Dissociating Neural Correlates of Action Monitoring and Metacognition of Agency

David B. Miele[1], Tor D. Wager[2], Jason P. Mitchell[3], and Janet Metcalfe[1]

Abstract

■ *Judgments of agency* refer to people's self-reflective assessments concerning their own control: their assessments of the extent to which they themselves are responsible for an action. These self-reflective metacognitive judgments can be distinguished from *action monitoring*, which involves the detection of the divergence (or lack of divergence) between observed states and expected states. Presumably, people form judgments of agency by metacognitively reflecting on the output of their action monitoring and then consciously inferring the extent to which they caused the action in question. Although a number of previous imaging studies have been directed at action monitoring, none have assessed judgments of agency as a potentially separate process. The present fMRI study used an agency paradigm that not only allowed us to examine the brain activity associated with action monitoring but that also enabled us to investigate those regions associated with metacognition of agency. Regarding action monitoring, we found that being "out of control" during the task (i.e., detection of a discrepancy between observed and expected states) was associated with increased brain activity in the right TPJ, whereas being "in control" was associated with increased activity in the pre-SMA, rostral cingulate zone, and dorsal striatum (regions linked to self-initiated action). In contrast, when participants made self-reflective metacognitive judgments about the extent of their own control (i.e., judgments of agency) compared with when they made judgments that were not about control (i.e., judgments of performance), increased activity was observed in the anterior PFC, a region associated with self-reflective processing. These results indicate that action monitoring is dissociable from people's conscious self-attributions of control. ■

INTRODUCTION

The present study examines the neurocognitive basis of action monitoring and metacognitive judgments of agency. Action monitoring involves the ability to detect whether action outcomes are in line with the expected consequences of our intended actions (i.e., with our action plans). It plays an important role in the way we coordinate our movements with objects and events in the environment, including other agents. Although action monitoring is a critical component of voluntary action, empirical evidence suggests that it may (at least part of the time) operate at a spontaneous or automatic level of processing. For example, a number of studies have demonstrated that we can respond to minor violations of our action plans (such as unpredicted jumps in target position) at both a neural and behavioral level without awareness of having done so (e.g., Pisella et al., 2000). Thus, it appears that we can (and sometimes do) use the output of our monitoring to successfully coordinate our actions without first consciously interpreting whether this output indicates the presence or absence of control. In contrast, to make an explicit judgment concerning our own agency, we do need to consciously assess whether or not we were in control. For instance, to determine our responsibility for a particular action, we must be able to metacognitively reflect on the output of our action monitoring, as well as other relevant cues, and then consciously infer the extent to which we were or were not the cause of the action in question (Haggard & Tsakiris, 2009; Synofzik, Vosgerau, & Newen, 2008; Wegner, Sparrow, & Winerman, 2004; Wegner, 2002, 2003). Therefore, when it comes to understanding human agency, the question is not simply how we monitor disturbances in the execution of our actions but also how we use the output of this monitoring to make metacognitive judgments about whether or not we were in control.

Although we as well as others (Pacherie, in press; Haggard & Tsakiris, 2009; Synofzik et al., 2008; Metcalfe & Greene, 2007; Georgieff & Jeannerod, 1998) have proposed that there may be a distinction between action monitoring and metacognition of agency, brain imaging studies have focused exclusively on identifying patterns of neural activity that are associated with action monitoring (see David, Newen, & Vogeley, 2008, for a review). Typically, the participants in these studies attempted to perform a specific action or behavior and then received sensory feedback concerning the spatial trajectory or temporal sequence of their movement that either coincides or conflicts with their action plan. When the feedback was discrepant with expectations, the imaging results have consistently shown increased activation in areas surrounding the right TPJ

[1]Columbia University, [2]University of Colorado, [3]Harvard University

(rTPJ; e.g., Nahab et al., 2011; Spengler, von Cramon, & Brass, 2009; Farrer et al., 2003; Leube et al., 2003). The TPJ is a region with boundaries that cannot be structurally defined and is often described as encompassing the supramarginal gyrus, the inferior parietal lobe/angular gyrus, caudal parts of the superior temporal gyrus, dorsal–rostral parts of the occipital gyri, and posterior parts of the STS (Decety & Lamm, 2007). We later report the results of a meta-analysis of recent agency studies that we conducted to define this broad ROI for the present study.

The fact that increases in TPJ activation are reliably associated with decreases in control has led many researchers to speculate that the TPJ receives input from a comparator mechanism that is responsible for determining whether the predicted consequences of an intended action (on the basis of the output of an internal "forward model") match sensory feedback about the actual trajectory of the action (Hohwy & Frith, 2004; Blakemore, Wolpert, & Frith, 2002; Wolpert & Ghahramani, 2000; cf. Synofzik et al., 2008). When the predicted and actual trajectory do not match, the TPJ is thought to receive a signal indicating that a mismatch exists and that corrective action may be required. On a somewhat less consistent basis, the results of previous action monitoring studies have sometimes shown increased activation in the cerebellum (Blakemore, Frith, & Wolpert, 2001), precuneus (David et al., 2007), and extrastriate body area (David et al., 2007) in response to a mismatch, and in the insula (Tsakiris, Longo, & Haggard, 2010; Farrer et al., 2003) and pre-SMA (Tsakiris et al., 2010) when actions proceed as planned.

Although participants in some of the studies on action monitoring provided judgments of agency after they completed the focal task (e.g., Nahab et al., 2011; Spengler et al., 2009; Tricomi, Delgado, & Fiez, 2004), these judgments were used only to identify neural activity during the task that was related to feeling in or out of control. Imaging data corresponding to the judgments themselves, though, were either not collected or not analyzed. Thus, as of yet, no studies have been published that examine the neural activity associated with metacognition of agency itself—that is, with reflecting on the output of one's action monitoring after an action has been performed.

In the present experiment, each trial consisted of both a *game phase*, in which participants' task was to move the cursor (via the trackball) along a horizontal track to touch each of the downward scrolling Xs, and a *judgment phase*, in which they judged either how much "in control" they felt during the game or how well they had performed (see Figure 1). In a manipulation somewhat similar to those used in previous action-monitoring studies (Farrer et al., 2003; Franck et al., 2001; Blakemore, Frith, & Wolpert, 1999), in which the trajectory of the person's apparent movement was altered, we disrupted participants' objective control by adding random noise or "turbulence" to the position of the cursor. Alternatively, we sometimes distorted control by having the Xs disappear "by magic" (i.e., without the cursor having actually touched the X). Our behavioral question was whether participants' judgments of agency would be sensitive to these disruptions of control. Our imaging question was whether similar patterns of brain activation would be observed during the game and judgment phases of the task or whether action monitoring and judgments of agency involve different sets of neural processes.

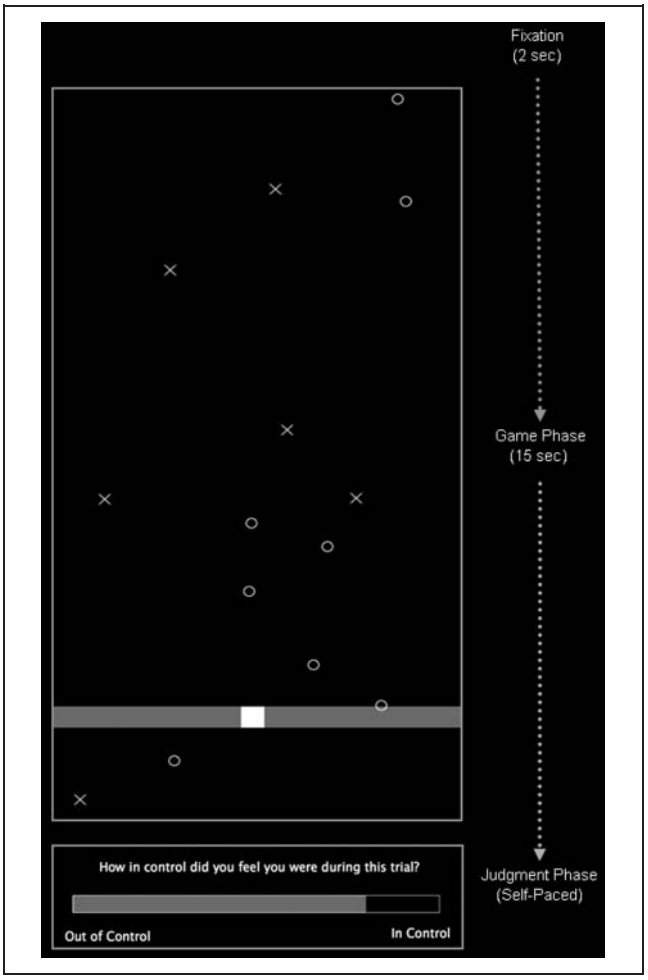

Figure 1. A schematic diagram of trials from the experiment. Participants viewed a fixation cross for 2 sec, played the game for 15 sec (during which they were exposed to the turbulence and magic manipulations), and then took as long as they needed to make either a judgment of agency or a judgment of performance (depending on the run). The turbulence manipulation involved adding pseudorandom noise to the spatial correspondence between the mouse and the cursor; the magic manipulation involved causing the Xs to disappear and the corresponding ping noise to sound whenever the cursor was moved within 10 pixels of the outer edge of a target (such that participants often received credit for Xs without actually touching them).

Because our previous research using a similar paradigm has shown that the judgments of agency made by college students, older adults, and even (to some extent) children tend to accurately reflect these manipulated disruptions of control (Metcalfe, Eich, & Castel, 2010; Kirkpatrick, Metcalfe, Greene, & Hart, 2008; Metcalfe &

Greene, 2007), we hypothesized that when participants experienced such disruptions during the game phase of the present experiment, they would exhibit increased neural activity in regions that are associated with action monitoring, particularly the rTPJ. In contrast, we hypothesized that when they reflected back on this disruption (i.e., when they made their agency judgments) compared with when they reflected back on their performance, there would be increased activation in areas that subserve key processes involved in metacognition of agency, such as self-reflection and self-attribution.

