PNAS
www.pnas.org

Supplementary Information for

Picture perception reveals mental geometry of 3D scene inferences

Erin Koch[1], Famya Baig[1], Qasim Zaidi[1]

1. Graduate Center for Vision Research, State University of New York, College of Optometry, 33 W 42nd St New York, NY 10036.

Qasim Zaidi

qz@sunyopt.edu

This PDF file includes:

Supplemental Methods: Derivation of 2D retinal orientations from 3D object poses

Figs. S1 to S4

www.pnas.org/cgi/doi/10.1073/pnas.1804873115

Derivation of 2D retinal orientations from 3D object poses:

This projection is derived for objects in 3-Space (XYZ – Space) lying on the ground plane (i.e. object elevation = 0°), and extending from the center of the scene $(0,0,0)$. The camera is centered at that point. Thus each object has one endpoint at $(0,0,0)$, and the other at $(x,0,z)$.

We first rotate the ground plane around the x-axis to account for the camera elevation angle, ϕ_c. The center point,$(0,0,0) \rightarrow (0,0,0)$. We compute the new coordinates of the other endpoint :

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} x \\ 0 \\ z \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\phi_c) & \sin(\phi_c) \\ 0 & -\sin(\phi_c) & \cos(\phi_c) \end{pmatrix} \quad \text{(S1)}$$

giving:

$$\begin{aligned} x' &= x \\ y' &= z\sin(\phi_c) \\ z' &= z\cos(\phi_c) \end{aligned} \quad \text{(S2)}$$

Next, we compute the projection of each endpoint onto the Picture Plane (UV-space). We let the distance from the camera to the center of the 3D scene be d_c , the focal length of the camera be f_c. The central endpoint is mapped as $(0,0,0) \rightarrow (0,0)$. We compute the picture plane coordinates of the other endpoint:

$$r = \frac{u'}{d_v + w'} f_v \quad \text{(S3)}$$

$$s = \frac{v'}{d_v + w'} f_v$$

The final projection is from the picture plane to the 2D retinal plane (RS-space). In our experiment the monitor remains static and the observer changes viewpoint by physically moving to different locations that form a semi-circle around the monitor. This is equivalent to fixing the location of the observer and rotating the monitor. Because the latter is more computationally intuitive, we will frame the derivation of the projection in this way. We first compute the new coordinates in 3-Space (UVW space), where the plane W=0 is defined by the fronto-parallel location of the monitor. Moreover, we add depth to the 2D –Space defined by the fronto-parallel picture plane. The central point is mapped as, $(0,0) \rightarrow (0,0,0)$. Computing the new coordinates for the other endpoint we have:

$$\begin{pmatrix} u' \\ v' \\ w' \end{pmatrix} = \begin{pmatrix} u \\ v \\ 0 \end{pmatrix} \begin{pmatrix} \cos(\phi_v) & 0 & \sin(\phi_v) \\ 0 & 1 & 0 \\ -\sin(\phi_v) & 0 & \cos(\phi_v) \end{pmatrix} \quad \text{(S4)}$$

And:

$$\begin{aligned} u' &= u\cos(\phi_v) \\ v' &= v \\ w' &= -u\sin(\phi_v) \end{aligned} \quad \text{(S5)}$$

We let d_v equal the observer's distance to the center of the picture plane, and f_v equal the observer's focal length. Projecting the endpoints from UVW-Space to 2D retinal space (RS-space), the central endpoint is mapped as, $(0,0,0) \rightarrow (0,0)$, and the other endpoint is defined by:

$$r = \frac{u'}{d_v + w'} f_v \qquad \text{(S6)}$$

$$s = \frac{v'}{d_v + w'} f_v$$

To geometrically recover the orientation in the retinal plane in terms of the original orientation on the 3D ground plane we use a simple trigonometric derivation and substitution from the above equations:

