

Research Report

Understanding Animate Agents

Distinct Roles for the Social Network and Mirror System

Thalia Wheatley, Shawn C. Milleville, and Alex Martin

Laboratory of Brain & Cognition, National Institute of Mental Health, Bethesda, Maryland

ABSTRACT—*How people understand the actions of animate agents has been vigorously debated. This debate has centered on two hypotheses focused on anatomically distinct neural substrates: The mirror-system hypothesis proposes that the understanding of others is achieved via action simulation, and the social-network hypothesis proposes that such understanding is achieved via the integration of critical biological properties (e.g., faces, affect). In this study, we assessed the areas of the brain that were engaged when people interpreted and imagined moving shapes as animate or inanimate. Although observing and imagining the moving shapes engaged the mirror system, only activation of the social network was modulated by animacy.*

The ability to detect and understand the actions of animate others is critical for survival. Although such detection is often aided by sensory cues (e.g., self-propelled motion, faces), the ability to impute animacy accurately in circumstances when such cues are absent or misleading (e.g., sleeping animals: Barrett & Behne, 2005; robots: Poulin-Dubois, Lepage, & Ferland, 1996) points to a conceptual representation of animacy that goes beyond sensory input. Recently, two hypotheses have emerged to explain the neural substrate underlying the understanding of animacy. One, the mirror-system hypothesis, posits that people understand the actions of an animate other by simulating the other's actions with their own motor programs (Gallese, Keysers, & Rizzolatti, 2004). Mirror neurons, the proposed neural substrate for such a mechanism (see Fig. 1), activate when a monkey initiates an action and when a monkey observes a similar action performed by others (Rizzolatti, Fogassi, & Gallese, 2001). A similar pattern of hemodynamic activity in human inferior parietal lobe and inferior frontal gyrus/ventral premotor cortices has led researchers to regard these two areas as the homologous mirror system in humans (Rizzolatti &

Craigheero, 2004). Most recently, the human mirror system was found to be associated with imitation and with understanding others' intentions and emotions (Carr, Iacoboni, Dubeau, Mazziotta, & Lenzi, 2003; Dapretto et al., 2006; Iacoboni et al., 2005), fueling the proposal that human social cognition is derived from human motor cognition.

A competing hypothesis claims that a more likely neural substrate for understanding the actions of animate agents is the social network, which consists of temporal and medial areas that consistently activate during performance of varied social cognitive tasks (see Fig. 1). This network includes areas active during the perception of biological motion (superior temporal sulcus, STS) and biological form (lateral fusiform gyrus), during mentalizing (medial prefrontal cortex, mPFC; posterior cingulate), and during the recognition and experience of affect (amygdala, insula; Adolphs, 2001; Mitchell, Heatherton, & Macrae, 2002; Saxe, 2005). The social-network hypothesis posits that these contributions are integrated within the network to achieve a unified representation of an animate being. Both theories' claims about how people achieve social understanding rest on the assumption of biological selectivity: preferential responding to animate relative to inanimate stimuli. A system that has evolved to understand the thoughts, feelings, and actions of others should be more sensitive to entities characterized by these things than to entities that are not.

Assessing which of the two neural networks underlies the understanding of animacy requires using stimuli that eliminate any low-level, animate-specific sensory cues. Toward this end, researchers have used abstract stimuli such as animations of simple geometric shapes that avoid such veridical cues as faces, eyes, and articulated joints (Castelli, Happé, Frith, & Frith, 2000; Martin & Weisberg, 2003). However, because these studies have used different stimuli in different conditions, they have not eliminated bottom-up interpretations entirely (e.g., self-propulsion: Premack, 1990; interactivity: Schultz, Friston, O'Doherty, Wolpert, & Frith, 2005). In the present functional magnetic resonance imaging (fMRI) study, we used contextual cues to bias interpretations of identical stimuli; this approach provides the most stringent test of the association between these neural networks and the concept of animacy.

Address correspondence to Thalia Wheatley, Department of Psychological and Brain Sciences, Moore Hall, Dartmouth College, Hanover, NH 03755, e-mail: thalia.p.wheatley@dartmouth.edu.

