Theory of Mind (Neural Basis)

Rebecca Saxe

Department of Brain and Cognitive Sciences, MIT

In Press at: Encyclopedia of Consciousness

The externally observable components of human actions carry only a tiny fraction of the information that matters. Human observers are vastly more interested in perceiving or inferring the mental states - the beliefs, desires and intentions - that lie behind the observable shell. If a person checks her watch, is she uncertain about the time, late for an appointment, or bored with the conversation? If a person shoots his friend on a hunting trip, did he intend revenge or just mistake his friend for a partridge? The mechanism people use to infer and reason about another person's states of mind is called a 'Theory of Mind' (ToM). One of the most striking discoveries of recent human cognitive neuroscience is that there is a group of brain regions in human cortex that selectively and specifically underlie this mechanism.

The importance of mental states for behavior prediction is especially clear when the person doing the behavior is misinformed: that is, when the actor has a false belief. False beliefs therefore figure heavily in the study of ToM. The original "False Belief task" was developed for use with pre-school aged children. In the basic design, a child watches while a puppet places an object in location A. The puppet leaves the scene and the object is transferred to location B. The puppet returns and the child is asked to predict where the puppet will look for the object. Three-year-olds say the puppet will look in location B, where the object actually is; older children say the puppet will look in location A, where the puppet last saw the object.

Following the tradition in developmental psychology, many of the early neuroimaging investigations of Theory of Mind required subjects to attribute false beliefs to people in stories or cartoons. Most of these studies have used functional magnetic resonance imaging (fMRI) to measure changes in cortical blood oxygenation. Across

these studies, a very reliable set of brain regions was implicated in the "false belief" condition of each study, including right and left temporo-parietal junction (TPJ), medial parietal cortex (including posterior cingulate and precuneus), and medial prefrontal cortex (MPFC). The same brain regions have also been identified by converging methods, including both lesion and transcranial magnetic stimulation (TMS) studies of Theory of Mind. This group of brain regions is sometimes collectively called the "ToM network."

A critical question that emerged from this work is whether the ToM network is a distinct and domain specific neural system for thinking about the mind. The alternative is that some or all of the regions in the ToM network are actually recruited for some other aspect of solving a false belief task. There is more to solving a false belief task than a concept of belief, and there is more to a concept of belief than passing the false belief task. In particular, attributing a false belief to another person depends heavily on two cognitive capacities that are not specific to Theory of Mind: language and inhibitory control. Plausibly, activation in the ToM network could reflect a combination of language processing and inhibitory control. This article thus first reviews evidence for the relationship between ToM, language, and inhibitory control in the brain, and then turns to more detailed investigations of the neural basis of ToM.

Theory of Mind and Language: fMRI and lesion studies

Mental states cannot be observed directly; beliefs and desires are invisible, abstract entities. Nor is there any simple correlation between mental states and behaviour. For example, there is no observable behaviour that is reliably diagnostic of whether a person currently believes that today is Tuesday. One invaluable way to learn about the elusive contents of the mind is therefore to listen to how other people talk about the mind.

During development, linguistic ability is correlated with false belief task performance. In a striking example, deaf children of hearing parents (that is, whose parents are non-native signers) are selectively delayed in passing the false belief task. These children have similar difficulty even on non-verbal tests of false belief understanding, suggesting that the delay does not reflect the language demands of the tasks themselves. Moreover, even after accounting for the child's own skill with sign

language, the child's performance on the false belief task is independently predicted by the mother's proficiency with sign language, and specifically her use of mental state signs. Deaf children of deaf parents (native signers), by contrast, are not delayed. Clearly, linguistic exposure influences Theory of Mind development, but the mechanism underlying this influence remains controversial.

