey (2019). 2 Map from Boccadoro et al. (2019) shows results from a multistudy analysis on spontaneous ToM (not a meta-analysis, based on original data), three studies, n = 68. 3 Map from Cheong et al. (2017) shows the most consistent overlaps across five published studies, summarized based on expanding reported coordinates into 15-mm-radius spheres. See the online article for the color version of this figure.
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 heteroge-
nity includes differences in stimuli and task instructions, which
might be unrelated to central processes linked to empathy and
toToM. We illustrate the role of stimuli and task instructions in our
additional color coding in Figure 5. For the sake of brevity, we
only map out two popular distinctions: verbal versus nonverbal
stimuli and instructed versus uninstructed tasks.16 Note that sev-
eral other (and partially related) categorizations have been dis-
cussed elsewhere but go beyond the scope of our exemplary
illustration (see, e.g., Fan et al., 2011; Mar, 2011; 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 3 contains tasks
from only one category (e.g., only verbal tasks).17 At lower levels,
parts of subclusterings are accompanied by separations in terms of
stimulus or stimulus formats. For example, the three task groups False
belief, trait judgments, and strategic games cluster together at a
higher level (three-cluster solution), but strategic games are sepa-
rated at a lower level (eight-cluster solution). These tasks differ in
that two of the former are verbal (sentences or words) whereas the
latter is nonverbal (strategic decision-making). Although 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 reading 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 ToM liter-
ature. 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 although stimulus and task format drives part of the variability
in meta-analytic activations, it does not provide a complete ac-
count of them. As we will argue in detail in our Conclusion section
(A Hierarchical Perspective), 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 eight- and 11-cluster
solutions (lower level) capture additional variability, which possi-
bly describes how central processes are applied in concrete con-
texts of particular tasks. For such task contexts, both the stimulus
format and the task instructions play important roles.

Relation to Other Meta-Analyses on Social Cognition

The present meta-analysis and review focused on processes
involved in inferring others’ unobservable mental states and thus
on the literature on empathy and ToM. 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”). Fur-
thermore, we did not include tasks which involved mental state
inference as a subcomponent employed alongside other diverse
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 focus-
ning on whole-body movement such as walking or dancing (Gros-
bras, Beaton, & Eickhoff, 2012), we found the highest overlap with
the intermediate cluster. Moreover, the map also showed overlaps
with 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 observa-
tion or mirroring (Hardwick, Caspers, Eickhoff, & Swinnen,
2018), we illustrate overlaps specifically for meta-analysis on (a)
all types of action observation, (b) observation of actions per-
formed with arms, and (c) observation of actions performed with
faces. For all three maps, we found the largest overlap with the
affective cluster and, to a 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., 2015), we again found preferential overlaps for the affective
and intermediate clusters.

Next, we overlaid our three-cluster solution to maps from rel-
ated 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 several other faces, overlap was
clearly highest with our affective cluster. For an additional meta-
analysis on emotion labeling (Dricu & Frühholz, 2020), we found
considerable overlaps with both the affective and intermediate
clusters. 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 pro-
cessing (Morelli, Sacchet, & Zaki, 2015; see also Apps, Rush-
worth, & 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 a pleasant touch.

We further included a meta-analysis on Social Exclusion tasks
(Vijayakumar, Cheng, & Pfeifer, 2017) in our overlap analysis.
Here, participants experienced being rejected mostly in the context
of a virtual ballgame (Cyberball), and some additional social
judgment/chatroom contexts. Overlaps for this meta-analysis were
mainly found with our cognitive cluster and, to a lesser extent, also
with our affective cluster. For a multistudy analysis (n = 68) on
Spontaneous ToM (Boccadoro et al., 2019), we found equal de-
grees of overlap with our three clusters, and thus, no preference for
affective, intermediate, or cognitive clusters became clear. Tasks

16 We define verbal stimuli as items of written or spoken language that
contain task-relevant information and go beyond trivial task cues (e.g. recur-
rning 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.

17 Figure 5 shows that at model order 3, the cognitive cluster contains
7 verbal tasks; the intermediate cluster contains all tasks that explicitly instruct participants to infer, judge, or think about the mental states (cognitive or affective) of others.
in this multistudy 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, thus developing a false belief about where a certain object has been placed (e.g., inside vs. outside a box). Somewhat similarly, for a meta-analysis on Visual Perspective Taking (Schurz, Aichhorn, Martin, & Perner, 2013), we found no prominent overlap with any one of our three clusters (but see Dumontheil, Küster, Apperly, & Blakemore, 2010; Santiesteban, Banissy, Catmurt, & Bird, 2012). Visual Perspective Taking tasks require participants to judge what another person can see or how another person sees an object. For a summary map of studies on the computational modeling of ToM (Cheong, Jolly, Sul, & Chang, 2017), that is, studies using game-theoretical 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 a 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 intermediate clusters and comparatively little for the affective cluster. Moral Cognition tasks featured scenarios with moral violations or dilemmas and required participants to make appropriate judgments on the actions of one individual toward others.

