Flaws in Data Binning for Population Receptive Field Analyses

Susanne Stoll*[1], Elisa Infanti[1], Benjamin de Haas[2], and D. Samuel Schwarzkopf[1,3]

[1]University College London | United Kingdom

[2]Justus-Liebig-Universität Gießen | Germany

[3]The University of Auckland | New Zealand

Abstract

Data binning can cope with overplotting and noise, making it a versatile tool for comparing many observations. However, it goes awry if the same observations are used for *binning* and *contrasting*. This creates an inherent circularity, leaving noise and *regression to the mean* insufficiently controlled. Here, we use population receptive field analyses – where data binning is commonplace – as an example to expose this flaw through simulations and empirical repeat data.

Main text

Data binning is often applied to large data sets in order to prevent overplotting and control noise. As such, it has become commonplace in population receptive field (pRF) modeling (Dumoulin & Knapen, 2018; Dumoulin & Wandell, 2008),

*Corresponding author | Email address: stollsus@gmail.com (Susanne Stoll)

where researchers are commonly interested in comparing visual field maps with thousands of observations between different (experimental) conditions. However, pRF modeling is only one out of several research areas where some form of differential data binning has been adopted, such as psychology (Gignac & Zajenkowski, 2020; Holmes, 2009; Preacher, MacCallum, Rucker, & Nicewander, 2005; Shanks, 2017), systems neuroscience (Holmes, 2009; Kriegeskorte, Simmons, Bellgowan, & Baker, 2009), epidemiology (Barnett, van der Pols, & Dobson, 2005), and presumably many more.

Although differential data binning can help us see an overall pattern in the face of an abundance of details, it goes awry if the same noisy observations are used for *binning* (selection) and *contrasting* (selective analysis). This is because dipping into noisy data more than once violates assumptions of independence, favoring some noise components over others, and eventually biasing descriptive and inferential statistics (Kriegeskorte et al., 2009). As such, double-dipping in differential data binning prevents us from – amongst other things – controlling for *regression to the mean* (e.g., Galton, 1886; Gignac & Zajenkowski, 2020; Holmes, 2009; Makin & De Xivry, 2019; Shanks, 2017). Regression to the mean is a statistical phenomenon operating when two variables are imperfectly correlated (e.g., due to random noise). In this case, extreme observations for one variable will on average be less extreme (closer to the mean) for the other variable (Campbell & Kenny, 1999; Cohen, Cohen, West, & Aiken, 2003; Shanks, 2017)[1]. The magnitude of regression to the mean tends to be higher the lower the correlation between the variables.

Regression to the mean and/or double-dipping are of particular concern in what is better known as *post hoc subgrouping* (Preacher et al., 2005), *post hoc data selection* (Shanks, 2017), and *extreme groups approach* (Preacher et al., 2005), all of which can be considered as subtypes of data binning. Post hoc subgrouping refers to collecting two measures, defining extreme subgroups post hoc using one measure (e.g., the lower and upper quantile), and then performing statistics on these mea-

[1] To be precise, regression to the mean refers to standard scores (z-scores; Campbell & Kenny, 1999; Kenny, 2005).

