Brief Communication

Abnormal moral reasoning in complete and partial callosotomy patients

Michael B. Miller [a,*], Walter Sinnott-Armstrong [b], Liane Young [c], Danielle King [a], Aldo Paggi [d], Mara Fabri [d], Gabriele Polonara [d], Michael S. Gazzaniga [a]

[a] *University of California, Santa Barbara, CA, United States*
[b] *Dartmouth College, Hanover, NH, United States*
[c] *Massachusetts Institute of Technology, Cambridge, MA, United States*
[d] *Ospedali Riuniti and Università Politecnica, Ancona, Italy*

ARTICLE INFO

ABSTRACT

Recent neuroimaging studies suggest lateralized cerebral mechanisms in the right temporal parietal junction are involved in complex social and moral reasoning, such as ascribing beliefs to others. Based on this evidence, we tested 3 anterior-resected and 3 complete callosotomy patients along with 22 normal subjects on a reasoning task that required verbal moral judgments. All 6 patients based their judgments primarily on the outcome of the actions, disregarding the beliefs of the agents. The similarity in performance between complete and partial callosotomy patients suggests that normal judgments of morality require full interhemispheric integration of information critically supported by the right temporal parietal junction and right frontal processes.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

Recent functional neuroimaging studies indicate that processes for ascribing beliefs and intentions to other people are lateralized to the right temporal parietal junction (TPJ) (Young, Cushman, Hauser, & Saxe, 2007; Young & Saxe, 2008). Specifically, Young and Saxe (2009) found that TPJ activity in the right hemisphere, but not the left, is correlated with moral judgments of accidental harms. These findings suggest that patients with disconnected hemispheres would provide abnormal moral judgments on accidental harms and failed attempts to harm, since normal judgments in these cases require information about beliefs and intentions from the right brain to reach the judgmental processes in the left brain.

The present study examines this hypothesis by comparing the performance of 22 normal subjects to 6 patients, 3 with the corpus callosum completely severed and 3 with only anterior portions severed (see Table 1 and Fig. 1). If normal moral judgments require transfer of information regarding an agent's beliefs from the rTPJ, full split-brain patients should be abnormal in their judgments. Partial split-brain patients with an intact splenium and isthmus, however, might show normal moral reasoning because the fibers connecting the right TPJ with the left hemisphere are intact.

Patients and controls made moral judgments about scenarios used in previous neuroimaging studies (Young et al., 2007). In each scenario the agent's action either caused harm or not, and the agent believed that the action would either cause harm or cause no harm. The crucial scenarios involved accidental harm (where the agent falsely believed that harm would not occur, but the outcome was harmful) and failed attempts (where the agent falsely believed that harm would occur but the outcome was not harmful). After each scenario was read, subjects were asked to judge the agent's action by vocally responding "permissible" or "forbidden". For the patients, this testing did not require any lateralized procedures as only the left hemisphere was assumed to be responding verbally.

