Flexible Reference Frames for Grasp Planning in Human Parietofrontal Cortex¹,²,³

Frank T. M. Leoné,¹ Simona Monaco,² Denise Y. P. Henriques,² Ivan Toni,¹ and W. Pieter Medendorp¹

DOI: http://dx.doi.org/10.1523/ENEURO.0008-15.2015

¹Donders Institute for Brain, Cognition and Behavior, Radboud University Nijmegen, 6525 HR, Nijmegen, The Netherlands, and ²Centre for Vision Research, York University, Toronto, Ontario, M3J 1P3 Canada

Abstract

Reaching to a location in space is supported by a cortical network that operates in a variety of reference frames. Computational models and recent fMRI evidence suggest that this diversity originates from neuronal populations dynamically shifting between reference frames as a function of task demands and sensory modality. In this human fMRI study, we extend this framework to nonmanipulative grasping movements, an action that depends on multiple properties of a target, not only its spatial location. By presenting targets visually or somaesthetically, and by manipulating gaze direction, we investigate how information about a target is encoded in gaze- and body-centered reference frames in dorsomedial and dorsolateral grasping-related circuits. Data were analyzed using a novel multivariate approach that combines classification and cross-classification measures to explicitly aggregate evidence in favor of and against the presence of gaze- and body-centered reference frames. We used this approach to determine whether reference frames are differentially recruited depending on the availability of sensory information, and where in the cortical networks there is common coding across modalities. Only in the left anterior intraparietal sulcus (aIPS) was coding of the grasping target modality dependent: predominantly gaze-centered for visual targets and body-centered for somaesthetic targets. Left superior parieto-occipital cortex consistently coded targets for grasping in a gaze-centered reference frame. Left anterior precuneus and premotor areas operated in a modality-independent, body-centered frame. These findings reveal how dorsolateral grasping area aIPS could play a role in the transition between modality-independent gaze-centered spatial maps and body-centered motor areas.

Key words: functional magnetic resonance imaging; grasping; multivariate; parietofrontal cortex; reference frames; transformations

Significance Statement

Computational models of sensorimotor control suggest that neuronal populations dynamically shift between reference frames dependent on task demands and sensory modality. This fMRI study distinguished the reference frames used to plan a reach-to-grasp visual or somaesthetic target. Using a novel analysis approach, combining multivariate classification and cross-classification evidence, we distinguished between gaze- and body-centered frames in dorsomedial and dorsolateral grasp circuits. While most parietal regions in these circuits use modality-independent, gaze-centered coordinates and premotor regions use modality-independent, body-centered coordinates, the anterior intraparietal area (aIPS) switches its reference frame dynamically, depending on the sensory modality of the grasp target. We conclude that aIPS operates as an important hub in the transition between modality-independent, gaze-centered spatial maps and body-centered grasp areas.
Introduction

Parietofrontal neurons involved in reaching to a location in space operate in a variety of reference frames (Battaglia-Mayer et al., 2003; McGuire and Sabes, 2011; Buchholz et al., 2013). Theoretical modeling and behavioral evidence has suggested that these reference frames can be dynamically weighted according to task demands and sensory input (Pouget and Snyder, 2000; Beurze et al., 2007; McGuire and Sabes, 2009).

Recent neurophysiological work (Bernier and Grafton, 2010) has shown how sensory input influences the contributions of different parietofrontal regions while subjects reached out, by rotating their wrist, to touch visual or proprioceptive targets with their right index finger. By changing the fixation point relative to the target, that study showed that the anterior precuneus encoded the motor goal for visual targets selectively in gaze-centered (GC) coordinates, while other parietofrontal areas showed a mixture of GC and body-centered (BC) encoding. In contrast, reaching to proprioceptive targets revealed negligible gaze-centered encoding but considerable body-centered encoding throughout the parietofrontal network. Building on this work, here we investigate how task demands affect these modality-dependent reference frames within the grasping network.

In the study by Bernier and Grafton (2010), “touching” served as the steering task demand. Another common task demand supported by a reaching movement is grasping an object. So far, reaching to grasp has been studied predominantly in the context of visually guided movements. This has led to the notion that reaching to grasp is guided through two visuomotor channels in the parietofrontal network, which are involved in specifying where to move the hand in space and how to shape the fingers around the grasping target (Jeannerod, 1988; Culham and Valyear, 2006; Filimon, 2010). These parietofrontal areas are anatomically organized in a dorsomedial and a dorsolateral circuit (Rizzolatti and Matelli, 2003; Grol et al., 2007; Gamborini et al., 2009), with different accounts emphasizing either the parallel or the hierarchical organizations of these circuits (Culham and Valyear, 2006; Glover, 2004; Cavina-Pratesi et al., 2010b; Grafton, 2010; Verhagen et al., 2013; Binkofski and Buxbaum, 2013). However, these accounts have largely ignored how grasp demands affect the frames of reference involved in specifying where to reach, which could be crucial in understanding the functional contributions of these circuits to the overall sensorimotor transformation (Crawford et al., 2011).

Using fMRI in human participants, we tested the dominant reference frames in the dorsomedial and dorsolateral pathways when preparing a reach to grasp. Toward this end, we manipulated the availability of visual or proprioceptive information about the grasping target (the task demand), and the position of a target in relation to the subject’s gaze and body midline. The manipulation of the sensory modality relies on the rationale that the acquisition of visual information is intimately linked to the direction of gaze, whereas proprioceptive information is linked to the relative body part in the early stages of processing.

Accordingly, this manipulation quantifies how cortical reaching areas adjust to the different frames of reference imposed by sensory information of the grasp demand. The gaze manipulation directly discriminates between activity for reaching to grasp independent from gaze (i.e., linked to the body) and activity linked to gaze direction. We combine these manipulations within a novel multivariate analysis framework, aggregating evidence from both classification and cross-classification measures in favor and against gaze- and body-centered reference frames. Using this novel approach, we unveil reach-to-grasp areas that determine reference frames flexibly, depending on sensory modality, and those that use a modality-independent code.