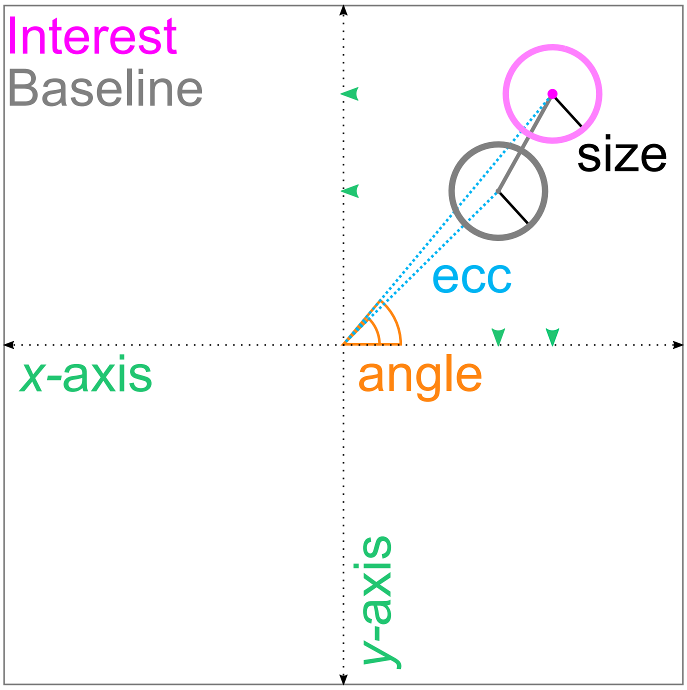


Figure 1. **Population receptive field estimates.** The two circles represent a pRF that changes its position (gray solid line) in an Interest (magenta) compared to a Baseline (gray) condition. The black solid square represents a cutout of the visual field and the black dashed arrows a Cartesian coordinate system. The pRF was modeled as a 2D Gaussian function. The center of the 2D Gaussian (tiny gray dot and small magenta dot) represents the position of the pRF. PRF position can be expressed in terms of x_0 and y_0 coordinates (green arrow heads) or eccentricity (blue dashed line) and polar angles (orange solid line). Eccentricity corresponds to the Euclidean distance between the center of gaze (origin) and the center of the 2D Gaussian. Polar angle corresponds to the counter-clockwise angle running from the positive x-axis to the eccentricity vector. The standard deviation of the Gaussian (1σ; black solid line) represents pRF size. Both pRF position and size are typically expressed in degrees of visual angle. Polar angles are typically expressed in degrees. Ecc = eccentricity. pRF = population receptive field.

sures for the extreme subgroups (Preacher et al., 2005). Post hoc data selection is similar but involves only one extreme subgroup (Shanks, 2017). Both of these practices are different from the extreme groups approach, where extreme subgroups are selected a priori based on one measure; that is, without collecting the whole range of the other measure (Preacher et al., 2005). Here, we focus on a post hoc scenario where essentially all subgroups are considered, not just the extreme ones (see also Gignac & Zajenkowski, 2020; Holmes, 2009). We label this procedure *post hoc binning analysis.*

Imagine we conduct a retinotopic mapping experiment (Dumoulin & Wandell, 2008), where we estimate pRF position and size for each voxel in the brain under a *Baseline* condition as well as a condition of *Interest* (see Figure 1 for a single pRF). We can think of the Interest and Baseline conditions as repeat data (e.g., Benson et al., 2018; van Dijk, de Haas, Moutsiana, & Schwarzkopf, 2016), different attention conditions (e.g, de Haas, Schwarzkopf, Anderson, & Rees, 2014; Klein, Harvey, & Dumoulin, 2014; van Es, Theeuwes, & Knapen, 2018; Vo, Sprague, & Serences, 2017), mapping sequences (e.g., Binda, Thomas, Boynton, & Fine, 2013; Infanti & Schwarzkopf, 2020), mapping stimuli (e.g., Alvarez, de Haas, Clark, Rees, & Samuel Schwarzkopf, 2015; Binda et al., 2013; Le, Witthoft, Ben-Shachar, & Wandell, 2017; Yildirim, Carvalho, & Cornelissen, 2018), scotoma conditions (e.g., Barton & Brewer, 2015; Binda et al., 2013; Haak, Cornelissen, & Morland, 2012; Prabhakaran et al., 2020), pRF modeling techniques (e.g., Carvalho et al., 2020) or uni- and multisensory conditions (Holmes, 2009) – to name but a few examples. As a pRF model, we adopt a 2D Gaussian, where pRF position represents the center of a pRF in visual space (the center of the Gaussian) and pRF size its spatial extent (the standard deviation of the Gaussian; see Figure 1). We then fit this model to the voxel-wise brain responses we measured in the retinotopic mapping experiment (Dumoulin & Wandell, 2008). To compare pRF positions in the Interest and Baseline condition voxel-by-voxel, we bin the pRF positions from both conditions according to the pRF positions from the Baseline condition. Subsequently, we quantify for each voxel the position shift from the Baseline to the Interest condition (see Figure 1 for a single pRF). Finally, we calculate the bin-wise mean shift. This is conceptually equivalent to calculating the bin-wise simple means for each condition and comparing them subsequently, be it descriptively or inferentially.

Either way, by adopting such a post hoc binning analysis, we essentially assume that the mean pRF position we quantify for each bin in the Baseline condition approximates the true mean pRF position. In particular, we presuppose that binning voxels according to pRF positions from the Baseline condition and aggregating them subsequently for this condition ensures that bin-wise noise components cancel out on average (see also Shanks, 2017). This, however, is not the case.

