Neurocognitive mechanisms of theory of mind impairment in neurodegeneration: a transdiagnostic approach

Cherie Strikwerda-Brown1–3
Siddharth Ramanan1–3
Muireann Irish1–3
1The University of Sydney, Sydney, NSW, Australia; 2The University of Sydney, School of Psychology, Sydney, NSW, Australia; 3ARC Centre of Excellence in Cognition and its Disorders, Sydney, NSW, Australia

Abstract: Much of human interaction is predicated upon our innate capacity to infer the thoughts, beliefs, emotions, and perspectives of others, in short, to possess a “theory of mind” (ToM). While the term has evolved considerably since its inception, ToM encompasses our unique ability to apprehend the mental states of others, enabling us to anticipate and predict subsequent behavior. From a developmental perspective, ToM has been a topic of keen research interest, with numerous studies seeking to explicate the origins of this fundamental capacity and its disruption in developmental disorders such as autism. The study of ToM at the opposite end of the lifespan, however, is paradoxically new born, emerging as a topic of interest in its own right comparatively recently. Here, we consider the unique insights afforded by studying ToM capacity in neurodegenerative disorders. Arguing from a novel, transdiagnostic perspective, we consider how ToM vulnerability reflects the progressive degradation of neural circuits specialized for an array of higher-order cognitive processes. This mechanistic approach enables us to consider the common and unique neurocognitive mechanisms that underpin ToM dysfunction across neurodegenerative disorders and for the first time examine its relation to behavioral disturbances across social, intimate, legal, and criminal settings. As such, we aim to provide a comprehensive overview of ToM research in neurodegeneration, the resultant challenges for family members, clinicians, and the legal profession, and future directions worthy of exploration.

Keywords: prefrontal cortex, social cognition, mentalizing, executive function, dementia, empathy

Introduction

The term “theory of mind” (ToM) was first coined in 1978 to refer to an individual’s capacity to attribute mental states to oneself and others.1 Since then, over 1,450 studies have been published on this topic, exploring the cognitive and neural mechanisms that support this capacity and its disruption in clinical populations (Figure 1). The definition has since been updated to incorporate a broader range of abilities, including understanding and inferring the thoughts, feelings, beliefs, and intentions of ourselves (ie, first-person ToM) and others (ie, third-person ToM).2 Intuitively, the conceptualization of ToM overlaps with the related construct of empathy, which involves the basic recognition and understanding of another person’s affective state, in addition to the sharing of this emotional experience.3 The precise manner in which these two processes are related, however, remains a matter of debate.3,4 One argument holds that ToM is a “domain-specific” ability, distinct from other cognitive functions,6 providing the foundational mechanism upon which other complex social processes, such as empathy,6 are built. By contrast, the “domain-general” theory proposes that ToM
involves the understanding of representations in general, invoking various component cognitive processes, some of which are shared with empathy\textsuperscript{3} and other social functions.\textsuperscript{7} Accordingly, affective ToM (ie, feelings and emotions) is separable from its cognitive counterpart (ie, thoughts, beliefs, and intentions).\textsuperscript{8}

As shall be demonstrated, this fractionation between cognitive and affective ToM has proven particularly useful in characterizing the nature of social cognitive dysfunction in neurodegeneration,\textsuperscript{9} as well as in differentiating between dementia syndromes.\textsuperscript{10} Nonetheless, parcellating dynamic and multifaceted social interactions into exclusive categories remains challenging,\textsuperscript{11} as cognitive and affective ToM are themselves multifaceted constructs, which are not always uniformly affected in neurodegeneration.\textsuperscript{12,13} Moreover, according to the domain-general account, multiple cognitive abilities support ToM capacity, including executive function,\textsuperscript{14,15} memory,\textsuperscript{16} language,\textsuperscript{17} and visuospatial skills,\textsuperscript{18} a pertinent issue to consider in the context of neurodegenerative disorders characterized by widespread cognitive impairment. Here, we provide an update on ToM research in neurodegeneration, considering how large-scale brain network dysfunction, and resultant cognitive impairment, impacts cognitive and affective expressions of ToM. In doing so, we aim to explicate the common and divergent neurocognitive mechanisms that sublend ToM dysfunction across neurodegenerative disorders and adjudicate between domain-specific vs domain-general accounts of ToM impairments in these syndromes.

**Why study ToM in neurodegenerative disorders?**

Functional neuroimaging studies in healthy individuals, complemented by lesion evidence from clinical populations, have accelerated our understanding of the neural circuitry supporting ToM performance. In the last decade, however, we have witnessed a shift from understanding the roles of localized brain regions to considering interactions between large-scale neural networks that support complex cognitive endeavors such as ToM (Figure 2). The study of neurodegenerative disorders offers compelling insights into the neurocognitive architecture of ToM, as these syndromes target large-scale brain networks implicated in ToM and many of its associated processes (Figure 3).\textsuperscript{19} Moreover, neurodegenerative disorders frequently present with co-occurring social cognitive, memory, and executive impairments, offering an opportunity to explore the intersection between ToM and cognitive function more broadly. In this review, we focus on the most common neurodegenerative syndromes that present with early impairments in social cognition, as well as emerging evidence of ToM dysfunction in syndromes not traditionally classified as disorders of social cognition.

---

**Figure 1** The exponential increase in number of published articles on theory of mind since its inception.

**Notes:** Asterisk (*) refers to the first appearance of the term “theory of mind”, coined by Premack and Woodruff in 1978. In total, a PubMed search returned 1,454 studies containing the term “theory of mind” in the title as on November 23, 2018. PubMed search was limited to empirical studies and review papers, published in English, on and after 1978.
Figure 2 A distributed brain network supporting ToM reasoning, highlighting the corresponding putative general cognitive function and ToM-specific roles of each region. Note: Lighter shades indicate medially located regions, whereas darker shades indicate laterally located regions. 

