

Attributions on the brain: Neuro-imaging dispositional inferences, beyond theory of mind[☆]

Lasana T. Harris, Alexander Todorov, and Susan T. Fiske*

Department of Psychology and Center for the Study of Brain, Mind, and Behavior, Green Hall 2-N-14, Princeton University, Princeton, NJ 08544-1010, USA

Received 3 February 2005; revised 26 April 2005; accepted 2 May 2005
Available online 19 July 2005

People need to predict what other people will do, and the other person's perceived disposition is the preferred mode of prediction. People less often use, for example, shared social norms to explain another person's behavior. Social psychology's last half-century of research on attribution theory offers precise, validated paradigms for testing how people think about other people's minds. Neuro-imaging data from one classic attribution paradigm shows the unique priority given to inferring chronic, idiosyncratic dispositions (unique attitudes, individual personality, idiosyncratic intent), compared to other kinds of mental contents. Specifically, sentences describing behavior that is low in consensus across actors, low in distinctiveness across entities, and high in consistency over time (compared with the other 7 low–high combinations) uniquely elicits (a) person attributions and (b) activation in the superior temporal sulcus. Ignoring consensus, both low-distinctiveness, high-consistency combinations (compared to 6 remaining combinations) also activate the MPFC, consistent with decades of behavioral data showing that general social cognition neglects consensus information. Thus, activated areas converge with prior neuro-imaging data on theory of mind and social cognition, but more precisely isolate the exact nature of the inferences that activate these areas.

© 2005 Elsevier Inc. All rights reserved.

Unlike objects, people are agents, that is, they originate action. Humans therefore focus on knowing other people's intentions and dispositions. People more often explain other people's behavior by their allegedly enduring dispositions and intentions than by other plausible accounts, for example the circumstances (Jones, 1979; Ross, 1977). People favor dispositions for several reasons: They have unrealistic beliefs that situations do not much influence behavior, they tend to start with dispositional explanations and inadequately correct for situational influences, and they have difficulty construing other people's situational constraints (Gilbert and Malone, 1995). People's

disproportionate focus on idiosyncratic dispositions stems from their search for consistency in other people's behavior (Heider, 1958).

Recent research indicates that primary, universal dimensions of social cognition do include social intent for good or ill and capability to enact those intentions (Fiske et al., 2002; Cuddy, Fiske, and Glick, unpublished; Cuddy et al., in press). Our social institutions prioritize intent also, as in the legal emphasis on state of mind (Fiske, 1989). Thus, people believe that other people's intentions are central, and they indeed use them in social cognition. Intentions are seen to follow from dispositions.

Social psychology's focus on disposition and intentions reflects their role in identifying people as causal agents. Attribution theory has examined how people infer another's disposition from a single behavior, that is, whether the behavior corresponds to a unique intent (Jones, 1979); attribution theory has also examined patterns of behavior over time (Kelley, 1972). In both cases, observers want to know what other people would do (social desirability and consensus) and what other targets of the behavior evoke from that person (unique effects and distinctive responses). For example, if the individual actor is the only person behaving that way and does so indiscriminately toward the entire class of target entities, then the behavior is caused by something about the person, particularly if it occurs consistently over time. Suppose John laughs at the comedian, but no one else does, he laughs at all comedians, and he does it every time, then something about John as a person apparently caused his behavior. Empirical tests generally support this model, varying consensus across actors, distinctiveness over entities, and consistency over time (e.g., McArthur, 1972; Ferguson and Wells, 1980; Orvis et al., 1975; Pruitt and Insko, 1980; Zuckerman, 1978). People use consensus information less than predicted (see Fiske, 2004, for a review), but otherwise the model predicts the attribution of intent and dispositions¹.

[☆] These data were presented at the Princeton seminar on the Neuroscience of Social Decision-Making, winter 2005, and at the Social Neuroscience Pre-conference preceding the meetings of the Society for Personality and Social Psychology, January 2005.

* Corresponding author. Fax: +1 609 258 1113.

E-mail address: sfiske@princeton.edu (S.T. Fiske).

Available online on ScienceDirect (www.sciencedirect.com).

¹ Although precise in many respects, attribution theories have not distinguished among types of dispositions (intent, personality, and attitude). Also, note that not all attribution theorists agree with Jones' and Kelley's reading of Heider. Malle (2004), for instance, suggests that Kelley's cube is faulty because it only addresses instances where intention is already known. Even with these criticisms, however, we still rely upon this paradigm because it allows us to explore unique dispositions, from which intent is inferred. Moreover, people use all types of dispositions to predict other people's behavior.