Given these results, it is possible that the brain regions implicated in Theory of Mind are instead involved in some aspect of language processing. One way to test this hypothesis is to present stimuli that imply the protagonists' false belief, but do not include any explicit statements of beliefs (and therefore do not include sentence complements). One study used verbal stories that merely described a sequence of actions by a character. No mental states were explicitly mentioned, but the sequence of actions could most readily be explained in light of the protagonist's belief. The control stories were sequences of mechanical or physical events that similarly required subjects to infer an unstated physical causal process. As predicted, all of regions in the "ToM network" - the right and left TPJ, medial parietal and MPFC - showed a stronger response during the implied-belief stories than during the implied-physical stories. Other studies have used non-verbal single-frame or animated cartoons that required subjects to attribute beliefs to the characters, but used no language at all. In all cases, the same brain regions were recruited when subjects attributed beliefs to the characters, independent of modality.

No brain regions in the "ToM network" appear to be associated with the linguistic demands of the task, at least in adults. These results raise the possibility that while language is necessary for ToM during development, once concepts of beliefs and desires are formulated, a mature ToM could function even in the absence of language. The critical test is therefore to investigate the consequences of late acquired aphasia (especially loss of grammatical skills) on an already mature theory of mind. In a series of studies, Rosemary Varley and Michael Siegal found that aphasic patients could be relatively spared in ToM.

Ian Apperly and colleagues provided strong new evidence for the same hypothesis. PH, a young man who had a left hemisphere stroke, was tested on a battery of language and ToM tests. Although PH was severely impaired on tests of syntax,

including specifically the syntax of embedded clauses, he showed no impairments on non-verbal tests of ToM, including 1st and 2nd-order (what X thinks that Y thinks) false belief tasks. Taken together, these studies provide clear evidence that adults with a mature ToM can formulate thoughts about other people's thoughts, even in the face of severe grammatical impairments, and thus that by adulthood, the neural bases of ToM and language are largely distinct.

Theory of Mind and Executive Control: fMRI and lesion studies

The other major cognitive resource that is implicated in passing a false belief task is inhibitory control. In the classic false belief task, the participant must be able to juggle two competing representations (the actual world and the world represented in the other person's head) and to inhibit responding based on the true location of the object. In development, performance on false belief tasks is correlated with the child's overall ability to select among competing responses. In a protocol closely matched to the false belief task, a photograph is taken (or a map is drawn) of the object in location A, the object is moved to location B, and children are asked where the object will be in the photograph. No mental state understanding is required but 3-year-olds fail the task. In fact, typically developing children find the false photograph and false map task harder than the false belief task; by contrast, children with autism solve the false photograph and false map tasks, but fail the false belief.

So are brain regions in the "ToM network" implicated specifically in reasoning about other people's mental states, or in the inhibitory control necessary to pass a false belief task? If these brain regions are involved specifically in reasoning about beliefs, they should not show a high response during other stories about false representations, like photographs or maps. Consistent with this prediction, Rebecca Saxe and Nancy Kanwisher initially reported that all of the brain regions in the ToM network are recruited significantly more when participants read about beliefs than about physical representations. In a careful subsequent study, Josef Perner and colleagues distinguished between stories describing photographs versus maps. Maps are more conceptually similar to beliefs because photographs are designed to represent the past, whereas both maps and beliefs are supposed to represent current reality. Using this tighter comparison, Perner et

al found that only the right TPJ showed the profile of a region selectively involved in ToM: a significantly higher response to stories about beliefs than about photographs or maps, and no difference between photographs or maps and a fourth control condition involving no representations at all. Taken together, these results suggest that at least the right TPJ is specifically recruited for attribution of beliefs, and not for the inhibitory demands of the task.

There is a challenge to this view, though. Lesion and imaging studies implicate a region of right TPJ in a different kind of task: changing the direction of attention in response to unexpected stimuli. Could this be the same brain region as the right TPJ implicated in studies of ToM? If so, it seems unlikely that it's a specific mechanism for thinking about mental states. However, two recent studies have found that the regions for belief attribution and exogenous attention are neighboring but distinct. Both studies revealed that the attention region is located approximately 10 mm superior to the region involved in theory of mind, close enough to be confusable, but far enough to probably be reflect a distinct group of neurons.