$$\theta_R = \operatorname{atan}\left(\frac{s}{r}\right) = \operatorname{atan}\left(\frac{\frac{v'}{d_v+w'}f_v}{\frac{u'}{d_v+w'}f_v}\right) = \operatorname{atan}\left(\frac{v'}{u'}\right) = \operatorname{atan}\left(\frac{v}{u\cos(\phi_v)}\right) \qquad \text{(S7)}$$

$$\Rightarrow \qquad \theta_R = \operatorname{atan}\left(\frac{\frac{y'}{d_c+z'}f_c}{\frac{x'}{d_c+z'}f_c\ \cos(\phi_v)}\right) = \operatorname{atan}\left(\frac{y'}{x'\cos(\phi_v)}\right) = \operatorname{atan}\left(\frac{z\ \sin(\phi_c)}{x\cos(\phi_v)}\right)$$

$$\Rightarrow \qquad \theta_R = \operatorname{atan}(\tan(\Omega_T)\cdot\ (\sin(\phi_c)/\cos(\phi_v)))$$

Lastly we can take the inverse of the above projection to get the geometric back projection from the retinal orientation to 3-Space orientation:

$$\Omega_T = \operatorname{atan}(\tan(\theta_R)\cdot\ (\cos(\phi_v)/\sin(\phi_c))) \qquad \text{(S8)}$$

Equations 5a, 5b, and 5c, for off-centered, elevated, and floating objects were derived similarly by using the proper coordinates for the object endpoints in Equation (S1).

