An exploratory fMRI study into inferences of self-agency

Robert A. Renes¹², Neeltje E. M. van Haren², Henk Aarts¹, Matthijs Vink²

¹Utrecht University, Utrecht, the Netherlands
²Brain Center Rudolf Magnus, Department of Psychiatry, University Medical Center Utrecht, Utrecht, the Netherlands

Running head: Neural correlates of self-agency inferences

Word count: 3082

Abstract: 124
Figures: 4
Tables: 1

Address correspondence to Robert A. Renes, Department of Psychology, Utrecht University, Heidelbergraan 1, 3584 CS Utrecht, The Netherlands; E-mail: R.A.Renes@uu.nl; Telephone +3130-253-9032.
Abstract

Building on the recent findings that the experience of self-agency over actions and corresponding outcomes can also rely on cognitive inferential processes, rather than motor prediction processes, this study aims to investigate the brain areas involved in agency inference processing in a setting where action and outcome are independent. Twenty-three right-handed subjects were scanned using functional MRI while performing an agency-inference task, in which action-outcomes matched or mismatched goals. The experience of self-agency was associated with increased activation in the inferior parietal lobule as well as bilateral (medial) superior frontal cortex and medial prefrontal cortex. These findings provide new and exciting insights in the processing of inferential self-agency, providing a first look at the neural correlates of self-agency processing independent of motor-prediction processes.

Keywords: Self-Agency; Inference; Matching; fMRI; Medial Prefrontal Cortex
**Introduction**

Considering yourself as the cause of your behavior is fundamental to self-awareness and social interaction. For instance, when you press a key on a piano and the intended sound is produced, you experience yourself as the author of this effect. This sense of agency over operant action – i.e., when an action is followed by a specific and anticipated consequence – has primarily been explained by comparator processes described in models on motor control (Blakemore et al. 2002; Farrer et al. 2003; Frith et al. 2000; Wolpert and Flanagan 2001). These models often rely on paradigms in which visual, tactile or auditory feedback of an individual movement is manipulated (e.g., Sperduti et al. 2011; David 2012). The volitional or goal-directed execution of the action is accompanied by the prediction of sensory action-outcomes based on internal copies of movement-predicting signals (i.e., efference copies) generated by the motor system. Because these internal motor predictions are generally very reliable, sensory outcomes are readily perceived as self-produced until this prediction no longer corresponds with the actual outcomes following one’s action (Wolpert et al 1995). This misprediction of motor processes in a self-agency context has been found to be associated with brain activity in various areas, including the superior temporal gyrus, the inferior parietal lobe, as well as motor regions such as the pre-supplementary motor area and the cerebellum (for an overview, see Sperduti et al. 2011).

While one’s sense of agency is decreased when efference copies do not predict sensory outcomes, there is research to suggest that this does not necessarily always occur. People can experience self-agency even when moving involuntarily (Dogge et al. 2012; Moore et al. 2009), when there is no clear causal relationship between an action and a following event (Moore and Haggard 2008), or when actions have
multiple causes and outcomes (Van der Weiden et al. 2013; Wegner 2002). For instance, when two persons press a separate light switch and one light bulb turns on, they cannot exclusively rely on motor predictions to generate the sense of agency over the event. This recent work points to the existence of an additional – cognitive – route to agency experiences that may result from non-motor prediction cues. Specifically, it has been proposed that agency experiences can also arise from the inferred correspondence between actual action-effects and pre-activated knowledge of these outcomes, even though action and outcome are independent (Moore and Haggard 2008; Wegner 2002). In contrast to motor predictions, research on the neural components of cognitive inferences as agentive cues has received little attention. The present study aims to offer an initial test to fill this void.

According to the inference model, people quickly and fluently infer whether or not a perceived event results from their behavior. Experienced self-agency emerges when the perception of an outcome corresponds with the outcome that one consciously intends to attain by performing an action, while a mismatch is generally ascribed to other causes (Wegner 2002; Synofzik et al. 2013).