So the current evidence suggests that regions in ToM network are involved in reasoning about minds, not in the inhibitory control necessary for performing false belief tasks. Importantly, there are other regions in the brain that *are* associated with inhibitory control, including intra-parietal sulcus and frontal eye-fields. Reasoning about beliefs does provoke robust activity in these brain regions as well - but no more than is provoked by false photograph stories. That is, domain-general mechanisms for inhibitory control, response selection, and so on are recruited by both false belief and false photograph stories, and there are additional brain regions that are recruited only for beliefs. These results suggest that belief reasoning depends on both domain-general and domain-specific mechanisms, while reasoning about false photographs depends only on the domain-general components.

A similar division between specific and general contributions to false belief task is suggested by a double dissociation observed in focal lesion patients. Samson and colleagues developed a series of elegant non-verbal false belief tasks. In one set of tasks (reality known), an object was moved without a character's knowledge, but the

participants themselves always knew the true location of the object. Passing these tasks required both the ability to represent the character's belief, and the ability to inhibit the participant's own knowledge (high inhibition). By contrast, in a second group of tasks (reality unknown), an object was moved without the character's knowledge, but the participant also did not know the true location of the object and therefore did not have to inhibit any alternative location (low inhibition).

WBA, a patient with left lateral frontal damage, was selectively impaired on the reality-known (i.e. high inhibition) false belief tasks. WBA also failed tests of inhibitory control that did not involve ToM, but performed normally on the reality-unknown (i.e. low inhibition) false belief tasks. By contrast, a group of patients with left TPJ damage failed both high and low inhibition false belief tasks, but did not fail general tests of inhibitory control outside of the domain of ToM.

In all, fMRI and lesion evidence converge on three conclusions about the relationship between inhibitory control and ToM in the adult brain. (1) Successful performance on many kinds of false belief task depends on both domain-general inhibitory control and on domain-specific mechanisms for representing mental states. (2) These contributions are supported by distinct neural mechanisms, which can be dissociated. (3) The domain-specific component of ToM is supported, at least in part, by the TPJ.

One important open question concerns the relative role of the right and left TPJ in ToM. The fMRI studies have pointed to right TPJ, while the left TPJ has been the focus of the lesion studies. Future studies should use the non-verbal tasks developed by Apperly and colleagues with patients who have lesions to right TPJ.

Theory of Mind and Person perception: fMRI and lesion studies

The research described above suggests that the "ToM network" reflects a domain-specific mechanism, or mechanisms, for reasoning about other people and other minds. Of course, identifying the brain regions involved in Theory of Mind is just the first step. The real challenge is to understand *how* these brain regions perform the relevant computations. A first step may be to understand the distinct contributions of each of these brain regions. Initial studies suggest that regions of MPFC are implicated in distinct

components of social cognition, and the right TPJ region is the most selectively recruited for ToM.

Some evidence for a division of labour between the TPJ and MPFC came from a study using two new kinds of stories, highlighting different aspects of reasoning about another person: (1) 'Appearance' stories described representing socially relevant information about a person that is visible from the outside, like the person's clothing and hair colour; and (2) 'Thoughts' stories described the contents of another person's thoughts or beliefs, specifically tapping the later-developing component of Theory of Mind.

A region involved in any general aspect of social cognition, including detecting the presence of a person in the story, or tracking information about people, would show a high response in both story conditions, since all included socially relevant information about a protagonist. The MPFC showed exactly this predicted response: an equally high response to all story conditions that describe people. By contrast, a region involved specifically in attributing and reasoning about mental states like beliefs and desires should show a high response only in the 'Thoughts' condition. Consistent with this prediction, the right and left TPJ showed a high response when subjects read about thoughts or beliefs, but a low response (no different from resting baseline) when participants read about false photographs or a person's physical appearance.

Generally, the fMRI literature suggests a division in the neural system involved in making social judgments about others, with one component (the RTPJ) specifically recruited for the attribution of mental states, while a second component (the MPFC) is involved more generally in the consideration of the other person. If so, there should be tasks that do involve reasoning about another person, but do not involve attribution of specific mental states, and that therefore recruit the MPFC but not the RTPJ. There are, and they do. One example is the attribution of personality traits. Personality traits may be conceived as a heuristic alternative to ToM, allowing observers to predict a person's "typical" behavior, without considering that person's specific mental states. In a series of studies, participants read single words describing human personality traits like "daring" or "shy", and judged either (a) whether each word is semantically positive or negative, or

(b) whether that trait applies to a specific person - either the self or a well-known other. As predicted, the MPFC, but not right or left TPJ, is recruited significantly more for trait-attribution than for semantic judgement.