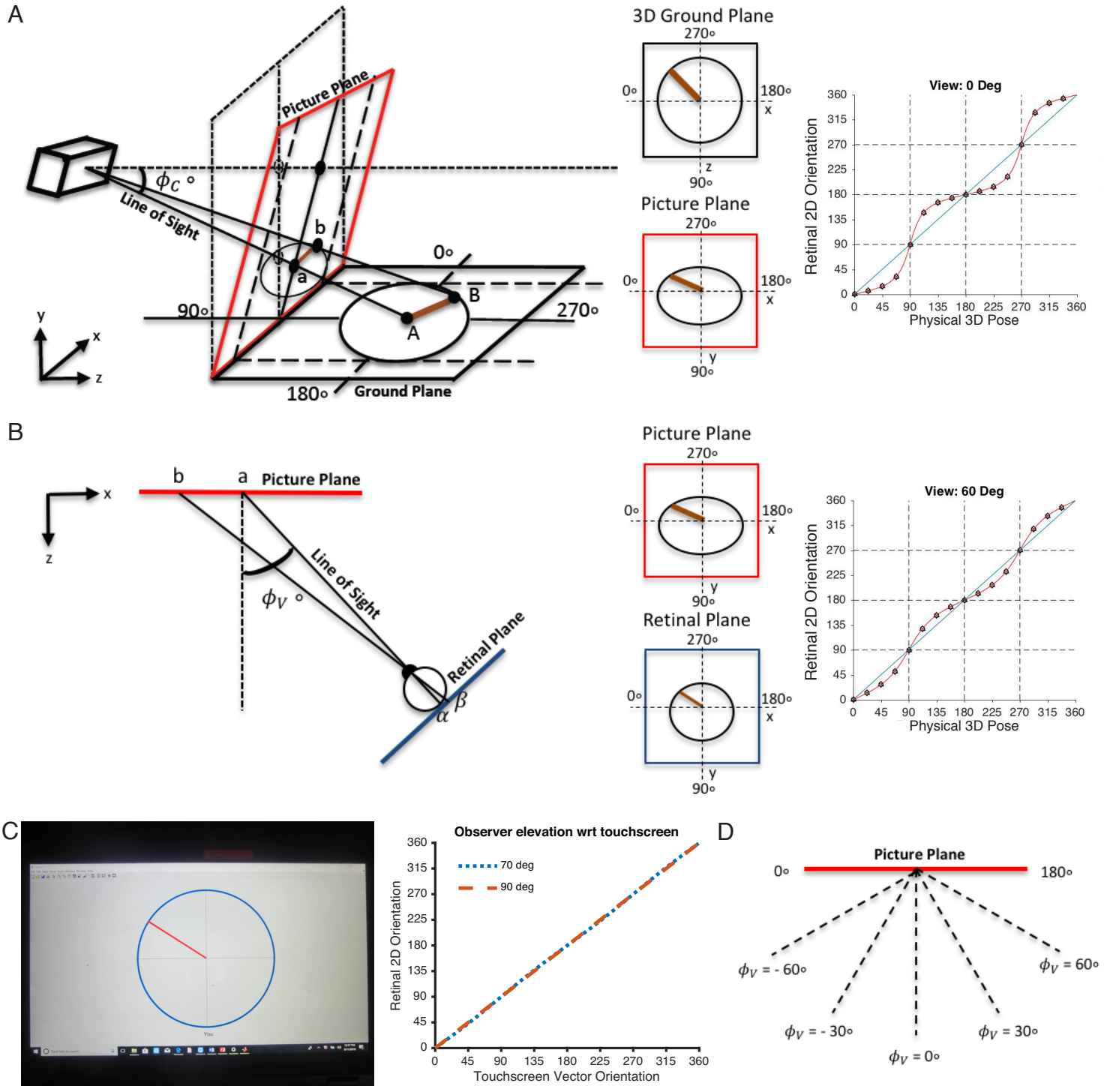

Figure S1: Refers to Figure 1
A. (Left) A depiction of the projection from the 3D scene to the picture plane. The picture plane is depicted in front of the camera for simplicity. (Center) Top-down view of the ground plane, above a frontal view of the picture plane. The brown line is the projected image of a stick at an oblique pose. These plots show how a circle on the ground plane would become vertically compressed in the picture plane, and how the orientation of a stick extending from the center of the circle on the ground plane becomes compressed towards the horizontal axis in the picture plane. (Right) Projected retinal orientation as a function of 3D pose of a stick lying on the ground, for the fronto-parallel viewing condition where retinal and picture orientations are identical. Black circles are projections of 16 stick poses, and the red curve is the continuous projection function (Equation 1). B. (Left) A depiction of the projection from the picture plane to the retinal plane for oblique viewpoints. (Center) The same frontal view of the picture plane as in (A) and a frontal view of the retinal plane (resulting from oblique viewing). The retinal plane becomes horizontally compressed relative to the picture plane giving the illusion of vertical elongation and a tilting of the ground plane. (Right) Forward projection from the +60 degree observer viewing location. Black circles are projections of 16 stick poses, and the red curve is the continuous projection function (Equation 4). C. (Left) Directed vector on horizontal touch screen used to report perceived 3D pose, set by touching a location along the blue circle and fine tuned using keyboard arrows. (Right) Observers viewed the touch screen at an elevation of 70-90 degrees, leading to almost no distortion of vector orientation in retinal image. D. Top-down view of the response coordinate space (0 to 180 degree axis is fronto- parallel to the observer), and 5 locations from which the observer viewed the screen.