In a test of this idea (Aarts et al. 2005; Renes et al. 2013), participants moved a single gray square in a counterclockwise direction on a computer screen while the computer moved another gray square in the opposite (clockwise) direction. Participants were instructed to stop their square with a key-press at one of eight possible locations. When they pressed a key, they saw one of the two squares stopped at a specific location. Thus, the participants or the computer could have caused the observed stop, and the observed stop could not be predicted by the key-press. Participants indicated whether they or the computer had caused the square to stop at that position. In actuality, the computer always determined where the square stopped,
so action and outcome were independent and participants had no control. Enhanced self-agency was reported when their intended outcome matched the actual outcome as compared to when it did not. Since the experimental set-up did not allow the motor prediction process to produce reliable input for establishing a sense of agency, the agency experiences have likely resulted from inference processes based on the congruency between the intended outcome and the actual outcomes (Synofzik et al. 2013).

Whereas research on neural processes of self-agency inferences in action-outcome independent settings is lacking, there is some prior research on other types of inferences that provides clues about these processes. A recent meta-analysis on the neural substrates involved in the observation of other people’s behavior suggests that the mirror system, consisting of the anterior intraparietal sulcus and the premotor cortex, is engaged when one perceives the motions of body parts of another person (van Overwalle and Baetens 2009). Furthermore, the mentalizing system, consisting of the medial prefrontal cortex and the temporo-parietal junction, seems to be implicated when the observation of another person’s behavior allows the observer to make inferences about goals and beliefs that might underlie the other agent’s behavior. Interestingly, unexpected motions of body parts (e.g., movements that do not fit with the predicted action-effects) seem to activate both the mirror and mentalizing system, suggesting that if the mirror system is unable to deal with the unexpected motion on inconsistency, the mentalizing system is activated to understand the occurrence of the error (van Overwalle and Baetens 2009).

The present study aims to explore brain activation related to agency inferences of an individual’s own behavior in a context where action and outcomes are independent and thus motor processes do not relate to or predict outcomes. Thus, we
address the experience of self-agency as they occur when intended action-outcomes or goals match or mismatch with the actual outcomes. In line with research on the neural foundations of mentalizing, there is research to suggest that the medial prefrontal cortex is also implicated in self-referential processing (Amodio & Frith 2006; Mason et al. 2007; Northhoff et al. 2006; Van Buuren et al. 2010). Accordingly, as agency experiences over one’s own behavior are self-referential, it is likely that the medial prefrontal cortex is also involved in self-agency inferences. Furthermore, as these inferences are likely to follow from a comparison of intended outcomes and actual outcomes, specific brain areas in the parietal cortex (such as the inferior parietal lobule) that are associated with such comparisons (Farrer and Frith 2002; Spence 2002; O’Connor et al. 2010) may also be involved in the neural base of self-agency experiences.

Based on this earlier work, the present study thus aims to explore whether medial prefrontal and parietal areas are recruited to infer self-agency over the consequences of action that either match or mismatch with a person’s goal to produce the outcome. For this purpose we used the agency inference task in an operant action performance context as described above. Simultaneously, brain activation related to the observation of the outcome is measured by using fMRI.

**Methods**

**Participants**

Twenty-three right-handed subjects (11 women; mean age = 21.7±2.6 SD) participated in the study. Participants were recruited from the University of Utrecht and received monetary compensation for participation. The study was performed at the University Medical Center Utrecht. Participants were excluded from participation...
in case of alcohol and drug abuse/dependence during the last six months, the presence of a psychiatric disorder in the participant or a first-degree relative of the participant, and in case of a serious medical condition. After detailed explanation of the study design and any potential risks, all subjects gave their written informed consent. The ethics committee of the University Medical Center Utrecht approved this study.