Previous studies show that normal subjects typically base their moral judgments on agents' beliefs even when these are inconsistent with the actions' outcomes (Young & Saxe, 2008). In the example in Fig. 2, if Grace believed the powder was sugar but it was really poison, normal subjects judge Grace's action to be morally permissible because of her neutral belief. In contrast, when Grace falsely believed the powder was poison, normal subjects judge Grace's action to be morally impermissible because of her negative belief even though no harm ensues. Given prior evidence that processing of agents' beliefs is supported by regions in the right hemisphere, whereas the left hemisphere is responsible for the verbalization of moral judgments, we hypothesized that agents' beliefs would have less impact on moral judgments when the hemispheres are disconnected in only full callosotomy patients.
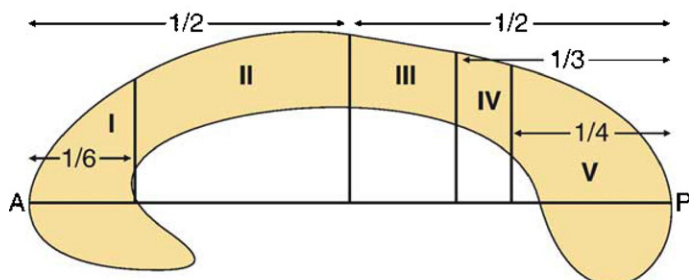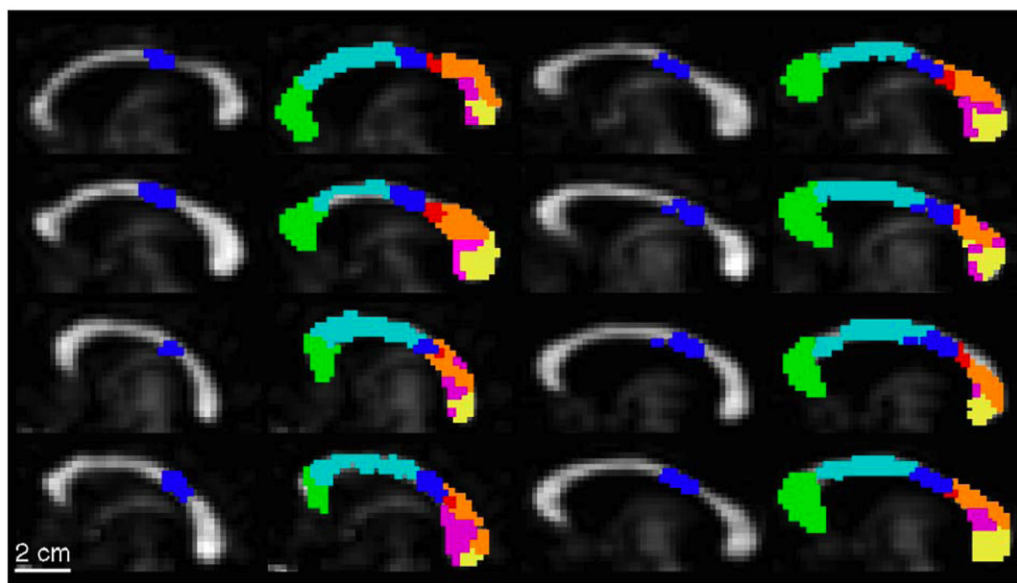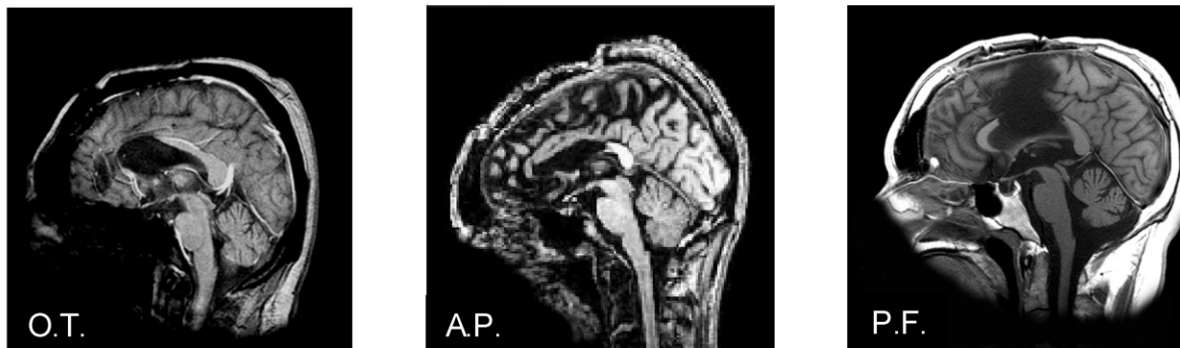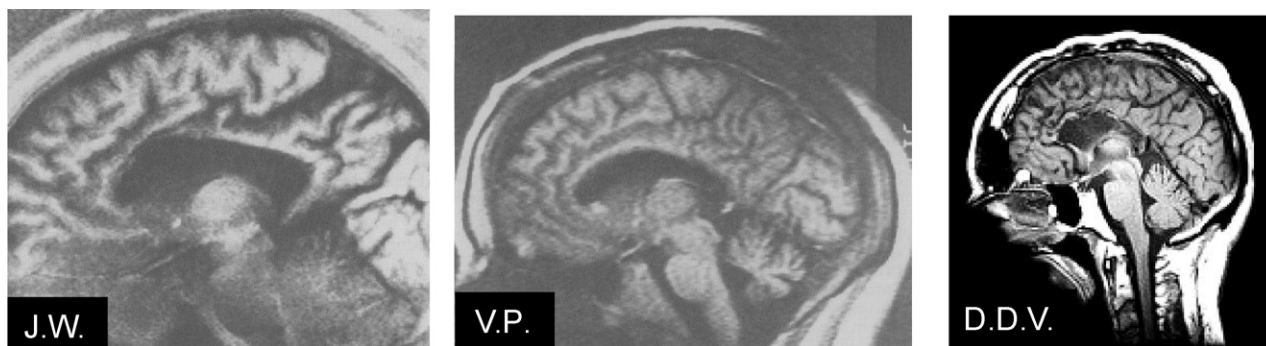
2. Methods