Materials and Methods

Nineteen healthy right-handed participants with normal or corrected-to-normal vision participated in this study. The data of one participant (male) were excluded from further analysis due to noncompliance with task instructions. The remaining 18 participants (5 female) were in the age range 18-42 years (mean age, 26.5 years). Participants gave their written consent in accordance with the local ethics committee.

Experimental setup

Participants were lying supine in the scanner in complete darkness. Their upper body was cushioned and strapped to minimize trunk movement. The head was stabilized with foam blocks and wedges inside a phased-array receiver head coil. Head and coil were tilted 30° above the horizontal plane to allow direct vision of the grasping device. The grasping device was supported by an arch placed above the hips, and consisted of three light-emitting diodes (LEDs; called fixation LEDs) and a rotating platform, which was aligned parallel to the longitudinal body axis (Fig. 1). The fixation LEDs were positioned to the left, at the center, and to the right of the platform, 10 cm above the center of the platform (at –11°, 0°, and 11° of visual angle from the mid-sagittal plane). One half (Fig. 1, top) of the platform supported two Plexiglass blocks [size, 10 × 10 × 5 cm (length × width × height)].
allowing for the screening of incorrect target selection or movement time, each movement started from and performed in accordance with the task. To record reaction time and movement time, participants were instructed by an auditory tone to plan the grasping movement to either one of the two visual objects, or to one of the two segments of the left hand, while fixating one of three fixation lights (the yellow cross marks one of the three possible locations). After a variable delay (middle column), an auditory instruction (right column) triggered the execution of the movement (to the one or the other visual object or hand segment, indicated by jagged images). The experiment was performed in the dark; only elements indicated in yellow were visible to the participant.

Figure 1 Stimulus setup and experimental paradigm. A rotating platform presented either two visual objects (yellow rectangles, top row) or the participant’s left hand (bottom row, the hand was not visible during the experiment). At the start of the trial (left column), participants were instructed by an auditory tone to plan the grasping movement to either one of the two visual objects, or to one of the two segments of the left hand, while fixating one of three fixation lights (the yellow cross marks one of the three possible locations). After a variable delay (middle column), an auditory instruction (right column) triggered the execution of the movement (to the one or the other visual object or hand segment, indicated by jagged images). The experiment was performed in the dark; only elements indicated in yellow were visible to the participant.

Experimental paradigm
Participants performed an instructed-delay grasping task to either visual or somaesthetic targets, which were presented to the left or right of the body midline (see description above), while fixating on one of the three fixation LEDs (Fig. 1B). Participants started each trial with their right hand on the button box and their gaze directed at one of the three fixation LEDs. An auditory cue (160 or 480 Hz) indicated the target of the ensuing movement (sound–target mapping was counterbalanced across subjects). After a variable delay period (range, 3-6 s; uniform distribution), an auditory spoken “Go” cue instructed the participant to execute the movement, grasping the target and then returning to the button box. The grasp was nonmanipulative (i.e., firmly touching the target with all fingers), using the grasp equivalent of the manipulation used in Bernier and Grafton (2010). Subsequently, the fixation point changed position twice, as follows: in first 2 s after the Go cue, or 500 ms after movement onset, whichever was later; and then again 1 s later. Finally, 500 ms after the second saccade, the next trial started with the presentation of a new auditory cue. The double change in gaze position was included for an independent analysis (beyond the scope of this article) involving repetition suppression analysis.

We focused the experiment on the four combinations of gaze and target position, which can be directly compared between a gaze-centered and body-centered reference frame, as follows: gaze left-target left; gaze center-target left; gaze center-target right; and gaze right-target right.
Configurations with the target further separated from the gaze line (gaze left-target right; gaze right-target left) were not tested, as these trial types had no counterpart in gaze-independent, body-centered coordinates, because in body-centered coordinates the target was always directly to the left or right. Because the head and body were immobilized and the arm had a fixed starting position during the planning phase, head-, hand-, body-, and space-centered reference frames can be treated as equivalent, and were referred to as a body-centered ref-
ence frame. Likewise, under the present conditions, retinocentric, eye-centered, and gaze-centered reference frames could be considered synonymous, and were referred to as a gaze-centered reference frame.

Trials were grouped in 16 runs of 22 trials each (on average). Trial order within and across runs was arranged such that each of the four trial types followed each other equally often, while appearing random to the participant (Brooks, 2012). The length of the breaks between runs was determined by the participant. Visual and somaesthetic runs were alternated. The condition (visual or somaesthetic) of the upcoming run was indicated at the end of each break by an auditory signal, which instructed the participant to position the grasping device accordingly. Each run started and ended with 20 and 10 s of fixation, respectively. These intervals served as baseline in the general linear model (GLM) analysis. The total duration of the experiment (with eight runs for each of the two conditions and 88 trials for each gaze–target combination) was 64 min.

**MRI settings and preprocessing**

MRI images were acquired using a Siemens Tim Trio 3 tesla MRI scanner with an eight-channel head coil. A multiecho sequence of two echoes (TE, 14 and 34 ms; TR, 1.63 s) was used. The sequence encompassed 28 slices, centered on the parietal and frontal motor areas (voxel size, 3.5 mm isotropic; FOV, 192 mm; flip angle, 80°). In addition, high-resolution anatomical images were acquired using T1-weighted MP-RAGE generalized autocalibrating partially parallel acquisition (176 sagittal slices; voxel size, 1 × 1 × 1 mm; TR, 2300 ms; TE, 3.93; FOV, 256 mm; flip angle, 8°). The first echo was used to estimate realignment parameters, as it gives the best estimate of head motion, after which the parameters were copied to the second echo, which was used to estimate the BOLD signal (Poser et al., 2006). Slices were temporally aligned to the center slice (14th) to accommodate for slice-timing differences. High-pass filtering (cutoff, 128 s) was applied to filter out low-frequency confounds. To retain maximal pattern information, no spatial smoothing was applied. Functional data were normalized to MNI space using the DARTEL normalization procedure (Ashburner, 2007), retaining the original dimensions of the data. To estimate the normalization of the flow fields, the structural images were segmented by tissue type. A high-resolution MNI152 template (Fonov et al., 2011) was used to reconstruct and inflate the cortical sheet, separately for the left and right hemisphere, using the FreeSurfer Toolbox (Dale et al., 1999; Fischl et al., 1999). All further processing and analysis steps were performed using SPM8 [Statistical Parametric Mapping (http://www.fil.ion.ucl.ac.uk/spm/)] and Matlab (MathWorks).

