Triggering the intentional stance

Raymond A. Mar and C. Neil Macrae*[1]

*Department of Psychology, University of Toronto, 100 St. George Street, Toronto, ON M5S 3G3, Canada, and *School of Psychology, University of Aberdeen, King's College, Aberdeen AB24 2UB, UK*

Abstract. While humans possess a ready capacity to view a target (biological or other-wise) as an intentional agent (i.e. the 'intentional stance'), the conditions necessary for spontaneously eliciting these mentalizing processes are less well understood. Although research examining people's tendency to construe the motion of geometric shapes as intentional has done much to illuminate this issue, due to methodological limitations (a reliance on subjective self-report) this work has not fully addressed the potentially auto-matic and obligatory nature of mentalizing. Acknowledging this problem, recent research using prelinguistic infants, neuroimaging technology and methods that avoid explicit self-report all provide unique paths to circumvent this shortcoming. While work of this kind has generally corroborated the results of previous investigations, it has also raised a number of new issues. One such issue is whether spontaneous mentalizing processes for abstract non-biological stimuli are instantiated in the same neural architecture as those for realistic representations of intentional biological agents. This question is con-sidered in the current chapter.

Spontaneous social perception versus controlled social judgements

The ability to comprehend the beliefs, emotions and intentions of others is both characteristic of, and necessary for, successful human interaction. Our richly social nature and complex societal hierarchies demand these skills, such that those who exhibit deficits in this domain experience considerable difficulty interacting with others (e.g., individuals with autism; Tager-Flusberg 2001). Known as possessing a theory-of-mind (ToM), mentalizing, or adopting the intentional stance, this capacity to view others as possessing mental states can be directed to targets other than conspecifics. Not only are we tempted to believe that a pet hamster is 'just like a little person,' we routinely view quite abstract nonliving representations as if they were intentional agents. Be it an animated movie populated by talking

---

[1] This paper was presented at the symposium by Neil Macrae to whom correspondence should be addressed

animals, the Sunday morning cartoon strip, or a novel whose characters are represented by mere words on a page, our enjoyment of all these media depend upon an ability to adopt the intentional stance toward clearly nonintentional (and non biological) objects (Mar 2004, Mar et al 2006). Moreover, such engagements do not appear to be cognitively taxing. The ease with which we can comprehend the sorry misadventures of poor old Charlie Brown in a *Peanuts* comic strip contributes to our enjoyment. In some cases, it certainly feels as if we cannot help but view a representation as intentional (e.g. characters in an animated film), although we certainly know that this perceived or felt agency is entirely illusory. Thus, while we are capable of making social judgements when prompted, we also appear to spontaneously and automatically perceive certain displays as representing intentional agents.

Isolating the specific qualities that evoke this illusion of intentionality has been the subject of active research for at least half a century (Heider & Simmel 1944) and this phenomenon has been distinguished as a form of social perception rather than social judgement (Allison et al 2000, Scholl & Tremoulet 2000). These two ideas—that of automatic and spontaneous perceptions of intentionality in contrast to effortful and controlled judgements of the same—appears to parallel a number of theoretical splits within the literature on mentalizing. Explanations for ToM have tended to group around two competing general ideas: theory-theory and simulation-theory (Carruthers & Smith 1996). The unfortunately named theory-theory proposes that mentalizing is achieved through propositional, rule-based thinking. Humans are viewed as amateur scientists with folk-psychological theories regarding the relation between mental states and behaviours, from which predictions are made. Developmental progress in attaining a ToM, then, is seen as the gradual construction of more accurate theories, wrought through experience in a complex social world.

Simulation-theory, in contrast, proposes that inferring the emotions and goals of other agents is achieved by imagining what our own feelings and aims would be were we placed in a similar situation—by simulating the experience of this other person. This viewpoint can be seen as resting on a form of embodied cognition, in which the understanding of another is built upon concurrent engagement of affective and motor systems in the self (see Blakemore & Frith 2005, Keysars & Perrett 2004). These two explanations are not mutually exclusive, and many researchers are beginning to propose theories that blend the two approaches (Carruthers & Smith 1996). Simulation-theory and theory-theory also appear part of a broader debate concerning the existence of both a cognitive, language-based and propositional form of mental inference (more akin to theory-theory) as well as an emotional and embodied form of empathic understanding (in keeping with simulation theory—Tager-Flusberg 2001, Preston & de Waal 2002). To be clear, we are by no means putting forth the argument that social perception should be

D2

equated with simulation-theory, nor with emotional empathy for that matter. Spontaneous mentalizing, however, can be seen as a fundamental contributing process for these broader categories. Answering questions regarding the former is absolutely necessary for understanding the latter. What are the conditions under which social perceivers spontaneously view others as intentional agents? Put differently, what triggers the adoption of the intentional stance?