Perhaps the most extensive line of research on self-reflection has focused on identifying brain regions that underlie the evaluation of one's own traits, abilities, and mental states (for reviews, see van der Meer, Costafreda, Aleman, & David, 2010; Schmitz & Johnson, 2007; Northoff et al., 2006). In these studies, participants are typically presented with trait or mental state adjectives and are asked to indicate whether the adjectives apply to them (self-attribution condition) or whether they apply to another individual, such as a friend, relative, or celebrity (other-attribution condition). Results have consistently shown stronger activation of cortical midline structures and their surrounding regions, including the middle and medial anterior PFC (aPFC; i.e., BA 10) during self-attribution trials compared with other-attribution trials (e.g., Jenkins & Mitchell, 2011; see also Vinogradov et al., 2006). A closely related line of research has shown increased activity in the aPFC when people reflect on their own feelings or intentions while forming judgments or making decisions (for reviews, see Mitchell, 2009; Amodio & Frith, 2006; Ramnani & Owen, 2004; Christoff & Gabrielli, 2000). Given the convergence of these findings, a number of investigators have theorized that the processing of self-relevant information (i.e., information about one's attributes, experiences, and mental states) emerges from cortical midline structures, including regions of the aPFC, that are tightly interconnected and functionally distinct from structures that underlie domain-general cognitive processing (van der Meer et al., 2010; Schmitz & Johnson, 2007; Northoff et al., 2006). In the present experiment then, it was our hypothesis that these structures, rather than the structures associated with action monitoring, would be particularly active when people made metacognitive judgments of agency compared with judgments of performance.

METHODS

Participants

The participants were 11 members of the Columbia community (7 women, 1 left-handed, mean age = 27 years, range = 18–36 years) who provided consent in a manner approved by the Columbia University Institutional Review Board and were treated in accordance with APA ethical guidelines. An outlier analysis determined that the left-handed participant's data did not differ significantly from the rest of the sample, and so they were included in all analyses.

Behavioral Procedure

The primary task, which was framed as a computer game, involved using a trackball to move a cursor left and right along a horizontal track at the bottom of the screen as approximately 20–30 stimuli (Xs and Os) that were randomly distributed from left to right scrolled down from the top of the screen to the bottom (see Metcalfe et al., 2010; Metcalfe & Greene, 2007). During each trial of the task, the Xs or Os disappeared as soon as they were "touched" by the cursor but continued to scroll past the horizontal track if they were not touched. In addition, a "ping" sound occurred each time an X was hit and a "pong" sound occurred each time an O was hit (see Figure 1).

Before entering the scanner, participants were instructed to use the track ball to touch as many Xs with the cursor as possible and to avoiding touching any of the Os. They were also told that, after completing each trial, they would be asked to make either a judgment of agency or a judgment of performance. For judgments of agency, participants were presented with a visual analogue scale and instructed to pull the slider to the left if they felt they were "not in control" or to the right if they felt they were "in control." For judgments of performance, participants were presented with the same scale but were instructed to pull the slider to the left if they felt that their performance was "very low" or to the right if they felt their performance was "very high." For both types of judgment, the left end of the scale was coded 0 and the right end was coded 100, with values in between being assigned a value based on the proportion of the distance between the two ends that was spanned by the slider.

Once in the scanner, each experimental session consisted of four to six runs (depending on the amount of scanner time available for that participant after instructions and structural scans had been completed), with each run consisting of a single set of 24 trials, for a total of 96–144 trials per participant. All trials consisted of two phases, a game phase that lasted 15 sec and a judgment phase that lasted from the end of the game phase (at which point the visual analogue scale immediately appeared) until the participant made an agency or performance rating and clicked the "OK" button (see Figure 1). The intertrial interval (during which a fixation cross appeared on the screen) was set to 2 sec. Within each run, we manipulated two factors (termed "turbulence" and "magic"), such that the game phase of each trial belonged to one of four conditions: a Control condition in which the participants experienced a direct temporal and spatial correspondence between their movement of the mouse and the movement of the cursor on the screen, a Turbulence condition in which pseudorandom noise was added to the spatial correspondence between the mouse and the cursor, a Magic

condition in which Xs disappeared and the corresponding ping noise sounded whenever the cursor was moved within 10 pixels of the outer edge of a target (such that participants often received credit for Xs without actually touching them), and a combined TurbMagic condition in which both manipulations occurred together. The pseudorandom noise in the Turbulence and TurbMagic conditions was generated each time the computer program updated the game's graphical display (every 33–83 msec) and was based on the following formula $\Delta x' = \Delta x + [6 \times \sin(2t\pi/45) + r]$, wherein Δx was the distance the participant actually moved the mouse, t was the number of updates that had been computed since the counter was last reset (the counter was always reset after 45 updates), r was a random integer between −4 and 4, and $\Delta x'$ was the resultant movement on the computer screen. The order of the 24 trials in each run was pseudorandom.

Across runs, we manipulated a single factor (judgment type), such that participants made only judgments of agency for all of the trials in half of the runs and made judgments of performance for all of the trials in the other half of the runs. Judgment type alternated from one run to the next, with the judgment type of the first run counterbalanced across participants. The purpose of manipulating judgment type across runs, as opposed to having participants make both types of judgment for each trial or on consecutive trials, was to avoid carryover effects. In addition to recording participants' judgments, the program recorded the number of Xs and Os participants touched during the game phase of each trial, the number of Xs and Os that appeared on the screen but were not touched, and the length of the judgment phase.

Behavioral Data Analysis

Performance

Trial-by-trial hit rates (i.e., percentage of Xs touched), false alarm rates (i.e., percentage of Os touched), and *d'* values were computed for each participant. Because analyses of these performance variables yielded similar patterns of results, only the analysis of hit rate (which was the primary measure of performance in previous studies; e.g., Metcalfe et al., 2010; Metcalfe & Greene, 2007) is reported. To determine the effectiveness of the turbulence and magic manipulations, we submitted hit rate to a 2 (Turbulence: turbulence vs. no turbulence) × 2 (Magic: magic vs. no magic) repeated measures ANOVA.

Judgments

Because there tends to be a strong correlation between participants' judgments of agency and their hit rate (Metcalfe et al., 2010; Metcalfe & Greene, 2007), the primary question of interest was whether or not participants correctly inferred their lack of agency in conditions under which they were not fully in control (i.e., the Turbulence, Magic, and TurbMagic conditions), regardless of how well they thought they had performed in these conditions. That is, when controlling for differences in perceived performance across trials, do we still find effects of the Turbulence and Magic manipulations on judgments of agency? Affirmative answers would suggest that any brain activity resulting from these manipulations was related to the feeling of agency itself. To address the question, we conducted several different analyses.

First, in keeping with our past experiments using this paradigm, we computed summary scores for each participant based on the following formula, $(JOP_C - JOA_C) - (JOP_E - JOA_E)$, where JOP refers to the mean judgment of performance, JOA refers to the mean judgment of agency, the subscript C refers to the Control condition, and the subscript E refers to the particular experimental condition of interest (i.e., Turbulence, Magic, or TurbMagic). These summary scores, which compress the interactions between judgments of agency and judgments of performance into a single measure (see Metcalfe et al., 2010), should have been negative, so long as participants realized that their performance in the experimental conditions (where agency was objectively decreased) was not entirely because of their own actions. Second, to take advantage of the factorial design used in the present study (and to be consistent with the contrast analyses of the imaging data reported below), we also computed another set of summary scores based on the main effects of the two manipulations. The formula used for these scores was $(JOP_A - JOA_A) - (JOP_P - JOA_P)$, where the subscript A refers to the absence of the particular experimental manipulation of interest (e.g., the average of the Control and Magic conditions when looking at Turbulence) and the subscript P refers to the presence of the manipulation (e.g., the average of the Turbulence and TurbMagic conditions when looking at Turbulence).

Finally, to analyze the behavioral results in a manner that is more consistent with our parametric analysis of the imaging data reported below, we submitted each participant's judgments of agency to a separate regression test in which hit rate (which served as a proxy for perceived performance), turbulence (dummy-coded: 0 = no turbulence, 1 = turbulence), and magic (dummy-coded: 0 = no magic, 1 = magic) served as simultaneous predictors. We then performed a series of one-sample *t* tests to determine whether the mean beta coefficients for the group differed significantly from zero (Lorch & Myers, 1990). Next, to determine whether the manipulations uniquely targeted participants' metacognition of agency or whether they had similar effects on participants' perceived performance, we repeated the previously described regression analyses and *t* tests with judgments of performance as the dependent variable. We completed the analysis by directly comparing the beta coefficients associated with the two types of judgments in a series of paired-samples *t* tests.

fMRI Data Acquisition

Images were acquired with a GE twin-speed 1.5-T scanner. Whole-brain functional data were acquired in 27 contiguous axial slices (4.0-mm thick, 3 × 3 mm in-plane resolution) parallel to the anterior–posterior commissure line with a T2*-weighted EPI sequence (TR =2000, TE = 38, flip angle = 90°, field of view = 192, array size = 64 × 64). Each run acquired 275 whole-brain volumes, the first five of which were discarded to allow the scanner to stabilize. Structural data were acquired with a high-resolution T1-weighted spoiled gradient-echo scan, which recorded 182 slices at a thickness of 1 mm and a resolution of 1 × 1 mm.

fMRI Data Analysis

Preprocessing

Spatial preprocessing and statistical analyses were performed with SPM5 (Wellcome Trust Centre for Neuroimaging, University College London, London, UK). First, each functional brain volume was corrected for differences in acquisition time between slices and realigned to correct for head movement. The data were then normalized to a standard anatomical space using the Montreal Neurological Institute (MNI) EPI template, resampled into 2-mm isotropic voxels, and spatially smoothed with an 8-mm FWHM Gaussian kernel.