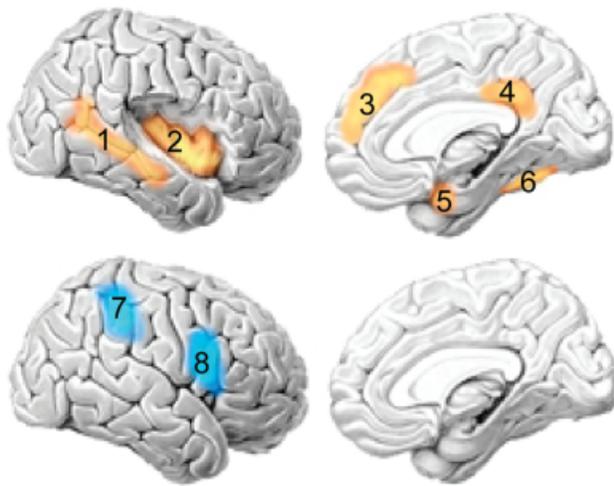


Fig. 1. Lateral and medial views of the social network (top, highlighted in yellow) and mirror system (bottom, highlighted in blue). The social network includes areas associated with biological motion (superior temporal sulcus, labeled “1”), biological form (lateral fusiform gyrus, labeled “6”), mentalizing (medial prefrontal cortex and posterior cingulate, labeled “3” and “4,” respectively), and affective processing (insula and amygdala, labeled “2” and “5,” respectively). The mirror system consists of the inferior parietal cortex (labeled “7”) and the ventral-premotor/inferior-frontal cortex (labeled “8”).

METHOD

Stimuli

We created 12 animations of moving shapes, each of which was displayed against two different backgrounds. One background biased the observer toward an animate interpretation of the moving shape, whereas the other background biased the observer toward an inanimate interpretation. Thus, animacy could

be attributed to the moving shape only on the basis of its context. We hypothesized that areas fundamental to inferring animate action would incorporate contextual information into animate representations even in the absence of animate-specific intrinsic cues.

Experimental Paradigm

For each shape, three trials were presented. On the first trial (*background* trial), the word “Look” appeared on the screen immediately before a background was displayed for 15 s. On the second trial (*motion* trial), the word “Watch” appeared immediately before a shape moved for 15 s on top of the previously viewed background. On the third trial (*imagery* trial), the word “Imagine” appeared just before the same background was presented, again for 15 s. The word “Imagine” cued subjects to “replay in their mind’s eye the motion previously paired with that background” (see Fig. 2). Participants viewed just one background for each moving shape: either the background that suggested an animate interpretation or the background that suggested an inanimate interpretation; different backgrounds were used for different shapes. As participants made only one interpretation per moving shape, we eliminated the possibility that initial interpretations would bias later interpretations. All backgrounds were represented equally often across subjects, allowing for between-subjects comparisons.

After each background trial, subjects chose which one of four object names matched an object in the scene (one correct, three foils). After the motion and imagery trials, subjects chose one of four plausible interpretations of the main figure: two animate (e.g., ice-skater, dancer) and two inanimate (e.g., crayon, spinning top). If their interpretation did not match any of the four options, subjects could select “none of the above.” All subjects

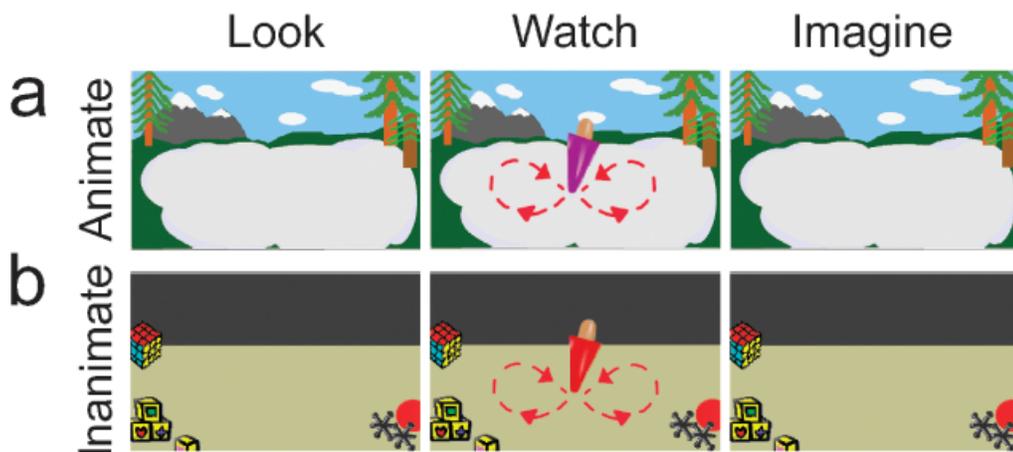


Fig. 2. The experimental paradigm. Subjects first saw a background alone (“Look”; background condition). Upon the second presentation of that background, a moving shape was overlaid (“Watch”; motion condition). The third time that background appeared, participants replayed in their mind’s eye the motion earlier paired with that background (“Imagine”; imagery condition). Each background suggested either an animate (a; e.g., ice-skater) or an inanimate (b; e.g., spinning top) interpretation. The red dashed line is included here to indicate the path of motion of the main figure, but was not visible to the subjects.