2.1. Participants

Twenty-eight participants (6 patients, 22 control subjects) provided prior informed consent and were treated according to APA ethical standards.

Fig. 1. The top panel shows MR images of midsagittal brain slices from the 6 patients with either full (J.W., V.P., and D.D.V.) or partial (O.T., A.P., and P.F.) corpus callosum resections. The bottom panel is from Hofer and Frahm (2006) showing fractional anisotropy maps of the midsagittal corpus callosum from 4 female (on the left side) and 4 male (on the right side) subjects and their classification scheme for originating brain regions of white matter projections.

Background

Grace and her friend are taking a tour of a chemical plant. When Grace goes over to the coffee machine to pout some coffee. Grace's friend asks for some sugar in hers. There is white powder in a container by the coffee.

<div style="text-align:center">

Negative　　　　**Neutral**

</div>

Foreground

Negative	Neutral
The white powder is *a very toxic substance left behind by a scientist, and deadly when ingested*.	The white powder is *just the regular sugar that the scientists have, and perfectly safe for coffee*.

Belief

The container is labeled "*toxic*", so Grace believes that the white powder is *a toxic substance*.	The container is labeled "*sugar*", so Grace believes that the white powder is *regular sugar*.

Outcome

Grace puts the substance in her friend's coffee. Her friend drinks the coffee and *dies*.	Grace puts the substance in her friend's coffee. Her friend drinks the coffee and *is fine*.

Judgment

Putting the substance in the coffee was: forbidden or permissible?

	Belief Neutral	Belief Negative
Outcome Neutral	consistent	**Failed Attempt**
Outcome Negative	**Accidental Harm**	consistent

Fig. 2. An example of the moral scenarios read to each patient and subject. Each scenario included either a negative or neutral belief and a negative or neutral outcome and foreground. The table at the bottom of the figure represents the four possible outcomes
Adapted from Young and Saxe (2008).

The six patients (four males, two females; ages 24–38) had undergone corpus callosum (CC) resection to treat drug-resistant epilepsy. Three patients (J.W., V.P., and D.D.V.) had complete CC resectioning, while the other three (O.T., A.P., and P.F.) had only partial CC resections (see Table 1). All patients were right-handed. Four resided in Italy, two in the United States.

Twenty-two neurologically normal control subjects participated in the study. Sixteen of these subjects (10 females, 6 males; ages 17–23) were recruited from an online subject pool at the University of California, Santa Barbara and paid $10 for participation. An additional six subjects (2 females, 4 males; ages 44–50) were recruited from the community in order to control for any differences that may be due to age and/or level of education. English was a second language in three of these older subjects, and they were matched to the patients by level of education (high school only or some college). All subjects were debriefed after the experiment.