Primary Statistical Analyses

Statistical analyses were performed to test two separate models using the general linear model framework implemented in SPM5 (Friston et al., 1995). In both cases, a high-pass filter (128 sec) was applied to the data to remove low-frequency drifts in signal changes. The first model (i.e., the *simple contrast model*) was designed to identify brain regions that responded to disturbances (i.e., Turbulence and Magic) in agency during the game and judgment phases, as well as regions that showed increased activation when making judgments of agency compared with judgments of performance. Thus, eight within-run conditions (i.e., the four game conditions crossed with the two task phases) were modeled as boxcar functions (such that each phase of each trial was treated as a separate epoch) and convolved with the canonical hemodynamic response function to create regressors of interest. For each participant, voxelwise statistical parametric maps were created for the game and judgment phase contrasts of *turbulence versus no turbulence* (Turbulence + TurbMagic vs. Control + Magic), *magic versus no magic* (Magic + TurbMagic vs. Control + Turbulence), and *turbulence versus magic* (i.e., Turbulence + Magic vs. TurbMagic + Control). In addition, a map was created for the judgment phase contrast of *judgment of agency versus judgment of performance*. Whereas the first three contrasts compared different regressors collapsing across all runs, the last contrast compared the same judgment phase regressor between two sets of runs.

The second model (i.e., the *parametric model*) was setup to identify brain regions that varied in activation during the game phase as a function of perceived agency. Because we were primarily interested in participants' *sense of agency* independent of their perceived performance on the game (i.e., their sense of being more or less in control than was indicated by the number of Xs they managed to touch) and because we did not collect judgments of performance on the same trials as judgments of agency (as that would have precluded the temporal separation necessary for computing the *judgment of agency vs. judgment of performance* contrast described above), we operationalized this construct as the difference between their judgment of agency and hit rate on each trial of the judgment of agency runs (i.e., judgment of agency $[-X_{hits}/X_{total} \times 100]$). As a control, we also computed the difference between participants' judgment of performance and hit rate on each trial of the judgment of performance runs.

For each set of runs (i.e., the judgment of agency or judgment of performance runs), the corresponding difference score for each trial (i.e., judgment of agency $[-X_{hits}/X_{total} \times 100]$ or judgment of performance $[-X_{hits}/X_{total} \times 100]$) was entered into the model as a parametric modulator of both task phases, each of which was modeled as a boxcar function and convolved with the canonical hemodynamic response function. In addition, because we wanted to control for potential confounds associated with performance, such as motor activity and affect (e.g., participants may have moved the mouse more or been less pleased with their performance on the difficult turbulence trials than on other trials), each difference score was entered into the model after we had included hit rate as a secondary modulator of the two task phases. That is, because SPM5's default option is the serial orthogonalization of parametric modulators, entering hit rate before the difference score ensured that none of the variance shared by the two regressors would contaminate activation that is uniquely associated with the difference score. On the basis of this model, subject-specific statistical parametric maps were created for the game phase regressors corresponding to each difference score. These maps represented the extent to which changes in the BOLD signal during the game phase were correlated with participants' subsequent sense of agency or performance at each voxel in the brain.

To explore each model, single-subject contrast images from the first level were entered into a second-level random effects analysis, which involved computing one-sample *t* tests across the contrast images of all subjects to create a series of group level statistical parametric maps (Mumford & Nichols, 2009). For whole-brain analysis, these maps were thresholded first at the voxel level ($p < .005$, uncorrected) and then at the cluster level ($p < .05$, family-wise error corrected) to protect against false-positive activations (Friston,

Table 1. Studies, Contrasts, and Coordinates Used in the Meta-analysis

Study	*Contrast*	*Imaging Method*	*Peak Coordinates (MNI)*			*Original Space*
Balslev, Nielsen, Lund, Law, & Paulson, 2006	*Asynchronous > Synchronous* for active movements	fMRI	−42	−51	45	MNI
			−54	−48	27	
			54	−42	33	
Chaminade & Decety, 2002	*Following the other > Acting at will* masked exclusively with *Leading the other > Acting at will*	PET	44	−58	54	MNI
			32	−50	52	
David et al., 2006	*Passive > Active*	fMRI	−50	−58	16	MNI
			58	−38	54	
			44	−52	54	
David et al., 2007	*Incongruent > Congruent*	fMRI	−40	−48	42	MNI
			60	−42	38	
Decety, Chaminade, Grezes, & Meltzoff, 2002	*Imitation of the other by the self > Self-action*	PET	−54	−48	24	MNI
			54	−52	40	
Farrer & Frith, 2002	*Other-attribution > Self-attribution*	fMRI	−48	−52	40	MNI
			44	−58	32	
Farrer et al., 2003	*25° deviation > 0°* in conjunction with *50° > 25°* and *Other-controlled > 50°*	PET	−64	−58	32	MNI
			56	−56	36	
Farrer et al., 2004	*25° deviation > 0°* in conjunction with *50° > 25°* and *Other-controlled > 50°* (normal subjects)	PET	56	−56	36	MNI
Farrer et al., 2008 (Study 1)	Awareness of action discrepancy	fMRI	−40	−58	36	MNI
			−48	−38	54	
			44	−54	38	
Farrer et al., 2008 (Study 2)	Perturbed agency	fMRI	−48	−46	56	MNI
			58	−46	48	
			−40	−58	36	
Leube et al., 2003	Correlated with extent of delay	fMRI	48	−42	18	MNI
Nahab et al., 2011	*Loss responsive regions*	fMRI	62	−50	13	TAL
			58	−45	19	
			45	−46	15	
			64	−50	40	
			−52	−49	51	
			39	−56	41	
			56	−55	27	
			52	−60	43	
			52	−53	56	
			−46	−48	44	
Ruby & Decety, 2001	*Third-person simulation > First-person simulation*	PET	44	−64	24	MNI
			50	−58	30	

Table 1. (*continued*)

Study	*Contrast*	*Imaging Method*	*Peak Coordinates (MNI)*			*Original Space*
Schnell et al., 2007	*Monitoring condition > Control condition*	fMRI	−59	−51	36	MNI
			−50	−53	47	
			62	−51	22	
			53	−45	24	
Spengler et al., 2009	*Correlated with increased discrepancy*	fMRI	−49	−56	15	TAL
			54	−52	20	
Tsakiris et al., 2010	*Main effect of asynchronous stimulation*	fMRI	40	−58	26	MNI
			52	−38	38	
Williams et al., 2006	*Imitation > Action execution*	fMRI	59	−26	27	MNI
			55	−33	38	
Yomogida et al., 2010	*Agency error*	fMRI	55	−33	38	MNI

Talairach coordinates were converted to MNI space using tal2icbm_spm.m as implemented in GingerALE v2.0.

The current meta-analysis was based on a previous meta-analysis of agency studies conducted by Decety and Lamm (2007). However, we excluded five studies (Kable & Chatterjee, 2006; Ramnani & Miall, 2004; Saxe, Xiao, Kovacs, Perrett, & Kanwisher, 2004; Fink et al., 1999; Spence et al., 1997) from the Decety and Lamm meta-analysis because the contrasts for which they reported increased TPJ activity did not fit with our definition of agency.

Worsley, Frackowiak, Mazziotta, & Evans, 1993). All significant activations were overlaid on sections of the MNI canonical brain, with peak voxels reported in MNI coordinate space. For the anatomical labeling of these activations, we used the automated anatomical labeling atlas (Tzourio-Mazoyer et al., 2002).

ROI Analyses

Brain regions that were identified on the basis of theory and literature review were analyzed using small volume corrections (SVCs) as implemented in SPM5 (Worsley et al., 1996). Specifically, when testing for increased activation in regions of the aPFC during the judgment phase of the agency task, we restricted the search volume to four 10-mm spheres that were centered at coordinates reported in a recent meta-analysis of self-reflection studies (van der Meer et al., 2010). The meta-analysis explored two sets of contrasts found in fMRI and PET studies of self-reflection: self-reflection versus baseline contrasts and self-reflection versus other-reflection contrasts. We based the ROI on the analysis of the latter set because, as the authors suggest, it identified regions "unique to self-reflective processing" as opposed to regions pertaining to "reflective processing on a broader scale." The analysis of the 17 studies that included self-reflection versus other-reflection contrasts yielded a large cluster with peaks in the pregenual part of ACC (pACC; BA 24, 32; 2, 42, 20) and the bilateral aPFC (0, 50, −2; −2, 54, 8; −18, 50, 16)—the coordinates of these peaks were used to create the four 10-mm spheres.

When testing for increased activity in the bilateral TPJ during the game phase of the task, we restricted the search volume to a 6745-voxel region derived from a meta-analysis of recent agency studies. The meta-analysis, which we conducted ourselves, was based on 21 TPJ coordinates from 10 agency studies originally included in a meta-analysis by Decety and Lamm (2007),[1] as well as 25 additional coordinates reported in seven recent agency studies (see Table 1). Analysis of the 46 activation peaks was conducted in MNI space using multilevel kernel density analysis (Wager, Lindquist, Nichols, Kober, & Van Snellenberg, 2009). As shown in Figure 3, the resultant activation map, which was thresholded first at the voxel level ($p < .05$) and then at the cluster level ($p < .05$), yielded large clusters in the left (−50, −50, 34; $K_E = 1738$) and right TPJ (54, −50, 32; $K_E = 5007$) that included portions of the angular gyus, supramarginal gyrus, inferior parietal lobule, superior temporal gyrus, and middle temporal gyrus. Multilevel kernel density analysis was used because it designates contrast maps (and not peaks) as the unit of analysis. This means that contrast maps are essentially treated as random effects, which ensures that no one contrast map can contribute disproportionately to the overall results, even if it contains numerous peaks in the same area. Unless otherwise indicated, all maps analyzed using SVCs were thresholded at a voxel level of $p < .005$ (uncorrected) and a cluster level of $p < .05$ (family-wise error corrected).