Table 1
Items constituting the experimental questionnaire (McArthur, 1972)

Verb category	Item
Emotions	1. John laughs at the comedian.
	2. Sue is afraid of the dog.
	3. Tom is enthralled by the painting.
	4. Mary is angered by the psychology department.
Accomplishments	5. George translates the sentence incorrectly.
	6. Henry gets a birdie on the fifth hole.
	7. While dancing, Ralph trips over Joan's feet.
	8. Linda receives three invitations to the church picnic.
Opinions	9. Jim thinks the building is very tall.
	10. Mark believes the operation will be a success.
	11. Bill thinks his teacher is unfair.
	12. Kay thinks the stream is polluted.

The burgeoning field of social cognitive neuroscience has begun to address attribution theory. Lieberman et al. (2002) discussed attribution theory from the perspective of causal attributions. The authors hypothesize a “behavioral identification” (p. 216) pathway in the temporal cortex, ascribing the role of inferring intention to the superior temporal cortex (STS). While the authors describe this as a reflexive “X-system” process, they posit that regions in the prefrontal cortex are a part of the reflective “C-system” process that is responsible for propositional thought and internally generated inferences.

Operating in parallel with attribution theory has been the developmental neuroscience of Theory of Mind (ToM). This research generally has separated people's understanding of other people's individual dispositions (hypothesized to be an early-developing process) and their beliefs (hypothesized to be later) (Saxe et al., 2004). ToM work on dispositions includes goals (inferred from bodily actions), attention (inferred from gaze direction), and emotion (inferred from others' expressions). Goals are close cousins to intentions, which ordinary people demonstrably believe predict behavior. Other prominent ToM work separates perceived intent and biological motion, episodic memory retrieval, and the decoupling of mental states from reality (e.g., Frith and Frith, 2001; Gallagher and Frith, 2002). Again, it is the unique intentions that are most relevant here.

Imaging predictions for the attribution of dispositions are clear from previous literature. The fMRI studies show that social cognition (thinking about people) generally activates the medial prefrontal cortex (MPFC) (Brunet et al., 2000; Castelli et al., 2000; Gallagher et al., 2000; Mitchell et al., 2002, 2004; Ochsner et al., 2004).

The fMRI measures on tasks specifically involving inferences of intentions consistently identify three core neural regions: medial prefrontal cortex (MPFC), superior temporal sulcus (STS), and anterior temporal poles (Abu-Akel, 2003; Blakemore et al., 2003; Brunet et al., 2000; Calarge et al., 2003; Castelli et al., 2000; Frith and Frith, 2001; Gallagher and Frith, 2002; Sabbagh, 2004; Saxe et al., 2004²). The MPFC region, in addition to activating to general “mentalizing” tasks, is particularly active for inferring agent-centered descriptions (Leslie, 1994). For the second major area highlighted in the ToM, studies have identified the STS as a region particularly sensitive to biological motion (Grezes et al., 1998; Hoffman and Haxby, 2000). This area is also involved in

² Most studies draw participants from patients with lesions and mental illness. This brief review does not cite these studies.

understanding the meaning of stories involving people (Gallagher et al., 2000), necessary, but not sufficient for person attributions. Finally, the Gallagher and Frith (2002) review mentions the temporal poles as another area recruited in the theory of mind tasks. This area responds to familiar characteristics of interpersonal interaction, such as voices, faces and social environments (Nakamura et al., 2000, 2001). However, written stimuli, such as those used here, might not activate the temporal poles.

Materials and methods

Twelve Princeton undergraduates participated for course credit. Participants reported no abnormal neurological condition, were right-handed and suffered no incidence of head trauma or brain lesions. All participants had normal or corrected vision, were native English speakers, and provided informed consent. The mean age was 20 years, with 11 women, one ethnic minority (Asian), and two multi-racial participants.

The method stayed as close to the McArthur (1972) paradigm as possible. Deviations are noted and explained. We used 12 of the original 16 sentences (see Table 1). The verb in each sentence belonged to the categories emotion, accomplishment, or opinion. The verb category action, a part of the original design, was left out because those original sentences were twice as long as the other three categories, and the topic was dated (automobile safety campaigns). The sentences were presented using the stimulus display package Eprime.

We employed repeated measures, in which all subjects rated all 12 sentences in all eight conditions (instead of the 16×16 Latin Square design used by McArthur), due to power advantages. All participants practiced the task on a dummy set of sentences before entering the scanner. The task consisted of responding to an attribution question about each target sentence, after seeing additional information about consensus, distinctiveness, and consistency. A fixation cross appeared on the screen at the beginning of each run for 6 s and between each sentence block for 2 s. The overall design was 2 (Consensus: high vs. low) $\times 2$ (Consistency: high vs. low) $\times 2$ (Distinctiveness: high vs. low; see Table 2), as was demonstrated in our “John laughs at the comedian” example above (see also Fig. 1). Altogether, participants performed 8 runs of 12 sentences each. All 8 conditions appeared in each run of our event-related design, and all

Table 2
Combinations of consensus across actors, distinctiveness over entities, and consistency over time