**Abbreviations:** ACC, anterior cingulate cortex; ATL, anterior temporal lobe; fN, frontoinsular; mPFC, medial prefrontal cortex; PCC, posterior cingulate cortex; TPJ, temporoparietal junction; ToM, theory of mind.

**Behavioral-variant of frontotemporal dementia (bvFTD)**

BvFTD represents the prototypical disorder of social cognition. Characterized by insidious changes in personality and behavior, patients with bvFTD display flagrant violation of social norms, lack of social comportment, and apparent loss of empathy for others. The considerable parallels in social cognitive difficulties between bvFTD and developmental disorders, such as autism, provided an early clue that ToM might be altered in this dementia syndrome (John R Hodges, personal communication). From a neuroanatomical perspective, this observation prior to the advent of modern day neuroimaging techniques was particularly astute, given the now well-established vulnerability of large-scale networks specialized for socioemotional processing in bvFTD.

The emergence of prominent ToM impairments as manifested in socially disruptive and inappropriate behaviors typically heralds the onset of bvFTD. Family members frequently report the affected individual to lack warmth, have an apparent disregard for others, and to display increasingly rigid and egocentric behavior. Collectively, these disturbances are posited to reflect a core, syndrome-specific difficulty in ToM, which in turn profoundly increases carer burden. The observation of significant cognitive and affective ToM disruption irrespective of modality of testing (ie, verbal or nonverbal) points to a primary ToM impairment in bvFTD. The evidence to date corroborates this position, with bvFTD patients presenting marked deficits across the broad spectrum of ToM tasks including first-order (ie, understanding others’ false beliefs about the world) and second-order false belief (ie, understanding others’ false beliefs about a third party’s mental state), detection of social faux pas, intention and emotion attribution on short video vignettes, and decoding emotion from visual cues (Reading the Mind in the Eyes Task; RMET).

**Mechanisms of ToM disruption in bvFTD**

Arriving at a precise understanding of the cognitive origins of ToM disruption in bvFTD has proven challenging, in part due to the widespread nature of cognitive impairment in this syndrome. As such, ToM impairments may arise simply as a consequence of global cognitive and executive dysfunction, rather than a primary socioemotional impairment per se. By this view, bvFTD patients would be predicted to perform poorly on ToM tasks that demand greater executive resources. Paradoxically, however, bvFTD patients...
demonstrate deficits on even simple first-order false belief tasks12 as well as nonverbal ToM tasks with limited cognitive and executive loading,16 suggesting a fundamental, domain-specific impairment. Crucially, ToM impairments in bvFTD emerge statistically independent of overall cognitive and executive status when data-driven clustering and prediction models are employed,14,37 suggesting that social cognitive deficits are dissociable from general cognitive decline in this syndrome. The observation of significant difficulties on ToM, but not on related control tasks,38 reinforces the presence of a core ToM impairment, independent of co-occurring executive, memory, language, and other general cognitive deficits in this syndrome.14,25,37

Could ToM impairments arise simply as a function of widespread socioemotional dysfunction characteristic of the bvFTD syndrome? A recent meta-analysis suggests that ToM and emotion processing impairments are present to a commensurate degree.25 It is possible that the deterioration of affective ToM abilities co-occurs with a general decline in emotion processing in bvFTD,27,39 reflecting the shared neural structures implicated in these abilities. Importantly, however, the presence of early deficits on cognitive, as well as affective, ToM tasks in bvFTD suggests that emotion processing is unlikely to entirely account for disrupted ToM in this syndrome. Other mechanisms, such as an inability to integrate social and contextual cues, likely also contribute to ToM dysfunction in bvFTD.40 Furthermore, it remains unclear whether the ToM impairment in bvFTD reflects a general incapacity to inhibit one’s own mental perspective41 or adopt any perspective beyond the “here and now”,41,42 resonating with recent reports of egocentric, rigid behavior, and environmental dependency in this syndrome.43,44 As such, while ToM impairments in bvFTD appear to manifest independent of memory and executive functions, the precise contribution of socioemotional processes remains to be fully elucidated.

Neuroanatomy of ToM impairments in bvFTD

A common assumption is that ToM dysfunction in bvFTD reflects early atrophy to the medial prefrontal cortex (mPFC), one of the core regions of the ToM brain network.
(Figure 2).

Reduced gray matter intensity of mPFC and neighboring prefrontal regions has been linked to performance impairments on the ToM component of the Frith–Happé animation task. Similarly, a reduction in coherent, resting state mPFC activity has been found to correlate with poorer performance on emotion attribution tasks employing short story vignettes. Finally, modulative neurostimulation via transcranial direct current stimulation over the mPFC has been demonstrated to improve accuracy on cognitive ToM (measured using intention attribution) tasks in bvFTD, suggesting a causal role for this region in ToM performance.

Neurodegenerative disorders, however, rarely target discrete regions in isolation, and progressive neural degradation in bvFTD reveals the importance of regions beyond the mPFC in modulating ToM performance. Emerging evidence suggests that in parallel with early atrophy of the mPFC, anterior cingulate cortices/frontoinsular (ACC/fIN) regions are particularly vulnerable to functional disruption and volumetric reduction in this syndrome. The importance of ACC/fIN regions in the context of ToM impairments in bvFTD deserves particular attention (Figure 3). A prominent hypothesis contends that accumulation of pathology around the ACC/fIN regions in bvFTD directly targets a specific population of neurons called von Economo neurons (VENs). VENs are an evolutionarily specialized set of cortical neurons suggested to play a critical role in ToM capacity, possibly through supporting the selection of quick and intuitive responses during uncertain social situations. Furthermore, the ACC/fIN cortices, along with the mPFC, are implicated in regulating and selecting context-specific emotional responses, and their degradation, in part, may contribute to co-occurring emotion dysregulation and ToM impairments in bvFTD. Finally, the ACC/fIN regions anchor the Salience Network of the brain, proposed to tag incoming salient information for further processing, thereby guiding appropriate behavioral and social responses. Degeneration of ACC/fIN regions reduces VEN counts, disrupts Salience Network functioning, and is directly implicated in ToM dysfunction in bvFTD. Atrophy of the ACC/fIN cortices in bvFTD is associated with compromised cognitive ToM on the Frith–Happé animations and affective ToM difficulties on emotion attribution tasks. Furthermore, lower emotion attribution scores on affective ToM tasks in bvFTD parametrically relate to reduced connectivity between ACC/fIN, PFC, and temporal cortices, as demonstrated using resting-state functional connectivity, suggesting that alterations in ACC/fIN circuitry disrupt ToM performance in bvFTD.