RESULTS

Behavioral Results

Performance

The results of the hit rate analysis revealed a significant main effect of Turbulence ($F(1, 10) = 57.35$, $p < .001$,

MSe = 707.42, η_p^2 = .85), such that participants were less accurate when turbulence was present (M = 64.40, SE = 1.66) than when it was absent (M = 72.42, SE = 1.61), as well as a significant main effect of Magic ($F(1, 10)$ = 1,146.72, $p < .001$, *MSe* = 12,087.35, η_p^2 = .99), such that participants were more accurate when magic was present (M = 84.99, SE = 1.76) than when it was absent (M = 51.84, SE = 1.47). These main effects were qualified by a significant interaction ($F(1, 10)$ = 148.80, $p < .001$, *MSe* = 543.84, η_p^2 = .94). Bonferroni tests revealed that, although turbulence lowered the mean hit rate when magic was absent ($t(10) = 11.42, p < .001$), it had no effect on hit rate when magic was present ($t(10) = .91, p = .38$). It should also be noted that, across all conditions, there were strong within-participants correlations between hit rate and judgments of performance (r_{mean} = .71), an important finding given that we used hit rate as a proxy for judgments of performance in some of the imaging analyses.

Judgments

As shown in the left panel of Figure 2, the results of the first analysis replicated the findings from our previous agency studies. More specifically, the summary agency scores for the Turbulence ($\bar{X} = -27.21$, $t(10) = 4.68$, $p < .001$), Magic ($\bar{X} = -13.35, t(10) = 3.79, p = .004$), and TurbMagic conditions ($\bar{X} = -37.45$, $t(10) = 5.78$, $p < .001$) were significantly less than zero, indicating that in all three instances people correctly recognized that they were not completely in control. Follow-up tests revealed that, although this effect was stronger in the Turbulence condition than in the Magic condition ($t(10) = 2.81, p = .02$), it was strongest when both manipulations were present in the same trial (i.e., in the TurbMagic condition; ts > 4.12, ps < .003). The results of the second analysis showed virtually the same pattern, as depicted in the right panel of Figure 2. The summary agency scores for the main effects of the Turbulence ($\bar{X} = -25.65$, $t(10) = 4.74$, $p < .001$) and Magic manipulations ($\bar{X} = -11.80$, $t(10) = 6.05$, $p < .001$) were both significantly less than zero, and this effect was again stronger for turbulence than for magic ($t(10) = 2.81$, $p = .02$).

For the final analysis, within-subject regression tests revealed significant effects of hit rate ($\beta_{mean} = .34, t(10) = 9.03$, $p < .001$) and turbulence ($\beta_{mean} = -.66, t(10) = -12.34$, $p < .001$) on participants' judgments of agency, but no effect of magic ($\beta_{mean} = -.05$, $t(10) = -.83$, $p = .43$). Thus, consistent with the results of first two analyses, decreases in control because of turbulence (though not magic in this case) led to decreases in participants' judgments of agency, even when controlling for task performance. Additional regression analyses revealed significant effects of hit rate ($\beta_{mean} = .64, t(10) = 10.69, p < .001$) and turbulence ($\beta_{mean} = -.17$, $t(10) = -4.11$, $p = .002$), but no effect of magic ($\beta_{mean} = .03$, $t(10) = .43$, $p = .78$) on participants' judgments of performance. However, a comparison of beta coefficients between the two types of judgments showed that turbulence had a significantly stronger negative effect on judgments of agency compared with judgments of performance ($t(10) = -9.28$, $p < .001$), whereas hit rate had a significantly weaker positive effect ($t(10) = -8.40, p < .001$). Once again, these results indicate that the main outcome of the turbulence manipulation was a reduction in participants' sense of control.

fMRI Results

Simple Contrast Model

Turbulence contrasts. To determine whether the turbulence manipulation increased activity in the brain region that is most directly associated with detecting disruptions of control, we examined neural responses in the TPJ ROI

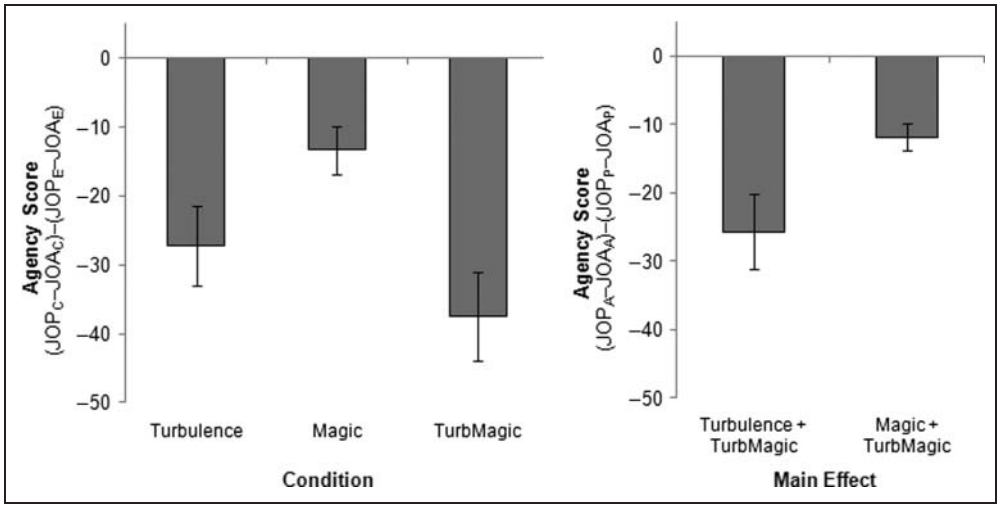

Figure 2. (Left) Participants' summary agency scores based on the formula $(JOP_C - JOA_C) - (JOP_E - JOA_E)$, wherein JOP refers to mean judgment of performance, JOA refers to mean judgment of agency, the subscript C refers to the control condition, and the subscript E refers to the particular experimental condition of interest (i.e., Turbulence, Magic, or TurbMagic). (Right) Agency scores based on the formula $(JOP_A - JOA_A) - (JOP_P - JOA_P)$, wherein the subscript A refers to the absence of the particular experimental manipulation of interest (e.g., the average of the Control and Magic conditions, when looking at Turbulence) and the subscript P refers to the presence of the manipulation (e.g., the average of the Turbulence and TurbMagic conditions, when looking at Turbulence). In both cases, a negative score indicates that participants correctly perceived themselves *not* to be in full control of the cursor. Error bars reflect *SEM*s.

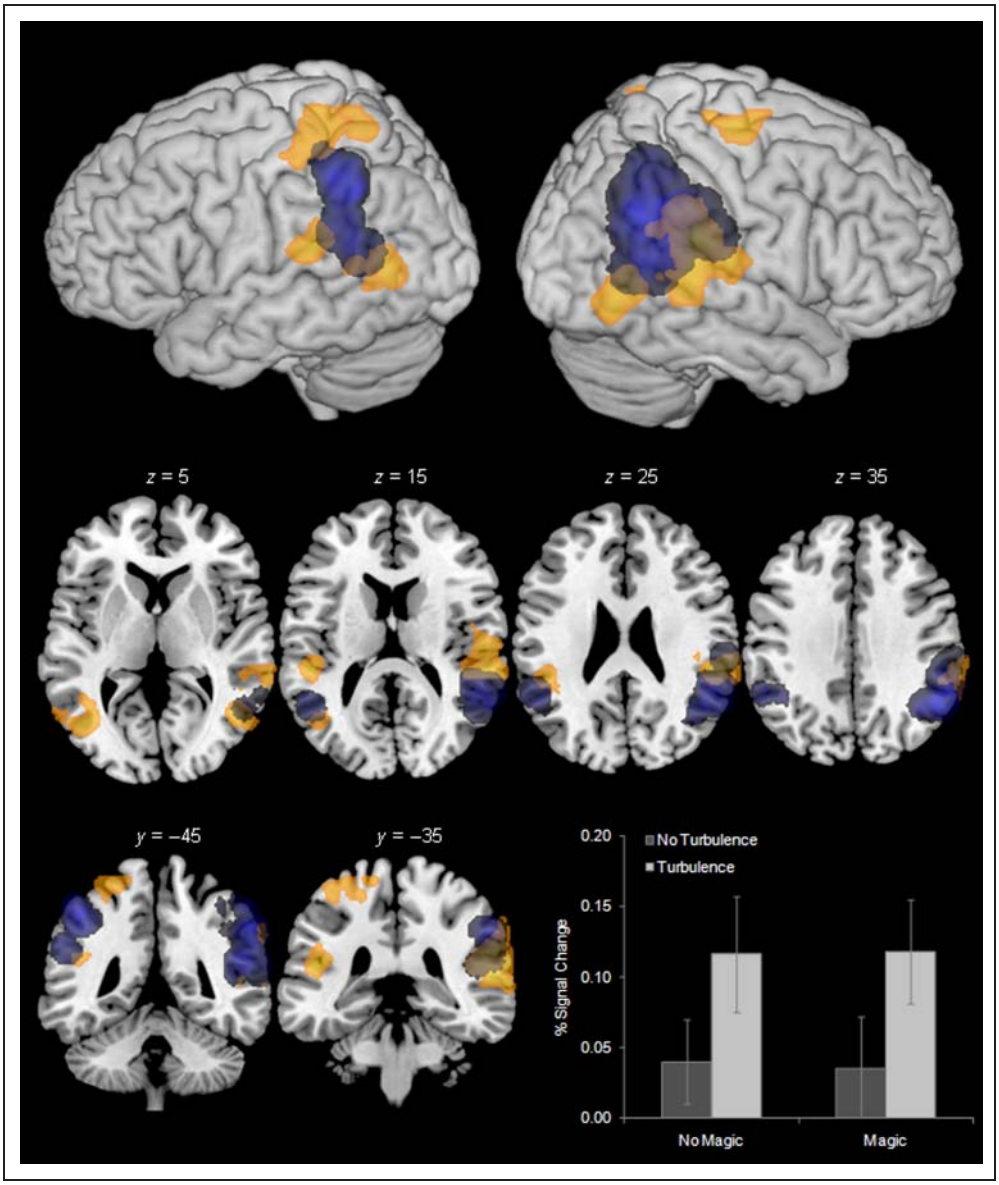

Figure 3. The resultant activation maps from our meta-analysis of recent agency studies (in blue) and from the game phase *turbulence* > *no turbulence* contrast (in yellow). An SVC analysis revealed clusters of overlapping activity in both the right (56, −34, 26; $Z = 4.90$; $K_E = 618$; $p < .05$, cluster corrected; 48, −62, 4; $Z = 3.96$; $K_E = 72$; $p < .05$, FDR corrected) and left TPJ (−44, −64, 12; $Z = 3.37$; $K_E = 65$; $p < .05$, FDR corrected; −50, −40, 24; $Z = 3.48$; $K_E = 21$; $p < .05$, FDR corrected). The bottom right graph shows parameter estimates for the large cluster in the rTPJ that met our primary threshold. Percent signal change was computed for individual participants using MarsBaR v0.42 (marsbar.sourceforge.net). Error bars reflect *SEM*s.