Agency inference task and procedure

The agency inference task (Fig. 1) was adapted from Aarts et al. (2005). Eight white tiles were presented on the screen. At the beginning of each trial, one of these eight tiles turned black, indicating this specific location as the outcome goal of this trial. Next, participants pressed a button on an MRI-compatible fiber-optic response box to cause a square to move along those white tiles in a counter-clockwise direction. The computer independently moved another square at the same speed, but in the opposite direction. When the stop cue appeared on the screen, participants had to press the button, thereby stopping the movement of the squares. However, as soon as the stop cue appeared, the squares disappeared. Participants were told that they continued to rotate invisibly until they pressed the button. Next, one of the eight white tiles turned black. This square represented either the participant’s square or the computer’s square thereby creating two potential causes of the outcome. Importantly, in half of the trials this outcome location corresponded with the location that was presented at the beginning of the trial (i.e. outcome goal), and in the other half of the trials it did not (i.e., the square stopped four locations farther away). Participants were asked to rate their level of self-agency experience on a nine-point horizontal visual analogue scale (not me (1) – me (9), see Fig. 2), reflecting the extent to which they felt they had caused the displayed square to stop on that particular position.
Prior to the fMRI session, participants practiced the task both outside and inside the scanner. The total task consisted of 64 trials: 32 trials in which the outcome matched the participant’s goal, 32 in which the outcome did not match the goal. For all participants, trials were presented in a single pseudo-random counterbalanced order. The four task blocks of 16 trials – each with 8 matching and 8 mismatching trials – were interleaved with 30s rest periods.

Image acquisition

The experiment was performed on a 3.0 T Philips Achieva MRI scanner (Philips Medical Systems, Best, the Netherlands) at the University Medical Center Utrecht. Head motion was restricted using a vacuum cushion and foam wedges. Images were acquired using an eight-channel sensitivity-encoding (SENSE) parallel-imaging head coil. Whole-brain T2*-weighted echo planar images (EPI) with blood-oxygen level-dependent (BOLD) contrast (625 volumes per task; 30 slices per volume; interleaved acquisition; repetition time, 1600 ms; echo time, 23.5 ms; field of view: 256×208 mm; flip angle = 72.5°; 64×51 matrix; 4×4 mm in-plane resolution; 4 mm slice thickness; SENSE-factor, 2.4 (anterior-posterior)) oriented in a transverse plane tilted 20° over the left-right axis were acquired in a single run. The first six images were dummy scans to allow for T1 equilibration effects. A whole-brain three-dimensional fast field echo T1-weighted scan (150 slices; repetition time =8.4 ms; echo time =3.8 ms; flip angle = 8°; field of view, 288×252×185 mm; voxel size: 1 mm isotropic) was acquired for within-subject registration purposes.

Imaging data pre-processing
Image preprocessing and analyses were carried out with SPM 5 (http://www.fil.ion.ucl.ac.uk/spm/). After realignment, the structural scan was co-registered to the mean functional scan. Next, using unified segmentation, the structural scan was segmented and normalization parameters were estimated. Subsequently, all scans were registered to a MNI T1-standard brain using these normalization parameters and a 3D Gaussian filter (8-mm full width at half maximum) was applied to all functional images.

**Data analysis**

For each subject, a model was generated describing event-related changes time-locked to the start of the trial (rotation of squares), the stop cue, the actual outcome, and the self-agency rating (self-agency experience\(^1\): rating > 5, no self-agency experience: rating < 5). Self-agency related activation was modeled as activation during the presentation of the actual outcome, based on the self-agency experience. This resulted in four conditions: outcome match and self-agency experience, outcome match and no self-agency experience, outcome mismatch and self-agency experience, outcome mismatch and no self-agency experience. To correct for head motion, the six realignment parameters were included in the design matrix as regressors of no interest. Two participants were excluded from further analysis due to excessive head movement (> 4mm) during the acquisition of fMRI scans (van Dijk et al. 2012). To correct for drifts in the signal, a high-pass filter (discrete cosine transform basis functions) was applied to the data with a cut-off frequency of 0.0039 Hz. For each

\(^1\) The middle of the 9-point scale is ambiguous as to the sense of agency. We therefore deemed it appropriate to not include these middle ratings in the fMRI analyses in order to create a clear contrast between self-agency experience and no self-agency experience.
individual subject, brain activation related to self-agency was calculated by contrasting match-agency trials with mismatch-no-agency trials (henceforth referred to as the Agency > No-Agency contrast). The resulting individual statistical maps were used to perform a group-wise whole-brain analysis. Maps resulting from this analysis were tested for significance using cluster-level inference (cluster-defining threshold, $p < 0.001$; critical cluster size: 27 voxels, cluster probability of $p < 0.05$, family-wise error corrected for multiple comparisons). These parameters were determined using SPM and a script (CorrClusTh.m, http://www2.warwick.ac.uk/fac/sci/statistics/staff/academic-research/nichols/scripts/spm), which uses estimated smoothness (estimated Full Width at Half Maximum (FWHM): 3.56x3.65x3.46 voxels) and Random Field Theory to find these corrected thresholds. We were unable to perform a full-factorial analysis as the factors match-no-agency and mismatch-agency contained almost no trials (see Fig. 3).