Looking beyond the prefrontal cortex

We next consider how damage to regions beyond the prefrontal cortex gives rise to ToM dysfunction in bvFTD. With disease progression, atrophy encroaches into the anterior and medial temporal lobes most noticeably on the right-hand side, including the hippocampus and amygdala, parietal cortices including temporoparietal junction (TPJ), precuneus, and posterior cingulate cortices (PCC) as well as subcortical regions including the thalamus and caudate nuclei. Importantly, atrophy to regions beyond the prefrontal cortex is associated with ToM dysfunction in bvFTD. For example, impairments on emotion attribution and faux-pas tasks in bvFTD have been shown to correlate with lateral temporal cortex integrity, including the right anterior temporal lobes (ATLs) and superior temporal cortices, as well as medial and lateral parietal regions such as the TPJ, precuneus, and PCC. Emergent evidence has further implicated cerebellar degeneration in modulating aspects of cognitive dysfunction as well as ToM difficulties on the Frith–Happé animations task. Longitudinal studies tracking how the evolution of ToM dysfunction in bvFTD relates to encroachment of pathology in temporoparietal cortices and the cerebellum are clearly warranted.

Alzheimer’s disease (AD)

At first glance, it seems somewhat counterintuitive to discuss AD in the context of ToM impairment, given its conception primarily as an amnestic disorder. Clinically, AD patients are noted to retain their warmth, affability, and social graces, with social cognitive impairments typically manifesting in later stages of the disease course. As a result, AD patients were historically included in ToM studies as disease control groups rather than syndromes of interest. When mounting evidence of episodic memory impairments in bvFTD, ToM performance was further suggested to hold high diagnostic utility in the discrimination of bvFTD from AD. The emerging picture, however, is less clear, with mixed reports of moderate impairments on social cognitive tasks in AD, modulated largely by the cognitive demands of the task itself.

Mechanisms of ToM disruption in AD

Two competing hypotheses have been advanced in relation to ToM disturbances in AD. One position holds that AD patients display a fundamental, domain-specific ToM impairment, in light of evidence for a degraded capacity to accurately infer and attribute beliefs, intentionality, and emotional
states to characters on some studies of cognitive (eg, false belief tasks\textsuperscript{70,71}) and affective (eg, RMET\textsuperscript{16}) ToM. Other findings reveal significant impairments in AD in inferring mental exchanges of interlocutors from simple conversational exchanges in vignettes\textsuperscript{72} as well as in real life.\textsuperscript{73} The observation that impairment in attributing physical causality to characters in first-order false belief tasks is independent of comprehension abilities and executive performance\textsuperscript{70} has been interpreted as reflecting fundamental cognitive and affective ToM impairments in AD even on tasks with relatively low cognitive demand.

On the other hand, a more prominent view holds that ToM impairments in AD reflect a global decline in cognitive processes including episodic and working memory and executive function.\textsuperscript{24,74} In support of this position, general cognitive, executive, and memory performances in AD have been shown to predict subsequent ToM capacity.\textsuperscript{14,74} For example, Ramanan et al\textsuperscript{14} found that between 50% and 70% of cognitive ToM performance in AD (as measured by the faux-pas test) could be explained by patients’ overall attention and executive performance. Similarly, Synn et al\textsuperscript{16} found significant associations between ToM performance on the Frith–Happé animations and episodic memory retrieval impairments in AD. Moreover, AD performance on ToM tasks with low cognitive demands, such as first-order false belief, is relatively spared in the majority of studies,\textsuperscript{32,34,75,76} whereas its more cognitively demanding counterpart (second-order false belief) is impaired.\textsuperscript{72} Interestingly, AD patients are impaired even on control items of second-order false belief tasks\textsuperscript{75} that rely on working memory and comprehension abilities, further suggesting that ToM deficits in the syndrome arise due to the cognitive complexity of the tasks.

Although the aforementioned studies suggest that a decline in executive and memory functions influences ToM performance in AD, the precise nature of their contributions remains unclear. These processes may support the integration and maintenance of relevant information on ToM tasks, as well as facilitating the inference of mental states in the presence or absence of situational cues.\textsuperscript{77,78} With disease progression, increasing executive, visuospatial, and general cognitive dysfunction likely impact ToM performance, especially on tasks that tax visuospatial abilities or lack explicit situational cues.\textsuperscript{16,79} In this regard, the study by Synn et al\textsuperscript{16} is notable in demonstrating comparable difficulties on the ToM and “random movement” conditions of the Frith–Happé animations task, suggesting that ambiguity of stimuli and increasingly abstract task demands contribute to ToM impairments in AD.\textsuperscript{16} By contrast, with the provision of situational and social contexts, ToM deficits in AD are mitigated,\textsuperscript{70} presumably due to attenuation of the executive and mnemonic task demands. In summary, the bulk of evidence points to a domain-general impairment in ToM, attributable to a primary decline in memory, executive, and general cognitive abilities.