2.2. Materials and design

The moral reasoning task presented 48 scenarios in a 2 × 2 × 2 design (see Fig. 2) for an example scenario). Each scenario began with background and facts foreshad-

owing the eventual outcome, followed by the agents' belief and the outcome. The first factor in the design was whether the agent's action caused harm or not (outcome: O− or O+); the second factor was whether the agent believed or did not believe that the action would cause harm (belief: B− or B+); and the third factor represented whether the agent's belief was revealed before or after the description of the outcome in the scenario (BO or OB). This design produced eight different versions of the task. Each version contained 6 trials of eight different trials types (O+B+, O+B−, O−B+, O−B−, B+O+, B+O−, B−O+, B−O−). The 6 split-brain patients and the 6 education-matched controls were randomly assigned to a unique version of the task. For the remaining 16 normal control subjects, the 8 different versions were counterbalanced across subjects. Scenarios were presented in the same order for all subjects.

In half of the scenarios, the belief and the outcome were consistent with each other (neutral belief and outcome, or negative belief and outcome). In the other half, the belief and outcome were inconsistent: neutral belief and negative outcome (accidental harm) or negative belief and neutral outcome (failed attempt to harm).

2.3. Procedure

Subjects were first read aloud instructions saying that they would hear a series of fictional scenarios, each of which would contain information about the actions of a person that would affect somebody else, and their task would be to judge whether the action was permissible or forbidden. "Permissible" was explained as meaning "morally acceptable," whereas "forbidden" was described as meaning "not morally acceptable." Experimenters emphasized that there was no right or wrong answer, and subjects should base their answers on their own moral judgment. They were also encouraged to ask for clarification if they were ever confused or needed something repeated. After the experimenter read aloud each scenario, subjects indicated vocally whether they judged the action to be morally permissible or forbidden.

Table 1
Split-brain patient information.

Patient	Sex	Hand	Born	IQ	Home	Surgery	Type	Year
J.W.	M	R	1954	96	USA	Full	2-stage	1979
V.P.	F	R	1953	91	USA	Full	2-stage	1980
D.D.V.	M	R	1964	81	Italy	Full	2-stage	1986/1994
O.T.	M	R	1958	80	Italy	Partial	1/3 anterior	1995
A.P.	M	R	1965	90	Italy	Partial	3/4 anterior	1993
P.F.	M	R	1984	??	Italy	Partial	1/3 central	2008