were successfully biased by the backgrounds (88% agreement with the bias). Only trials on which the interpretation was successfully biased were considered in the analyses. A 15-s control condition in which subjects counted the number of times a blue dot appeared at random spatial locations served as a baseline condition that controlled for eye movement. The word “Count” appeared immediately before each of these trials. Each imaging run consisted of two trials of each type (background, motion, imagery, and control) interleaved with fixation intervals varying from 2 to 6 s in duration. Stimuli were presented in an event-related design that was pseudorandomized with the caveat that for each triplet of trials, the imagery trial could not precede the motion trial, which could not precede the background trial.

As each background was viewed alone before its associated motion and imagery trials, hemodynamic activity generated by viewing the different backgrounds could be contrasted with activity generated by observing and imagining the moving shapes on those backgrounds. As the form and motion of each shape were identical across the two biasing backgrounds, any differences in activation between the motion trials using the different backgrounds would reflect different interpretations of the same observed stimulus. Similarly, the imagery trials for a given object afforded a direct comparison of neural activation for animate versus inanimate imagery (simulation) of the same moving shape.

Image Acquisition

Twenty-two subjects viewed the stimuli while being scanned in a 3.0-T fMRI machine (General Electric, Fairfield, CT). For each participant, a high-resolution, spoiled-gradient-recall anatomical scan (124 sagittal slices, 1.2 mm thick, field of view = 24 cm, acquisition matrix = 256×256) preceded the functional runs (gradient-echo, echoplanar imaging sequence; repetition time = 2,000 ms, echo time = 30 ms, flip angle = 90° , 24 contiguous 5-mm axial slices, voxel size = $3.75 \text{ mm} \times 3.75 \text{ mm} \times 5.00 \text{ mm}$).

fMRI Analyses

Functional images were motion corrected and smoothed with a 4.5-mm full-width/half-maximum Gaussian filter. Individual subjects' maps, both anatomical and functional, were normalized to the standardized space of Talairach and Tournoux (1988). Data were analyzed with AFNI software (Cox, 1996). Blood-oxygenation-level-dependent (BOLD) responses, to stimulus events of each of the conditions, were estimated with deconvolution. This approach does not constrain the shape of the hemodynamic response. A random-effects analysis of variance was performed using the mean activations for each condition across a window of time consistent with peak stimulus-related activity (6–12 s after the beginning of a trial). The resulting group activation map was thresholded to show only voxels in which any experimental condition (backgrounds, motion, and imagery) or combination of conditions differed significantly from the control condition ($p < .0001$).

We then identified clusters of voxels with different patterns of activation in different conditions (relative to the pattern of activation in the backgrounds condition). We identified regions of the brain that were more active when subjects inferred motion as animate than when they inferred motion as inanimate ($p < .05$). In addition, we identified regions that were more active when subjects imagined motion they had earlier deemed animate than when they imagined motion they had earlier deemed inanimate ($p < .05$). Given that concepts generalize across modality, the critical analysis was determining which areas were more active for animate than for inanimate interpretations during both motion observation and imagery. Therefore, we performed a conjunction analysis identifying regions that exhibited overlapping activation across the previous two analyses. That is, these conjunction areas were more active both when subjects inferred motion as animate rather than inanimate and when they imagined motion as animate rather than inanimate.

Further analyses collapsed across interpretation (animate, inanimate) to define areas associated with motion and imagery in general. We identified regions that were more active when subjects saw and interpreted the moving shapes than when they saw the backgrounds alone ($p < .0001$). In addition, we identified regions that were more active when subjects imagined the moving shapes than when they saw the backgrounds alone ($p < .0001$). A conjunction analysis identified the areas with overlapping activation across the previous two analyses. That is, these conjunction areas were more active both when subjects saw and when they imagined the moving shapes compared with when they observed the backgrounds alone. A more stringent threshold was required for the overall motion and imagery analyses than for the interpretation-based (animacy-inanimacy) analyses because of the amount of data represented (twice that of the interpretation-based analyses). All identified regions of interest were overlaid on individual subjects' data to extract the hemodynamic response associated with each condition.