**Neuroanatomy of ToM impairments in AD**

Efforts to clarify the neuroanatomical signature of ToM disruption in AD have been limited as social cognitive impairments emerge relatively late in the disease course.\textsuperscript{56} Moreover, the diffuse gray and white matter neural damage at these later stages limits our capacity to attribute emergent impairments exclusively to the degeneration of the ToM network. Two recent studies, however, warrant discussion. First, Le Bouc et al\textsuperscript{35} found ToM performance impairments on a false belief task in a combined cohort of AD and bvFTD patients to correlate with hypometabolism of the left TPJ/ inferior parietal cortex; however, the admixture of bvFTD patients in this analysis precludes any interpretation of these results as specific to AD. Second, Synn et al\textsuperscript{16} revealed significant associations between impaired overall performance on the Frith–Happé animations task and reduced gray matter integrity of the right hippocampus and bilateral cerebellum in AD. In addition to their key role in the ToM network, the TPJ, hippocampus, and cerebellum have also been implicated in general cognitive processes such as memory and executive function (Figure 2). As such, whether the involvement of these regions in ToM impairments in AD reflects a domain-specific or domain-general process remains unclear. For example, the TPJ and neighboring parietal regions play a well-established role in supporting the representation of mental states,\textsuperscript{50} but are also involved in the integration of information pertinent to successful semantic\textsuperscript{81} and episodic memory retrieval.\textsuperscript{52} Although typically discussed in relation to episodic memory dysfunction in AD,\textsuperscript{68} the hippocampus may facilitate ToM by allowing us to draw upon past experiences to anticipate social intentions and reactions\textsuperscript{52} and supporting cognitive flexibility by updating our knowledge of social structure through acquired information.\textsuperscript{84} Finally, cerebellar involvement has been reported during ToM tasks requiring high levels of abstraction, such as thinking about other’s traits,\textsuperscript{85} yet it is also implicated in executive control and working memory capacity.\textsuperscript{86} As such, further research delineating the precise role of these regions in ToM will enable us to adjudicate between domain-specific and domain-general accounts of ToM disruption in AD.
Amyotrophic lateral sclerosis (ALS)
ALS is a neurodegenerative disease involving the degradation of motor neurons in the brainstem, spinal cord, and motor cortex, manifesting in a progressive loss of motor function. Until recently, ALS was considered primarily a motor disease, although cognitive and behavioral changes are now accepted as common features of the condition. While a proportion of ALS patients will also be diagnosed with frontotemporal dementia (FTD), up to 50% of the individuals without dementia develop impairments in executive functioning, language, memory, and behavior, leading to the conceptualization of ALS and FTD as lying on a disease continuum. Accumulating evidence reveals ToM deficits in ALS, not unlike those found in bvFTD, leading to inclusion of deficits in “social cognition” in the revised diagnostic criteria for the ALS-FTD spectrum disorder. The neuroanatomical profiles of ALS and FTD, however, are not identical, with ALS spreading from its origin in the brain stem or spinal cord into motor cortex and subsequently frontotemporal, anterior cingulate, and basal ganglia regions. The cognitive and neurobiological processes underlying impaired ToM in ALS are thus unlikely to precisely mirror those implicated in FTD.

ToM impairments have been uncovered across the majority of tasks in ALS. This incorporates basic affective ToM tasks, involving the attribution of emotions to characters based on vignettes (emotional attribution tasks) or facial expressions (RMET). Performance on more complex measures of cognitive and affective ToM is also compromised, including false belief tasks, understanding thoughts and feelings of cartoon characters (Happé cartoons), detecting faux pas, and attributing intentions to characters in stories (but see Ref. 98). Findings are less consistent for making social inferences from dynamic videos (The Awareness of Social Inference Test; TASIT) and judging others’ preferences based on eye-gaze (Judgment of Preference task), with intact performance revealed by some, but not other, studies. On balance, though, ToM impairments appear relatively pervasive in ALS.

Mechanisms of ToM disruption in ALS
Unsurprisingly, the cognitive mechanisms underlying ToM impairments in ALS vary according to the task in question. The so-called “pure” affective tasks (eg, RMET, emotion attribution) typically do not correlate with executive function in ALS, indicating that degraded affective ToM in ALS arises independently from executive dysfunction. Similarly, executive function is not associated with performance on more complex ToM tasks combining affective and cognitive components, such as faux pas and intention attribution, further suggesting a domain-specific impairment of ToM in ALS. By contrast, executive functions have been consistently implicated in performance on false belief and the Happé cartoon in ALS. This suggests that these latter tasks impose greater executive demands and as such may be less sensitive in detecting “pure” ToM impairment independent of other cognitive confounds. Studies examining the contribution of cognitive processes beyond executive function are limited; however, preliminary findings hint at a role for verbal memory and nonabstract reasoning in emotional attribution and semantic naming in Judgment of Preference.

Neuroanatomy of ToM impairments in ALS
The neural substrates of ToM disruption in ALS appear to be task dependent. Emotion attribution performance is linked to gray matter intensity decrease in the ACC/fIN regions, as well as reduced resting state functional connectivity between PCC and mPFC and between left supramarginal gyrus and right frontoparietal regions. As such, impairments in affective ToM in ALS may result from disruption to a distributed frontotemporo–parietal network. A separate network, however, has been implicated in the more complex process of detecting false beliefs. Poor performance on this task in ALS is associated with reduced PET glucose consumption in bilateral dorsomedial and dorsolateral prefrontal cortex and supplementary motor area, potentially reflecting the executive demands of the test. In addition, in a task-based fMRI study, poor Judgment of Preference performance in ALS was related to reduced activation in the right precentral gyrus extending into the inferior frontal gyrus. Intriguingly, the same study revealed that reduced ToM in ALS was associated with increased brain activity in the right postcentral gyrus. Increased brain activity on PET imaging has also been documented in relation to the false belief task in ALS, with hypermetabolism in the left fusiform gyrus correlating with reduced performance. Together, these findings potentially reflect compensatory activity in the face of widespread degeneration of the ToM network.