during the game phase of the task for *turbulence* trials (Turbulence + TurbMagic) compared with no turbulence trials (Control + Magic). As shown in Figure 3, the results of the analysis yielded a large cluster of activity in the right TPJ (56, −34, 26; $Z = 4.90$; $K_E = 618$; $p < .05$, cluster corrected) as well as a smaller cluster in the right TPJ (48, −62, 4; $Z = 3.96$; $K_E = 72$; $p < .05$, false discovery rate [FDR] corrected) and two smaller clusters in the left TPJ (−44, −64, 12; $Z = 3.37$; $K_E = 65$; $p < .05$, FDR corrected; −50, −40, 24; $Z = 3.48$; $K_E = 21$; $p < .05$, FDR corrected) that did not meet the primary threshold. To explore the pattern of this activity across all four game conditions, we used MarsBar v0.42 software (marsbar.sourceforge.net) to extract percent signal change for each participant from the portion of the large right-sided cluster that overlapped with the TPJ ROI. As depicted in the bottom right inset of Figure 3, there was significantly more activity in the Turbulence condition than in the Magic or Control conditions (ts > 4.46, ps < .002); furthermore, these effects were not qualified by a significant Turbulence × Magic interaction ($F(1, 10) = .11$, $p = .75$, $MSe < .001$, $\eta_p^2 = .01$). Thus, it appears that (independent of any effects of Magic) our turbulence manipulation had a similar effect on participants' neural activity as the manipulations of perceived control used in previous studies of agency.

The corresponding whole-brain analysis revealed that much of the activity within the TPJ ROI belonged to a significant cluster of right-sided activation that included regions of the supramarginal gyrus, the posterior superior temporal gyrus/STS, and the rolandic operculum (see Table 2). The analysis also revealed a similar but smaller cluster of activation on the left side as well as clusters in the bilateral posterior middle temporal gyrus, the left postcentral gyrus, and the right precentral gyrus. The

Table 2. Suprathreshold Clusters for the Turbulence Manipulation

Side	*Anatomical Region of Cluster*	*Cluster Size*	*Cluster Maxima (MNI)*			*Z Score*
Game Phase: Turbulence (Turbulence + TurbMagic) > No Turbulence (Control + Magic)						
R	TPJ (supramarginal gyrus, posterior superior temporal gyrus/sulcus, rolandic operculum)	1386	56	−34	26	4.90
			68	−34	18	4.54
			58	−34	14	4.51
R	Posterior middle temporal gyrus	356	48	−62	4	3.96
R	Precentral gyrus, SMA	541	26	−16	68	3.78
			14	−28	66	3.75
			12	−8	68	3.54
L	TPJ (posterior superior temportal gyrus)	359	−48	−34	18	4.42
			−40	−48	24	2.54
L	Posterior middle temporal gyrus	409	−46	−66	8	3.85
			−46	−56	10	3.21
			−60	−64	6	2.80
L	Postcentral gyrus, superior parietal lobule	637	−16	−38	70	3.75
			−34	−42	70	3.35
			−28	−58	64	3.19
Game Phase: No Turbulence (Control + Magic) > Turbulence (Turbulence + TurbMagic): ns						
Judgment Phase: Turbulence (Turbulence + TurbMagic) > No Turbulence (Control + Magic): ns						
Judgment Phase: No Turbulence (Control + Magic) > Turbulence (Turbulence + TurbMagic):						
R	Lingual gyrus, calcarine sulcus, cuneus	2117	18	−86	6	4.59
			12	−78	2	4.29
			14	−78	20	4.10
R	Frontal operculum, precentral gyrus	358	44	12	38	4.10
			40	14	22	3.20
			32	18	20	3.11

R = right; L = left.

reverse contrast (i.e., *no turbulence > turbulence* trials) did not reveal any suprathreshold brain regions during the game phase.

During the judgment phase, the *turbulence > no turbulence* contrast did not reveal any suprathreshold brain regions. However, the reverse contrast revealed a large cluster of right-sided activity that extended across regions of the lingual gyrus, calcarine sulcus, and cuneus, as well as a smaller cluster that included regions of the right frontal operculum and precentral gyrus.

Magic contrasts. To determine whether the magic manipulation also increased activity in the TPJ, we examined neural responses in the TPJ ROI during the game phase of the task for magic trials (Magic + TurbMagic) compared with no magic trials (Control + Turbulence). The results revealed several small clusters of activation in the left and right TPJ; however, these clusters did not survive correction. The corresponding whole-brain analysis showed increased activation in the bilateral putamen/piriform cortex and the right precuneus. The reverse

contrast (i.e., performance during the *no magic* > *magic* trials), did not reveal any suprathreshold brain regions.

For the judgments phase of the task, the *magic* > *no magic* contrast revealed a significant cluster of activation in the right lateral PFC. The reverse contrast did not reveal any suprathreshold brain regions. The weak effects of Magic, relative to Turbulence, may correspond to the relatively weak behavioral effects that we observed for this variable (see above).

Turbulence × Magic contrasts. To determine whether there were any regions of the TPJ that responded more strongly to the combined effects of Turbulence and Magic than to either manipulation alone, we examined neural responses in the TPJ ROI during the game phase of the task for the TurbMagic and Control trials compared with the Turbulence and Magic trials. The results of this analysis and of the corresponding whole-brain analysis did not reveal any significant clusters. The reverse contrast (which would indicate regions that exhibited interference between the manipulations) revealed several minute clusters of activation in the right TPJ; however, these clusters did not survive correction. In addition, the corresponding whole-brain analysis did not reveal any suprathreshold brain regions. For the judgments phase of the task, neither of the two interaction contrasts revealed significant clusters of activation.

Judgment of agency versus judgment of performance contrasts. To determine whether there was increased activity in brain regions associated with self-reflection when participants made judgments of agency compared with when they made judgments of performance, we examined neural responses in the aPFC ROI during the judgment phase of the task for judgment of agency trials compared with judgment of performance trials. The results of the analysis revealed a significant area of activation (−20, 50, 20; $Z = 3.09$; $K_E = 100$; $p < .05$, cluster corrected) that was part of a larger cluster ($K_E = 261$) in the left aPFC (see Figure 4). In addition, a cluster was found in the right OFC (i.e., BA 11) slightly outside the aPFC ROI (18, 46, −12; $K_E = 97$; $p < .0001$, uncorrected). These activations suggest that judgments of agency may involve a greater degree of self-referential thought than typical metacognitive judgments. The whole-brain analyses for the contrasts of *judgment of agency* > *judgment of performance* and *judgment of performance* > *judgment of agency* did not reveal any suprathreshold brain regions.

Parametric Model

Regions during the game phase that were negatively correlated with sense of agency. The analysis examining the particular brain regions whose activity during the game phase was negatively correlated with participants' reported *sense of agency* (i.e., judgments of agency minus hit rate) while controlling for actual performance did not yield any suprathreshold clusters of activation. However, small clusters of activity were found in the right (62, −36, 18; $K_E = 26$; $p < .005$, uncorrected) and left sides (−50, −40, 24; $K_E = 12$; $p < .005$, uncorrected) of the TPJ ROI. In addition, a conjunction between the uncorrected map from the whole-brain analysis ($p < .005$) and the results of the *turbulence* > *no turbulence* contrast revealed that these clusters were part of common activity in the right (64, −36, 16; $K_E = 52$; $p < .005$, uncorrected) and left TPJ (−46, −38, 18; $K_E = 85$; $p < .005$, uncorrected). Thus, in accordance with previous findings, it appears that activity in the TPJ is associated with disruptions of control.

As a control, a similar analysis of the regions that were negatively correlated with the difference between participants' judgments of performance and their hit rate yielded a significant cluster in the medial precuneus/posterior cingulate cortex. And although a small cluster of activity in the TPJ ROI was also found (62, −20, 30, $K_E = 41$; $p < .005$, uncorrected), a conjunction with the *turbulence* > *no turbulence* contrast showed that this cluster was not part of common activity in the TPJ (the conjunction did, however, show a small cluster of

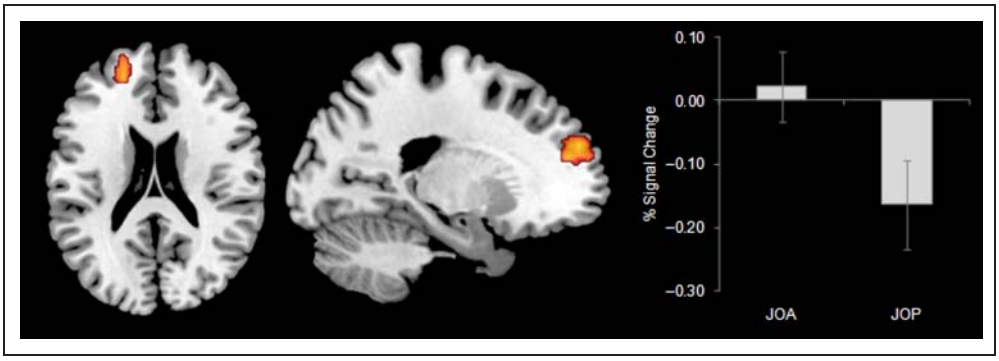

Figure 4. Cluster of activation during the judgment phase of the task that overlapped with the aPFC ROI when participants made judgments of agency, relative to when they made judgments of performance (−20, 50, 20, $Z = 3.09$; $K_E = 100$; $p < .05$, cluster corrected across the small volume). Percent signal change was computed for individual participants using MarsBaR v0.42 (marsbar.sourceforge.net). Error bars reflect *SEM*s.