Results

Behavioral self-agency experiences

Self-agency ratings are presented in Fig. 3. A repeated-measures ANOVA yielded a significant main effect of matching ($F(1,22) = 71.50; p < 0.001, \eta_p^2 = .765$), indicating that participants experienced more self-agency when the actual outcome matched the outcome goal presented at the start of the trial ($M = 6.75, SD = 0.99$) as opposed to when it did not ($M = 3.31, SD = 1.19$).

Reaction times to the stop-cue
In order to check whether response time might influence the self-agency inference effects, we compared reaction times for the stop-cues in the match and mismatch condition (see Fig. 2). A paired-samples t-test revealed no difference ($t(22) = 0.896, p = .380$) between matching trials ($M = 411, SD = 149$) and mismatching trials ($M = 402, SD = 119$). Accordingly, participants did not react differently on matching trials as compared to mismatching trials.

**fMRI**

Due to the nature of self-agency inferences, we analyzed brain activations during the presentation of the outcome. Self-agency related activation (as calculated with the Agency > No-Agency contrast) was found in the parietal cortex (inferior parietal lobule), and in three clusters in the frontal cortex: the left (medial) superior frontal cortex, right (medial) superior frontal, and medial prefrontal cortex (see Fig. 4 and Table 1). No regions were associated with the opposite contrast (No-Agency > Agency).

**Discussion**

The present study used fMRI to reveal brain areas involved in self-agency inferences, i.e., when self-agency is retrospectively established in a situation where motor processes do not predict the outcomes. Our main finding is that self-agency inferences were associated with activation in the medial prefrontal cortex, bilateral (medial) superior frontal gyrus and the left inferior parietal lobule. Interestingly, these findings followed from the contrast test between Agency > No Agency, indicating that self-agency inferences are associated with the activation of these brain areas when the actual outcome matches the predicted outcome.
In contrast to findings in motor-prediction self-agency research, the present study shows bilateral (medial) superior frontal gyrus and medial prefrontal cortex activation while experiencing self-agency. These medial frontal regions have been found to be associated with thinking about self-agency judgments (Miele et al. 2011), self-referential processing (van Buuren et al. 2010) and broader inferential processing, such as trait inferences (Ma et al. 2011) and social inferences (van Overwalle and Baetens 2009). Moreover, lateral areas in the cortex have been related to higher-order processing of the self, such as awareness of the self (Northoff et al. 2006).

Furthermore, the left inferior parietal lobule has been associated with the integration of sensory information (Joseph 1982; Seghier 2013), lending credence to the idea that it can compare expected and actual outcomes. Interestingly, in the context of motor-prediction activation of the inferior parietal lobule has often been associated with a mismatch between the expected outcome and the actual outcome based on motor-sensory information (for a review, see Sperduti et al. 2011). However, in the present study the inferior parietal lobule showed more activation when the outcome matched participants’ goals. This implies a reliance on either a different network or a different process in the same network, based on the signaling of a cognitive match between the participant’s outcome goal and the actual outcome, without involvement of motor-sensory processes. This notion of cognitive matching is consistent with literature describing the involvement of this region in signaling a match between cognitive expectations of a specific outcome and subsequent memory retrieval experiences when this outcome occurs – a process similar to self-agency inferences (e.g., O’Connor et al. 2010).