Huntington’s disease (HD)
HD is a dominantly inherited neurodegenerative disease affecting motor, psychiatric, and cognitive functioning. Although formal diagnosis of HD is based on the presence of an extrapyramidal movement disorder, the advent of genetic testing has permitted assessment of symptoms in premanifest individuals harboring the CAG-repeat expansion in the Huntington gene, with mounting evidence to suggest...
that cognitive and psychiatric deficits can actually predate motor symptom onset.\(^{105}\) HD pathology primarily infiltrates subcortical regions including the striatum, putamen, and caudate, and related frontostriatal networks, though also affects amygdala, insula, and thalamus, with extension into wider cortical areas with disease progression.\(^{106}\) Given this neuroanatomical signature, it is unsurprising that deficits in ToM are a cardinal feature of the manifest disease.\(^{107}\)

**Mechanisms of ToM disruption in HD**

In patients with manifest HD, cognitive and affective ToM impairments have been documented across the majority of studies, irrespective of task employed. This includes basic affective tasks of decoding emotions from visual cues (RMET),\(^ {108,109}\) attributing emotions to story characters,\(^ {108}\) as well as low demand cognitive tasks such as ascribing mental states to animated shapes (Frith–Happé animations)\(^ {110}\) and spatial perspective taking.\(^ {111}\) More complex measures encompassing both cognitive and affective elements are also impaired in HD, including faux pas detection,\(^ {111,112}\) making social inferences (TASIT),\(^ {108,111}\) understanding the thoughts and feelings of cartoon characters (Happé cartoons),\(^ {29}\) and attributing intentions to others.\(^ {114,115}\) Unlike other neurodegenerative disorders, ToM performance is closely linked to executive function in HD, with consistent correlations between ToM impairments and executive function emerging across all the aforementioned studies. More intriguingly, deficits in intention attribution were ameliorated when cognitive flexibility impairments were controlled for.\(^ {115}\) Taken together, these findings suggest that the pervasive ToM impairments in HD are largely attributable to executive dysfunction, and as such, domain-general in nature. One anomaly is the Judgment of Preference task, which does not correlate with executive function in HD.\(^ {29,116}\) Performance on this measure is not consistently impaired in HD,\(^ {29}\) suggesting that it may capture slightly different aspects of ToM vs other tasks. Deficits in ToM have also been documented prior to motor symptom onset in premanifest HD (ie, individuals carrying the CAG-repeat expansion in the Huntingtin gene),\(^ {29}\) while intact at the whole-group level in premanifest HD, are positively associated with time to symptom onset, suggesting that alterations in ToM ability may be an early feature of the disease.

**Neuroanatomy of ToM impairments in HD**

To date, no neuroimaging studies to our knowledge have directly related ToM performance in manifest HD to underlying brain changes. Based on the progression of pathology, ToM impairments in manifest HD are speculated to reflect widespread cortical atrophy of frontoparietal regions supporting higher-level cognition; a proposal that meshes well with the executive contribution to these deficits. Evidence from premanifest cases, however, suggests that alterations in functional brain networks likely play a mediating role. For example, in premanifest HD, reduced task-based functional connectivity between the left amygdala and the right fusiform face area during viewing of facial expressions was correlated with poorer attribution of emotions on a separate measure of RMET.\(^ {118}\) By contrast, a task-based fMRI study of cognitive ToM, involving the attribution of intentions and beliefs to cartoon characters, failed to discriminate between premanifest and control cases on the behavioral level, with no significant differences in fMRI activation patterns emerging during task performance.\(^ {119}\) As such, early impairments in affective ToM in HD may arise due to dysfunction to regions supporting emotional processing, such as the amygdala, with more extensive ToM deficits developing across the course of the disease, as a result of progressive neurodegeneration into frontoparietal cortices.\(^ {107}\) The genetic nature of this condition offers a unique opportunity to test this hypothesis, by identifying individuals in advance of marked cognitive decline and tracking longitudinal structural and functional network changes that may give rise to ToM disruption.

**Parkinson’s disease (PD)**

PD is a progressive extrapyramidal movement disorder, involving the degeneration of dopaminergic neurons in substantia nigra, striatum, thalamus, and subthalamic nuclei, which secondarily affects frontostriatal loops. Although traditionally conceptualized as a motor disorder, nonmotor symptoms are common, with the majority (80%) of PD patients developing mild cognitive impairment followed by frank dementia.\(^ {120}\) The characteristic cognitive complaints in PD include executive, visuospatial, memory, and processing speed impairments,\(^ {121}\) with mounting evidence of significant social cognitive dysfunction.\(^ {122}\)

**Mechanisms of ToM disruption in PD**

Impairments in ToM are well established in PD, evident across the majority of tests employed including basic affective tasks of ascribing emotions to faces on the RMET (eg, reference 123–125 but see reference 18), determining the affective state of cartoon characters in stories (emotion attribution task),\(^ {126}\) as well as low cognitive demand tasks such as Judgment of Preference,\(^ {127–129}\) and more complex cognitive and affective measures of false belief\(^ {30,131}\) (but see reference 125), faux pas detection,\(^ {18,132,133}\) understanding of nonliteral utterances...
(Advanced Test of Theory of Mind), attribution of intentions to characters in comic strips, and other story-based tasks. It has been suggested that ToM impairments in PD may be at least partly attributable to executive dysfunction. Significant associations between executive function and ToM performance in PD have been demonstrated on complex measures combining cognitive and affective elements, such as the faux pas detection, Advanced Test of Theory of Mind, intention attribution, and other story tasks.

Mixed findings, however, have been reported on less cognitively demanding tasks, such as Judgment of Preference and RMET, with ToM correlating with executive function only at later disease stages. Collectively, this suggests the presence of an independent, domain-specific ToM deficit in early stages of PD, with executive dysfunction playing a modulating role in more complex tasks and with disease progression. The modality of testing also warrants consideration given prominent visuospatial impairments in PD. Although a recent meta-analysis found no evidence for a disproportionate impairment on visual vs verbal ToM tasks, impaired RMET performance has been associated with poor visuospatial function and reduced nonverbal reasoning in PD. Determining the precise role of visuospatial dysfunction in ToM impairment in PD is required, particularly given the heavy visual demands of many commonly used measures.