RESULTS AND DISCUSSION

When the same moving shapes were interpreted and imagined as animate beings rather than inanimate objects, activity across the entire social network increased (see Figs. 3a and 3b). Furthermore, this pattern of animate-specific activity was strikingly concentrated within that network (see Table 1). It is somewhat surprising that animacy activated the whole social network rather than a subset of it. Clearly, it would be untenable to suggest that social cognition requires no additional processing beyond that required to infer animacy. A more reasonable hypothesis is that animacy serves as a systemwide alert, readying the network to process socially relevant information. Activity of specific areas within this network would then be expected to increase or decrease depending on the particular computations required (e.g., emotion recognition, theory of mind).

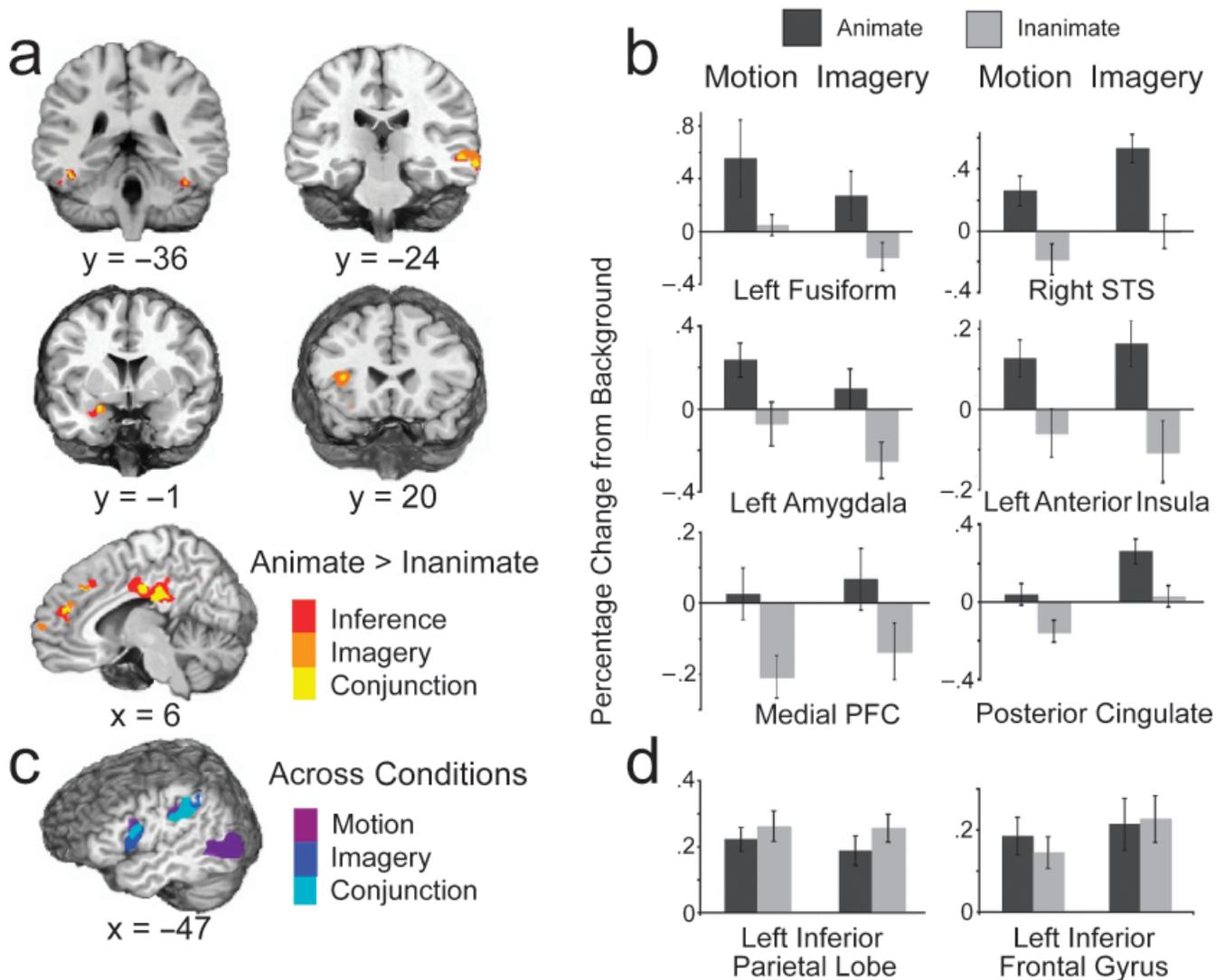


Fig. 3. Experimental results. The brain slices in (a) depict areas of the social network that were more active when moving shapes were inferred (red) or imagined (orange) as animate than when they were inferred or imagined as inanimate. Yellow areas were more active for both animate inference and imagery ("conjunction"). The graph in (b) displays the average hemodynamic responses within the conjunction areas as a function of animacy (animate, inanimate) and condition (motion, imagery). (Results are not shown for the posterior insula, although this was also a conjunction area.) The illustration in (c) shows areas of the mirror system that were more active when subjects watched and made inferences about the moving shapes (purple) and when they imagined (dark blue) the moving shapes relative to when they viewed the backgrounds alone; light-blue areas were more active during both the motion and imagery conditions ("conjunction") than in the background condition. The graph in (d) shows the average hemodynamic responses of the conjunction mirror areas as a function of animacy and condition. For purposes of illustration, all group data are presented on the N27 (AFNI software) brain. Error bars represent standard errors. STS = superior temporal sulcus; PFC = prefrontal cortex.

Although all social-network areas showed significantly increased activity when subjects inferred animacy, regardless of modality (motion, imagery), some areas were more extensively activated during one modality than the other. As seen in previous research demonstrating overlapping but stronger activity for perception rather than imagery (Ishai, Ungerleider, & Haxby, 2000; O'Craven & Kanwisher, 2000), the lateral region of the fusiform gyrus was more extensively engaged when animacy was inferred while observing the stimuli than when animacy was imagined. In contrast, the STS and mPFC both showed the reverse pattern. Implicated in the perceptual and the conceptual pro-