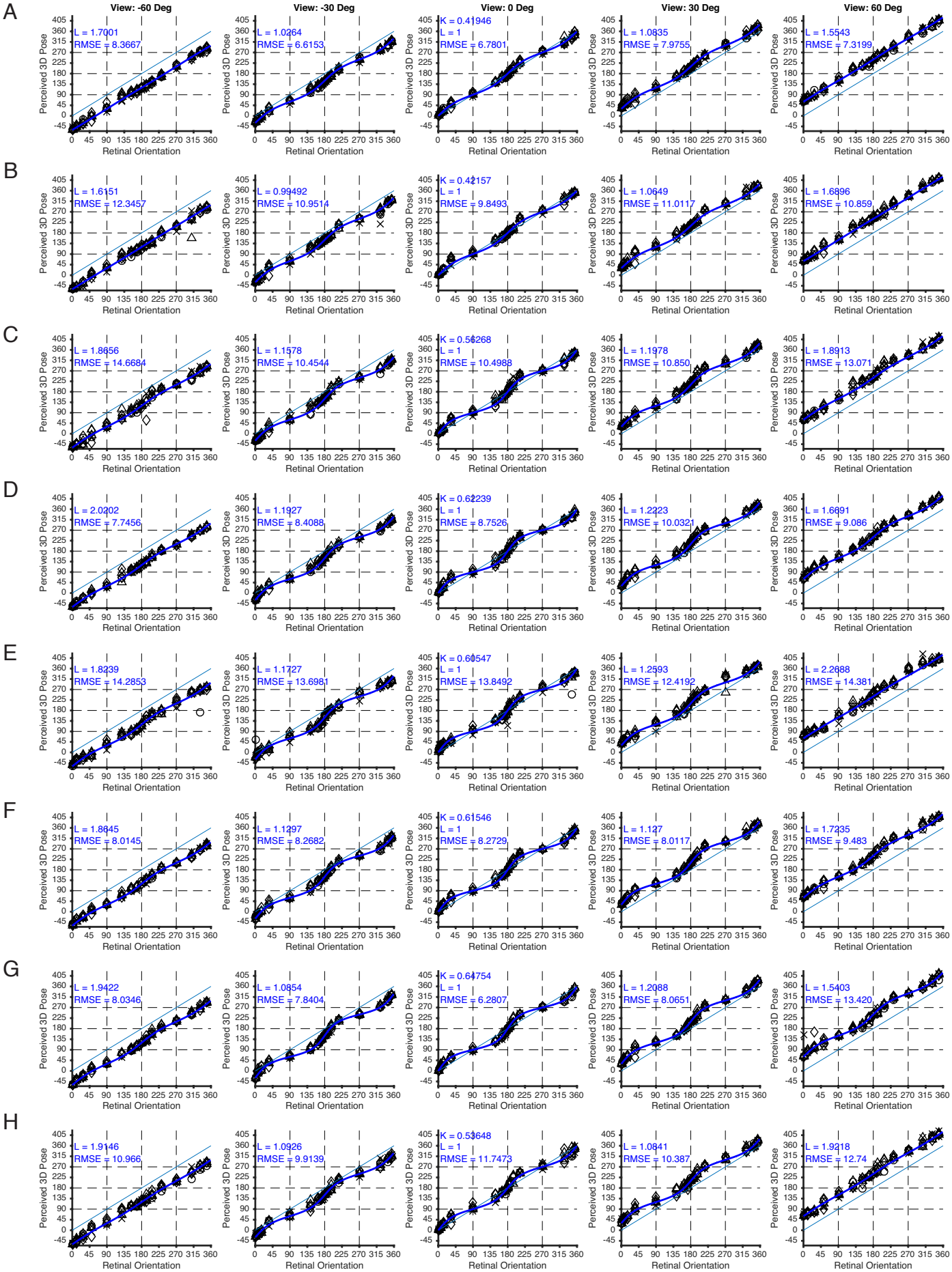

Figure S2: Refers to Figure 2

A-H. Pose estimates for each observer in main experiment. Each observer made 14 judgments for each physical pose, (symbols are as in Figure 1), once binocularly and once monocularly. Blue curves show the best fitting model for each individual at each viewpoint. Parameter value and RMSE are reported in the inset.