**fMRI analysis**

As the basis of the analysis, a GLM was run to estimate the number of responses per voxel, after which the searchlight-based evidence combination method was applied, combining classification and cross-classification measures (Fig. 2D).

GLM

In each of the 16 runs, there were four regressors of interest (i.e., square waves encompassing the delay period). These regressors captured variance related to planning movements in the four tested spatial configurations (gaze left-target left, gaze center-target left, gaze center-target right, and gaze right-target right). The four hemodynamic regressors of interest were convolved with a standard hemodynamic response function (Friston et al., 2011).

Additional regressors were used to constrain the variance explained by the planning regressors. First, we included seven hemodynamic regressors. Four of those seven regressors were square waves encompassing the movement period, from presentation of the Go cue until the return of the hand on the home key (for details, see Experimental setup). These were used to capture variance related to movement execution separately for each of the four trial configurations. Two other regressors were spikes, time locked with stimulus presentation and with the saccade cues. These captured transient stimulus- and saccade-specific effects. A seventh spike regressor, time locked at the onset of a run, accounted for transient effects related to task onset. The seven regressors of noninterest were convolved with the same standard hemodynamic response function as the regressors of interest. Second, we included 17 nuisance regressors. Twelve movement regressors (translation and rotation, as well as their derivatives) captured signal variance caused by head movements. Five additional regressors accounted for the variability in overall image intensity in five compartments that are not expected to hold task-related activity (white matter, cerebrospinal fluid, skull, fat, and out of brain; Verhagen et al., 2008).

Runs were modeled separately in the design matrix. Each run contained 28 regressors and, on average 149 scans, resulting in an average of 2383 scans in total. We used the t values of the contrast between planning regressors and baseline as the basis for the evidence aggregation analysis, which is described below. We chose t values over β values because t values have been shown to provide more information in classification analyses (Misaki et al., 2010).

**Searchlight analysis**

All analyses were performed within local searchlight spheres (Kriegeskorte et al., 2006) with a radius of two voxels (7 mm), moved across the cortex. On average, the sphere size was 30 voxels (or 1286 mm³), with smaller search spheres at the outer cortical borders. Instead of ascribing the classification values to the center voxel of a sphere, we averaged, for each voxel, all classification results of the spheres containing the particular voxel. For example, if a voxel was included in 30 search spheres, the ascribed classification result for that particular voxel was the average of the 30 search spheres. This procedure allows for smooth searchlight maps and a better impression of the contribution of single voxels (Björnsdotter et al., 2011; Etzel et al., 2013).
Reference frame analysis

Rationale
The goal of the reference frame analysis was to distinguish between gaze-centered and body-centered coding of visual and somaesthetic targets for grasping. We detected shifts in the weighting of sensory evidence by comparison of the reference frames in fMRI activation patterns between modalities. Modality-independent coding was tested by comparing the cortical spatial code explicitly for similarity across modalities.

The basis of the analyses were pairs of the four trial types (gaze left-target left, gaze center-target right, gaze center-target left, and gaze right-target right), together defining six unique pairs in relation to the location of the target (Fig. 2A,D), as follows: left w.r.t. gaze, right w.r.t. gaze, left w.r.t. body, right w.r.t. body, outer sides, and center sides. For the first four pairs (Fig. 2A), each reference frame makes a prediction on whether the cortical representation should be distinct or common (Fig. 2A).

For example, in an area using gaze-centered coding, gaze left–target left should invoke a different pattern of activity than gaze center–target left, as the target is at opposite sides of the gaze. In contrast, body-centered coding predicts those patterns of activity to be similar, because the target is on the same side of the body midline (thus, the pair is called the left w.r.t. body). The latter two pairs (Fig. 2A, right side) do not lead to reference frame–specific predictions and are instead used to define regions of interest (ROIs).

While previous analytical approaches have focused either on distinct (Beurze et al., 2010) or on common representations (e.g. repetition-suppression analyses; Bernier and Grafton, 2010), here we combine the evidence provided by distinct and common representations. We do so by using searchlight-based classification and cross-classification, aggregating the respective accuracies into an evidence score per the hypothesis. That is, we regard the outcome of classification tests, often referred to as information (Kriegeskorte et al., 2006), as evidence for a particular hypothesis or reference frame. The method first specifies, for the reference frame tested, whether the representations for each pair of trials should be common or distinct (Fig. 2B, step 1). Next, per search sphere, classification accuracy is calculated when a distinct representation is predicted, and cross-classification accuracy when a common representation is predicted (Fig. 2B, step 2). Finally, we combined the evidence from classification and cross-classification for either a body-centered or a gaze-centered reference frame (Fig. 2B, step 3). We will now explain the details of the procedure (Fig. 2C).