In all, one critical topic of future investigation will be to differentiate the specific role of right and left TPJ, PC and MPFC in Theory of Mind. In the meantime, though, the pattern response of brain regions in the ToM network can already be used to to test cognitive hypotheses about the nature of Theory of Mind. For example, one important question arising from cognitive and developmental psychology is whether ToM function is automatic, or controlled.

Theory of Mind: Automatic or Controlled

When do people formulate thoughts about thoughts? One possibility is that belief attribution is automatic, and occurs spontaneously whenever the situation *affords* it. This model predicts that ToM is almost always engaged. Any observation of a human action affords the attribution of many desires and an infinite number of true beliefs. At the other end of the spectrum, mental state attribution may be applied only sparingly, when the situation specifically demands it. There is some recent behavioural evidence consistent with this latter option. In one study, participants watched another person act based on a false belief. In one condition, that person's actions and beliefs were specifically relevant to the participant's own task at that moment; in the other condition, the person's actions and beliefs were not specifically relevant. Then, the participants were occasionally and unpredictably probed, to see whether they were automatically holding onto a representation of the person's false belief. The results (using reaction times) suggested the opposite: participants only kept track of the other person's false beliefs when those beliefs were specifically relevant to the participant's own task.

These two theories about ToM also make different predictions for the pattern of activation in the ToM network. If belief attribution is automatic, then whenever participants watch someone else act based on a belief, there will be activation in the ToM network. By contrast, if belief attribution is controlled and task dependent, then activation in the ToM network will depend on whether the observed person's belief is specifically relevant to the participant's current task.

Converging with the behavioural studies, some recent fMRI results support the controlled view of Theory of Mind. Sommer and colleagues had participants in the scanner watch characters who had either a true or a false belief, and then make a prediction about their next action. The false belief trials thus required the participants to use the observer's false beliefs to predict the observer's behavior (e.g., looking in the wrong place). By contrast, the true belief trials do not require belief attributions at all; participants simply have to respond based on the true location of the object. The response in the ToM network brain regions was higher on false belief trials, when belief attribution is required, than on true belief trials, when belief attribution is optional for task performance.

There is another possible interpretation of Sommer and colleagues results, though. The authors themselves concluded that the ToM network regions (and especially the right TPJ) were recruited specifically for *false* beliefs. However, in other previous studies, the response in the ToM network (and especially the right TPJ) has been equivalent for true and false belief conditions. More generally, one would expect ToM to be used for both true and false belief attributions: in our daily lives, we have to expect other people to have mostly true beliefs, or their actions would becomes completely unpredictable.

So what explains the conflicting results? The difference between those previous studies, and the task used by Sommer and colleagues, is that the other studies used explicit verbal statements of the beliefs (e.g. "Sarah believes that her shoes are under the bed"). When the belief is stated explicitly in the story, then participants have no choice about whether to think about thoughts while reading the story: the decision is out of their hands. Stories that explicitly state beliefs thus cannot be used to test the automaticity of ToM. On the other hand, those stories can be used to distinguish between the two interpretations of Sommer and colleagues results. Since explicitly stated true and false beliefs elicit equal responses in the ToM network, it seems likely that the difference observed by Sommer and colleagues was due, not to truth and falseness, but to the optional versus required belief attributions. If so, then those results are also evidence that thinking about other people's thoughts does not occur automatically.

More broadly, predicting which tasks or stimuli will elicit Theory of Mind - and thus elicit activation in the ToM network - is a critical challenge for cognitive neuroscience. Research on mental state attribution outside of the false belief task is thus beset by a chicken-and-egg problem. In order to know that a brain region is truly the neural substrate of ToM, we must establish that it is recruited for all and only the tasks that involve ToM – including going beyond the false belief task. However, exactly when people do think about thoughts, outside the false belief task, has not yet been established by independent methods. In some cases, the best evidence that a task does differentially invoke ToM may be the level of activation in the ToM network (a 'reverse' inference)! Fortunately, this problem is temporary. Different theories of ToM and different theories of the function of the ToM network can be tested simultaneously. The correct resolution will be determined by the consistency of the data that emerge from these tests, and the richness of the theoretical progress those data support.