cessing of biological motion (Allison, Puce, & McCarthy, 2000), the STS is robustly activated by the observation of the fluid and articulated motion vectors associated with animate agents (Beauchamp, Lee, Haxby, & Martin, 2002). It is possible that encoding the moving shapes in the present study as animate resulted in imagining more fluid and articulated motion (i.e., more animacy) than was present in the original nonarticulated percepts, commensurately engaging the STS. The broad mPFC activity during animate imagery is consistent with reports linking this region to similar metacognitive processes involving animate agents (e.g., self- and other-reflection; Ochsner et al., 2004).

TABLE 1
Areas of the Brain That Were More Active for Animate Than for Inanimate Inference and Imagery

Area	Side	Brodmann's area	Coordinates		
			<i>x</i>	<i>y</i>	<i>z</i>
Frontal cortex					
Medial frontal gyrus ^a	Right	9	6	39	28
Parietal cortex					
Precuneus	Right	7	13	-66	31
	Left	7	-8	-62	31
Postcentral gyrus	Left	40	-63	-18	15
Temporal cortex					
Superior temporal gyrus	Right	22	53	-10	8
Superior temporal sulcus ^a	Right	21	56	-24	-2
	Right	21	66	-23	-7
Fusiform gyrus ^a	Left	20	-41	-36	-14
Insular cortex					
Posterior insula ^a	Right	13	45	-26	15
	Right	13	29	-15	15
	Left	13	-39	-25	13
Anterior insula ^a	Left	13	-33	21	10
Cingulate cortex					
Posterior cingulate ^a	Right	23	5	-21	31
	Right	23	4	-33	25
Cingulate gyrus	Right	32	12	16	38
	Left	24	-13	6	29
Amygdala ^a	Left		-21	0	-12
Caudate	Right		19	-36	16

Note. For all the areas listed, both motion and imagery trials showed greater activation for animate than for inanimate interpretations of the same objects ($p < .05$).

^aThese areas are considered to be part of the social network.

The mirror system, comprising inferior parietal and inferior frontal cortices, was bilaterally engaged (though more strongly on the left) during both motion observation and imagery, a finding consistent with previous research (Rizzolatti et al., 2001; see Fig. 3c). It is important to note, however, that even at a greatly reduced threshold, mirror-system activation was not modulated by the interpretation of animacy during either motion observation or imagery ($ps > .27$; see Fig. 3d). Thus, the mirror system does not appear to be selective for biological actions. This is an important point because the hypothesized link between mirror neurons and social cognition is predicated on the assumption of biological selectivity.