An important question following from these observations is how the identified regions communicate to allow self-agency experiences in different contexts. Both
motor prediction and inference processes rely on a comparison between predicted outcomes and actual outcomes that has to be translated into a conscious experience (cf., Dehaene and Naccache 2001). Depending on the context (i.e., depending on the predictability of action-effects), the parietal region seems to inform conscious agency experiences in a different way. To date, only a few studies on the neural basis of agency have considered interactions between brain regions (David 2012). In one study investigating motor predictions, incongruent visual feedback was associated with a leading network, consisting primarily of the left anterior inferior parietal lobe, the right supramarginal gyrus, the right temporo-parietal junction and the anterior insula, which was sending information to a lagging network consisting mainly of the cingulate, posterior inferior parietal lobe and the pre-frontal lobe (Nahab et al. 2011). The authors speculated that the leading network is likely to be involved in mismatch detection between motor predictions and sensory action effects, whereas the lagging network translates the outcome of this comparison into a conscious agency experience.

A recent study also explored cortical interaction in an agency inference context similar to the present study (Dogge et al. 2013). Here, electroencephalography recordings assessed phase synchronization of neural oscillations (Sauseng and Klimesh 2008), which has been proposed as the mechanism underlying neural communication (Buszáki and Draghun 2004; Fries 2005; Sauseng and Klimesh 2008; Valera et al. 2001). In line with the present study, they showed that agency inferences involve fronto-parietal regions, and that the inferences depend heavily on the connectivity between these regions. Furthermore, the connectivity patterns analysis in the Dogge et al., study showed directional flow from parietal to frontal regions,
indicating that indeed the parietal regions inform frontal regions leading to self-agency experiences.

To conclude, the present study provides new and exciting insights in the processing of inferential self-agency, providing a first look at the neural correlates of self-agency processing independent of motor-prediction processes. The insights may help develop an initial brain model of inferential self-agency experiences, a process that is most likely relevant for efficient and high-quality social interactions.
**Figure Legends**

**Fig. 1.** Illustration of the experimental task showing how the squares move in opposite direction. Start and stop commands are presented in the middle of the square. After the stop command, the participant pressed a key, thereby stopping the movement of the squares. This action turned one of the eight white tiles black.

**Fig. 2.** Outline of the agency inference task.

**Fig. 3.** Frequency histograms of the self-agency ratings, separated by Matching condition.

**Fig. 4.** Brain areas activated in the Agency > No-Agency contrast. Cluster defining threshold of $p < 0.001$ with a $p < 0.05$ FWE-corrected cluster size of 27 voxels.
### Table 1

Peak voxel, number of voxels and T-value of the brain areas activated in the Agency > No-Agency contrast, all at threshold p < .001 (p < 0.05 FWE-corrected cluster size of 27 voxels).

<table>
<thead>
<tr>
<th>Brain region</th>
<th>x</th>
<th>y</th>
<th>z</th>
<th>Voxels</th>
<th>T value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left Inferior Parietal Lobule</td>
<td>-52</td>
<td>-68</td>
<td>32</td>
<td>37</td>
<td>5.08*</td>
</tr>
<tr>
<td>Left (Medial) Superior Frontal Gyrus</td>
<td>-20</td>
<td>52</td>
<td>40</td>
<td>35</td>
<td>5.13*</td>
</tr>
<tr>
<td>Right (Medial) Superior Frontal Gyrus</td>
<td>20</td>
<td>36</td>
<td>52</td>
<td>43</td>
<td>5.67*</td>
</tr>
<tr>
<td>Medial Prefrontal Cortex</td>
<td>8</td>
<td>64</td>
<td>4</td>
<td>104</td>
<td>4.88*</td>
</tr>
</tbody>
</table>

* p < .001
References


Illustration of the experimental task showing how the squares move in opposite direction. Start and stop commands are presented in the middle of the square. After the stop command, the participant pressed a key, thereby stopping the movement of the squares. This action turned one of the eight white tiles black.
Outline of the agency inference task.

243x159mm (96 x 96 DPI)
Frequency histograms of the self-agency ratings, separated by Matching condition.
108x114mm (96 x 96 DPI)
Brain areas activated in the Agency > No-Agency contrast. Cluster defining threshold of $p < 0.001$ with a $p < 0.05$ FWE-corrected cluster size of 27 voxels.

113x93mm (300 x 300 DPI)