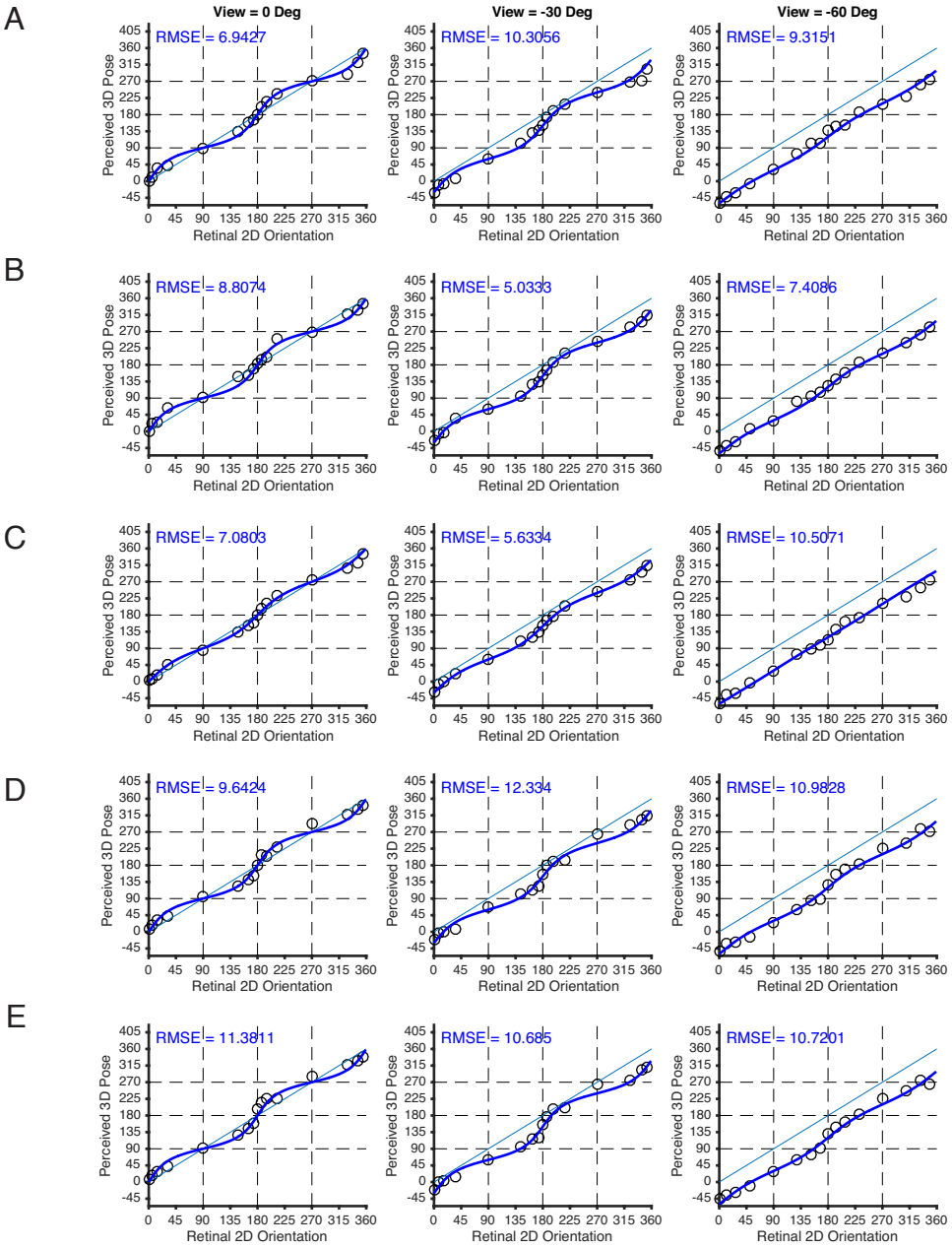

Figure S3: Refers to Figure 3
A-E. Pose estimates for each observer in experiment controlling for frame effects. Open circles blue curves represent the individual model fit from the main experiment (Figure S2). RMSE is reported on the plots.