**Neuroanatomy of ToM impairments in PD**

The prefrontal cortex has long been implicated in ToM dysfunction in PD, given secondary disruption to frontostriatal loops as a consequence of dopaminergic degeneration. Indeed, recent studies point to a distributed frontoparietal network mediating ToM impairments in PD, suggesting that these deficits reflect damage beyond dopaminergic pathways. Specifically, understanding nonliteral utterances on the Advanced Test of Theory of Mind is associated with reduced gray matter volume in lateral frontal and parietal regions and reduced white matter connectivity in tracts adjacent to frontal and parietal regions (eg, superior longitudinal fasciculus). Interestingly, correlations with these gray and frontal white matter areas are negated when executive function performance is controlled for, with integrity of the superior longitudinal fasciculus adjacent to the parietal lobe remaining significant. These findings further imply a domain-specific ToM impairment in PD, associated with the disruption of frontoparietal tracts.

**Other neurodegenerative disorders**

Comparatively less attention has been paid to ToM impairments in the primary progressive aphasias (eg, semantic dementia [SD], progressive nonfluent aphasia [PNFA], and logopenic aphasia [LPA]), Parkinson-related syndromes (eg, dementia with Lewy Bodies [DLB], progressive supranuclear palsy [PSP], and corticobasal syndrome [CBS]), and vascular dementia (VD). Emerging evidence, however, suggests significant ToM deficits across the majority of these syndromes, reflecting disruption of key nodes of the canonical ToM network. While current consensus criteria emphasize aphasia as the most prominent clinical feature, mounting evidence of social cognitive dysfunction in primary progressive aphasia suggests that the cognitive profile extends beyond the domain of language. For example, ToM impairments have been characterized in SD on cartoon tasks designed to circumvent the semantic deficits of this syndrome.

Importantly, a recent study revealed that impaired ToM capacity in SD persisted despite controlling for overall semantic comprehension, reflecting atrophy in right anterior temporal cortices, as well as the bilateral amygdala and temporal poles, left orbitofrontal, and insular cortices. Similarly, ToM impairments on RMET in PNFA have been shown to relate to atrophy in the insula, temporal pole, and amygdala bilaterally. These studies add to a growing body of evidence implicating the right temporal lobe in the origin of social cognitive dysfunction in neurodegenerative disorders. While ToM has not been directly examined in LPA, early atrophy to the TPJ, along with reports of empathy loss, converges to suggest that ToM is likely to be affected.

With regard to Parkinson-related syndromes, deficits on false belief, faux pas, and RMET tasks have been uncovered in DLB. Interestingly, only RMET performance was correlated with executive function, supported on the neural level by an association with bilateral superior and middle frontal gyrus atrophy in DLB. By contrast, poor performance on the faux-pas test in DLB correlated with atrophy to key nodes of the ToM network, including mPFC, TPJ, precuneus, and insula, suggesting that this test may capture core aspects of ToM impairment in this syndrome. In PSP, impairments in social inference-making have been documented on the TASIT, reflecting atrophy in mPFC. Finally, RMET impairments in a single case of CBS have been posited to reflect mPFC hypometabolism. The status of ToM capacity in VD remains unclear, with intact comprehension of insincere speech (TASIT), but impaired RMET, faux pas, and false belief performance in a single case study.

These equivocal findings likely reflect the diffuse neuroanatomical predilection of this disease, which can target subcortical and/or cortical regions.

Research on ToM remains in its infancy in these disorders, yet the findings to date converge to suggest transdiagnostic impairments in the primary progressive aphasias (eg, semantic dementia [SD], progressive nonfluent aphasia [PNFA], and logopenic aphasia [LPA]), Parkinson-related syndromes (eg, dementia with Lewy Bodies [DLB], progressive supranuclear palsy [PSP], and corticobasal syndrome [CBS]), and vascular dementia (VD). Emerging evidence, however, suggests significant ToM deficits across the majority of these syndromes, reflecting disruption of key nodes of the canonical ToM network. While current consensus criteria emphasize aphasia as the most prominent clinical feature, mounting evidence of social cognitive dysfunction in primary progressive aphasia suggests that the cognitive profile extends beyond the domain of language. For example, ToM impairments have been characterized in SD on cartoon tasks designed to circumvent the semantic deficits of this syndrome.

Importantly, a recent study revealed that impaired ToM capacity in SD persisted despite controlling for overall semantic comprehension, reflecting atrophy in right anterior temporal cortices, as well as the bilateral amygdala and temporal poles, left orbitofrontal, and insular cortices. Similarly, ToM impairments on RMET in PNFA have been shown to relate to atrophy in the insula, temporal pole, and amygdala bilaterally. These studies add to a growing body of evidence implicating the right temporal lobe in the origin of social cognitive dysfunction in neurodegenerative disorders. While ToM has not been directly examined in LPA, early atrophy to the TPJ, along with reports of empathy loss, converges to suggest that ToM is likely to be affected.

With regard to Parkinson-related syndromes, deficits on false belief, faux pas, and RMET tasks have been uncovered in DLB. Interestingly, only RMET performance was correlated with executive function, supported on the neural level by an association with bilateral superior and middle frontal gyrus atrophy in DLB. By contrast, poor performance on the faux-pas test in DLB correlated with atrophy to key nodes of the ToM network, including mPFC, TPJ, precuneus, and insula, suggesting that this test may capture core aspects of ToM impairment in this syndrome. In PSP, impairments in social inference-making have been documented on the TASIT, reflecting atrophy in mPFC. Finally, RMET impairments in a single case of CBS have been posited to reflect mPFC hypometabolism. The status of ToM capacity in VD remains unclear, with intact comprehension of insincere speech (TASIT), but impaired RMET, faux pas, and false belief performance in a single case study.