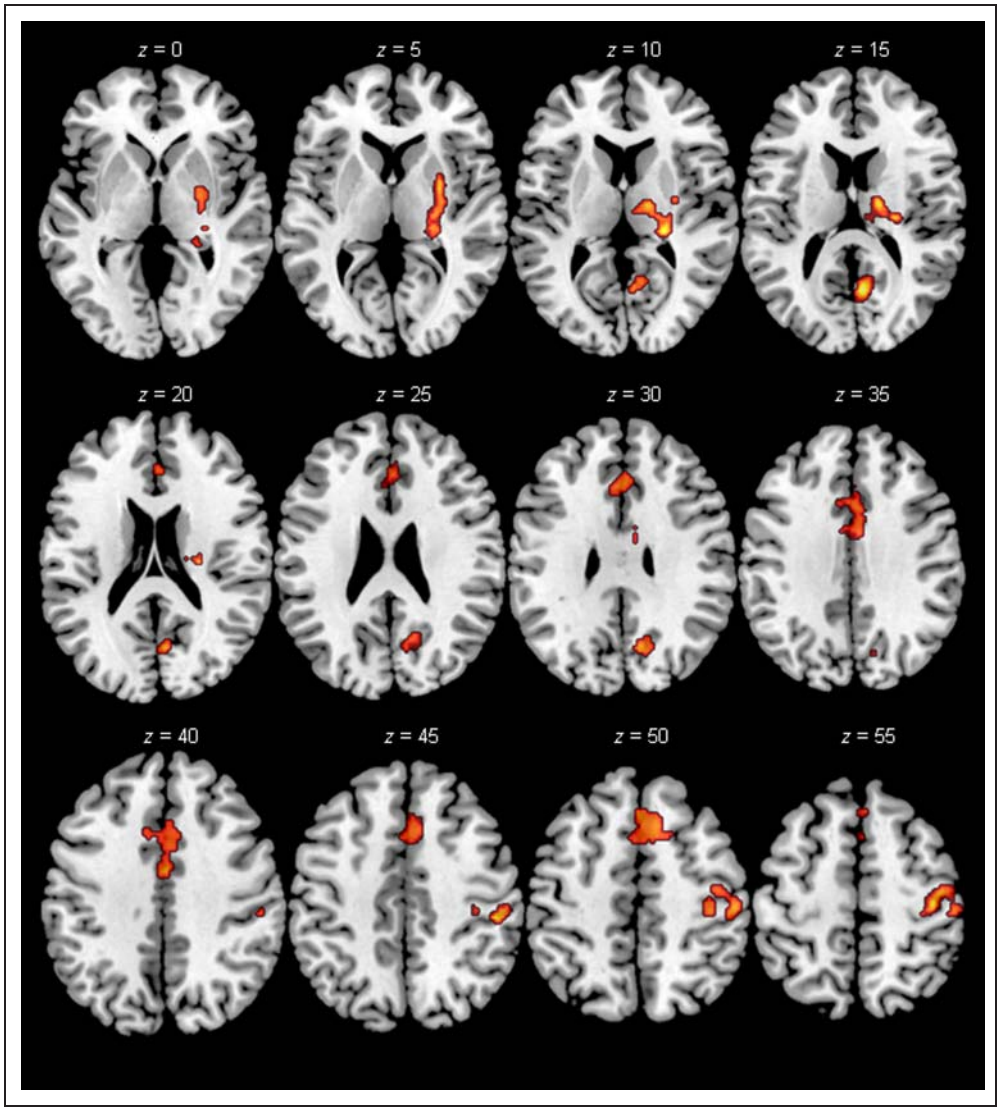

Figure 5. Clusters of activation during the game phase of the task that were positively correlated with participants' reported sense of agency. See Table 3 for a list of activation peaks.

common activity in a neighboring region; 64, −36, 12; $K_E = 21$; $p < .005$, uncorrected).

Regions during the game phase that were positively correlated with sense of agency. The analysis examining the particular brain regions whose activity during the game phase was positively correlated with participants' reported *sense of agency* while controlling for actual performance yielded significant clusters of activation in the right putamen/thalamus, bilateral rostral cingulate zone/pre-SMA, right pre/postcentral gyrus, and right calcarine/precuneus (see Figure 5 and Table 3). A similar analysis of the regions that were positively correlated with the difference between judgments of performance and hit rate (which again served as a control) did not yield any suprathreshold clusters.

DISCUSSION

In the present study, we dissociated the neural correlates of two distinct components of people's sense of agency: the action monitoring that people engage in as they perform an action and the metacognitive judgments of agency that they make when retrospectively assessing the extent of their control. These results are consistent with a model of agency (Metcalfe et al., 2010; Haggard & Tsakiris, 2009; Synofzik et al., 2008; Metcalfe & Greene, 2007; Wegner et al., 2004; Wegner, 2002, 2003; Georgieff & Jeannerod, 1998) that says that people form their metacognitive judgments of agency by reflecting on the output of their action monitoring, as well as other relevant cues, and then consciously inferring the extent to which they caused the action outcomes in question. Previous studies, like the present study, used judgments of agency to identify neural activity that is associated with action monitoring; however, unlike the present study, they did not attempted to determine whether there are specific patterns of neural activity associated with the judgments of agency themselves. The present study addressed this gap in the literature by using an agency paradigm that not only allowed us to examine the neural

Table 3. Suprathreshold Clusters for the Parametric Analyses of Game Phase Activity

Side	*Anatomical Region of Cluster*	*Cluster Size*	*Cluster Maxima (MNI)*			*Z Score*
Negatively Correlated with Sense of Agency: ns						
Negatively Correlated with Sense of Performance						
R&L	Medial precuneus, posterior cingulate cortex	402	−10	−42	8	3.45
			−6	−48	16	3.39
			2	−42	14	3.13
Positively Correlated with Sense of Agency						
R	Putamen, thalamus	566	24	−28	10	3.89
			14	−12	12	3.81
			28	0	4	3.74
R&L	Rostral cingulate zone, pre-SMA	634	2	−2	40	3.25
			8	18	48	3.17
			2	26	50	3.10
R	Pre/postcentral gyrus	301	48	−26	46	3.71
			40	−20	54	3.63
			44	−12	64	3.23
R	Calcarine, precuneus	252	6	−62	14	4.11
			10	−68	28	4.36
			14	−60	26	2.98
Positively Correlated with Sense of Performance: ns						

activity associated with the presence or absence of disruptions of control during the performance of a motor task but that also enabled us to determine which brain regions were active after the motor task had been completed and participants reflected back on how much control they had had.

Disruptions of Control

The results of both the simple contrast and parametric analyses showed increased activity in bilateral regions of the TPJ when the consequences of participants' actions were not in line with what they intended or expected them to be (e.g., when turbulence caused the cursor to move to the left after a participant moved the mouse to the right). An ROI analysis showed that these regions overlapped with brain areas identified in a meta-analysis of similar contrasts from previous agency studies, as shown in Figure 3.

One explanation for the role of the TPJ in agency processing comes from the comparator model of action monitoring, which proposes a mechanism by which the predicted consequences of an intended action are compared with sensory feedback about the actual consequences of the action (Hohwy & Frith, 2004; Blakemore et al., 2002; Wolpert & Ghahramani, 2000; cf. Synofzik et al., 2008). When the predicted and actual consequences do not match, the comparator sends a signal to the control mechanism responsible for executing motor corrections indicating that there has been a disruption of control. Because activity in the TPJ and its surrounding areas tends to be associated with disruptions of control, it is possible that this region is responsible for either monitoring the comparator signal or for executing corresponding motor corrections. Research on motor planning in monkeys (e.g., Mulliken, Musallam, & Andersen, 2008) indicates that the posterior parietal cortex (which is adjacent to the TPJ) uses multisensory input (including visual, auditory, and vestibular information) and efference copies of motor commands to predict the outcomes of self-initiated actions (e.g., the movement of a limb or the expected position of a cursor). Furthermore, TMS and lesion studies in humans (e.g., MacDonald & Paus, 2003) demonstrate that the

posterior parietal cortex plays a causal role in detecting unexpected outcomes of self-generated movements. These and other findings provide converging evidence that increased TPJ activity may be associated with detecting mismatches between forward estimates and sensory feedback. However, the possibility that TPJ activity is also associated with executing motor corrections cannot be entirely ruled out. Although a number of previous agency studies showed that TPJ activation was present during monitoring when control was held constant (e.g., when participants merely observed disruptions of control and were not given the opportunity to perform corrective actions; e.g., Spengler et al., 2009), none of these studies attempted to determine whether this TPJ activity was absent during control when monitoring was held constant. Furthermore, because the game phase of each trial in the present experiment involved a continuous stream of disruptions and corrective movements, our study is also unable to distinguish between these two interpretations of the observed TPJ activity.

Being in Control

Although previous agency studies (e.g., Tsakiris et al., 2010; Farrer et al., 2003) have occasionally identified patterns of brain activity that are associated with increases in control, these patterns have not been consistent across studies—suggesting, perhaps, that feeling in control is really just the absence of feeling out of control (Hohwy, 2007). However, the lack of a consistent neural signature for "being in control" might have been due to the fact that prior studies did not control for variation in task performance (which tends to be strongly correlated with perceived agency). The results of the parametric analysis we conducted, which did control for task performance, yielded significant clusters of activation in the right putamen/thalamus, bilateral rostral cingulate zone/pre-SMA, right pre/postcentral gyrus, and right calcarine/precuneus (see David et al., 2007, for a similar pattern of results). It is of interest and, perhaps, even intuitive that these are the same regions that have been indentified in studies of smoothly executed, self-initiated actions (Boecker, Jankowski, Ditter, & Scheef, 2008; Soon, Brass, Heinze, & Haynes, 2008; Walton, Devlin, & Rushworth, 2004; Cunnington, Windischberger, Deecke, & Moser, 2002; see Nachev, Kennard, & Husain, 2008, for a review).