??: The IQ score of this patient is unknown.
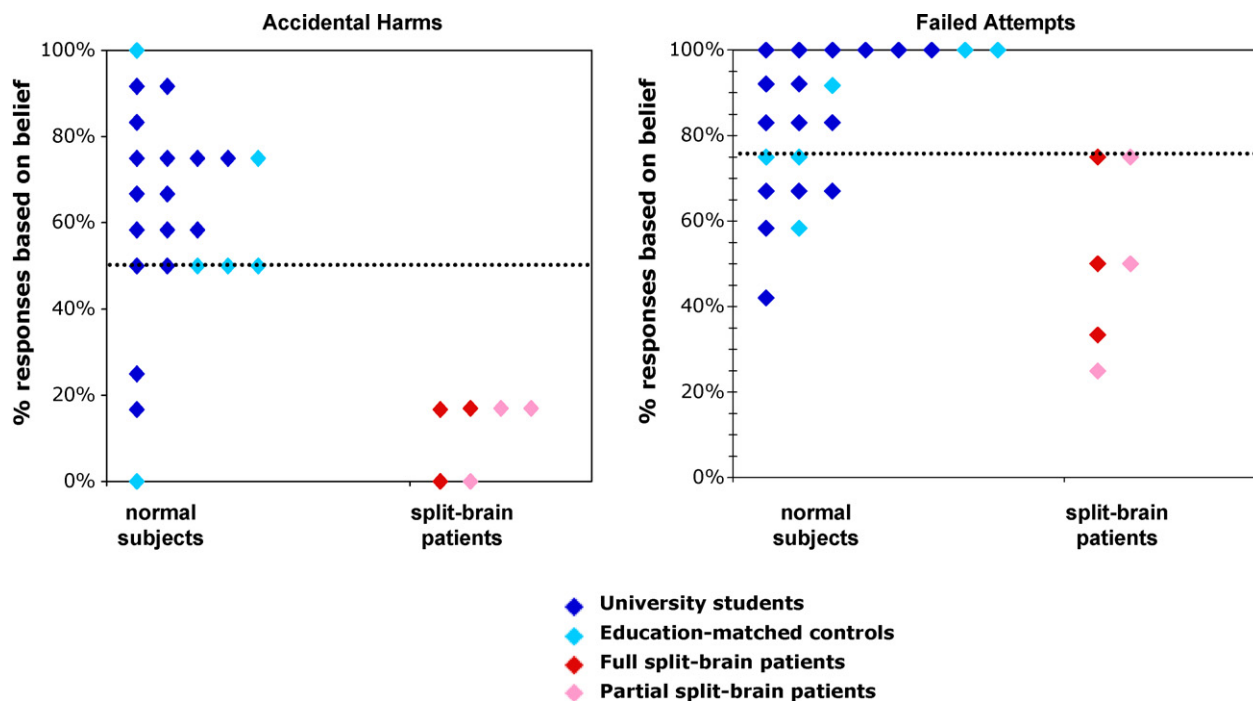
Fig. 3. A plot of accidental harms (where the agent falsely believed that harm would not occur, but the harmful outcome did occur) and failed attempts (where the agent falsely believed that harm would occur but the harmful outcome did not occur). All the split-brain patients fall below the 95% confidence interval (the dotted line).

2.4. The faux pas task

To determine whether any differences in performance represented a general deficit in belief attribution, two of the split-brain patients (V.P. and J.W.) and the six education-matched control subjects were also tested in a social faux task previously used to test theory of mind (Stone, Baron-Cohen, & Knight, 1998). Experimenters read aloud twenty scenarios, half of which involved a social faux pas (FP), such as telling somebody how much they dislike a bowl while forgetting that that person had given them that bowl as a wedding present. Subjects were then asked, "Did anyone say something they shouldn't have said or something awkward?" If their response was "yes," then they were asked five more questions about the faux pas, and two control questions to assess their comprehension of the story. If they responded "no," they were asked only the two control questions. All responses were recorded by the experimenter.

3. Results

3.1. The moral reasoning task

All 6 patients based their moral judgments primarily on the outcomes of the scenarios and not the beliefs of the agents (see Fig. 3 and Table 2). When beliefs and outcomes were inconsistent, the proportion of judgments that the patients based on agents' beliefs (mean = 31%) fell well below the 95% confidence interval level of the normal subjects (mean = 72%). A repeated measures ANOVA using the percentage of responses based on the agents' beliefs as the dependent variable showed a significant main effect for patients versus normal controls ($F(1,26) = 29.95$, MSE = .04, $p < .001$), a significant main effect for consistency between beliefs and outcomes or not ($F(1,26) = 105.04$, MSE = .02, $p < .001$), and a significant interaction between the consistency of the scenarios and the group ($F(1,26) = 16.89$, MSE = .02, $p < .001$).