• HHH	• LLH
-High consensus	-Low consensus
-High distinctiveness	-Low distinctiveness
-High consistency	-High consistency
• LHH	• LHL
-Low consensus	-Low consensus
-High distinctiveness	-High distinctiveness
-High consistency	-Low consistency
• HLH	• HLL
-High consensus	-High consensus
-Low distinctiveness	-Low distinctiveness
-High consistency	-Low consistency
• HHL	• LLL
-High consensus	-Low consensus
-High distinctiveness	-Low distinctiveness
-Low consistency	-Low consistency

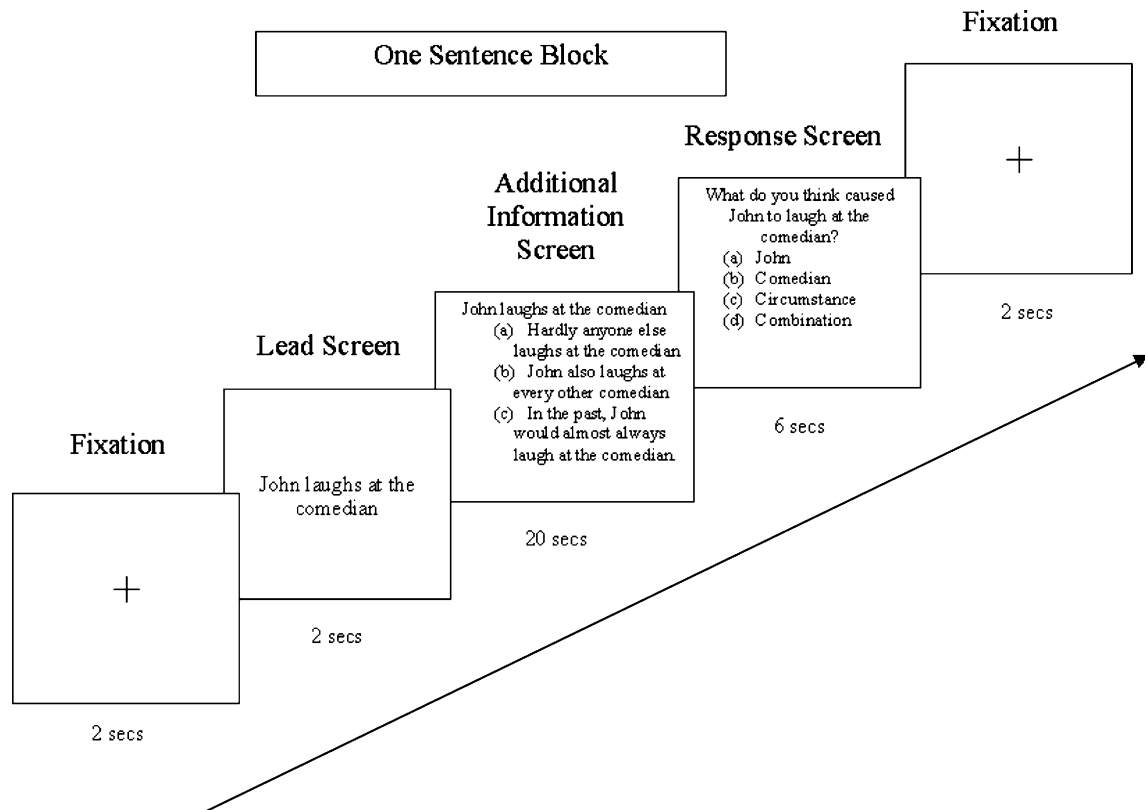


Fig. 1. One sentence block.

12 sentences appeared in each run, according to a partial Latin square design³.

Each trial operated as follows (see Fig. 1): The lead sentence, such as “Sue is afraid of the dog”, was centered on the first or “lead screen” for 2 s. A second screen, the “additional information screen”, displayed the lead sentence again, along with sentences that manipulated consensus, distinctiveness and consistency. This additional information screen remained for 20 s. Finally, the “response screen” appeared, asking participants to make an attribution. Thus, for our example, “Sue is afraid of the dog”, the response screen read, “What do you think caused Sue to be afraid of the dog?” The four options were listed below the question. Again, for our example, the four options were “(a) Sue (person), (b) The dog (stimulus), (c) Circumstance, (d) Combination”. Note that for each sentence, we tailored the options to the information in the sentence, to keep our participants from forming a heuristic about people-versus-stimuli attributions, and also to force them to think about the person as a person. The response screen remained for 6 s,⁴ followed by the fixation cross, which remained for 2 s,⁴ and then the next sentence block. The participants were instructed and trained to choose one of the four options by pressing a button on a response box while in the scanner. The response box was connected to a separate computer that

recorded the responses. We always kept the assignments fixed, that is, the person option was always labeled (a), the stimulus (b), circumstance (c), and combination (d).