One key dimension of task design is whether we think of ToM as predominantly (1) the process of initially forming a representation of someone's thought, or (2) the process of using that representation to answer questions and make predictions. This question has important implications for studies of ToM that use methods with high temporal resolution, like event-related potentials (ERPs) from an electroencephalogram (EEG).

ERP studies of Theory of Mind

The studies described above, using fMRI, lesions and TMS, help to establish *where* Theory of Mind processing happens in the brain. In order to eventually understand *how* this processing happens, another group of methods must be brought into the picture: methods with high temporal resolution, that can help reveal *when* the relevant processing is happens. High temporal resolution has two benefits. First, knowing when a certain kind of processing occurs can be useful in its own right. For example, the speed of a process can constrain the relative contributions of bottom-up (faster) and top-down (slower) input to that process. Second, once we know both where and when to look for a specific process, then we can begin to measure the actual firing of neurons that is responsible for that process. At the moment, recording for individual neurons involved in ToM would be

extremely hard because we don't know when, during the task, ToM is likely to be happening.

As a first approach to these challenges, Mark Sabbagh, David Liu, Henry Wellman and colleagues have conducted a series of studies of ToM using event-related EEG potentials. ERPs have very high temporal resolution (it's possible to distinguish between signals that are separated by 10 ms) but are not very good for spatial resolution (establishing where those signals come from).

Critically, ERP analyses require that there is a specific and predictable start time for the events of interest. In this case, the event of interest is thinking about another person's thoughts. When would that start? As hinted above, the answer depends on whether we think of ToM as the process of initially forming a representation of someone's thought, or as the process of using that representation to answer questions and make predictions. The timing of initially forming a representation is very hard to predict. In a typical false belief task, the participant might begin to consider the character's mental states when the character turns his back and/or leaves the room, when the object is removed from it's original location or placed into the novel location, or when the character returns to the room.

Sabbagh and colleagues avoided this uncertainty by focusing on the process involved in answering the critical ToM question. In their version of the false belief (or false photograph) task, there are always two objects in the scene: A and B. While the character is outside of the room (or after the photograph is taken), object A is moved to a new location, and object B remains in its original location. Participants are then asked "According to [the character/ the photograph], where is the [Object A / Object B]?" Retrieval of the character's specific belief about the target object must therefore begin at the onset of the final word of the question – the first moment at which participants know which object is the target.

Using this paradigm, Sabbagh and collagues compared the brain activation when participants answered questions about false beliefs versus false photographs. They did find a specific difference: a left-lateralised frontal late slow wave was larger on false belief relative to false photograph trials. However, this ERP pattern did not come from

any of the brain regions in the "ToM network." In fact, none of the brain regions in the ToM network were identified by the ERP method.

How should we explain the discrepancy between ERP and other methods? One possibility is that the ERP study focussed on the processes for answering questions about ToM, whereas the fMRI, lesion and TMS studies may have been mostly tapping the process for forming a representation of someone's belief, in the first place. In Sabbagh and colleague's task, since participants are asked about the content of a belief or photograph on every trial, they may preemptively formulate a representation of all the relevant beliefs or photographs as the trial is evolving, long before the critical question.

Consistent with this idea, the ToM network revealed by fMRI appears to be involved mostly in the initial inference of beliefs. Although fMRI lacks the temporal resolution of ERPs, false belief stories are typically presented for 10 – 20 seconds, followed by a question presented for 4 – 6 seconds. The response in the TPJ, MPFC and medial parietal regions occurs almost exclusively in the first time period, while the participants are reading the story and forming their initial inferences and representations.

These results suggest a distinction between the processes supporting initially inferring someone's belief, and then using that belief to answer questions and make predictions. If this is right, then it will be important to devise new tasks and methods for studying the elusive inference process with high temporal resolution.