Classification and cross-classification
Classification analysis tests for differences between representations (Cox and Savoy, 2003; Kriegeskorte and Bandettini, 2007); more dissimilar representations show higher classification. Classification was performed separately per the modality, that is, across 8 of the 16 runs. Before classification, we $z$-scored the $t$ values per voxel (Misaki et al., 2010). Then, for each pair of trial configurations (six in total, as shown in Fig. 2A), we trained and tested a binary linear support vector machine classifier as implemented in Donders Machine Learning toolbox (https://github.com/distrep/DMLT)]. Leave-one-out cross-validation was applied to avoid overfitting of the data. Each run contained four patterns (i.e., the $t$ values for the four planning regressors), corresponding to the four trial configurations. Per binary distinction, two of four patterns were used for each of the eight runs, generating 16 patterns. For each cross-validation fold, the classifier was trained on seven of the eight runs, or 14 patterns, and tested on the 2 patterns of the remaining run. The average classification performance across folds constituted the classification score for one pair, that is, the evidence about the dissimilarity of two representations.

Cross-classification analysis tests whether representations are similar (Etzel et al., 2008; Zhang et al., 2013; also called generalization analysis, see Barany et al., 2014); higher cross-classification scores indicate more similar representations. The per-sphere cross-classification scores were based on tests of the six trained binary classifiers (trained as described in the previous paragraph) on the two nontrained conditions. For each cross-validation fold, the tested conditions were taken from the runs not actually trained on. Critically, as the cross-classification analysis tests for generalization, only pairs not used to train the original classifier received a cross-classification score in this way.

For example, consider a classifier trained on the left w.r.t. body pair (i.e., gaze left-target left vs gaze center-target left; Fig. 2C, top, first column). For cross-classification, the trained classifier is tested on a comparable distinction, but with one of the two trial configurations replaced (gaze center-target left is replaced by gaze right-target right, second column). Critically, this replacement is different in body-centered coordinates (from left to right), but is equal in gaze-centered coordinates (still left from gaze). Hence, if the cross-classification returns high performance, this can be taken as evidence for gaze-centered coding. The cross-classification evidence is ascribed to the pair describing the replacement, left w.r.t. gaze (third column). Such a replacement can be made in four ways, hence four pairs receive cross-classification evidence from one classification pair. After cross-classification was performed for all six binary classifiers, the resulting cross-classification scores were averaged per receiving pair. This average cross-classification score indicates the similarity of a pair of representations.

Combination of evidence
The classification and cross-classification scores constitute separate evidence for the two reference frames, which were combined into evidence measures per reference frame. The rationale for combining the two measures is that classification and cross-classification are complementary measures, giving evidence on two separate sides of a hypothesis: which representations are predicted to be distinct, and which are predicted to be equal. In methodological terms, the measures together disambiguate chance-level performance on either measure. For example, a classification score can be at chance level for the following two reasons: either patterns are too noisy/incon-
sistent or patterns are equal. In the former case, cross-
classification scores will be low; in the latter, cross-
classification scores will be high. In addition, when an
area shows a high cross-classification score for a pair of
patterns, a classifier specifically trained on the distinction
between the two seemingly similar patterns could still be
able to distinguish them. This would mean the patterns
share particular characteristics, but differ in other as-
pects, allowing the detection of overlapping distinct
and common representations. Last, a region could show nei-
ther a classification nor a cross-classification effect when
no consistent signal is present or two effects in opposite
directions conflict.

We combined the evidence across pairs using two
approaches. First, we considered whether both sides of
the evidence (classification and cross-classification)
were in line with the hypothesized reference frame. For
gaze-centered coding, left or right w.r.t. gaze cross-
classification and left w.r.t. or right w.r.t. body classifica-
tion needed to be significant (for details on significance,
see Group analysis). For body-centered coding, a left or
right w.r.t. gaze classification and a left or right w.r.t. body
cross-classification were required to be significant. This
combination of significant evidence in classification and
cross-classification analyses was labelled as conjunction.
It allows for detecting the presence of either or both
reference frames, leading to binary maps (see Figs. 4, top
row, 6, top row, 7). Second, we considered whether the
average across classification and cross-classification
scores were in line with the hypothesized reference frame:
for gaze-centered coding, the cross-classification values
for left and right w.r.t. gaze and the classification val-
ues for left and right w.r.t. body. For body-centered cod-
ing, we averaged the classification values for left and right
w.r.t. gaze, and the cross-classification values for left and
right w.r.t. body. This combination of evidence was la-
belled as aggregation, and is depicted as scalar values on
surface maps (beled as aggregation, and is depicted as scalar values on
right w.r.t. body. This combination of evidence was la-
w.r.t. gaze, and the cross-classification values for left and
right w.r.t. body. For body-centered cod-
ing, we averaged the classification values for left and right
w.r.t. gaze, and the cross-classification values for left and
right w.r.t. body. This combination of evidence was la-
belled as aggregation, and is depicted as scalar values on
surface maps (Figs. 3, 4, bottom row, 5, 6, bottom row, 7)
and in ROI-specific measures (Figs. 5B, 6B). It allows for
detecting relative dominance in the reference frame with
high sensitivity, but low specificity. Note that both gaze-
centered and body-centered measures have an equal
number of classification and cross-classification scores,
which cancels out any possible imbalance between classi-
fication and the more stringent cross-classification
scores.

To illustrate the low sensitivity of the aggregation
method, body-centered coding predicts high classifica-
tion on two of the four main pairs and high cross-
classification on two other pairs (Fig. 2B). Critically,
the two classification pairs differ not only in body-centered
cordinates, but also in the direction of the gaze (Fig. 2A,
compare the left and right w.r.t. bodies). This means that
gaze-direction coding also predicts high classification
scores for the two pairs, and thus a high average score.
In other words, using the aggregate measure, body-
centered coding is confounded by gaze-direction coding
(as is gaze-centered coding; compare left and right w.r.t.
gaze). The conjunction analysis, however, also explicitly
requires the two cross-classification pairs to be signifi-
cant, which are not predicted to be significant for gaze
direction, separating the two explanations.