Scanning parameters

All fMRI scanning was conducted at Princeton’s Center for the Study of Brain, Mind, and Behavior, which uses a 3.0 T Siemens Allegra head-dedicated MR scanner. A Dell computer projecting to a screen mounted at the rear of the scanner bore, which participants viewed while prone, through a series of mirrors, presented the

Table 3
Regions of brain activity associated with attributing behavior to the characteristics of a person

Anatomical label	x	y	z	Cluster size	t value
Right superior temporal sulcus	60	-44	18	20	3.25
Left medial prefrontal cortex (BA 9)	-2	46	28	22	3.74
Right medial prefrontal cortex	5	50	0	27	2.85
Right frontal cortex (BA 10)	25	50	19	22	2.86
Right middle temporal gyrus	58	-51	6	209	3.97
Right precentral frontal gyrus (BA 13)	50	-10	12	63	3.52
Right middle temporal gyrus	49	-66	22	122	3.63
Right middle occipital gyrus (BA 19)	46	-71	7	647	4.42
Right precentral frontal gyrus (BA 4)	45	-14	47	124	3.79
Left insula (BA 13)	45	10	-2	7	3.24
Right precuneus	5	-63	19	14	3.13
Left cingulate gyrus (BA 24)	-3	6	36	10	3.09

Coordinates are from the Talairach and Tournoux (1988) atlas. The displayed t values are for the area’s peak hemodynamic response relative to the other seven conditions at $P < 0.025$, corrected for multiple comparisons.

³ The original McArthur Latin Square was 16 sentence stimuli \times 16 conditions (8 stimuli combinations, $(2 \times 2 \times 2) \times 2$ replications). Because we used only 12 of the 16 sentences, but did use the Latin Square, our design is a partial replication.

⁴ The Inter Trial Interval (ITI) was, in fact, a lot longer than 2 s because in all likelihood, the participants thought of the response before the response screen was displayed. That leaves us with 8 s, including the 6 s when the response screen was displayed.

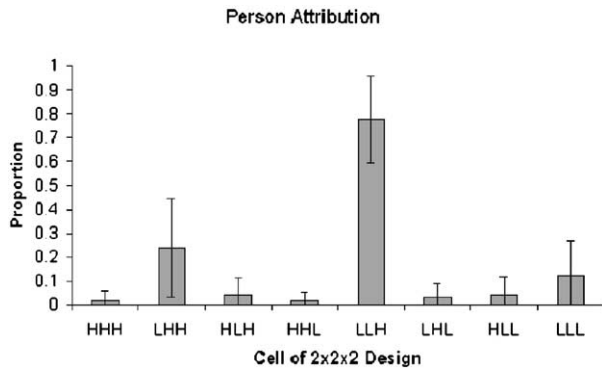


Fig. 2. Behavioral data: proportion of person attributions, by $2 \times 2 \times 2$ condition⁵.

stimuli. Responses were recorded using bimanual fiber-optic response pads (Current Designs Inc., url: <http://www.curdes.com/response>). Prior to the functional echo planar image (EPI) acquisitions, structural MRI scans were acquired to allow for subsequent functional localization. These scans took approximately 15 min and included: (1) a brief scout for landmarking and (2) a high-resolution whole-brain MPRAGE sequence for later localization and intersubject registration. Functional imaging then proceeded using an EPI sequence that allowed for whole-brain coverage in a relatively short period of time (thirty-two 3-mm axial slices; 1 mm gap; TR: 2 s; TE: 30 ms). In-plane resolutions were 3 mm \times 3 mm (196 mm FOV, 64 \times 64 matrix).

Preprocessing

Both image preprocessing and statistical analysis used Brain Voyager 2000, version 4.9 (Brain Innovation, Maastricht, The Netherlands, www.BrainVoyager.com). Before statistical analysis, image preprocessing consisted of: (1) slice acquisition order correction; (2) 3D rigid-body motion correction; (3) voxel-wise linear detrending across time; and (4) temporal bandpass filtering to remove low- and high-frequency (scanner and physiology related) noise. Distortions of EPI images were corrected with a simple affine transformation. Functional images were registered to the structural images and interpolated to cubic voxels. After coregistering the participants' structural images to a standard image using a 12-parameter spatial transformation, their functional data were similarly transformed, along with a standard moderate degree of spatial smoothing (Gaussian 8 mm FWHM).

Data analysis

Data were analyzed using the general linear model available on the Brain Voyager software package. Functional image data were analyzed on a voxel-wise basis with multiple regression. A series of regressors examined blood oxygen level-dependent (BOLD) brain activity to each of the eight possible combinations resulting from Kelley's cube. The beta weights for these regressors were converted to indices of response strength, expressed in units of percent signal change relative to a baseline—the fixation cross. The maps of response magnitudes for each participant were converted to Talairach space (Talairach and Tournoux, 1988) to allow for a group analysis of the data. Using the regressors of interest, we tested the significance of an a priori defined contrast for low consensus,

high consistency, low distinctiveness versus the other 7 experimental cells.⁶ Participants were treated as a random factor in the analysis, providing a conservative test of the hypothesis. Significant clusters were defined as contiguous voxels with $P < .025$, corrected for multiple comparisons using the false discovery rate (FDR) correction (Genovese et al., 2002). The maximum t value for each cluster is reported in the relevant figure and in Table 3.