The finding that the mirror system is not modulated by the interpretation of animacy challenges its candidacy as the origin of general social cognition. However, it is possible that this network plays a more narrow role in the understanding of action goals. Although one could argue that the animate interpretations in this study were more likely to refer to goal-directed action (e.g., running over a hill) than were inanimate interpretations

(e.g., a sun rising and setting), we did not test this hypothesis directly. Thus, it is possible that a role of the mirror system is to represent a target's action goals, regardless of the animacy of that target.

The finding that motion observation and imagery engaged the mirror system regardless of animacy interpretation is compatible with two conclusions. First, in keeping with a broad view of mirror function, these areas may play a general role in action comprehension. For example, the mirror system may underlie simulation, as other researchers have suggested, but in the more broadly defined sense of mentally re-creating an action or motor sequence rather than the more animate-centric view associated with mind reading. This broader definition casts the mirror system as domain-general simulation machinery that may operate in tandem with domain-specific systems to elucidate the actions of *all* objects. Such a definition is consistent with previous research associating inferior frontal cortex with imitation, a cognitive process that requires mentally re-creating (simulating) a motor sequence, and with observing an action with the intent to imitate it later (Decety et al., 1997; Heiser, Iacoboni, Maeda, Marcus, & Mazziotta, 2003; Iacoboni et al., 1999). Correspondingly, in the present study, the mirror system showed increased activation when participants mentally re-created motion (imagery condition) and observed motion with the intent to simulate it later (motion condition). Although studies investigating the mirror system have focused almost exclusively on understanding animate action, the results obtained with inanimate conditions here and in previous work (e.g., wooden blocks touching each other—Keysers et al., 2004) suggest that its role in understanding action is not limited to the animate domain.

Second, it is instead possible that the engagement of the inferior frontal and parietal cortices reflects processes critical to interpretation and imagery that are not specific to action comprehension. Thompson-Schill, Wagner, and other researchers have demonstrated the role these areas play in the selection and retrieval of information from memory—processes necessitated by any interpretation or imagery task (Thompson-Schill, Bedny, & Goldberg, 2005; Wagner, Shannon, Kahn, & Buckner, 2005). Whether these areas subserve mirror or memory processes, or both, they appear to play an important but domain-general role in the understanding of objects, including, but not limited to, animate entities. Thus, a functional or anatomical deficit in the mirror system would be expected to compromise social cognition (Dapretto et al., 2006; Hadjikhani, Joseph, Snyder, & Tager-Flusberg, 2005) because of a disruption of process rather than content.

The present findings suggest that the understanding of animacy is the domain of the social network. That this basic biological concept shares the same neural footprint as higher-order social cognition is consistent with the view that inferring animacy is the developmental precursor to complex social understanding (Legerstee, 1992). Once conceptually established, animacy may continue to serve as an alert, engaging the entire network in a state of readiness to process incoming social information.

Acknowledgments—This work was supported by the Division of Intramural Research, National Institute of Mental Health. We thank Alfonso Caramazza, Patrick Bellgowan, Josh Greene, Andrea Heberlein, and two reviewers for useful comments and discussions.