Theory of Mind and Mirror Neurons

In the domain of "understanding human actions," there is an intuitive distinction between one's own actions, and actions executed by others. The neural mechanisms necessary for executing one's own goal directed action are fairly concrete, including sensory perception of the local environment, motor planning and control. Understanding someone else's action may seem by contrast like a highly abstract -- if not semi-miraculous -- achievement. Recently, though, many researchers have proposed that this abstract higher cognitive function could have concrete sensori-motor foundations. That is, an observer might understand someone else's action using the same cognitive and neural mechanisms that she uses to plan her own. The idea is sometimes called the "motor theory of social cognition".

One advantage of the motor theory of social cognition is its parsimony. Action prediction and understanding could be achieved with the same cognitive and neural mechanisms that the observer already uses for her own action planning and execution; she doesn't need a whole extra system for ToM.

As described above, though, recent neuroscientific evidence undermines this view. There are brain regions specifically implicated in attributing mental states, and these brain regions are not part of the observer's own motor system. The ToM network is completely distinct, anatomically, from the brain regions implicated in action execution or action perception. Many neuroimaging studies have focused on the overlapping activation during action perception and action execution, of ventral premotor cortex, inferior frontal gyrus and right inferior parietal cortex. By contrast, the regions implicated in ToM have no known role in motor planning or action execution.

Although the two groups of regions are clearly anatomically segregated, their functional properties have not yet been investigated within a common task. Future work should investigate tasks in which both ToM and the mirror system are involved, to allow for a direct functional dissociation between these mechanisms.

Conclusions

In all recent studies are beginning to shed light on the brain regions involved when human adults reason about one another's minds - that is, in Theory of Mind. One surprising result of these early studies is that a specific group of cortical regions is reliably implicated in Theory of Mind, the so-called "ToM network", including the right and left TPJ, MPFC and PC. The brain regions involved in Theory of Mind are incredibly robust. These regions can be identified in 90% of individual subjects, after just 20 minutes of scan time; the same regions have been reported by labs on different continents, using different procedures and different stimuli. The same group of regions have been identified as relevant for ToM by lesion and TMS studies. Similarly reliable patterns of activation are routinely observed for perceptual mechanisms, like primary sensory and motor cortices, but rarely for dimensions of cognition as abstract and complex as Theory of Mind. The 'Theory of Mind' regions thus offer a rare window through the brain to the mind.

Nevertheless, critically important questions remain open: What are the specific and distinct roles of the brain regions that make up the ToM network? How does the ToM network interact with other brain systems underlying language, inhibitory control, and action perception (e.g. the mirror system)? How do these brain regions develop? Are these brain mechanisms universal to all human beings? Are they uniquely human? Lots of important research remains to be done.

Recommended reading

Apperly, I. A., D. Samson, et al. (2006). "Intact first- and second-order false belief reasoning in a patient with severly impaired grammar." Social Neuroscience 1(3-4): 334-348.

Bloom, P. and T. P. German (2000). "Two reasons to abandon the false belief task as a test of theory of mind." Cognition 77(1): B25-31.

Jacob, P. and M. Jeannerod (2005). "The motor theory of social cognition: a critique." Trends Cogn Sci 9(1): 21-5.

Leslie, A. M. and L. Thaiss (1992). "Domain specificity in conceptual development: neuropsychological evidence from autism." Cognition 43(3): 225-51.

Perner, J., M. Aichorn, et al. (2006). "Thinking of mental and other representations: the roles of left and right temporo-parietal junction." Social Neuroscience 1(3-4).

Rizzolatti, G., L. Fogassi, et al. (2001). "Neurophysiological mechanisms underlying the understanding and imitation of action." Nat Rev Neurosci 2(9): 661-70.

Sabbagh, M. A. and M. Taylor (2000). "Neural correlates of theory-of-mind reasoning: an event-related potential study." Psychol Sci 11(1): 46-50.

Samson, D., I. A. Apperly, et al. (2004). "Left temporoparietal junction is necessary for representing someone else's belief." Nat Neurosci 7(5): 499-500.