Results and discussion

Behavioral data

For each participant, we computed the proportion of responses indicating person attribution. As predicted by the attribution model of Kelley (1972), and replicating McArthur (1972), participants were most likely to attribute causality to the person when consensus and distinctiveness were low and consistency was high (77.69%, where chance is 25%). Consistent with this finding, a 2 (Consensus) \times 2 (Distinctiveness) \times 2 (Consistency) repeated measures analysis of variance revealed a significant three-way interaction, $F(1,11) = 24.05$, $P < .05$, with an effect size as indicated by partial $\eta^2 = .69^7$.

As Fig. 2 shows, attributions of causality to the person occurred only when consensus and distinctiveness were low and consistency was high. This was the only condition in which the proportion of person attribution was significantly higher than chance (25%), $t(11) = 10.06$, $P < .05$. The next highest proportion (.24) was at chance and occurred when low-consensus, high-distinctiveness and high-consistency information (LHH, in the figure) was presented.

Imaging data

The analysis compared the event-related BOLD responses of the 20 s when the information was on the screen. We decided to use the entire 20 s while the additional information screen was displayed because the precise moment when the subject made the attribution is unknown. Furthermore, the BOLD signal change curves suggested that different subjects made the attribution for different sentences at different times, although the average was between 9 s and 15 s. We employ this strategy for all reported BOLD activity.

We contrasted the condition in which participants were most likely to attribute causality to the person—low distinctiveness, low consensus, high consistency (LLH)—with the other 7 conditions. The condition that implied person attribution evoked activity in superior temporal sulcus (STS) (see Fig. 3), $t(11) = 3.25$, $P < .008$, with an effect size as indicated by $\eta^2 = .47^8$.

⁶ Note that two runs were lost each for two participants, due to equipment malfunctions.

⁷ Attribution to the stimulus also revealed a significant triple interaction, $F(1,11) = 24.32$, $P < .05$, with an effect size as indicated by partial $\eta^2 = .69$; stimulus attributions were most likely (.83, all others below chance), when consensus, distinctiveness, and consistency all were high (HHH), as predicted by Kelley's model. The combination attribution, also predicted by Kelley, showed a significant three-way interaction, $F(1,11) = 23.05$, $P < .05$, effect size as indicated by partial $\eta^2 = .68$. The highest cell was (HLH), at .60, but 4 other cells were close. The circumstance attribution revealed only a marginally significant interaction, $F(1,11) = 3.49$, $P = .09$, but it occurred most often in cell (LHL), as Kelley would predict.

⁸ All reported effect sizes for imaging data were calculated using the formula: $\eta^2 = t^2 * df_{(between)} / (t^2 * df_{(between)} + df_{(within)})$.

⁵ See Table 2 for an explanation of the uppercase lettering.

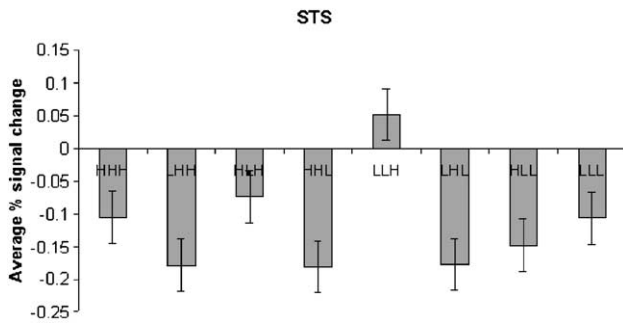


Fig. 3. Superior temporal sulcus: average signal change.

This was predicted by our reading of the previous literature on ToM tasks specifically related to intent. It is particularly remarkable because in the other 7 conditions, participants are also thinking about the person’s patterns of responses. But this condition reflects an attribution to the target’s unique disposition or intent, relative to other people, an evocative entity, or unique circumstances. The STS activity differentiated this condition from the others about 9 s after exposure to the 4 sentence-screen, allowing for hemodynamic delay, and lasted another 4–5 s (see Fig. 4). In addition, this area is anatomically close to