To illustrate this flaw, we generated a simplified contrast scenario with a null ef-

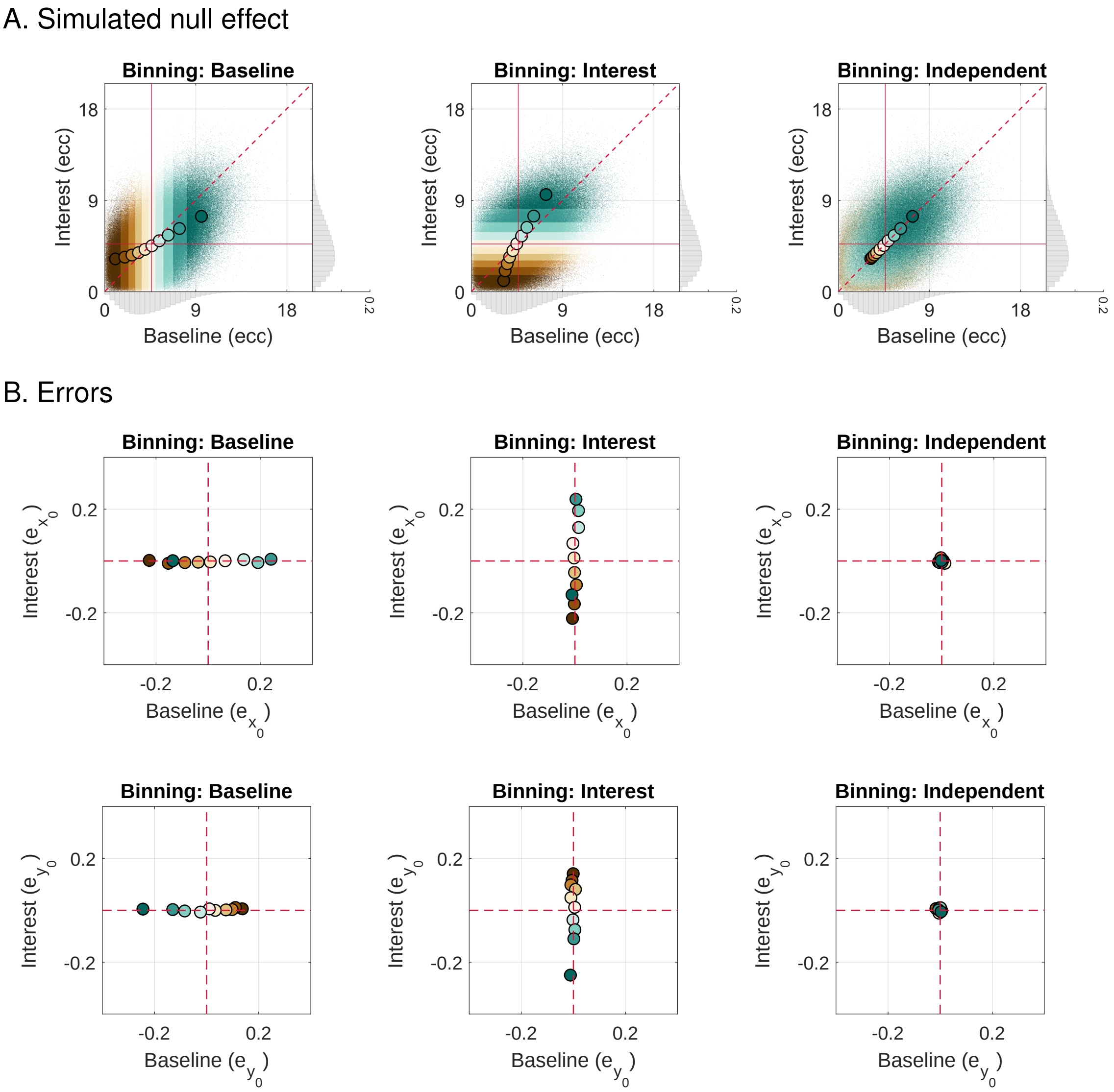


Figure 2. **Simulated 1D post hoc binning analysis on eccentricity | Null effect. A.** Bin-wise eccentricity values and means in the Interest and Baseline condition for a simulated null effect and different data binning scenarios. Eccentricity values in the Baseline and Interest condition were either binned according to eccentricity values in the Baseline (1st column), Interest (2nd column), or an Independent condition (equivalent to repeat data of the Baseline condition; 3rd column). The gray marginal histograms (bin width = 0.5 dva; y-axis: relative frequency) show the simulated eccentricity distributions for each condition, obtained by repeatedly disturbing the x_0 and y_0 values of an empirical visual field map with random Gaussian noise ($sd = 2$ dva) and subsequently converting them to eccentricity values. Note that the range of the marginal y-axis is the same for all histograms. The red crosshair indicates the location of the overall mean for the Interest and Baseline condition. The red dashed line corresponds to the identity line. **B.** Bin-wise mean errors for the x_0 and y_0 values in the Interest and Baseline condition for the same binning scenarios as in A. The dashed red lines reflect the zero error line. Dark brown colors correspond to lower and dark blue-green colors to higher decile bins. The maximal eccentricity of the stimulated visual field area subtended 8.5 dva. Dva = Degrees of visual angle. Ecc = Eccentricity.

fect. In particular, we used random Gaussian noise to repeatedly disturb voxel-wise x_0 and y_0 coordinates (Figure 1) of a V1 visual field map from a single participant ($N_{repeat} = 200$; $sd_{noise} = 2$ degrees of visual angle, dva). We did this twice to generate a Baseline and an Interest condition. We then converted the voxel-wise x_0 and y_0 samples to eccentricity values (Figure 1), as is often done in the pRF literature (see Figure 1-figure supplement 1 for interpretational difficulties with eccentricity when it comes to position shifts). This resulted in a gamma-like eccentricity distribution. Lastly, we binned the eccentricity values in both conditions according to the eccentricity values in the Baseline condition using deciles and calculated the bin-wise means for each condition[2].