Contrasts between aggregate evidence values for the
reference frames are unaffected by gaze direction, as an
equal number of classification and cross-classification
scores are included on both sides, such that the gaze-
direction effects cancel out. Rather, conjunction and con-
trasts between aggregate values, as used in the overlay
images, are likely to give similar results, as the reference
frames are the mirror images of each other (Fig. 2B), and
the classification and cross-classification for a pair are, on
average, negatively correlated.

As an additional control measure, we explicitly tested
for gaze-direction effects by testing for significant classi-
ification on all four main pairs, combined with significant
cross-classification on the center-sides pair. The latter
pair is the only pair predicted to be common in regions
coding gaze direction, but distinct in either reference
frame, allowing direct estimation of a possible gaze-
direction effect.

Within- versus across-modality coding
We tested both the dominant reference frame within a
modality and the consistency of coding across modalities:
that is, the grasp areas that determine reference frames
depending on the sensory modality, and those that code
representations using modality-independent codes.

To test reference frame evidence within a modality,
both training and test data were from the same modality
(Fig. 2C, left). To test reference frame-specific common
coding across modalities, the training and test data were
from different modalities (Fig. 2C, right), averaging the two
directions (from visual to somaesthetic and vice versa).
Thus, for across-modality coding, we applied cross-
modal classification (Etzel et al., 2008).

Group analysis
For group analysis, we ran a $t$ test on the evidence maps
for gaze-centered, body-centered, and gaze-direction
coding across subjects. The GLM was implemented in the
GLM Flex toolbox (http://mrtools.mgh.harvard.edu/index.
php/Main_Page). All group $t$ tests were run using a GLM
with 432 scans (18 subjects $\times$ 2 modalities $\times$ 6 trial pairs
$\times$ 2 tests) and 42 regressors (18 subjects + 2 modalities
$\times$ 6 trial pairs $\times$ 2 tests), or a df of $432 - 41 = 391$. For
surface results, results were thresholded at $\alpha = 0.05$, cluster size $>250$ ($\alpha = 0.1$, cluster-level corrected),
to display the whole range of effects in the data. Moreover,
full cluster correction on average searchlight results has
proven to be conservative (Stelzer et al., 2013).

Regions of interest
Because we had clear hypotheses on the cortical path-
ways involved in planning right-handed grasping move-
ments, we focused our analyses on five ROIs in the left
hemisphere. These ROIs characterized areas along the
dorsolateral and dorsomedial pathways that have been
implicated in reach and grasp control (Connolly et al.,
2003; Verhagen et al., 2008; Fillimon et al., 2009; Bernier
and Grafton, 2010; Fabbri et al., 2014). Each ROI was
determined on the basis of the local maxima in evidence
for the independent pairs (center sides and outer sides; Fig. 2A; Kriegeskorte et al., 2009) closest to the average coordinates reported in the study by Bernier and Grafton (2010). The local maximum was determined on the basis of the group t values for the combined classification score for the two pairs, as these two pairs are predicted to be distinct in both reference frames. The peak could be in visual coding, somaesthetic coding, or in the sum of their scores. Peaks were restricted to the cortical surface, and the ROI was defined as a 7-mm-radius sphere surrounding the peak.

Along the dorsolateral pathway, we considered anterior intraparietal sulcus (aIPS) and ventral premotor cortex (PMv). The area of the aIPS (reference, (−35, −47, 50; centered at, −32, −42, 46 in MNI coordinates)) has been shown to be involved in planning grasping movements (Grafton et al., 1996; Binkofski et al., 1998; Culham et al., 2003; Dinstein et al., 2008; Verhagen et al., 2008). PMv (reference, (−52, −1, 31; centered at, −56, −3, 32) has been implicated in movement preparation, including the preshaping of the hand during grasping (Toni et al., 2001; Davare et al., 2006; Hoshi and Tanji, 2007).

Along the dorsomedial pathway, we selected the superior parieto-occipital cortex (SPOC), anterior precuneus (aPCu), and dorsal premotor cortex (PMd). The SPOC region (reference, (−16, −77, 40; centered at, −17, −74, 46) has been implicated in reaching and grasping movements (Connolly et al., 2003; Culham and Valyear, 2006; Verhagen et al., 2008; Bernier and Grafton, 2010). As a region, SPOC has also been referred to as parieto-occipital junction (Prado et al., 2005; Bernier and Grafton, 2010) and as the human homolog of macaque V6A (Pitzalis et al., 2013). The aPCu region (reference, (−9, −55, 63; centered at, −4, −56, 63) has been involved in reaching movements, independent of visual feedback (Filimon et al., 2009; Wenderoth et al., 2006), using a reference frame that is dependent on the sensory modality of the target (Grafton, 2010). This region falls into the probability maps of Brodmann areas 7A and 5L (Eickhoff et al., 2005). Dorsal premotor cortex (reference, (−21, −10, 56; centered at, −28, −4, 56) has been involved in planning and controlling both reaching and grasping (Davare et al., 2006; Raos et al., 2006; Hoshi and Tanji, 2007).

**Results**

We investigated the reference frames in which parietal and frontal regions operate during the planning of grasping movements toward visual and somaesthetic targets. In the experiment, we manipulated the position of a grasping target (left or right from the body midline) and the direction of gaze (left, center, right). We focused our analyses on four pairs of trials (left or right w.r.t. gaze or body; Fig. 2A), which could either be different or equal in GC and/or BC coordinates. We used a novel searchlight-based evidence combination method to take into account the following two types of evidence: classification tested for distinct representations (i.e., where coordinates are predicted to be different in the tested reference frame); and cross-classification for common representations (i.e., where coordinates are predicted to be equal; for details, see Materials and Methods; Fig. 2). Based on the specific predictions of each reference frame, the two pieces of evidence were combined into an evidence score. We used comparisons between the modalities and cross-classification across modalities to test which regions code a representation linked to the reference frame of the sensory modality and which code a common, modality-independent representation.