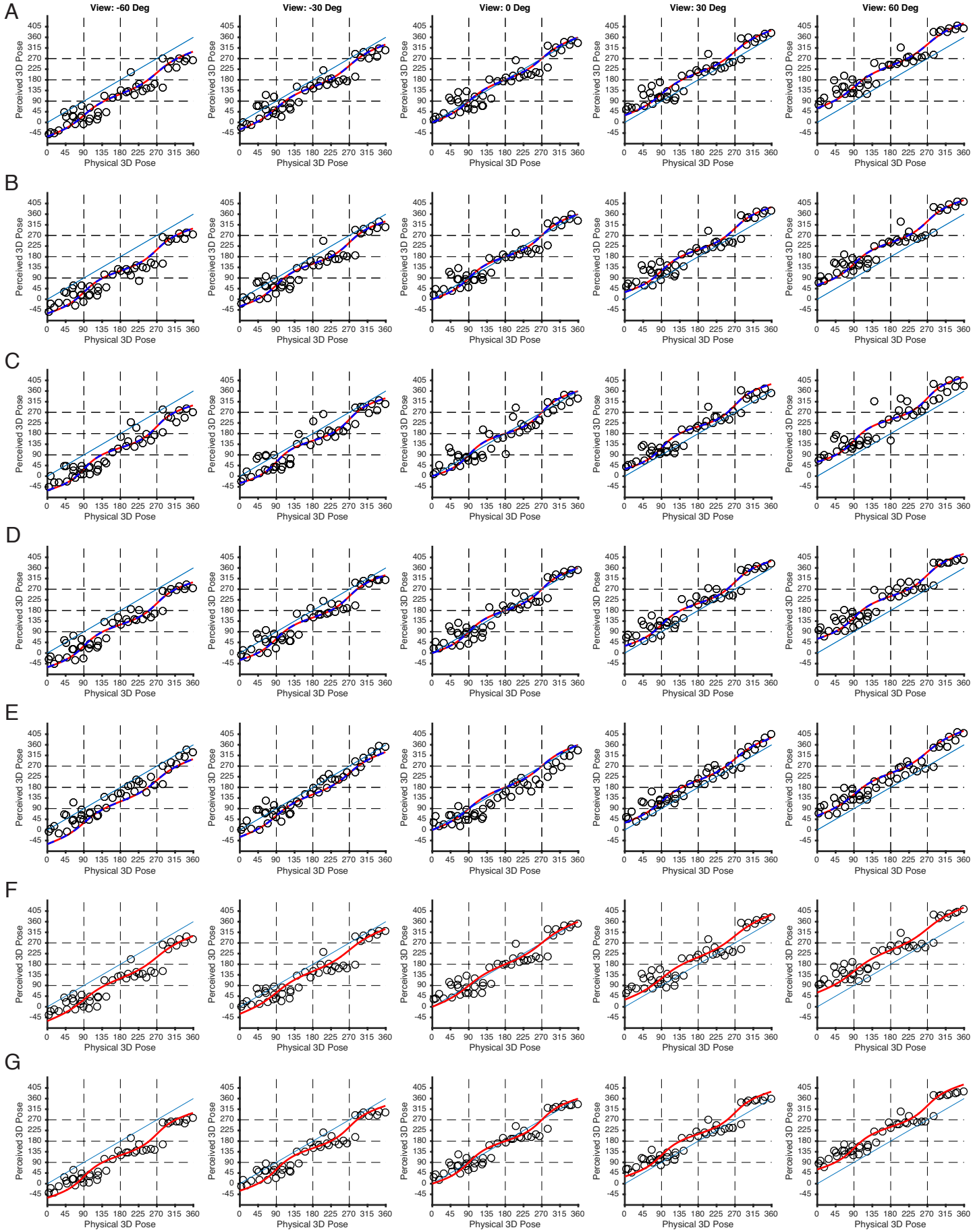

Figure S4: Refers to Figure 4

A-G. Pose estimates for each observer in complex scene experiment. Observers (A-E) were also observers in the main experiment. Observers (F-G) only took part in the complex scene experiment. Blue dashed curves are derived from the individual best fitting models from Figure S2, where applicable. Red curves are derived from the best model fits for the displayed data.