These equivocal findings likely reflect the diffuse neuroanatomical predilection of this disease, which can target subcortical and/or cortical regions.
alterations in ToM capacity in neurodegeneration. Future work uncovering the nature and extent of these impairments in the lesser researched syndromes is essential not only to validate this hypothesis but also to assist in identifying and managing these diagnostically challenging syndromes and providing feedback to clinicians and carers about the expected deficits. The careful study of ToM disruption in these syndromes will further refine our understanding of the neurocognitive architecture of ToM more broadly.

**Summary: Toward a refined understanding of ToM impairment in neurodegeneration**

Taken together, the evidence suggests that ToM impairments, and their respective neurocognitive mechanisms, vary in a syndrome-specific manner. Domain-specific impairments in ToM are present in bvFTD, ALS, and PD, reflecting pathological insult targeting, but not exclusive to, multiple nodes of the ToM network (Figure 3). These deficits, however, are exacerbated with encroachment of disease pathology into brain regions supporting ancillary cognitive processes implicated in ToM, such as executive function. By contrast, the bulk of the behavioral evidence in AD points to domain-general ToM impairments, attributable to disruption of brain regions supporting memory and executive processes. Finally, ToM impairments in manifest HD patients are strongly related to executive dysfunction, suggesting a domain-general impairment. Findings of amygdala involvement in ToM dysfunction in premanifest HD, however, raise the possibility of a domain-specific impairment of affective ToM in the earliest stages of the disease. Determining how this variability across syndromes manifests on the behavioral level is essential to arrive at a comprehensive understanding of ToM and its expression in everyday functioning. For example, it may be that neurodegenerative disorders with domain-specific ToM impairments (eg, bvFTD, ALS, PD, and HD) display significantly greater disruption to real-world social functioning than those with domain-general deficits (eg, AD), an issue we tackle in the next section.

**Understanding the real-world impact of ToM disturbances in neurodegeneration**

The literature reviewed thus far highlights ToM disruption as a potential transdiagnostic feature across neurodegenerative disorders, mediated by common and divergent neurocognitive mechanisms. The majority of these studies, however, have employed laboratory-based measures of ToM, which arguably lack many of the contextual and situational cues that inform social behavior in the real world. While more recently developed tasks involving dynamic videos of social interactions, such as the TASIT,11 offer improved ecological validity, the participant nevertheless remains an observer, rather than an active agent in the social milieu. As such, it is unclear to what extent ToM performance in the laboratory reflects real-world social functioning in its many permutations and settings. A more thorough picture of ToM deficits, and its resultant impact on interpersonal function, may be gleaned from carer observations and reports of patients’ behavior in daily life. This approach reveals syndrome-specific patterns of social functioning that do not always neatly map onto ToM performance as assessed formally on experimental tasks. For example, although significant impairments on ToM tasks have been documented in AD, such patients are widely reported to uphold appropriate social functioning, at least in the mild-to-moderate stages of the disease. In particular, social graces, interpersonal warmth, and the ability to form and maintain interpersonal relationships remain intact in AD,63–65 as does empathic concern and behavior.132 In fact, social–emotional sensitivity may be enhanced with AD onset, attributable to increased functional connectivity between the ACC/fIN and striatum.53 We suggest that this discrepancy between compromised task performance and intact real-world behavior in AD reflects the often ambiguous and contextually devoid nature of ToM tasks. For example, many of the tasks in which significant ToM impairments in AD have been documented rely upon interpretation of decontextualized stimuli, such as cartoons or photographs. Social function, however, hinges upon multifaceted interpersonal and contextual cues enabling us to interpret and understand other’s mental states (see also Ref. 79). Individuals with AD may find abstract tasks of ToM function overly cognitively demanding,16 yet if assessed within familiar and contextually rich social settings may display appropriate behavior. Our brief survey of the literature highlights the need for a new breed of ToM tasks to better capture this dimension of social cognitive functioning in AD.

Despite limited translation of laboratory-based ToM studies to daily functioning in AD, emerging evidence suggests that ToM impairments closely mirror real-world social dysfunction in other neurodegenerative disorders. This is exemplified by bvFTD, where carer reports of severe behavioral disruption in everyday life corroborate the stark ToM impairments displayed by patients on formal testing. For example, bvFTD patients are reported to display an array of inappropriate social behaviors such as loss of empathy,132
Reduced tact, and failure to acknowledge the presence of others.153 Furthermore, ToM deficits on formal measures are strongly correlated with carer ratings of behavioral disturbance in bvFTD,32 suggesting that performance on laboratory measures accurately captures functionally relevant impairments in this syndrome. We note further parallels between ToM impairments and carer reports of behavioral dysfunction in SD. Although typically viewed as a language disorder, patients with SD can present with a similar profile of social behavioral disturbances to bvFTD,154 although to a milder degree. Such disturbances include reduced empathy,145,155 and increased egocentric behavior manifesting in a failure to consider other’s viewpoints.156 In line with their ToM impairments, reductions in empathy are also reported by carers across the primary progressive aphasias, namely LPA and PNFA.147 Difficulties adopting or considering the mental state of others in daily life in neurodegenerative disorders result in profound carer stress and negatively impact the patient–carer relationship,157 underscoring the importance of studying real-world manifestations of ToM in these patients.

Emerging evidence also suggests an association between ToM deficits and real-world social functioning in PD, ALS, and HD. In PD, carer-reported reductions in empathy occur in parallel with impairments on ToM tasks.127 Furthermore, ToM impairments in PD are associated with disturbances in social behaviour27 and increased apathy.126 This relationship between ToM and apathy has also been uncovered in ALS,101,102 which may reflect reduced motivation for social engagement in the face of difficulties adopting other’s perspectives. The directionality of this relationship, however, is yet to be explored. In addition, carer reports reveal altered empathy and impaired social functioning in a subset of ALS patients,99,138 although whether this subset overlaps with those demonstrating ToM deficits remains to be determined. While the direct link between ToM and social functioning has not been explored in HD, alterations in social behavior such as empathy are characteristic of the disease,29 suggesting that impairments on ToM measures in HD extend to real-world dysfunction. By contrast, carers of PSP and CBS report intact empathy in these patients,145 and no study, to our knowledge, has explored empathic behavior in DLB or VD. Further examination of the potential relationship between ToM impairment and social functioning across each condition is critical to establish the ecological validity of formal ToM measures.