Behavioral analysis showed that participants performed the task effectively, with virtually no errors in target selection (0.05% error; range, 0–3.05% across subjects), and matched performance across modalities, both in reaction time (reaction time to visual targets: 630 ms; SD, 189 ms; reaction time to somaesthetic targets: 626 ms; SD, 181; $t_{34} = -0.07, p > 0.05$) and movement time (movement time to visual targets: 2.465 s; SD, 504 ms; movement time to somaesthetic targets: 2.574 s; SD, 551 ms; $t_{34} = 0.62, p > 0.05$).

In the following, we start by describing the cortical distribution of planning-related information that is distinct in both gaze- and body-centered reference frames. This information, which is based on the classification of centersides and outer-sides pairs (Fig. 2A, right side) served as the basis for the independent definition of the ROIs. Next, we use the remaining trial types to decode the reference frames involved in processing target information in visual and somaesthetic modalities, followed by a comparison between modalities and an examination of regions that use modality-independent codes. We end with an analysis of the cortical topography of gaze direction, based on cross-classification of the center-sides pair (i.e. trials with identical gaze position).

**Definition of ROIs**

To allow for independent ROI definition, we tested for information on grasp planning to the two targets with gaze either at the center or on the outer sides of the target (Fig. 2A, right side), because in this case the target is distinct in both gaze- and body-centered reference frames. Figure 3 shows that the cortical distribution of the classification information is similar when the target was presented visually or somaesthetically. The indicated ROIs were defined based on the evidence peaks in either modality.

**Reference frames per modality**

Figure 4 shows the reference frame scores across the surface, which were based on the combination of evidence across four of the six pairs of trial types, separately for each reference frame and modality.

Planning grasping movements to visual targets evoked gaze-centered representations only in occipital and parietal cortices (Fig. 4A, top). Body-centered coding was present in the left frontal and right parietofrontal regions. The contrast between references frames values confirmed that tuning was predominantly gaze centered in occipitoparietal cortex and predominantly body-centered in frontal regions (Fig. 4A, bottom).

For somaesthetic targets, we found gaze-centered information in bilateral occipitoparietal cortex only, as for visual targets (Fig. 4B, top). Body-centered information was present in frontal regions, extending into bilateral occipitoparietal cortex.
rostral parietal cortex. The contrast between the reference frames showed that gaze-centered tuning was dominant in small parts of occipital and parietal cortex, while more frontal regions, and rostral parietal cortex, were biased to body-centered coordinates (Fig. 4B, bottom). The strongest peak was found in right somatosensory and motor cortex, probably reflecting the representation of the somesthetic targets, provided by the left hand.

The predefined ROIs further characterize these visual and somesthetic grasping gradients, with some predominantly gaze-centered (SPOC), others predominantly body-centered (aPCu and the frontal ROIs). The area aIPS appears to change its dominant reference frame according to the sensory input modality. This effect is further explored below.

Modality-dependent reference frames

We further characterized how the relative contribution of these two reference frames depends on input modality, by examining the differences between modalities. Figure 5 shows the results of this analysis.

Gaze-centered tuning was stronger for visual targets in a large caudal bilateral cluster ranging from the occipital cortex, along the left IPS, to aIPS at the junction with postcentral sulcus (Fig. 5A, top row). No regions showed a bias for gaze-centered tuning for the processing of somesthetic targets, relative to visual targets. Body-centered coding was dominant for somesthetic targets in right S1 and M1, probably reflecting information on the left hand (which provides the somesthetic targets). Importantly, in the dorsolateral parietofrontal pathway, the left aIPS also showed body-centered dominance for somesthetic targets (Fig. 5A, middle row). The consistency analysis (Fig. 5A, bottom row) showed a clear switch, consistent with the sensory reference frame, specifically in left aIPS.

ROI analysis confirmed the findings of the whole-brain analysis (Fig. 5B). Of the ROIs, only left aIPS showed significantly higher gaze-centered tuning in the visual condition ($p < 0.05$, uncorrected), significantly higher body-centered tuning in the somesthetic condition ($p < 0.05$), as well as a significant difference in relative tuning (GC-BC) between visual and somesthetic targets.
the visual and somaesthetic conditions \( (p < 0.005) \). Other ROIs did not show significant evidence for a switch in reference frame across modalities.

Thus, aIPS was the only region in which it was possible to identify a modality-dependent reference frame for grasp planning.

**Modality-independent reference frames**

We next test whether the similarities in reference coding between modalities (Fig. 4) also imply modality-independent cortical coding (Fig. 6); that is, whether the same pattern in fMRI activations is used to represent the targets in a particular reference frame, irrespective of target modality.

Results showed extensive modality-independent coding (Fig. 6A). The gaze-centered representations in bilateral occipitoparietal cortex as well as part of the body-centered representations in frontal regions were shared across modalities.

The ROIs (Fig. 6B) show a strong division between body- and gaze-centered codes. Body-centered coding is shared across modalities in almost all ROIs (all BC, \( p < 0.01 \); significant conjunctive evidence in PMv and PMd). Only in aPCu, PMd, and PMv, the strength of common body-centered coding was significantly higher than that of common gaze-centered coding (GC-BC, \( p < 0.05 \)). In contrast, gaze-centered coding is specifically shared across modalities in SPOC (GC, \( p < 0.001 \), no difference between GC and BC). Importantly, even though SPOC appears to code both reference frames using a multimodal code when tested using the aggregate measure in the ROIs (Fig. 6B), only the gaze-centered coding is significant in the conjunction test (Fig. 6A,B, compare transparent dot, opaque dot).

In sum, areas PMd, PMv, and aPCu contain a modality-independent, body-centered code while area SPOC comprises a gaze-centered modality-independent code during grasp planning.

**Effect of gaze direction**

Finally, because there may be gaze-direction effects in the data (for details, see Materials and Methods), we specifically tested for such effects in Figure 7.