We plotted the bin-wise eccentricity means in the Baseline and Interest condition against one another along with individual observations per bin and marginal histograms (bin width = 0.5 dva) reflecting the simulated distributions[3] (Figure 2, A., 1st column). Importantly, since there was no true difference between conditions, the bin-wise means should lie on the identity line. Contrary to this prediction, the bin-wise means systematically diverged from the identity line. Strikingly, when using the Interest instead of the Baseline condition for binning, the systematic pattern of divergence flipped (Figure 2, A., 2nd column). This bidirectionality is a typical sign of regression to the mean (Campbell & Kenny, 1999; Shanks, 2017) and due to circularity that leads to asymmetric bins (see bin-wise ranges of observations for Baseline and Interest, Figure 2, A., 1st and 2nd columns) and biases bin-wise noise components. In particular, for the condition that was used for contrasting and binning (henceforth *circular* condition), the bin-wise noise components of the x_0 and y_0 values were skewed on average. For the other condition (henceforth *non-circular* condition), however, the bin-wise noise components cancelled out on average (Fig-

[2] Note that when evaluating data distributions with unequal means, variances, or non-linearity, z-standardization might be necessary to detect regression to or away from the mean (Campbell & Kenny, 1999; Shanks, 2017). In particular, z-standardization makes data distributions directly comparable. As such, bin-wise means should regress to wherever they intersect the identity line. Here, we always display data in native space, as this is typically done in the pRF literature. However, we use crosshairs to indicate the location of the mean and thus provide a visual guideline.

[3] Note that apart from the visualizations provided here, it might be beneficial to additionally look at Galton squeeze diagrams to detect regression to or away from the mean (Campbell & Kenny, 1999; Shanks, 2017).

ure 2, B., 1st and 2nd columns).

The skew in average noise renders the bin-wise eccentricity means of the circular condition more extreme, especially for lower and higher decile bins. As a result, the bin-wise eccentricity means for the non-circular condition regress – by statistical necessity – to the overall mean[4] for this condition (red crosshair); that is, they are less extreme (see different ranges of bin-wise means for the circular and non-circular conditions in Figure 2, A., 1st and 2nd columns). If the Interest condition is then contrasted to the Baseline condition, a mean increase in eccentricity for lower deciles and a mean decrease for higher deciles or vice versa occurs, depending on whether the data are binned on the Baseline or Interest condition (Figure 2, A., 1st and 2nd column). This artifact arises because we did not use independent conditions for binning and contrasting; that is, conditions with independent noise components.

Importantly, how the artifact manifests can change when data are thresholded across conditions (i.e., corresponding observations are deleted in a pair-wise fashion; Figure 2-figure supplement 1-2, A. and B., 1st and 2nd columns) and/or noise scales with eccentricity (heteroskedasticity; Figure 2-figure supplement 3, A. and B., 1st and 2nd columns; see also Holmes, 2009). In fact, in the event of cross-thresholding, noise components are modified and might not necessarily cancel out for the non-circular condition (Figure 2-figure supplement 1-2, B., 1st and 2nd columns). The case of eccentricity-scaled noise furthermore shows that the artifact can include some clear regression away from the mean[5] (egression; Figure 2-figure supplement 3, A., 1st and 2nd columns; e.g., Campbell & Kenny, 1999; Schwarz & Reike, 2018).

Condition cross-thresholding is common practice in the pRF literature where data are cleaned across conditions according to eccentricity, goodness-of-fit (R^2), pRF size, missing data or other criteria from one or multiple conditions. Eccentricity-scaled noise is an equally likely scenario that might arise from fitting errors due to

[4]Note that for skewed distributions (such as the gamma-like distribution here), the regression effect might be actually towards the mode and away from the mean of the overall distribution (Schwarz & Reike, 2018). If the location of the overall mode and mean are sufficiently close, our visualizations would be unable to distinguish these two cases.

[5]Note that the regression was presumably towards the nearest modes of the simulated bimodal distribution (see marginal histograms in Figure 2-figure supplement 3, A., 1st and 2nd columns; Schwarz & Reike, 2018).