The difference between the patient group and the control group was larger for accidental harm scenarios (neutral belief and negative outcome) than for failed attempt scenarios (negative belief and neutral outcome). Along with a significant main effect for negative versus neutral beliefs ($F(1,26) = 43.47$, MSE = .02, $p < .001$), there was also a significant interaction between the group and whether the belief was neutral or negative ($F(1,26) = 7.04$, MSE = .01, $p = .013$)

(the three-way interaction was not significant). This might be due either to the emotional salience of negative versus neutral outcomes or to qualitative differences between neutral versus negative beliefs. It has been suggested that the right TPJ mediates ascriptions of belief, whereas the ventral medial prefrontal cortex mediates ascriptions of desire and intent (Ciaramelli, Muccioli, Ladavas, & di Pellegrino, 2007; Koenigs, Young, Adolphs, Tranel, & Cushman, 2007; Young & Saxe, 2009). Judgments of accidental harm depend only on the former, whereas judgments of failed attempts may also involve the latter. Our results indicate that both attributions are affected, but that hemispheric disconnection might affect judgments of accidental harms more than judgments of failed attempts.

Also, the patients tended to respond "forbidden" more than normal controls, even when the belief and outcome were consistently neutral (see Table 1; $t(26) = 3.66$, $p = .001$). Spontaneous verbalizations by the split-brain patients indicated that they often based judgments on safety (e.g. judging an act forbidden because "kids shouldn't be playing near a swimming pool" even though the agent's belief and the action's outcome were both neutral). However, as noted above, the difference between the split-brain patients and the normal subjects in the proportion of responses based on agents' beliefs was much more pronounced for the inconsistent scenarios than for the consistent scenarios.

Although previous studies have not shown a relationship between moral judgments and intelligence (Cima, Tonnaer, & Hauser, in press), there may be concern that the low average IQ (87.6) of the callosotomy patients may be the critical factor in their tendency to not respond to the accidental harms and failed attempts based on beliefs. For example, are they confused by the task, leading to a lower percentage of responses based on beliefs? We believe this is an unlikely explanation for three reasons. First, all of the patients performed normally on the control tasks including the consistent scenarios and they expressed no confusion about the task during informal debriefing sessions. Second, although we only have IQ scores for 5 out of the 6 patients, there appears to be no relationship between IQ score and the percentage of times that their response is based on the beliefs

Table 2
Percentage of responses based on agents' beliefs (standard error in parentheses).

	Consistent		Inconsistent	
	Negative belief and outcome	Neutral belief and outcome	Negative belief and neutral outcome (failed attempts)	Neutral belief and negative outcome (accidental harms)
Normal subjects	98% (1%)	90% (2%)	83% (4%)	61% (5%)
Split-brain patients	92% (4%)	74% (3%)	51% (8%)	11% (4%)

for the inconsistent scenarios. For example, patient JW had the highest IQ score (96) yet he had the lowest percentage of belief responses (0% for accidental harms). And, third, six of the control subjects were selected in order to be matched with the patients for level of education (high school only or some college). Although level of education is not as reliable a measure of intelligence as IQ scores, the difference between the two groups was quite significant. As shown in Fig. 3, the 6 education-matched controls responded based on beliefs significantly more than the 6 patients ($t(10) = -3.51$, $p = .006$). Further, the 6 education-matched controls responded no differently than the 16 controls that were currently university students ($t(20) = -0.53$, n.s.).

Also, in order to discount the possibility that patients base their judgments on the last bit of information received, we counterbalanced when beliefs were presented in the scenarios (half before outcomes, half after). When the factor of order of belief and outcome was included in the analysis, we found no main effect for order ($F(1,26) = .25$, MSE = .03, n.s.) nor any significant interaction with other factors, including group characteristics.

Unexpectedly, all of the results reported above were similar for both the full and partial callosotomy patients. According to MRI scans of the partially disconnected patients, their surgical cut of the corpus callosum did not extend posterior enough to transect any crossing fibers from the right TPJ (see Fig. 1). While we predicted abnormal performance in the full callosotomy patients, we did not anticipate a similar effect in partial callosotomy patients. The observed effect in the partial patients might be explained in three ways: (1) despite the MRI evidence, the cut extended far enough to affect callosal fibers crossing from the TPJ; (2) the true site for decoding beliefs is anterior to the TPJ; or (3) the TPJ is necessary for decoding beliefs but must work together with more anterior modules to function properly.