Gaze-direction coding was found primarily in the occipital cortex, extending into the parietal cortex (Fig. 7A,B), irrespective of modality (Fig. 7C). Note the striking overlap between gaze information and gaze-centered coding in Figures 4 and 6, meaning gaze information and gaze-centered coding occupy similar regions. Of the predefined ROIs, we found that only SPOC coded gaze direction consistently within modalities (both \( p < 0.01 \), no difference between the visual and somaesthetic conditions) and across modalities (\( p < 0.05 \)). The latter could also explain the discrepancy between the body-centered tuning in SPOC when using the aggregate measure (Figs. 5B, 6B) and the absence of such tuning in the conjunction measure (Figs. 4, 6A, top row). Note that the conjunctive measure and the gaze-direction measure are independent, reaffirming the conclusion that gaze-direction and gaze-centered tuning are present in similar areas.
Discussion

This study investigated the reference frames used to plan grasping movements toward visual and somaesthetic targets. We focused on the manifestations of gaze- and body-centered frames of reference in the dorsomedial and dorsolateral grasping-related circuits. There are three main observations. First, a parietal node of the dorsolateral circuit (aIPS) encoded the grasping target flexibly, in either a body-centered or a gaze-centered reference frame, depending on whether the movement was planned toward a somaesthetic or a visual target. Second, posterior regions, including SPOC, coded the grasping target in gaze-centered coordinates, while anterior regions, including PMd and PMv, operated in a body-centered reference frame. Third, gaze- and body-centered coding of the grasping target was largely modality independent. These observations are based on a novel analysis method, combining multivariate classification and cross-classification evidence. Each of these points will now be discussed in detail.

Flexible reference frame in aIPS

In the dorsolateral circuit, we find a modality-dependent reference frame in aIPS, which switches between gaze and body coordinates according to the sensory modality used to spatially define the grasping target. This observation is consistent with recent ideas predicting reference frame switches alongside the native reference frame of the sensory signal (Pouget and Snyder, 2000; McGuire and Sabes, 2009). For visual targets, input is coded with respect to gaze; for somaesthetic targets, input is coded with respect to the body. Thus, aIPS appears to code targets in a statistically optimal fashion, favoring coding in the presented sensory reference frame (Pouget et al., 2002).

Area aIPS is well positioned to play such an integrative role in planning grasping movements, both in functional profile and connectivity. Functionally, aIPS is necessary for appropriate planning of grasping movements (Gallese et al., 1994; Binkofski et al., 1998; Culham et al., 2003; Verhagen et al., 2012), and it represents a wide range of

![Figure 6 Modality-independent reference frames (GC and BC). A, Binary maps based on the conjunction analysis (top: blue, BC; red, GC; magenta, both) and the difference in aggregate evidence between the reference frames (bottom: cool colors, BC; warm colors, GC). Only relevant clusters are shown (p < 0.05, uncorrected; cluster size >250). Occipitoparietal cortex shows a modality-independent gaze-centered reference frame. A shared body-centered reference frame is seen in part of the frontal regions. Circles indicate ROIs. B, Comparison of t values for within-modality coding (sum of visual and somaesthetic, x-axis) versus across-modality coding (sum of both directions of cross-classification, y-axis) for the two reference frames (red and blue dots; transparency indicates that conjunction results are not significant for the respective reference frame and ROI, as shown in A, top). Dashed lines highlight significant differences between reference frames (p < 0.05). Areas highlighted by the triangular dark zone, marked as “Shared code,” show evidence for a shared, modality-independent coding for the indicated reference frame.

![Figure 7 A–C, Gaze-direction coding for planning grasping movements to visual targets (A), somaesthetic targets (B), and irrespective of modality (C). Voxels within the binary map are significant for classification on all four main pairs and cross-classification on the center-sides pair, as determined using the conjunction test. Compare to Figure 4A (top row) for an overview of all tuning. Circles indicate ROIs.](image-url)
shown that there is a cortical dissociation between hand and finger movements, represented in aIPS, and whole-limb movements, represented in aPCu (Heed et al., 2011; Leoné et al., 2014; Sereno and Huang, 2014) and other areas of the dorsomedial circuit (Cavina-Pratesi et al., 2010a,b). Accordingly, switching between gaze- and body-centered reference frames might be implemented in the parietal region, which is more directly involved in representing the combination of limb segments controlled during task performance. For example, reference frame switches for foot/leg movements are predicted to peak in more medial regions than reported here (Leoné et al., 2014).

### Gaze- versus body-centered networks

The present study indicates that areas aIPS and aPCu fall between posterior gaze-centered coding and anterior body-centered responses. Specifically, occipitoparietal cortex, including SPOC, uses a predominantly gaze-centered code for both visual and somaesthetic targets, which is compatible with previous findings on a gaze-centered dominance in parietal cortex (Stricanne et al., 1996; Crawford et al., 2011), and the visual and somaesthetic inputs reported in these regions (Fattori and Gambini, 2001; Fattori et al., 2005, 2009). In addition to these gaze-centered codes, we found evidence for gaze-direction coding in parietal regions (Galletti et al., 1995; Marzocchi et al., 2008; Hadjidimitrakis et al., 2011; Rossit et al., 2013). Access to this information may allow the regions to play a role in reference frame transformations (Zipser and Andersen, 1988; DeSouza et al., 2000; Barany et al., 2014). In fact, we believe that a caudal–rostral gaze-centered/body-centered gradient could be underlying our results for posterior parietal cortex (PPC; McGuire and Sabes, 2011). In frontal regions, including areas PMd and PMv, we instead found evidence for body-centered coding, which is in line with the assumed role of premotor regions in motor preparation (Graziano and Gross, 1998) and the implementation of joint-based motor commands (Beurze et al., 2010).