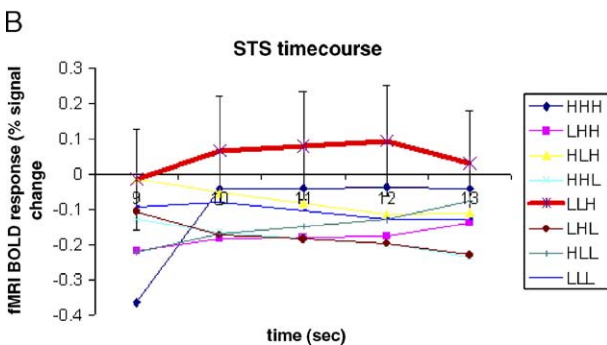
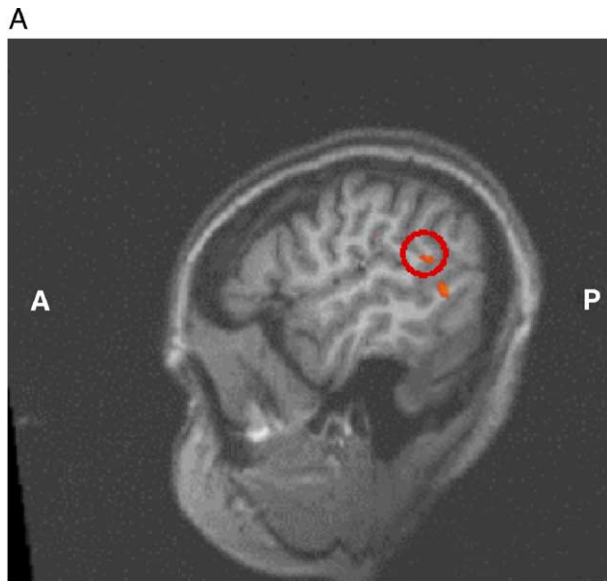


Fig. 4. (A and B) Superior temporal sulcus: location and time course of signal change.

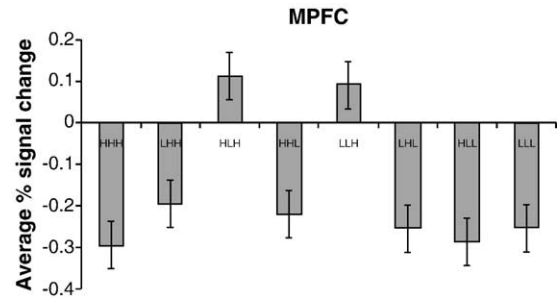


Fig. 5. Left medial prefrontal cortex: average signal change.

the area reliably found by ToM researchers (see Gallagher and Frith, 2002).

We predicted that the medial prefrontal cortex (MPFC) would activate under conditions focused on social cognition in general. As implied by the social psychological literature suggesting that people neglect consensus information, 2 of 8 conditions activated an area of MPFC in the left hemisphere: the same condition that implied person attributions (LLH), plus the low-distinctiveness, high-consistency combination with the opposite consensus information (HLH; everybody else does it, too) (Fig. 5), $t(11) = 3.74$, $P < .003$, with an effect size as indicated by $\eta^2 = .54$. The time course for these 2 cells (LLH and HLH) indicates similar timing for their differentiation from the other 6, starting about 9 s after the stimulus screen (shown in Fig. 6). In addition, the contrast excluding the LLH cell, comparing only the HLH cell against the remaining 6, was significant for this area, $t(11) = 2.63$, $P < .023$, with an effect size as indicated by $\eta^2 = .37$.

Some of the other areas activated in the LLH person attribution condition (see Table 3) have also been implicated in social cognition tasks. For example, the precuneus is activated in perceiving faces of significant others and is possibly involved in the retrieval of long-term affective memories (Gobbini et al., 2004). Our task did not require such retrieval, so this suggests a more general role for the precuneus in social cognition tasks⁹.

Conclusions

The findings provide initial support for the idea that person attributions do recruit the same areas that theory of mind tasks recruit. However, they also show that not all thinking about other people’s minds is equivalent. Unique dispositional attributions have a special status in neural activation patterns. Moreover, this study supports the idea that validated behavioral paradigms from social psychology are applicable and useful to cognitive neuroscience. Although both attribution research and ToM research address the same fundamental questions of social cognition, few attempts have integrated insights from both fields. By providing evidence for common regions underlying ToM and attribution tasks, we hope that this research will stimulate more

⁹ It should be noted that a test of the residual contrasts revealed a number of interesting findings. There was MPFC activation in the HHL and LLL conditions. However, these activations were in anatomically different locations from the area reported above (Talairarch: $x: 2, y: 44, z: 39$ and $x: 11, y: 37, z: 26$, respectively). There was also STS activation in the HLL condition, but again, in a different anatomical location (Talairarch: $x: 43, y: -53, z: 10$).