3.2. The faux pas task

To determine whether the observed effects were isolated to a particular moral reasoning task, we also tested two of the full callosotomy patients on a social faux pas task. The patients could not identify the faux pas. Out of ten faux pas, patient V.P. successfully detected only six, and patient J.W. correctly identified only four. Both patients responded correctly to both control questions in all ten of these scenarios, demonstrating intact comprehension. The control subjects in our study all recognized 100% of the faux pas, as did previously reported patient and control samples, including dorsal frontal patients ($M = 100\%$, SD = 0, $n = 4$), orbital frontal patients (82.5%, SD = 12.9, $n = 5$), anterior temporal patients ($M = 100\%$, SD = 0, $n = 1$) and normal controls ($M = 100\%$, SD = 0, $n = 10$) (Stone et al., 1998). Although our sample size was small, we observed a significant relationship between the performance on the moral reasoning task and on the faux pas task across the 6 education-matched control subjects and the 2 callosotomy patients ($r(7) = .65$, $p = .041$, one-tailed). This finding indicates that the impairment in the callosotomy patients reflects a specific deficit in belief attribution resulting in a deficit in this specific aspect of moral reasoning.

4. Discussion

We found that full and partial callosotomy patients based their moral judgments primarily on actions' outcomes, disregarding agents' beliefs. Young children make similar judgments on comparable tasks (e.g. Baird & Astington, 2004; Piaget, 1965/1932), often judging the wrongness of an action by the outcome and not by the intention of the actor. However, the patients appear to have no difficulty comprehending the scenarios, given their near-perfect responses to control questions and their explanations during debriefing sessions following testing.

It seems paradoxical that no real-life effects of impaired belief processing have been reported in the literature on callosotomy patients. There are three possible explanations: (1) profound effects have gone unreported. This seems unlikely, though a few anecdotal but unpublished reports from spouses and family members describe personality changes after surgery. However, most reports from family members suggest no changes in mental functions or personality and early studies that thoroughly tested patients pre- and post-operatively reported no changes in cognitive functioning (LeDoux, Risse, Springer, Wilson, & Gazzaniga, 1977). (2) The tasks used in this study have little relevance to the real world. This also seems unlikely given that the effect was found in tasks commonly used for probing social cognition. (3) The impairment is compensated for by other mechanisms when patients have more time to think and process information. Some have postulated two separate systems for deriving moral judgments (Greene, 2007) and for belief reasoning (Apperly & Butterfill, 2009; Apperly, Riggs, Simpson, Chiavarino, & Samson, 2006): a fast, automatic, implicit system and a slow, deliberate, explicitly reasoned system. If the patients do not have access to the fast, implicit systems for ascribing beliefs to others, their initial, automatic moral judgments might not take into account beliefs of others.

The third explanation seems most likely to us, given the spontaneous verbalizations and rationalizations offered by some patients during testing. The spontaneous drive to interpret and rationalize events in the world is a well-known phenomenon in split-brain research (Gazzaniga, 2000), as is the rationalization of moral judgments in the general population (Haidt, 2001). In this study, it appears the patients sensed their judgments were not quite right, and they often offered rationalizations for why they judged the act as they did without any prompting. For example, in one scenario a waitress served sesame seeds to a customer while falsely believing that the seeds would cause a harmful allergic reaction. Patient J.W. judged the waitress's action "permissible." After a few moments, he spontaneously offered, "Sesame seeds are tiny little things. They don't hurt nobody," as though this justified his judgment. Out of the 24 scenarios read to patient J.W. in this testing session, he offered 5 spontaneous rationalizations for his judgments (see the supplementary material). Although these examples are not necessarily evidence of a slow, deliberate system of moral reasoning that is adjusting for a fast, automatic moral judgment, they are illustrative of how these previously hypothesized systems may operate (Greene, 2007; Haidt, 2001). The proposed explanation is also supported by evidence that high-functioning autistic

children use cognitive processes like language and logical reasoning to compensate for deficits in belief reasoning or theory of mind (Tager-Flusberg, 2001).