Simulated null effect

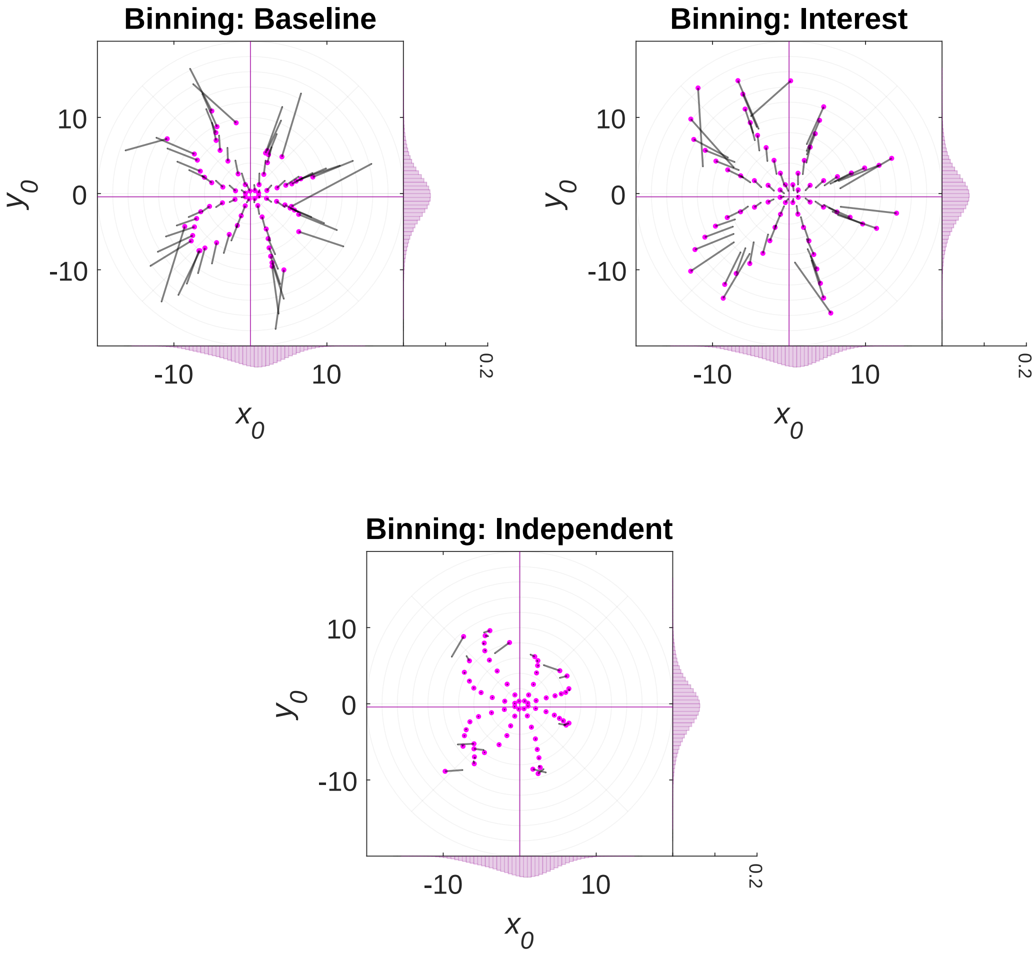


Figure 3. **Simulated 2D post hoc binning analysis on x_0 and y_0 | Null effect.** Bin-wise x_0 and y_0 means in the Interest and Baseline condition for a simulated null effect and different data binning scenarios. X_0 and y_0 values in the Baseline and Interest condition were either binned according to eccentricity and polar angle values in the Baseline (1st column, 1st row), Interest (2nd column, 1st row), or an Independent condition (equivalent to repeat data of the Baseline condition; 2nd row). The marginal histograms (bin width = 0.5 dva; y-axis: relative frequency) show the simulated x_0 and y_0 distributions for each condition, obtained by repeatedly disturbing the x_0 and y_0 values of an empirical visual field map with random Gaussian noise ($sd = 2$ dva). Magenta histograms correspond to the Interest condition and gray histograms to the Baseline condition. Note that the range of the marginal y-axis is the same for all histograms. The large magenta dots (arrow tip) correspond to the means in the Interest condition and the tiny gray dots (arrow knock) to the means in the Baseline condition. The gray line (arrow shaft) depicts the shift from the Baseline to the Interest condition. The magenta crosshair indicates the location of the overall x_0 and y_0 means for the Interest condition and the gray crosshair the location of the overall means for the Baseline condition. Please note that if there is no systematic difference between the Baseline and Interest condition, the histograms and crosshairs coincide (as is the case here). The light gray radar grid demarks the bin segments. Polar angle bins ranged from 0° to 360° with a constant bin width of 45° and eccentricity bins from 0 to 20 dva with a constant bin width of 2 dva. The maximal eccentricity of the stimulated visual field area subtended 8.5 dva. Dva = Degrees of visual angle.

partial stimulation of pRFs (especially near the edge of the stimulated mapping area), higher variability in pRF position estimates for wider pRFs as well as fluctuations in the signal-to-noise ratio of brain responses due to central fixation and/or manipulating attention across visual space[6].