The present results could be taken as inconsistent with other reports showing mixed coding in parietal and frontal regions (Galletti et al., 2003; Avillac et al., 2005; Mullette-Gilliman et al., 2005; Pesaran et al., 2006; Batista et al., 2007; Chang and Snyder, 2010), as well as important theoretical work (Pouget and Snyder, 2000). However, caution with such an interpretation is needed. Differently from previous reports, this study did not manipulate the body-centered reference frame by using a different start position for the right hand, leaving open the possibility that gaze-centered responses also become consistent with body-centered coding when both reference frames are manipulated. In addition, dominant gaze-centered coding could obscure body-centered coding and vice versa, as the expected effects in cortical patterns are opposite. Last, gaze-direction effects in parietal regions might have obscured body-centered effects or induced apparent body-centered effects in the current and other studies. For example, in our study SPOC appears to also code body-centered coordinates; however, the conjunc-

### Gaze- versus body-centered networks

The present study indicates that areas aIPS and aPCu fall between posterior gaze-centered coding and anterior body-centered responses. Specifically, occipitoparietal cortex, including SPOC, uses a predominantly gaze-centered code for both visual and somaesthetic targets, which is compatible with previous findings on a gaze-centered dominance in parietal cortex (Stricanne et al., 1996; Crawford et al., 2011), and the visual and somaesthetic inputs reported in these regions (Fattori and Gambini, 2001; Fattori et al., 2005, 2009). In addition to these gaze-centered codes, we found evidence for gaze-direction coding in parietal regions (Galletti et al., 1995; Marzocchi et al., 2008; Hadjidimitrakis et al., 2011; Rossit et al., 2013). Access to this information may allow the regions to play a role in reference frame transformations (Zipser and Andersen, 1988; DeSouza et al., 2000; Barany et al., 2014). In fact, we believe that a caudal–rostral gaze-centered/body-centered gradient could be underlying our results for posterior parietal cortex (PPC; McGuire and Sabes, 2011). In frontal regions, including areas PMd and PMv, we instead found evidence for body-centered coding, which is in line with the assumed role of premotor regions in motor preparation (Graziano and Gross, 1998) and the implementation of joint-based motor commands (Beurze et al., 2010).

The present results could be taken as inconsistent with other reports showing mixed coding in parietal and frontal regions (Galletti et al., 2003; Avillac et al., 2005; Mullette-Gilliman et al., 2005; Pesaran et al., 2006; Batista et al., 2007; Chang and Snyder, 2010), as well as important theoretical work (Pouget and Snyder, 2000). However, caution with such an interpretation is needed. Differently from previous reports, this study did not manipulate the body-centered reference frame by using a different start position for the right hand, leaving open the possibility that gaze-centered responses also become consistent with body-centered coding when both reference frames are manipulated. In addition, dominant gaze-centered coding could obscure body-centered coding and vice versa, as the expected effects in cortical patterns are opposite. Last, gaze-direction effects in parietal regions might have obscured body-centered effects or induced apparent body-centered effects in the current and other studies. For example, in our study SPOC appears to also code body-centered coordinates; however, the conjunc-
tion and gaze-direction analysis shows that this is probably an effect of gaze direction.

Modality-independent reference frames

We found the coding in both the gaze- and body-centered networks to be largely modality independent. This observation extends previous work on modality-independent spatial tuning (McGuire and Sabes, 2011), and on multisensory integration in both parietal and premotor regions (Bremmer et al., 2002; Macaluso et al., 2003; Todorov, 2007) by showing that the coding is not only in the same region, but shows the same within-region cortical pattern (as tested by cross-modal classification; Etzel et al., 2008).

The extent of the modality-independent coding is probably linked to the generality of the function served by the regions. PPC is also believed to code a general saliency map, which generalizes across tasks (Jerde et al., 2012). Within this notion, we show that it also generalizes across modalities. Such modality independence would fit the preserved role of PPC in the congenitally blind (Lingnau et al., 2012). The modality-independent, body-centered code in premotor regions fits the coding of the impending action, which is equal across the two modalities and is believed to be coded in body-centered coordinates. The widespread modality-independent tuning could have been influenced by the small number of sides (two per reference frame); further studies should include more spatial locations to further specify the modality-independent coding. Moreover, other studies should manipulate the properties of the grasping target to examine whether these properties influence the modality dependence (or independence) of tuning.

Two-sided evidence combination

We introduced a two-sided evidence combination method. This method allows one to combine evidence from both predicted distinctions between cortical representations, using classification, and predicted common representations, using cross-classification. This methodological approach provides access to overlapping representations in gaze- and body-centered frames. Previous methods, which were focused on classification effects or univariate differences, would not have detected common representations. Similarly, repetition suppression approaches would not be able to detect distinct representations.

We used the following two implementations of the method: one qualitative, aimed at explicitly determining regions that are significant for a conjunction of both sides of the evidence; and one quantitative, aimed at differences in the degree of evidence by means of averaging. The method can be further extended to accommodate more complex hypotheses by using GLMs to fit evidence scores rather than calculating averages. The method also combines naturally with the pattern-activation method by Leoné et al. (2014) as well as representational similarity analysis (Kriegeskorte et al., 2008), replacing correlations with cross-classification and classification.

Conclusion

This study shows that dorsolateral grasping area aIP switches reference frame depending on the sensory modality: visual targets are predominantly processed in a gaze-centered reference frame, and somaesthetic targets are coded in a body-centered reference frame. In contrast, other parietofrontal regions respond mainly in a single reference frame, with caudal parietal areas in the dorsomedial pathway code grasping targets in a modality-independent, gaze-centered reference frame, while premotor areas code targets in a modality-independent, body-centered reference frame. The modality-independent nature of the parietal and frontal clusters could reflect their roles in coding both saliency and motor preparation. Area aIP rather serves a potentially fundamental role as an in-between conversion hub when coding grasping movements.

References