**ToM and moral reasoning**

Alterations in moral behavior, or transgressions from societal values about how its members should behave, have been documented in neurodegenerative disorders. Specifically, patients with bvFTD and PD have been shown to make abnormal responses to moral dilemmas, which involve a choice between two morally conflicting actions159,160 (but see Ref. 161). Such difficult moral decisions are proposed to require ToM, as the thoughts and feelings of other people in the scenario are often considered when choosing how to respond.7 The overlap between ToM and moral reasoning is supported at the neural level, with mPFC, ACC/fIN, and PCC implicated across both processes.162–164 Altered moral reasoning in bvFTD correlates with impaired performance on cognitive, but not affective, ToM tasks,165 and morality and ToM can also be simultaneously affected in these patients.166 No such relationship between ToM and moral reasoning, however, is apparent in PD.160,161 Furthermore, despite impaired ToM performance in AD, moral reasoning appears intact in this syndrome.159 As such, the evidence to date suggests that ToM dysfunction may contribute to altered moral behavior in some neurodegenerative disorders (ie, bvFTD), potentially attributable to atrophy in regions such as the mPFC; however, the extent to which this relationship is present across other syndromes requires further exploration.

**ToM and criminality**

Taken to the extreme, impairments in moral reasoning can translate into real-world transgressions, such as criminal behavior. Crimes such as theft, violence, hypersexuality, traffic violations, and even homicide have been reported in neurodegeneration.147 In fact, for some individuals, law-breaking represents the first-noticed symptom of the disease. Estimates of criminality range from 37% to 54% in bvFTD and 4.5% to 12% in AD, as well as 20%–40% in HD.167,168 Despite commonalities in the neural substrates of these processes (eg, mPFC, PCC, TPJ),169 no study, to our knowledge, has explored the relationship between ToM dysfunction and criminal behavior in neurodegeneration. The nature of crimes committed in each syndrome, however, suggests potential syndrome-specific associations between ToM and criminality. In bvFTD, for example, crimes are often of an antisocial disinhibited nature, such as harassing strangers, sexual advances, theft from other’s purses, and urinating in public.167 Similarly, in HD, antisocial behaviors of violence and reckless driving represent the most frequently committed crimes.170 The socially disruptive nature of criminal behavior in bvFTD and HD may reflect an inability to consider or comprehend how behavior affects others, mediated, in part, by impaired ToM. In AD, however, criminal behavior is less likely to be of a violent or aggressive nature.
ToM and gullibility

Patients with neurodegenerative disorders are particularly vulnerable to financial exploitation and abuse, often falling victim to scams or being easily misled.\(^\text{171,172}\) Avoiding this kind of deception requires complex social processes,\(^\text{173}\) including the advanced ToM function of correctly interpreting the speaker’s intention as insincere.\(^\text{174}\) The intersecting network of brain regions implicated across ToM and the avoidance of deception. bvFTD patients are widely known to fall victim to scams, make irrational financial decisions, and engage in excessive gambling. Their concomitant difficulty in interpreting the sincerity of speech, related to their broader ToM deficits,\(^\text{27}\) combined with an inability to incorporate social contextual information to guide economic decisions,\(^\text{175}\) may explain the high levels of gullibility displayed by bvFTD patients in daily life.\(^\text{176,177}\) Misinterpretation of insincere speech is also found in AD\(^\text{180–183}\) (but see Ref. 27), along with an increasing tendency to believe things that are untrue.\(^\text{179}\) Financial exploitation in AD, however, likely reflects episodic memory dysfunction and overall cognitive decline.\(^\text{179}\) Collectively, these findings suggest that ToM underlies gullibility in some neurodegenerative syndromes, although further research directly testing this hypothesis is required.

Concluding remarks: beyond the clinic

Translating converging findings of ToM disruption from the cognitive neuroscience literature into socially relevant behaviors such as empathy, morality, criminality, and gullibility is essential to understand the real-world implications of these impairments and their underlying neurocognitive mechanisms to inform targeted interventions. In turn, this enhanced understanding will better equip carers, clinicians, and the legal system to cope with aberrant socioemotional behavior, particularly when viewed in light of underlying neuropathology. Moreover, the study of neurodegenerative syndromes in which ToM difficulties do not manifest in disruptive behaviors may offer compelling insights regarding potential protective factors for behavioral change. These findings have important implications for legal responsibility in neurodegeneration. In summary, our review highlights how extensive neural network degeneration can impair understanding of the effect of one’s behaviors on other people, as well as detection of other’s intentions, raising important questions regarding culpability for criminal activity and accountability for financial loss in neurodegenerative disorders. Moving forward, increased awareness and further rigorous investigation of these issues are required to ensure appropriate ethical treatment, care, and quality of life for individuals living with neurodegenerative disorders and their families.

Acknowledgments

This work was supported in part by an Australian Research Council (ARC) Centre of Excellence in Cognition and its Disorders (CE110001021). CS-B is supported by the National Health and Medical Research Council (NHMRC) Postgraduate Scholarship (APP1132764). SR is supported by the Faculty of Science PhD Research Scholarship from the University of Sydney. MI is supported by an ARC Future Fellowship (FT160100096) and an ARC Discovery Project (DP180101584). The authors would like to thank Ms Heidi Cartwright for assistance with preparation of Figures 2 and 3.

Disclosure

The authors report no conflicts of interest in this work.

References

44. Strikwerda-Brown C, Grilli MD, Andrews-Hanna J, Irish M. “All is not lost”: episodic and semantic contributions to continuity of self in dementia. 2019; [preprint]: 1–33.