The artifact also replicated when simulating a true effect (i.e., a radial shift of 2 dva in the Interest condition; Figure 2-figure supplement 4, A. and B., 1st and 2nd columns). The same was true for equidistant binning (Figure 2-figure supplement 5, A. and B., 1st and 2nd columns), which is frequently applied in the pRF literature. However, unlike decile binning, equidistant binning resulted in a lower number of observations for higher equidistant bins (due to the gamma-like eccentricity distribution; Figure 2-figure supplement 5, A., 1st and 2nd columns). Consequently, for higher equidistant bins, the skew in average noise for the circular condition was generally larger here. Similarly, for higher equidistant bins, noise components did not always cancel out for the non-circular condition (see all Figure 2-figure supplement 5, B., 1st and 2nd columns). This is because for random noise to cancel out on average, the number of observations needs to be sufficiently large.

For all presented simulation cases, the artifact likewise manifested for another kind of binning analysis, namely, when binning the x_0 and y_0 values according to both eccentricity and polar angle (i.e., 2D segments) and computing shift vectors (Figure 1 as well as Figure 3 and Figure 3-figure supplement 1-4, 1st row). Here, the bin-wise means regressed towards and away from the overall means of the x_0 and y_0 distribution.

Notably, for empirical repeat data from the Human Connectome Project (Benson et al., 2018, 2020), both kinds of binning analyses produced patterns consistent with the artifact (Figure 4-5 and Figure 4-figure supplement 1-3 and Figure 5-figure supplement 1-3, A.-C.). This establishes its practical relevance. Moreover, some of us recently retracted an article on attention-induced differences in pRF position

[6]Note that floor/ceiling effects (due to physiological and methodological constraints on the minimum and maximum observable value) and/or the calculation of absolute (raw) vs proportional (%) differences are further factors influencing the artifact's appearance (de Haas et al., 2014; de Haas, Schwarzkopf, Anderson, & Rees, 2020; Holmes, 2009).