REFERENCES

- Adolphs, R. (2001). The neurobiology of social cognition. *Current Opinion in Neurobiology*, *11*, 231–239.
- Allison, T., Puce, A., & McCarthy, G. (2000). Social perception from visual cues: Role of the STS region. *Trends in Cognitive Sciences*, *4*, 267–278.
- Barrett, H.C., & Behne, T. (2005). Children's understanding of death as the cessation of agency: A test using sleep versus death. *Cognition*, *96*, 93–108.
- Beauchamp, M.S., Lee, K.E., Haxby, J.V., & Martin, A. (2002). Parallel visual motion processing streams for manipulable objects and human movements. *Neuron*, *34*, 149–159.
- Carr, L., Iacoboni, M., Dubeau, M.-C., Mazziotta, J.C., & Lenzi, G.L. (2003). Neural mechanisms of empathy in humans: A relay from neural systems for imitation to limbic areas. *Proceedings of the National Academy of Sciences, USA*, *100*, 5497–5502.
- Castelli, F., Happé, 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.
- Cox, R.W. (1996). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research*, *29*, 162–173.
- Dapretto, M., Davies, M.S., Pfeifer, J.H., Scott, A.A., Sigman, M., Bookheimer, S.Y., & Iacoboni, M. (2006). Understanding emotions in others: Mirror neuron dysfunction in children with autism spectrum disorders. *Nature Neuroscience*, *9*, 28–30.
- Decety, J., Grèzes, J., Costes, N., Perani, N., Jeannerod, M., Procyk, E., et al. (1997). Brain activity during observation of actions: Influence of action content and subject's strategy. *Brain*, *120*, 1763–1777.
- Gallese, V., Keysers, C., & Rizzolatti, G. (2004). A unifying view of the basis of social cognition. *Trends in Cognitive Sciences*, *8*, 396–403.
- Hadjikhani, N., Joseph, R.M., Snyder, J., & Tager-Flusberg, H. (2005). Anatomical differences in the mirror neuron system and social cognition network in autism. *Cerebral Cortex*, *16*, 1276–1282.
- Heiser, M., Iacoboni, M., Maeda, F., Marcus, J., & Mazziotta, J.C. (2003). The essential role of Broca's area in imitation. *European Journal of Neuroscience*, *17*, 1123–1128.
- Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J.C., & Rizzolatti, G. (2005). Grasping the intentions of others with one's own mirror neuron system. *Public Library of Science: Biology*, *3*, 1–7.
- Iacoboni, M., Woods, R.P., Brass, M., Bekkering, H., Mazziotta, J.C., & Rizzolatti, G. (1999). Cortical mechanisms of human imitation. *Science*, *286*, 2526–2528.
- Ishai, A., Ungerleider, L.G., & Haxby, J.V. (2000). Distributed neural systems for the generation of visual images. *Neuron*, *28*, 379–390.
- Keysers, C., Wickers, B., Gazzola, V., Anton, J.-L., Fogassi, L., & Gallese, V. (2004). A touching sight: SII/PV activation during the observation and experience of touch. *Neuron*, *42*, 335–346.
- Legerstee, M. (1992). A review of the animate-inanimate distinction in infancy: Implications for models of social and cognitive knowing. *Early Development and Parenting*, *2*, 59–67.
- Martin, A., & Weisberg, J. (2003). Neural foundations for understanding social and mechanical concepts. *Cognitive Neuropsychology*, *20*, 575–587.
- Mitchell, J.P., Heatherton, T.F., & Macrae, C.N. (2002). Distinct neural systems serve person and object knowledge. *Proceedings of the National Academy of Sciences, USA*, *99*, 15238–15243.
- Ochsner, K.N., Knierim, K., Ludlow, D., Hanelin, J., Ramachandran, T., & Mackey, S. (2004). Reflecting upon feelings: An fMRI study of neural systems supporting the attribution of emotion to self and other. *Journal of Cognitive Neuroscience*, *16*, 1746–1772.
- O'Craven, K., & Kanwisher, N. (2000). Mental imagery of faces and places activates corresponding stimulus-specific brain regions. *Journal of Cognitive Neuroscience*, *12*, 1013–1023.
- Poulin-Dubois, D., Lepage, A., & Ferland, D. (1996). Infants' concept of animacy. *Cognitive Development*, *11*, 19–36.
- Premack, D. (1990). The infant's theory of self-propelled objects. *Cognition*, *36*, 1–16.
- Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, *27*, 169–192.
- Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews Neuroscience*, *2*, 661–670.
- Saxe, R. (2005). Against simulation: The argument from error. *Trends in Cognitive Sciences*, *9*, 174–179.
- Schultz, J., Friston, K.J., O'Doherty, J., Wolpert, D.M., & Frith, C. (2005). Activation in posterior superior temporal sulcus parallels parameter inducing the percept of animacy. *Neuron*, *45*, 625–635.
- Talairach, J., & Tournoux, P. (1988). *A co-planar stereotaxic atlas of a human brain*. Stuttgart, Germany: Thieme-Verlag.
- Thompson-Schill, S.L., Bedny, M., & Goldberg, R.F. (2005). The frontal lobes and the regulation of mental activity. *Current Opinion in Neurobiology*, *15*, 219–224.
- Wagner, A.D., Shannon, B.J., Kahn, I., & Buckner, R.L. (2005). Parietal lobe contributions to episodic memory retrieval. *Trends in Cognitive Sciences*, *9*, 445–453.

(RECEIVED 8/6/06; REVISION ACCEPTED 10/11/06;
FINAL MATERIALS RECEIVED 12/21/06)