Saxe, R. and N. Kanwisher (2003). "People thinking about thinking people. The role of the temporo-parietal junction in "theory of mind"." Neuroimage 19(4): 1835-42.

Saxe, R. and L. J. Powell (2006). "It's the thought that counts: specific brain regions for one component of theory of mind." Psychol Sci 17(8): 692-9.

Sommer, M., K. Dohnel, et al. (2007). "Neural correlates of true and false belief reasoning." Neuroimage 35(3): 1378-84.

Varley, R. and M. Siegal (2000). "Evidence for cognition without grammar from causal reasoning and 'theory of mind' in an agrammatic aphasic patient." Curr Biol 10(12): 723-6.

Wellman, H. M., D. Cross, et al. (2001). "Meta-analysis of theory-of-mind development: the truth about false belief." Child Dev 72(3): 655-84.

Figure Legend

Brain regions showing higher metabolism during false belief than false photograph stories. Clockwise from bottom left: axial, coronal and saggital slices, and a rendering of the lateral surface. Clearly visible are left and right TPJ, and the medial parietal (precuneus) regions.

Table 1. Verbal stimuli eliciting a robust response in the right TPJ.

1. (False Beliefs). Anne made lasagna in the blue dish. After Anne left, Ian came home and ate the lasagna. Then he filled the blue dish with spaghetti and replaced it in the fridge. Anne thinks the blue dish contains (choose one) spaghetti / lasagna.

2. (Beliefs - not false). Kate knew her colleague was very punctual and always came to work by 9. One day, he still wasn't in at 9:30. Since he was never late, Kate assumed he was home sick.

3. (False beliefs - no mental state verbs). A boy is making a papier mache project for his art class. He spends hours ripping newspaper into even strips. Then he goes out to buy flour. His mother comes home and throws all the newspaper strips away.

4. (Mental states that are unexpected, given other information about the protagonist). [Background: Your friend Carla lives in San Francisco. She has a top position at a large computer company there. She has been working at the same corporation for over ten years.] Carla has always told you that she wants a husband who would expect her to stay at home as a housewife, instead of having her own career.

5. (Mental states described in the second person). You went to London for a weekend trip and you would like to visit some museums and different parks around London. In the morning, when you leave the hotel, the sky is blue and the sun is shining. So you do not expect it to start raining. However, walking around in a big park later, the sky becomes gray and it starts to rain heavily. You forgot your umbrella.

6. (Morally relevant beliefs) […] Because the substance is in a container marked "toxic," Grace thinks that it is toxic.

7. (Induced false beliefs) The path to the castle leads via the lake, but the children tell the tourists, "The way to the castle goes through the woods". The tourists now think the castle is?

Table 2. Verbal stimuli not eliciting a robust response in the right TPJ.

1. (False physical representations). This map shows the ground floor plan. A photocopy was sent to the architect yesterday. The map initially had a flaw: the kitchen door was missing. It was added to the map this morning. The architect's photocopy (choose one) includes / doesn't include the kitchen door.

2. (Physical inferences). The night was warm and dry. There had not been a cloud anywhere for days. The moisture was certainly not from rain. And yet, in the early morning, the long grasses were dripping with cool water.

3. (Descriptions of visible facts about people). Harry looks just like a math professor. He wears dark old cardigans with holes in the elbows, corduroy trousers and brown loafers over green argyle socks. The shoes Harry wears are (choose one) brown / green.

3b. (Visible facts that convey social information). Joe was a heavy-set man, with a gut that fell over his belt. He was balding and combed his blonde hair over the top of his head. His face was pleasant, with large brown eyes.

4. (Social background information about a person). Your friend Carla lives in San Francisco. She has a top position at a large computer company there. She has been working at the same corporation for over ten years.

5. (Bodily sensations - subjective but not representational states) Marcus had been sick for three days. He had felt weak and had a high fever. On the fourth day his fever broke, and he woke up feeling cool and alert.

6. (Morally relevant facts). […] The white powder by the coffee is not sugar, but a toxic substance left behind by a scientist.

7. (False signs). The sign to the monastery points to the path through the woods. While playing the children make the sign point to the golf course. According to the sign, the monastery is now?