Spontaneous adult mentalizing and subjective self report

Similar to how Michotte (1946/1963) revealed spontaneous perceptions of causality using very simplistic animations of moving shapes, Heider & Simmel (1944) demonstrated that inferences of intentionality can be drawn from quite abstract, non-biological representations. These researchers presented a series of short animations, each involving two triangles (one large, one small) and a circle, all moving around an empty rectangle. Observers readily attributed personality traits to the shapes and described their movements in terms of mental states such as goals and emotions, a finding replicated by subsequent researchers (for a review see Scholl & Tremoulet 2000). In general, this work has supported the idea that it is the spatiotemporal characteristics of the animations that trigger animacy descriptions (such as changes in path, moving in response to other objects and self-propelled movement) and not the featural properties of the interacting shapes (Scholl & Tremoulet 2000).

Animacy, however, while likely a necessary cue for intentionality, is not an equivalent construct (for a discussion see Gergely et al 1995). Furthermore, there are concerns about whether this method truly addresses the spontaneous and perceptual quality of these inferences. Observers report their subjective percepts in response to the animations, and it is possible that higher-order cognitive processing is engaged in order to produce these descriptions (cf. Scholl & Tremoulet 2000). Individuals may not be perceiving intentionality, but merely reporting the observation of intentional behaviour as a result of other factors such as demand characteristics and calculated inference. Thankfully, there are ways to circumvent this problem, such as employing: (1) designs that do not require explicit prompts for judgement; (2) prelinguistic infants as participants; and (3) neuroimaging approaches.

Spontaneous mentalizing in infants

Infant participants are unique in that their responses are relatively uncontaminated by cultural experience, experimental demands and language-based cognition, making them ideal subjects for examining the question of automatic social perception (Johnson 2000). Naturally, prelinguistic infants cannot self-report their per-

D2

cepts of intentionality while viewing animations—precluding explicit prompts for judgement—so researchers must capitalize upon their tendency to attend longer to novel displays. After an infant has been habituated to a certain visual stimulus, new displays can be shown and if they are perceived to be different from the earlier stimulus, infants will tend to look longer (see Johnson 2000, Box 1). Gergely and colleagues (1995), for example, found that 12-month old babies were more surprised and attended longer when a moving shape behaved in a seemingly non-rational manner as opposed to when it moved in a rational and goal-based fashion, even if the latter display was more perceptually dissimilar to the habituation stimulus. This finding appears to demonstrate that even prelinguistic infants attribute intentionality to abstract shapes based solely upon spatiotemporal variables. Similarly, Luo & Baillargeon (2005) have convincingly demonstrated that infants as young as 5 months old readily attribute goals to novel non-human objects (such as a box), provided that it possesses a cue of agency, such as self-propelled movement. Other evidence, however, indicates that featural properties are also important. Infants are more likely to imitate the failed goal-attempts of an adult over those of robotic pincers, are more likely to view the grasping actions of a hand as goal-directed compared to a perceptually similar rod and are more likely to follow the 'gaze' of a novel object when it has a face (for reviews see Johnson 2000, 2003). Similarly, Guajardo & Woodward (2004) have shown that infants view a bare-handed grasp as goal-directed, but do not appear to do so when this hand is covered by a glove. Importantly, when the infants had an opportunity to associate this gloved hand with a human agent, they were more likely to then attribute the grasping actions as intentional.

Johnson (2003) has reviewed the cues that are thought to trigger the intentional stance, including: (1) facial features; (2) an asymmetry along one axis; (3) non-rigid transformations such as expansion and contraction; (4) self-propelled movement; and (5) the capacity for reciprocal and contingent behaviour. It remains unclear, however, which of these cues are either necessary and/or sufficient. We are unlikely to view all objects with an asymmetry as intentional, for example, and many objects that possess the property of self-propulsion are not viewed as intentional. Her own work has shown that morphological cues (such as those possessed by a mechanical orangutan) and movement cues (such as a furry blob that behaves contingently in response to an experimenter), either individually or in combination, can signal to infants the presence of an intentional, and not just animate, agent (Johnson 2003). There are questions, of course, as to whether these behaviours by infants reflect mentalizing processes identical to those undertaken by adults. Gergely and colleagues (e.g. Gergely & Csibra 2003), for example, have argued that studies such as those reviewed above are evidence for a non-mentalistic 'teleological stance' on the part of infants, that acts as a precursor to the mentalistic intentional stance adopted by older children and adults.