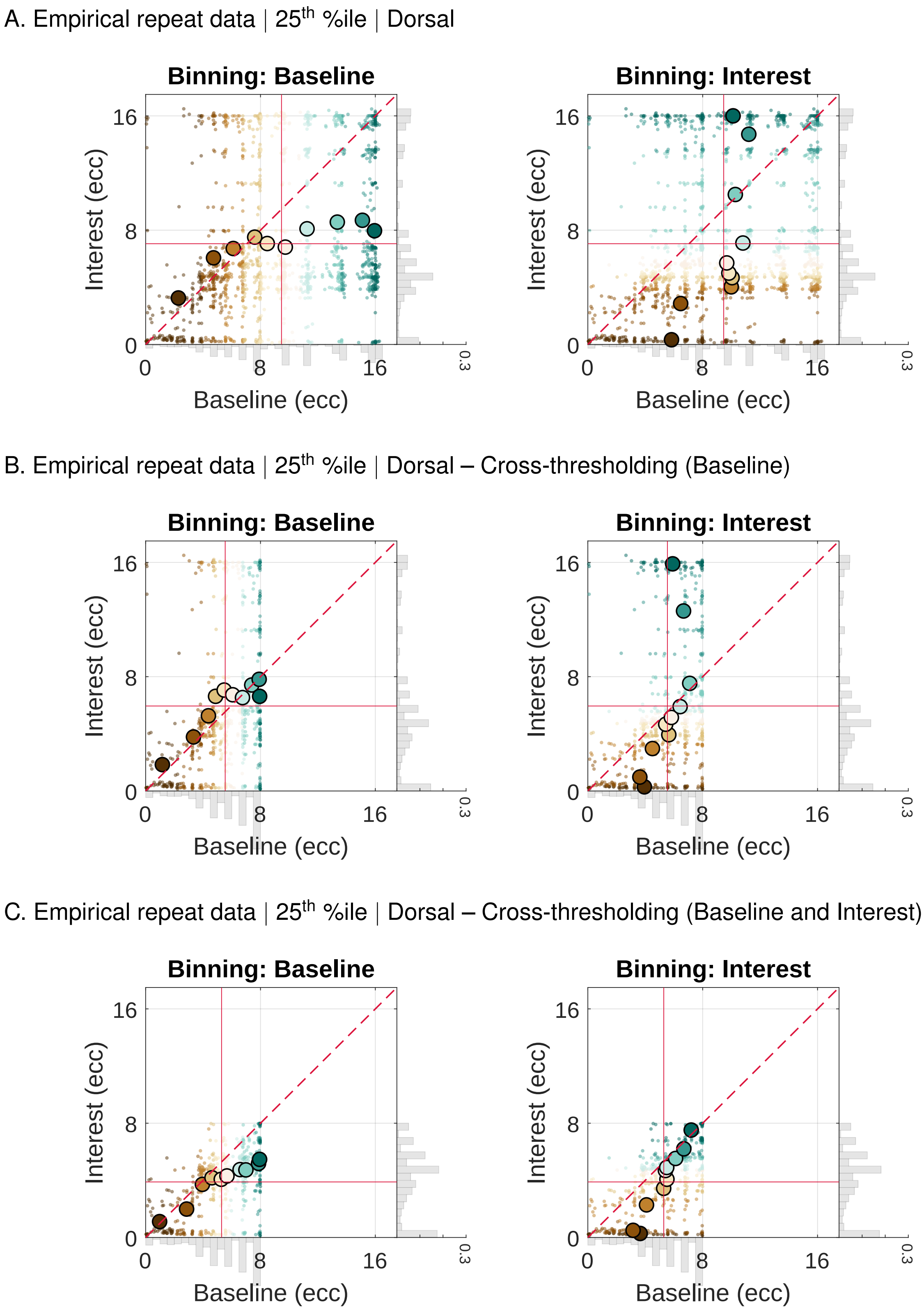


Figure 4. Caption on next page.

Figure 4. **Empirical 1D post hoc binning analysis on eccentricity | Repeat data | 25th %ile participant | Dorsal.** Bin-wise eccentricity values and means in the Interest and Baseline condition for repeat data from the HCP belonging to the 25th %ile participant of the median R^2 distribution and different data binning scenarios. **A.** Data from the dorsal complex (V3A/B and IPS0–5) without condition cross-thresholding. **B.** Same as A., but with condition cross-thresholding. To this end, eccentricity values falling outside a certain eccentricity range ($\geq$ 0 and $\leq$ 8 dva) and below a certain R^2 cut-off ($\leq$ 2.2%) in the Baseline condition were removed from both conditions. **C.** Same as B., although here, condition cross-thresholding was based on both the Baseline and Interest condition. Eccentricity values in the Baseline and Interest condition were either binned according to eccentricity values in the Baseline (1st column in A.-C.) or Interest (2nd column in A.-C.) condition. The gray marginal histograms (bin width = 0.5 dva; y-axis: relative frequency) show the eccentricity distributions for each condition. Note that the range of the marginal y-axis is the same for all histograms. The red crosshair indicates the location of the overall mean for the Interest and Baseline condition. The red dashed line corresponds to the identity line. Dark brown colors correspond to lower and dark blue-green colors to higher decile bins. The maximal eccentricity of the stimulated visual field area subtended 8 dva. HCP = Human connectome project. Dva = Degrees of visual angle. Ecc = Eccentricity. %ile = percentile.

and size in V1-V3 (de Haas et al., 2014) because an in-house reanalysis suggested that post hoc binning along with condition cross-thresholding and heteroskedasticity yielded artifactual (or artifactually inflated) results in the form of egression from the mean (de Haas et al., 2020). In this case, the apparent significant effect was an increase in eccentricity and pRF size in the Interest vs Baseline condition for eccentricity bins in the middle of the tested range.

Taken together, the heterogeneity in manifestation we exposed here makes it hard to spot the artifact by visual inspection alone and highlights its dependency on exact distributional properties of the data at hand (see also Campbell & Kenny, 1999; Holmes, 2009; Schwarz & Reike, 2018, for similar points).

How can we omit double-dipping and control for regression to the mean? We could, for instance, use an Independent condition for binning (such as repeat data or odd or even runs for the Baseline condition; Figure 2 and Figure 2-figure supplement 1-5, A., 3rd column as well as Figure 3 and Figure 3-figure supplement 1-4, 2nd row) or an anatomical criterion (Kriegeskorte et al., 2009), such as cortical distance. This way, noise components should nullify on average in both the Baseline and Interest condition (Figure 2 and Figure 2-figure supplement 1-5, B., third column), albeit not necessarily for sparsely populated bins (Figure 2-figure supplement 5, B., 3rd column as well as Figure 3 and Figure 3-figure supplement 1-3, 2nd row). Similarly, given that cross-thresholding reshapes noise components, they might not