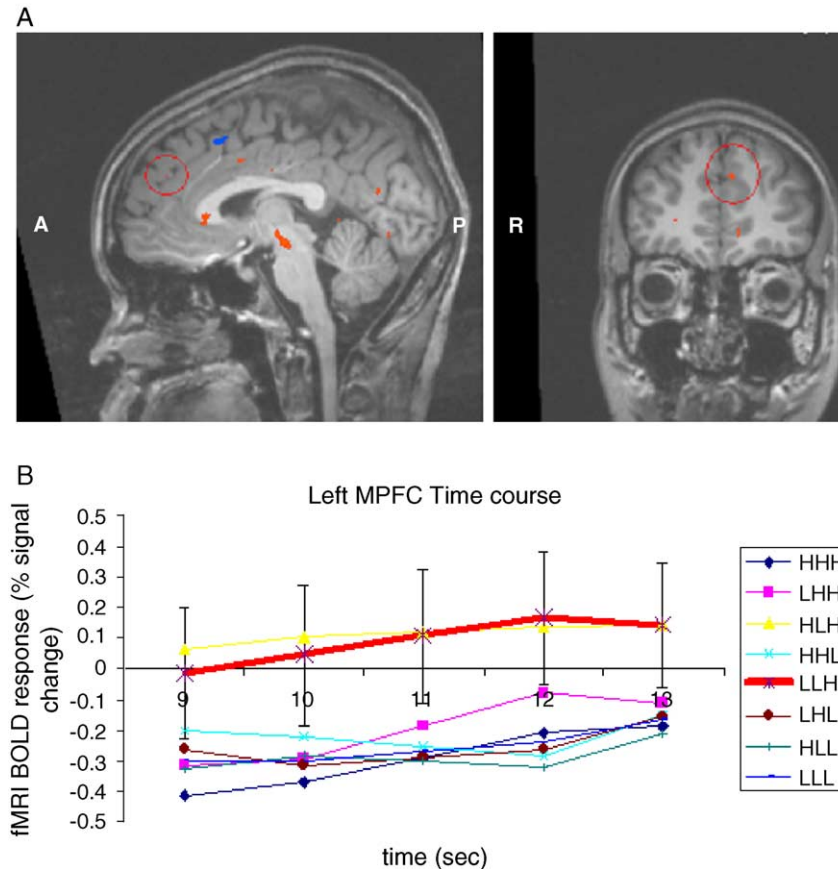


Fig. 6. (A and B) Left medial prefrontal cortex: locations and time course of signal change.

attempts to integrate knowledge about social and cognitive neuroscience.

Acknowledgments

The work was supported by the Princeton Center for the Study of Brain, Mind, and Behavior. We thank Jonathan Cohen, Ida Gobbini, Joshua Greene, James Haxby, and members of the Princeton Neuroscience and Social Decision-Making seminar for their comments.

References

- Abu-Akel, A., 2003. A neurobiological mapping of theory of mind. *Brain Res. Rev.* 43, 29–40.
- Blakemore, S.J., et al., 2003. The detection of contingency and animacy from simple animations in the human brain. *Cereb. Cortex* 13, 837–844.
- Brunet, E., Sarfati, Y., Hardy-Bayle, M.C., Decety, J., 2000. A PET investigation of the attribution of intentions with a nonverbal task. *NeuroImage* 11, 157–166.
- Calarge, C., Andreasen, N.C., O'Leary, D.S., 2003. Visualizing how one brain understands another: a PET study of theory of mind. *Am. J. Psychiatry* 160, 1954–1964.
- Castelli, F., Happe, F., Frith, U., Frith, C., 2000. Movement and mind: a functional imaging study of perception and interpretation of complex intentional movement patterns. *NeuroImage* 12, 314–325.
- Cuddy, A.J.C., Fiske, S.T., Kwan, V.S.Y., Glick, P., Demoulin, S., Leyens, J.-Ph., Bond, M.H., Croizet, J.-C., Ellemers, N., Sleebos, E., Htun, T.T., Yamamoto, M., Kim, H.-J., Maio, G., Perry, J., Petkova, K., Todorov, V., Rodríguez-Bailón, R., Morales, E., Moya, M., Palacios, M., Smith, V., Perez, R., Vala, J., Ziegler, R., in press. Is the stereotype content model culture-bound? A cross-cultural comparison reveals systematic similarities and differences. *Br. J. Soc. Psychol.*
- Ferguson, T.J., Wells, G.L., 1980. Priming of mediators in causal attribution. *J. Pers. Soc. Psychol.* 38, 461–470.
- Fiske, S.T., 1989. Examining the role of intent: toward understanding its role in stereotyping and prejudice. In: Uleman, J.S., Bargh, J.A. (Eds.), *Unintended Thought*. Guilford Press, New York, pp. 253–283.
- Fiske, S.T., 2004. *Social Beings: A Core Motives Approach to Social Psychology*. Wiley, New York.
- Fiske, S.T., Cuddy, A.J., Glick, P., Xu, J., 2002. A model of (often mixed) stereotype content: competence and warmth respectively follow from perceived status and competition. *J. Pers. Soc. Psychol.* 82, 878–902.
- Frith, U., Frith, C., 2001. The biological basis of social interaction. *Curr. Dir. Psychol. Sci.* 10, 151–155.
- Gallagher, H.L., Frith, C.D., 2002. Functional imaging of 'theory of mind'. *Trends Cogn. Sci.* 7, 77–83.
- Gallagher, H.L., Happe, F., Brunswick, N., Fletcher, P.C., Frith, U., Frith, C.D., 2000. Reading the mind in cartoons and stories: an fMRI study of 'theory of mind' in verbal and nonverbal tasks. *Neuropsychologia* 38, 11–21.
- Genovese, C.R., Lazar, N.A., Nichols, T.E., 2002. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *NeuroImage* 15, 870–878.
- Gilbert, D.T., Malone, P.S., 1995. The correspondence bias. *Psychol. Bull.* 117, 21–38.
- Gobbini, M.I., Leibenluft, E., Santiago, N., Haxby, J.V., 2004. Social and emotional attachment in the neural representation of faces. *NeuroImage* 22, 1628–1635.