D2

Neuroimaging and the intentional stance

One of the greatest strengths of neuroimaging approaches such as functional magnetic resonance imaging (fMRI) and positron emission tomography (PET) is its potential to covertly examine mental processes without the confound of explicit probes for self-report. There have been a handful of studies that have examined the neural correlates of intentionality cues, using animations like those created by Heider & Simmel (1944). Castelli and colleagues (2000) presented three types of animations while collecting PET scans, shapes that moved: (1) randomly; (2) in a goal-directed way; and (3) in ways that implied complex mental states (such as deception). In half the cases, participants were told what sort of animation they would see (the cued condition), and the order of these blocks was counterbalanced. For half the participants then, the uncued condition followed the cued, allowing for the possibility that awareness of the animation-types biased their attention and inferences during the former condition. This may explain why no differences were found during cued and uncued animations, prompting the researchers to combine the data across conditions for analysis.

Animations meant to elicit theory-of-mind attributions elicited more activation in several areas relative to the random animations: (1) the bilateral temporal parietal junction (TPJ); in the posterior superior temporal sulcus (STS); (2) bilateral basal temporal regions, including the temporal poles; (3) the bilateral extrastriate cortex: and (4) the medial prefrontal cortex (MPFC). A follow-up study involving both autistic patients and normally developing controls replicated this finding, and also found that autistics had less activation in the basal temporal area, the STS and TPJ, and the MPFC (Castelli et al 2002). Behavioural data have shown that those with autism are less likely to report percepts of intentionality when viewing these types of stimuli, allowing for the inference that these brain areas are responsible for engaging the intentional stance (Abell et al 2000, Castelli et al 2002). Corroborating this supposition, the TPJ and STS, temporal poles and MPFC have all been implicated in mentalising processes in numerous studies, using a variety of methods and tasks (Gallagher & Frith 2003).

Blakemore et al (2003) performed an influential fMRI study that employed simple shape animations to parse perceptions of animacy and contingency, as well as examine the effect of drawing attention to the contingency relations (via an explicit probe). Different parts of the brain were associated with viewing animate compared to contingent stimuli, and when both qualities were present activation was observed in superior parietal areas. Notably, when participants were cued to attend to the contingent movements, activation was observed in the right middle frontal gyrus (MFG) and left STS. It is worth noting that the stimuli in this study were much less complex than those in the previous studies, and did not involve imbuing shapes with complicated mental states.

D2

Similar areas of activation were found in a study by Shultz et al (2004), who presented animations of one circle chasing another. In half the animations, the chasing circle predicted the end-state of the other circle in order to catch it, and in the other half this circle merely followed the other. When comparing the predicting condition to the following condition, activations were observed around the STS in both hemispheres. In conjunction, similar to the findings of Blakemore et al (2003), explicitly drawing attention to the 'strategy' of the circle was associated with the left STS. The STS and TPJ have proven very important for the discussion of basic cues for intentionality; a number of researchers have argued that the STS is implicated in the understanding of biological motion, specifically with respect to intentions and goals (e.g. Allison et al 2000, Pelphrey et al 2004, Saxe et al 2004). [1]

One question that arises from the work reviewed thus far is whether the STS also codes for featural cues that trigger the intentional stance, along with motion cues. There is some evidence for this. Activation in the STS has been observed for static images of features that cue intentionality, such as eyes, mouths, hands, and faces; in some cases, however, these static images may have implied motion (for a review see Allison et al 2000). Because activations in this area are observed both when abstract shapes and realistic portrayals of biological agents are used as stimuli, it has been conjectured that the STS codes for intentional movement regardless of form (Shultz et al 2004). Direct comparisons of cartoon and realistic motion, however, are methodologically difficult to achieve. Pelphrey et al (2003) found that the STS did not appear to respond differentially to the movements of a computer-animated person compared to a similarly rendered 'robot' constructed of cylinders and spheres. In contrast, a separate fMRI study found that although the STS responded to very abstract representations of biological motion (point-light displays), it demonstrated a slightly stronger response to videos of real people in motion (Beauchamp et al 2003). A PET study, involving observations of grasping actions by a real hand compared to a 3D virtual reality hand, found the right TPJ near the posterior STS was more activated by the real hand; the right temporal pole also showed a similar preference (Perani et al 2001).