Metacognitive Judgments of Agency: Self-reflective Processing

The primary new finding of the present study is that the left aPFC and right OFC increased in activation when participants made judgments of agency compared with when they made judgments of performance. Prior research on self-attribution and self-reflection indicates that these areas are part of cortical midline structures that subserve self-reflective processing. According to recent meta-analyses and theoretical reviews of published fMRI and PET studies on self-reflection, these structures can be segregated into several functionally distinct regions (van der Meer et al., 2010; Schmitz & Johnson, 2007; Northoff et al., 2006). The most ventral region, which includes portions of the aPFC, OFC, and the pACC, receives projections from a number of sensory and emotional processing centers such as the limbic system and the striatum (Ongur & Price, 2000) and, thus, is thought to be involved in the appraisal of sensory and affective stimuli in terms of their self-relevance (Ochsner et al., 2004; Zysset, Huber, Ferstl, & von Cramon, 2002; see Amodio & Frith, 2006, for a review). The more dorsal midline structures,[2] including the dorsomedial PFC (dMPFC), are tightly connected to areas of the lateral PFC that are involved in high-level reasoning and decision-making and, therefore, are believed to underlie the conscious formation of inferences and judgments based on information that may have been tagged as self-relevant (or not) by the aPFC/OFC. Finally, the posterior midline structures are considered part of a hippocampal network that is implicated in the encoding and retrieval of autobiographical memories and, thus, are thought to be responsible for putting self-relevant information within a temporal context that is informed by prior experience. The location of the brain activity observed for judgments of agency compared with judgments of performance suggests that judgments of agency may involve more self-related sensory-affective processing compared with other types of metacognitive judgments. This is consistent with a number of theories of agency (e.g., Haggard & Tsakiris, 2009; Synofzik et al., 2008; Metcalfe & Greene, 2007).

An alternative explanation for the present results could be that activity in the aPFC/OFC reflects a more general process of evaluating information about individuals, regardless of whether or not the information pertains to the self (see Amodio & Frith, 2006). Although a number of studies have shown that inferences about other people's mental states and personality traits are typically associated with activity in the dMPFC, some of these studies have shown that such *non*-self-related inferences occasionally activate the aPFC/OFC as well (see Mitchell, Macrae, & Banaji, 2006, for a review). It is, however, important to note that in many of the studies that reported aPFC/OFC activation, participants were asked to mentalize about friends and similar others as opposed to strangers. A recent study (Mitchell et al., 2006) that specifically examined this distinction suggests that mentalizing about close or similar others is informed by knowledge about one's own thoughts and feelings, which implies that the medial aPFC/OFC is indeed dedicated to processing *self*-relevant information. Additional support for this conclusion comes from a recent meta-analysis of studies in which judgments about the self were contrasted with judgments about others (van der Meer et al., 2010), which showed that activity in the aPFC/OFC is more strongly associated with judgments about the self.

Finally, a study by Powell, Macrae, Cloutier, Metcalfe, and Mitchell (2009) seems, at first glance, to be inconsistent

with the present results. Participants in this study completed a self-reflection task, in which they judged their own or another person's personality traits, as well as an agency task in which they freely chose words to study or watched passively as words were selected for them by the computer. Although the results showed increased activation in the aPFC when participants judged their own traits as opposed to another's, no ventral activity was found when they actively chose words to study (compared with when the computer chose for them). Instead, freely choosing words was associated with increased activity in the intraparietal sulcus. However, because choosing is quite different from retrospectively assessing how much control one had (i.e., it is an act and not a reflection), the fact that the choice task was not associated with aPFC/OFC activity may actually be consistent with the present findings. In fact, considering that the intraparietal sulcus overlaps with the TPJ, it appears that making voluntary choices may have more to do with action and action monitoring than with metacognition of agency.

Conclusion

Although a number of studies have attempted to identify the brain regions that underlie the on-line monitoring of action, ours is the first to dissociate such low-level action monitoring from metacognitive assessments of agency. Although the results presented here were consistent with previous results on action monitoring, it is not action monitoring, but rather metacognition of agency, that is implicated in the attributions that underlie our feelings of responsibility for our own actions. The neural underpinnings of these metacognitions of agency have not been previously demonstrated. We found—by contrasting the neural activity associated with judging agency to the activity associated with a different type of metacognitive judgment (i.e., judging performance)—that there is a specific area (an area distinct from the regions involved in action monitoring and related instead to self-referential processing) that is central to people's conscious, retrospective judgments of their own agency.

Acknowledgments

This research was supported by an Academic Quality Fund Grant from the Provost of Columbia University and by grant 220020166 from the James S. McDonnell Foundation to J. M. We would like to thank Matt Greene, Spiro Pantazatos, Stephen Dashnaw, and Joy Hirsch.

Reprint requests should be sent to Janet Metcalfe, Department of Psychology, Columbia University, New York, NY 10027, or via e-mail: jm348@columbia.edu.

Notes

1. We excluded five studies (see Table 1) that were originally included in the meta-analysis by Decety and Lamm (2007) because they did not fit with our definition of agency (e.g., we did not include studies that involved perceiving agency in the actions of others). In addition, we included activation peaks from the left side of the brain, whereas Decety and Lamm did not, because most of the studies included in the meta-analysis reported bilateral TPJ activation.
2. Because the boundary between the ventral and dMPFC is not strictly defined, it is worth noting that we use the term "dorsal" to refer to areas that are above a z coordinate of 20, which roughly corresponds to BA 9. This usage is in keeping with cited reviews of self-reflective processing.

REFERENCES

Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: The medial frontal cortex and social cognition. *Nature Reviews Neuroscience, 7,* 268–277.

Balslev, D., Nielsen, F. A., Lund, T. E., Law, I., & Paulson, O. B. (2006). Similar brain networks for detecting visuo-motor and visuo-proprioceptive synchrony. *Neuroimage, 31,* 308–312.

Blakemore, S. J., Frith, C. D., & Wolpert, D. M. (1999). Spatio-temporal prediction modulates the perception of self-produced stimuli. *Journal of Cognitive Neuroscience, 11,* 551–559.

Blakemore, S. J., Frith, C. D., & Wolpert, D. M. (2001). The cerebellum is involved in predicting the sensory consequences of action. *NeuroReport, 12,* 1879–1884.

Blakemore, S. J., Wolpert, D. M., & Frith, C. D. (2002). Abnormalities in the awareness of action. *Trends in Cognitive Science, 6,* 237–242.

Boecker, H., Jankowski, J., Ditter, P., & Scheef, L. (2008). A role of the basal ganglia and midbrain nuclei for initiation of motor sequences. *Neuroimage, 39,* 1356–1369.

Chaminade, T., & Decety, J. (2002). Leader or follower? Involvement of the inferior parietal lobule in agency. *NeuroReport, 13,* 1975–1978.

Christoff, K., & Gabrielli, J. D. E. (2000). The frontopolar cortex and human cognition: Evidence for a rostrocaudal hierarchical organization within the human prefrontal cortex. *Psychobiology, 28,* 168–186.

Cunnington, R., Windischberger, C., Deecke, L., & Moser, E. (2002). The preparation and execution of self-initiated and externally-triggered movement: A study of event-related fMRI. *Neuroimage, 15,* 373–385.

David, N., Bewernick, B. H., Cohen, M. X., Newen, A., Lux, S., Fink, G. R., et al. (2006). Neural representations of self versus other: Visual-spatial perspective taking and agency in a virtual ball-tossing game. *Journal of Cognitive Neuroscience, 18,* 898–910.

David, N., Cohen, M. X., Newen, A., Bewernick, B. H., Shah, N. J., Fink, G. R., et al. (2007). The extrastriate cortex distinguishes between the consequences of one's own and others' behavior. *Neuroimage, 36,* 1004–1014.

David, N., Newen, A., & Vogeley, K. (2008). The "sense of agency" and its underlying cognitive and neural mechanisms. *Consciousness and Cognition, 17,* 523–534.

Decety, J., Chaminade, T., Grezes, J., & Meltzoff, A. N. (2002). A PET exploration of the neural mechanisms involved in reciprocal imitation. *Neuroimage, 15,* 265–272.

Decety, J., & Lamm, C. (2007). The role of the right temporoparietal junction in social interaction: How low-level computational processes contribute to meta-cognition. *Neuroscientist, 13,* 580–593.

Farrer, C., Franck, N., Frith, C. D., Decety, J., Georgieff, N., d'Amato, T., et al. (2004). Neural correlates of action attribution in schizophrenia. *Psychiatry Research, 131,* 31–44.