- Grezes, J., et al., 1998. Top-down effect of strategy on the perception on human biological motion: a PET investigation. *Cogn. Neuropsychol.* 15, 553–582.
- Heider, F., 1958. *The Psychology of Interpersonal Relations*. Wiley, New York.
- Hoffman, E.A., Haxby, J.V., 2000. Distinct representations of eye gaze and identity in the distributed neural system for face perception. *Nature Neurosci.* 3, 80–84.
- Jones, E.E., 1979. The rocky road from acts to dispositions. *Am. Psychol.* 34, 107–117.
- Kelley, H.H., 1972. Attribution in social interaction. In: Jones, E.E., Kanouse, D.E., Kelley, H.H., Nisbett, R.E., Valins, S., Weiner, B. (Eds.), *Attribution: Perceiving the Cause of Behavior*. Lawrence Erlbaum and Associates, Hillsdale, NJ, pp. 1–26.
- Leslie, A.M., 1994. Pretending and believing: issues in the theory of mind. *Cognition* 50, 211–238.
- Lieberman, M.D., Gaunt, R., Gilbert, D.T., Trope, Y., 2002. Reflexion and reflection: a social cognitive neuroscience approach to attributional inference. *Adv. Exp. Soc. Psychol.* 34, 199–249.
- Malle, B.F., 2004. *How the Mind Explains Behavior: Folk Explanations, Meanings, and Social Interactions*. MIT Press, Cambridge, MA.
- McArthur, L.Z., 1972. The how and what of why: some determinants and consequences of causal attribution. *J. Pers. Soc. Psychol.* 22, 171–193.
- Mitchell, J.P., Heatherton, T.F., Macrae, C.N., 2002. Distinct neural systems subserve person and object knowledge. *Proc. Natl. Acad. Sci.* 99, 15238–15243.
- Mitchell, J.P., Macrae, C.N., Banaji, M.R., 2004. Encoding-specific effects of social cognition on the neural correlates of subsequent memory. *J. Neurosci.* 24, 4912–4917.
- Nakamura, K., et al., 2000. Functional delineation of the human occipito-temporal areas related to face and scene processing: a PET study. *Brain* 123, 1903–1912.
- Nakamura, K., et al., 2001. Neural substrates of familiar voices: a PET study. *Neuropsychologia* 39, 1047–1054.
- Ochsner, K.N., et al., 2004. Reflecting upon feelings: an fMRI study of neural systems supporting the attribution of emotion to self and other. *J. Cogn. Neurosci.* 16, 1–27.
- Orvis, B.R., Cunningham, J.D., Kelley, H.H., 1975. A closer examination of causal inference: the roles of consensus, distinctiveness, and consistency information. *J. Pers. Soc. Psychol.* 32, 605–616.
- Pruitt, D.J., Insko, C.A., 1980. Extension of the Kelley attribution model: the role of comparison-object consensus, target-object consensus, distinctiveness, and consistency. *J. Pers. Soc. Psychol.* 39, 39–58.
- Ross, L., 1977. The intuitive psychologist and his shortcomings: distortions in the attribution process. In: Berkowitz, L. (Ed.), *Adv. Exp. Soc. Psychol.*, vol. 10. Academic Press, New York, pp. 174–221.
- Sabbagh, M.A., 2004. Understanding orbitofrontal contributions to theory of mind reasoning: implications for autism. *Brain Cogn.* 55, 209–219.
- Saxe, R., Carey, S., Kanwisher, N., 2004. Understanding other minds: linking developmental psychology and functional neuroimaging. *Ann. Rev. Psychol.* 55, 87–124.
- Talairach, J., Tournoux, P., 1988. *Co-Planar Stereotaxic Atlas of the Human Brain*. Thieme, New York.
- Zuckerman, M., 1978. Actions and occurrences in Kelley's cube. *J. Pers. Soc. Psychol.* 36, 647–656.