In a recent study (Mar et al 2006), we examined whether the brain responds differently to complex dynamic videos of social interactions presented in either a cartoon or realistic fashion. Footage for the film *Waking Life* (Linklater 2001) was shot using real actors, and later transformed by computer animators into a cartoon. Motion kinematics from the real footage were thus preserved in the animated version, and although both versions had numerous cues for animacy and intentionality (e.g. self-propelled movement, faces and other biological features) one was obviously realistic while the other was a cartoon (see Fig. 1). Equivalent content was shown in both versions, and shots within scenes alternated between cartoon and real. Participants were not prompted to make any social judgement,

D2

FIG. 1.   Screenshot from the video stimuli employed by Mar et al (2006); cartoon version above real version.

but only instructed to watch the videos closely, which were presented without sound. We found that the right STS and TPJ were more activated while participants watched realistic scenes compared to cartoon scenes (see Fig. 2A). Even though the cartoon version was closely matched to the real version, the latter appears to have preferentially engaged brain areas known to be involved in mentalizing and the inference of intentions from behaviour.

Interestingly, the right MFG was also more activated during the real condition (see Fig. 2B), and others have found similar activations to be associated with attending to contingency relations in the presence of animacy (Blakemore et al 2003), making judgements regarding persons (Mason et al 2004) and inferences of intentionality when perceiving actions (Pelphrey et al 2004). Moreover, because ⌷2⌷ this study did not employ explicit prompts for social judgement, this appears to be evidence for the spontaneous triggering of the intentional stance based upon perceptual cues (cf. German et al 2004). The right STS, TPJ and MFG thus appear to be highly sensitive to subtle cues signalling intentionality, moving beyond
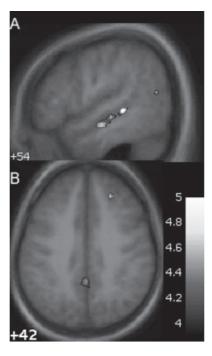
FIG. 2.   Activation in the right STS, TPJ (A) and MFG (B) for the Real > Cartoon contrast from Mar et al (2006). Activations superimposed on an average of all T1-weighted structural scans for participants. Legend indicates t-value.

animate motion and biological features (present in both versions of the stimuli) to the perception of targets as belonging to the real world. Previous work has also demonstrated such sensitivity to subtle cues of intentionality in the STS, such as a preference for mutual gaze as opposed to averted gaze from a dynamic computer-animated person (Pelphrey et al 2004). It appears that activation in the STS and TPJ may be modulated by the number of cues present that signal intentionality. While basic motion cues (e.g. animated shapes) as well as static featural cues (e.g. faces and eyes) can result in engagement of these superior lateral temporal regions, combining these cues appears to result in greater activity.

Conclusions

The ability to infer intentions from the behaviours of others is clearly important (Baldwin & Baird 2001). By understanding the basic cues that cause us to treat a target as intentional, we can begin to explore the very foundations of social cognition. From one perspective, our tendency to innately, automatically, and

3

D2

spontaneously view a broad variety of different targets as holding goals and mental states seems to fly in the face of parsimony and pragmatism. Why can our intentional stance be triggered so easily, and by so many stimuli that are clearly *not* intentional? How useful is a system when it often renders conclusions that do not reflect reality? We have no clear way of knowing that other agents, even other humans, are truly intentional (i.e. the solipsistic conundrum). Therefore, it may be that a low threshold for triggering the intentional stance—a bias toward viewing agents as having goals, beliefs, and desires—provides us with an adaptive heuristic for understanding our world and all its inhabitants.

References

Abell F, Happé F, Frith U 2000 Do triangles play tricks? Attribution of mental states to animated shapes in normal and abnormal development. Cognitive Devt 15:1–16

Allison T, Puce A, McCarthy G 2000 Social perception from visual cues: role of the STS region. Trends Cogn Sci 4:267–278

Baldwin DA, Baird JA 2001 Discerning intentions in dynamic human interaction. Trends Cogn Sci 5:171–178

Beauchamp MS, Lee KE, Haxby JV, Martin A 2003 fMRI response to video and point-light displays of moving humans and manipulable objects. J Cognitive Neurosci 15:991–1001

Blakemore S-J, Frith C 2005 The role of motor contagion in the prediction of action. Neuropsychologia 43:260–267

Blakemore S-J, Boyer P, Pachot-Clouard M, Segebarth C, Decety J 2003 The detection of contingency and animacy from simple animations in the human brain. Cereb Cortex 13: 837–844