Farrer, C., Franck, N., Georgieff, N., Frith, C. D., Decety, J., & Jeannerod, M. (2003). Modulating the experience of

agency: A positron emission tomography study. *Neuroimage, 18,* 324–333.
Farrer, C., Frey, S. H., Van Horn, J. D., Tunik, E., Turk, D., Inati, S., et al. (2008). The angular gyrus computes action awareness representations. *Cerebral Cortex, 18,* 254–261.
Farrer, C., & Frith, C. D. (2002). Experiencing oneself vs. another person as being the cause of an action: The neural correlates of the experience of agency. *Neuroimage, 15,* 596–603.
Fink, G. R., Marshall, J. C., Halligan, P. W., Frith, C. D., Driver, J., Frackowiak, R. S., et al. (1999). The neural consequences of conflict between intention and the senses. *Brain, 122,* 497–512.
Franck, N., Farrer, C., Georgieff, N., Marie-Cardine, M., Dalery, J., d'Amato, T., et al. (2001). Defective recognition of one's own actions in patients with schizophrenia. *American Journal of Psychiatry, 158,* 454–459.
Friston, K., Worsley, K., Frackowiak, R., Mazziotta, J., & Evans, A. (1993). Assessing the significance of focal activations using their spatial extent. *Human Brain Mapping, 1,* 210–220.
Friston, K. J., Holmes, A. P., Poline, J. B., Grasby, P. J., Williams, S. C., Frackowiak, R. S., et al. (1995). Analysis of fMRI time-series revisited. *Neuroimage, 2,* 45–53.
Georgieff, N., & Jeannerod, M. (1998). Beyond consciousness of external reality. A "Who" system for consciousness of action and self-consciousness. *Consciousness and Cognition, 7,* 465–477.
Haggard, P., & Tsakiris, M. (2009). The experience of agency. *Current Directions in Psychological Science, 18,* 242–246.
Hohwy, J. (2007). The sense of self in the phenomenology of agency and perception. *Psyche, 13.* Retrieved from www.theassc.org/files/assc/2668.pdf.
Hohwy, J., & Frith, C. (2004). Can neuroscience explain consciousness? *Journal of Consciousness Studies, 11,* 180–198.
Jenkins, A. C., & Mitchell, J. P. (2011). Medial prefrontal cortex subserves diverse forms of self-reflection. *Social Neuroscience, 6,* 211–218.
Kable, J. W., & Chatterjee, A. (2006). Specificity of action representations in the lateral occipitotemporal cortex. *Journal of Cognitive Neuroscience, 18,* 1498–1517.
Kirkpatrick, M., Metcalfe, J., Greene, M., & Hart, C. (2008). Effects of intranasal methamphetamine on metacognition of agency. *Psychopharmacology, 197,* 137–144.
Leube, D. T., Knoblich, G., Erb, M., Grodd, W., Bartels, M., & Kircher, T. T. (2003). The neural correlates of perceiving one's own movements. *Neuroimage, 20,* 2084–2090.
Lorch, R., & Myers, J. (1990). Regression analyses of repeated measures data in cognitive research. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 149–157.
MacDonald, P. A., & Paus, T. (2003). The role of parietal cortex in awareness of self-generated movements: A transcranial magnetic stimulation study. *Cerebral Cortex, 13,* 962–967.
Metcalfe, J., Eich, T. S., & Castel, A. D. (2010). Metacognition of agency across the lifespan. *Cognition, 116,* 267–282.
Metcalfe, J., & Greene, M. J. (2007). Metacognition of agency. *Journal of Experimental Psychology: General, 136,* 184–199.
Mitchell, J. P. (2009). Social psychology as a natural kind. *Trends in Cognitive Science, 13,* 246–251.
Mitchell, J. P., Macrae, C. N., & Banaji, M. R. (2006). Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. *Neuron, 50,* 655–663.
Mulliken, G. H., Musallam, S., & Andersen, R. A. (2008). Forward estimation of movement state in posterior parietal cortex. *Proceedings of the National Academy of Sciences, U.S.A., 105,* 8170–8177.
Mumford, J. A., & Nichols, T. (2009). Simple group fMRI modeling and inference. *Neuroimage, 47,* 1469–1475.
Nachev, P., Kennard, C., & Husain, M. (2008). Functional role of the supplementary and pre-supplementary motor areas. *Nature Reviews Neuroscience, 9,* 856–869.
Nahab, F. B., Kundu, P., Gallea, C., Kakareka, J., Pursley, R., Pohida, T., et al. (2011). The neural processes underlying self-agency. *Cerebral Cortex, 21,* 48–55.
Northoff, G., Heinzel, A., de Greck, M., Bermpohl, F., Dobrowolny, H., & Panksepp, J. (2006). Self-referential processing in our brain—A meta-analysis of imaging studies on the self. *Neuroimage, 31,* 440–457.
Ochsner, K. N., Ray, R. D., Cooper, J. C., Robertson, E. R., Chopra, S., Gabrieli, J. D., et al. (2004). For better or for worse: Neural systems supporting the cognitive down- and up-regulation of negative emotion. *Neuroimage, 23,* 483–499.
Ongur, D., & Price, J. L. (2000). The organization of networks within the orbital and medial prefrontal cortex of rats, monkeys and humans. *Cerebral Cortex, 10,* 206–219.
Pacherie, E. (in press). Sense of agency: Many facets, multiple sources. In H. S. Terrace & J. Metcalfe (Eds.), *Agency and joint attention*. Oxford, UK: Oxford University Press.
Pisella, L., Grea, H., Tilikete, C., Vighetto, A., Desmurget, M., Rode, G., et al. (2000). An "automatic pilot" for the hand in human posterior parietal cortex: Toward reinterpreting optic ataxia. *Nature Neuroscience, 3,* 729–736.
Powell, L. J., Macrae, C. N., Cloutier, J., Metcalfe, J., & Mitchell, J. P. (2009). Dissociable neural substrates for agentic versus conceptual representations of self. *Journal of Cognitive Neuroscience, 22,* 2186–2197.
Ramnani, N., & Miall, R. C. (2004). A system in the human brain for predicting the actions of others. *Nature Neuroscience, 7,* 85–90.
Ramnani, N., & Owen, A. M. (2004). Anterior prefrontal cortex: Insights into function from anatomy and neuroimaging. *Nature Reviews Neuroscience, 5,* 184–194.
Ruby, P., & Decety, J. (2001). Effect of subjective perspective taking during simulation of action: A PET investigation of agency. *Nature Neuroscience, 4,* 546–550.
Saxe, R., Xiao, D. K., Kovacs, G., Perrett, D. I., & Kanwisher, N. (2004). A region of right posterior superior temporal sulcus responds to observed intentional actions. *Neuropsychologia, 42,* 1435–1446.
Schmitz, T. W., & Johnson, S. C. (2007). Relevance to self: A brief review and framework of neural systems underlying appraisal. *Neuroscience & Biobehavioral Reviews, 31,* 585–596.
Schnell, K., Heekeren, K., Schnitker, R., Daumann, J., Weber, J., Hesselmann, V., et al. (2007). An fMRI approach to particularize the frontoparietal network for visuomotor action monitoring: Detection of incongruence between test subjects' actions and resulting perceptions. *Neuroimage, 34,* 332–341.
Soon, C. S., Brass, M., Heinze, H. J., & Haynes, J. D. (2008). Unconscious determinants of free decisions in the human brain. *Nature Neuroscience, 11,* 543–545.
Spence, S. A., Brooks, D. J., Hirsch, S. R., Liddle, P. F., Meehan, J., & Grasby, P. M. (1997). A PET study of voluntary movement in schizophrenic patients experiencing passivity phenomena (delusions of alien control). *Brain, 120,* 1997–2011.
Spengler, S., von Cramon, D. Y., & Brass, M. (2009). Was it me or was it you? How the sense of agency originates from ideomotor learning revealed by fMRI. *Neuroimage, 46,* 290–298.
Synofzik, M., Vosgerau, G., & Newen, A. (2008). Beyond the comparator model: A multifactorial two-step account of agency. *Conscious and Cognition, 17,* 219–239.
Tricomi, E. M., Delgado, M. R., & Fiez, J. A. (2004). Modulation of caudate activity by action contingency. *Neuron, 41,* 281–292.
Tsakiris, M., Longo, M. R., & Haggard, P. (2010). Having a body versus moving your body: Neural signatures of agency and body-ownership. *Neuropsychologia, 48,* 2740–2749.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., et al. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage, 15,* 273–289.

van der Meer, L., Costafreda, S., Aleman, A., & David, A. (2010). Self-reflection and the brain: A theoretical review and meta-analysis of neuroimaging studies with implications for schizophrenia. *Neuroscience & Biobehavioral Reviews, 34,* 935–946.

Vinogradov, S., Luks, T. L., Simpson, G. V., Schulman, B. J., Glenn, S., & Wong, A. E. (2006). Brain activation patterns during memory of cognitive agency. *Neuroimage, 31,* 896–905.

Wager, T. D., Lindquist, M. A., Nichols, T. E., Kober, H., & Van Snellenberg, J. X. (2009). Evaluating the consistency and specificity of neuroimaging data using meta-analysis. *Neuroimage, 45,* S210–S221.

Walton, M., Devlin, J., & Rushworth, M. (2004). Interactions between decision making and performance monitoring within prefrontal cortex. *Nature Neuroscience, 7,* 1259–1265.

Wegner, D. M. (2002). *The illusion of conscious will*. Cambridge, MA: MIT Press.

Wegner, D. M. (2003). The mind's best trick: How we experience conscious will. *Trends in Cognitive Sciences, 7,* 65–69.

Wegner, D. M., Sparrow, B., & Winerman, L. (2004). Vicarious agency: Experiencing control over the movements of others. *Journal of Personality and Social Psychology, 86,* 838–848.

Williams, J. H., Waiter, G. D., Gilchrist, A., Perrett, D. I., Murray, A. D., & Whiten, A. (2006). Neural mechanisms of imitation and "mirror neuron" functioning in autistic spectrum disorder. *Neuropsychologia, 44,* 610–621.

Wolpert, D. M., & Ghahramani, Z. (2000). Computational principles of movement neuroscience. *Nature Neuroscience, 3(Suppl.),* 1212–1217.

Worsley, K. J., Marrett, S., Neelin, P., Vandal, A. C., Friston, K. J., & Evans, A. C. (1996). A unified statistical approach for determining significant signals in images of cerebral activation. *Human Brain Mapping, 4,* 58–73.

Yomogida, Y., Sugiura, M., Sassa, Y., Wakusawa, K., Sekiguchi, A., Fukushima, A., et al. (2010). The neural basis of agency: An fMRI study. *Neuroimage, 50,* 198–207.

Zysset, S., Huber, O., Ferstl, E., & von Cramon, D. Y. (2002). The anterior frontomedian cortex and evaluative judgment: An fMRI study. *Neuroimage, 15,* 983–991.