Summary and Conclusion

A Hierarchical Perspective

More than two decades of neuroimaging and behavioral research have produced substantial 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 an increasing awareness of disagreement concerning the concepts and taxonomy underlying social processes (Happé et al., 2017; Schaafsma et al., 2015; Spunt & Adolphs, 2017). This meta-analysis aimed to support the development of a coherent and balanced theoretical model of major cognitive factors underlying the ability to understand other people’s minds. We have argued that neuroimaging data provides a good starting point for such a model, as other sources—such as behavioral data—have yet to provide a complete picture of task-by-task interrelations. As we have reviewed in the Behavioral Separability of Tasks From the Three-Cluster Solution section, behavioral intercorrelations were only studied for a limited number of tasks, some of which suffer from range restrictions in scores (e.g., nonimpaired 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 multilayered or multilevel phenomenon (De Waal, 2012; Preston & De Waal, 2002; Schaafsma et al., 2015; Singer, 2006). 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 the data were divided into a three-cluster solution. However, further local peaks in performance were found when the data were split into eight and 11 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 modeling social cognition by multilevel theories, similar to models in other fields such as intelligence research. There, some accounts feature a central construct which reflects a latent variable that indicates a broader, more abstract neurocognitive function (e.g., dynamic control via a multiple-demand network; Duncan, 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 trying to guess a person’s mental state based on their facial expression). Modeling our data as a multilevel construct accounts for the diversity in how social abilities manifest as well as for the specific demands placed by its particular instances (e.g., experimental tasks). To illustrate, several accounts have described mechanisms targeted at explaining false-belief understanding, which is assumed to be a hallmark of human social cognition (e.g., Premack & Woodruff, 1978; Saxe & Kanwisher, 2003; Wimmer & Perner, 1983). 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 (a) a mechanism for decoupling false beliefs is not perfectly applicable to other task contexts, but (b) a broader functional class of self-generated thought in social cognition forms the basis for belief decoupling (but does not completely specify it).

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. Whereas 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. The cooccurrence of cognitive and affective processes has been linked to more naturalistic forms of social cognition. For instance, 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 discussing emotional autobiographical events. After filming, the person shown in the video had been asked to rate their own emotional states during the clip on a moment-to-moment basis. In addition, participants observing 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 assessment 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 Evaluating Situated Emotions task group, which ended up in the intermediate cluster of our analysis.

Another central case of coactivation of cognitive and affective processes are everyday social interactions, which are essential for understanding other minds (see Schilbach et al., 2013). For example, Schilbach, Eickhoff, Mojszisch, and Vogeley (2008) recorded facial mimicry-related brain activity during online social interactions and found conjoint activation in the cortical midline (e.g., precuneus) and motor areas (e.g., precentral gyrus). This pattern of coactivated areas resembles our intermediate cluster, that is, the combination of cognitive and affective processes. Also, other studies presenting scenes or videos of social interactions reported a coactivation of cognitive and affective processes (e.g., Deuse et al., 2016; Wolf, Dziolek, & Heekeren, 2010). Note, however, that both naturalistic social cognition (see Zaki & Ochsner, 2012) and social interaction (Schilbach et al., 2013) were argued to engage additional processes not covered by the cognitive and affective clusters (e.g., reward-related areas during social interactions). Another group of tasks where coactivation of cognitive and affective clusters was observed is altruistic decisions (Hare et al., 2010; Tüsche, Böckler, Kanske, Trautwein, & Singer, 2016; see also Hein, Silani, Preuschlof, Batson, & Singer, 2010). Tüsche et al. (2016) found that the level of generosity in donations for charitable organizations was predicted by brain activity patterns in both the right temporoparietal junction (TPJ) and right anterior insula during that task, again indicating a cooccurrence of cognitive and affective processes.

Taken together, the reviewed studies suggest that the coactivation 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) engages cognitive and affective processes concurrently. 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 the cognitive (e.g., strange stories) and affective clusters (e.g., facial emotion recognition) tasks. This has implications for selecting tasks when measuring different aspects of social cognition. In fact, tasks linked to the intermediate cluster often show good clinical discrimination, such as the Faux Pas test (e.g., for autism spectrum: Baron-Cohen, O’Riordan, et al., 1999; frontotemporal dementia: Gregory et al., 2002; schizophrenia: Konstantakopoulos et al., 2014) or social animations (e.g., for autism spectrum: White, Conston, Rogers, & Frith, 2011; schizophrenia: Bliksted et al., 2019). Although these findings put the spotlight on tasks from the intermediate cluster with respect to ecological and clinical relevance, they also trigger a novel question regarding their interpretation. Because such tasks contain both cognitive and affective processes, their interindividual differences could reflect different sources—a difference in cognitive processes only, affective processes only, or both (combined). Therefore, we suggest to select tasks according to the three-cluster solution identified in this meta-analysis and ideally present a combination of them (cognitive, affective, and intermediate), because this may increase the ability to evaluate separate as well as conjoint social processes.

Outlook

The capacity to understand what other people think and feel remains one of the most elusive mental faculties. We have summarized neuroimaging research from more than two decades (188 studies, 4,207 subjects), based on which we propose to model social cognition as a hierarchical multilevel construct. Thereby, we are seeking to accommodate the diversity of processes and features of social cognition (on a lower level), while making explicit a level of coherence among them (i.e., a higher level). Based on our research focus and the evidence available to date, this meta-analysis has produced a first sample of processes, levels, and mechanisms that such a model might have. At the highest level of our model, we suggest to capture social-cognitive processes in terms of two overarching networks which are flexibly combined and relate to more sensory-affective versus more abstract and decoupled representations of others’ mental states.

References

References marked with an asterisk indicate studies included in the meta-analysis.


Singer, T. (2006). The neuronal basis and ontogeny of empathy and mind


Spreng, R. N., & Grady, C. L. (2010). Patterns of brain activity supporting


Toward A Hierarchical Model of Social Cognition

Received July 10, 2019
Revision received July 17, 2020
Accepted July 21, 2020