Carruthers P, Smith PK (eds) 1996 Theories of theories of mind. Cambridge University Press, Cambridge, MA

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

Castelli F, Frith C, Happé F, Frith U 2002 Autism, Asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes. Brain 125:1839–1849

Dennet D 1987 The intentional stance. MIT Press, Cambridge, MA

Gallagher HL, Frith CD 2003 Functional imaging of 'theory of mind'. Trends Cogn Sci 7: 77–83

Gergely G, Csibra G 2003 Teleological reasoning in infancy: the naïve theory of rational action. Trends Cogn Sci 7:287–292

Gergely G, Nádasy Z, Csibra G, Bíró S 1995 Taking the intentional stance at 12 months of age. Cognition 56:165–193

German TP, Niehaus JL, Roarty MP, Giesbrecht B, Miller MB 2004 Neural correlates of detecting pretense: Automatic engagement of the intentional stance under covert conditions. J Cognitive Neurosci 16:1805–1817

D2

Guajardo JJ, Woodward AL 2004 Is agency skin deep? Surface attributes influence infants' sensitivity to goal-directed action. Infancy 6:361–384

Heider F, Simmel M 1944 An experimental study of apparent behavior. Am J Psychol 57: 243–249

Johnson SC 2000 The recognition of mentalistic agents in infancy. Trends Cogn Sci 4:22–28

Johnson SC 2003 Detecting agents. Philos Trans R Soc Lond B Biol Sci, Series B, 358: 549–559

Keysars C, Perrett DI 2004 Demystifying social cognition: a Hebbian perspective. Trends Cogn Sci 8:501–507

Linklater R 2001 Waking Life. Fox Searchlight Pictures, Los Angeles

Luo Y, Baillargeon R 2005 Can a self-propelled box have a goal? Psychol Sci 16:601–608

Mar RA 2004 The neuropsychology of narrative: story comprehension, story production and their interrelation. Neuropsychologia 42:1414–1434

Mar RA, Kelley WM, Heatherton TF, Macrae CN 2006 Detecting agency from biological motion, in preparation 4

Mar RA, Oatley K, Hirsh J, dela Paz J, Peterson JB 2006 Bookworms versus nerds: Exposure to fiction versus non-fiction, divergent associations with social ability, and the simulation of fictional worlds. J Res Pers ••, ••–••, in press 5

Mason MF, Banfield JF, Macrae CN 2004 Thinking about actions: The neural substrates of person knowledge. Cereb Cortex 14:209–214

Michotte A 1946/English Trans 1963 The perception of causality. Basic Books, New York

Pelphrey KA, Mitchell TV, McKeown MJ, Goldstein J, Allison T, McCarthy G 2003 Brain activity evoked by the perception of human walking: Controlling for meaningful coherent action. J Neurosci 23:6819–6825

Pelphrey KA, Morris JP, McCarthy G 2004 Grasping the intentions of others: The perceived intentionality of an action influences the activity in the superior temporal sulcus during social perception. J Cognitive Neurosci 16:1706–1716 6

Pelphrey KA, Viola RJ, McCarthy G 2004 When strangers pass: Processing of mutual and averted social gaze in the superior temporal sulcus. Psychol Sci 15:598–603 7

Perani D, Fazio F, Borghese NA et al 2001 Different brain correlates for watching real and virtual hand actions. NeuroImage 14:749–758

Preston SD, de Waal FBM 2002 Empathy: Its ultimate and proximate bases. Behav Brain Sci 25:1–20

Saxe R, Xiao D-K, Kovacs G, Perrett DI, Kanwisher N 2004 A region of right posterior superior temporal sulcus responds to observed intentional actions. Neuropsychologia 42:1435–1446

Scholl BJ, Tremoulet PD 2000 Perceptual causality and animacy. Trends Cogn Sci 4:299–309

Shultz J, Imamizu H, Kawato M, Frith CD 2004 Activation of human superior temporal gyrus during observation of goal attribution by intentional objects. J Cognitive Neurosci 16: 1695–1705

Tager-Flusberg H 2001 A re-examination of the theory-of-mind hypothesis of autism. In: Burack J, Charman T, Yimiya N, Zelazo P (eds) The development of autism: Perspectives from research and theory. Lawrence Erlbaum Associates, Mahwah, NJ p 173–193

DISCUSSION

C Frith: It might be useful for us to discuss the uncanny valley. It is a concept developed in Japan and relates to animation. This is supposed to explain why some recent animated films didn't work. Animations are getting better and better in D2