The present study demonstrates that full and partial callosotomy patients fail to rely on agents' beliefs when judging the moral permissibility of those agents' actions. This finding confirms the hypotheses that specialized belief-ascription mechanisms are lateralized to the right hemisphere and that disconnection from those mechanisms affects normal moral judgments. Moreover, the neural mechanism by which interhemispheric communication occurs between key left and right hemisphere processes seems complex. Since the partial anterior callostomy patients also showed the effect, it would appear the right TPJ calls upon right frontal processes before communicating information to the left speaking hemisphere.

Acknowledgments

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.neuropsychologia.2010.02.021.

References

Apperly, I. A., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review*, *116*(4), 953–970.

Apperly, I. A., Riggs, K. J., Simpson, A., Chiavarino, C., & Samson, D. (2006). Is belief reasoning automatic? *Psychological Science*, *17*(10), 841–844.

Baird, J. A., & Astington, J. W. (2004). The role of mental state understanding in the development of moral cognition and moral action. *New Directions for Child and Adolescent Development*, *103*, 37–49.

Ciaramelli, E., Muccioli, M., Ladavas, E., & di Pellegrino, G. (2007). Selective deficit in personal moral judgment following damage to ventromedial prefrontal cortex. *Social, Cognitive, and Affective Neuroscience*, *2*, 84–92.

Cima, M., Tonnaer, F., & Hauser, M.D. (in press). Psychopaths know right from wrong but don't care. *Social, Cognitive, and Affective Neuroscience*.

Gazzaniga, M. S. (2000). Cerebral specialization and interhemispheric communication: Does the corpus callosum enable the human condition? *Brain*, *123*, 1293–1326.

Greene, J. D. (2007). Why are VMPFC patients more utilitarian?: A dual-process theory of moral judgment explains. *Trends in Cognitive Sciences*, *11*(8), 322–323.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, *108*, 814–834.

Hofer, S., & Frahm, J. (2006). Topography of the human corpus callosum revisited—Comprehensive fiber tractography using diffusion tensor magnetic resonance imaging. *Neuroimage*, *32*, 989–994.

Koenigs, M., Young, L., Adolphs, R., Tranel, D., Cushman, F., et al. (2007). Damage to the prefrontal cortex increases utilitarian moral judgements. *Nature*, *446*, 908–911.

LeDoux, J. E., Risse, G. L., Springer, S. P., Wilson, D. H., & Gazzaniga, M. S. (1977). Cognition and commissurotomy. *Brain*, *100*, 87–104.

Piaget, J. (1965/1932). *The moral judgment of the child*. New York: Free Press.

Stone, V. E., Baron-Cohen, S., & Knight, R. T. (1998). Frontal lobe contributions to theory of mind. *Journal of Cognitive Neuroscience*, *10*(5), 640–656.

Tager-Flusberg, H. (2001). A reexamination of the Theory of Mind hypothesis of autism. In J. A. Burack, T. Charman, N. Yirmiya, & P. R. Zelazo (Eds.), *The development of autism: Perspectives from theory and research*. Lawrence Erlbaum Associates.

Young, L., Cushman, F., Hauser, M., & Saxe, R. (2007). The neural basis of the interaction between theory of mind and moral judgment. *Proceedings of the National Academy of Science*, *104*, 8235–8240.

Young, L., & Saxe, R. (2008). The neural basis of belief encoding and integration in moral judgment. *Neuroimage*, *40*, 1912–1920.

Young, L., & Saxe, R. (2009). Innocent intentions: A correlation between forgiveness for accidental harm and neural activity. *Neuropsychologia*, *47*, 2065–2072.