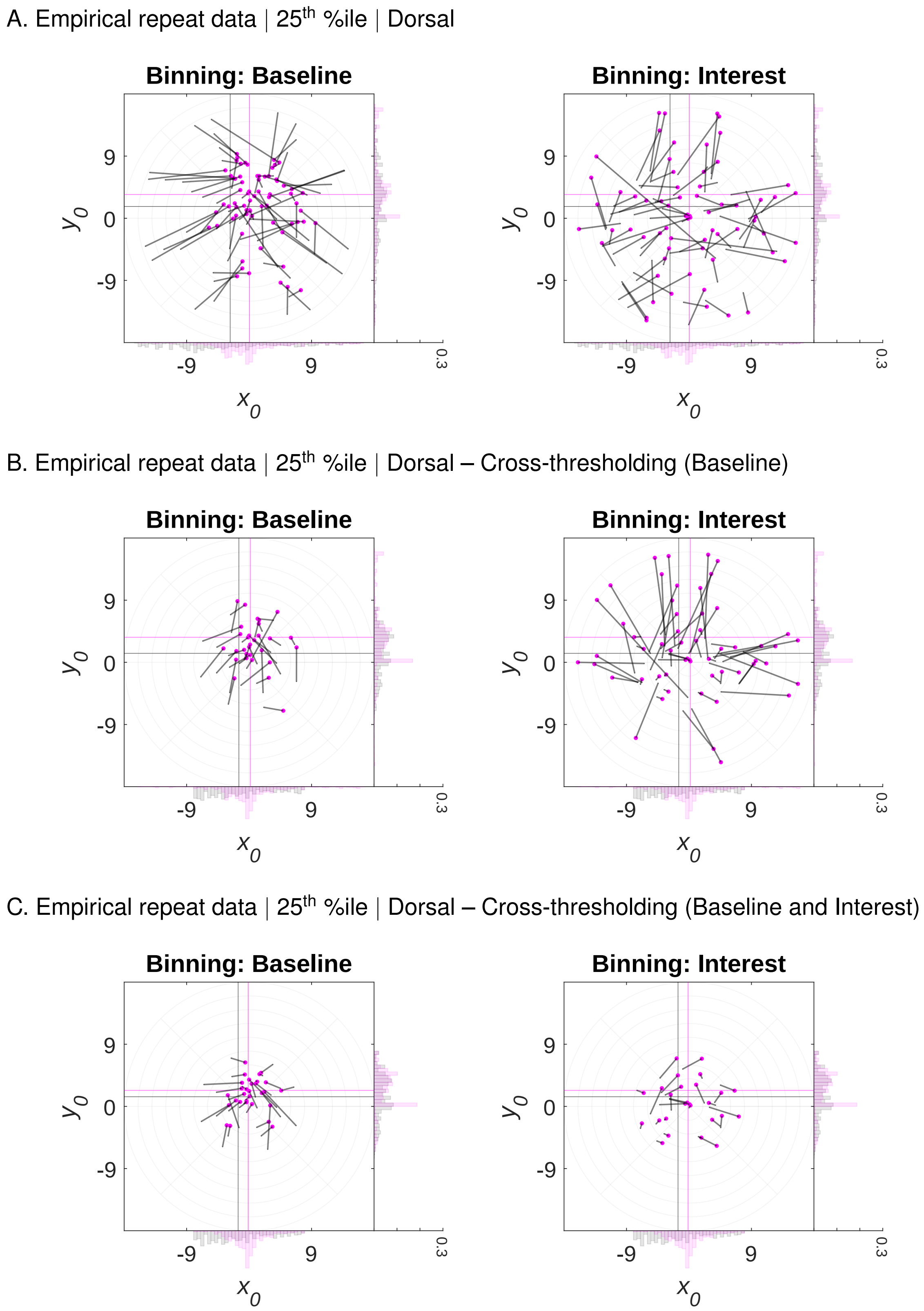


Figure 5. Caption on next page.

Figure 5. **Empirical 2D post hoc binning analysis on x_0 and y_0 | Repeat data | 25th %ile participant | Dorsal.** Bin-wise x_0 and y_0 means in the Interest and Baseline condition for repeat data from the HCP belonging to the 25th percentile participant of the median R^2 distribution and different data binning scenarios. **A.** Data from the dorsal complex (V3A/B and IPS0–5) without condition cross-thresholding. **B.** Same as A., but with condition cross-thresholding. To this end, eccentricity values falling outside a certain eccentricity range (≥ 0 and ≤ 8 dva) and below a certain R^2 cut-off ($\leq 2.2\%$) in the Baseline condition were removed from both conditions. **C.** Same as B., although here, condition cross-thresholding was based on both the Baseline and Interest condition. X_0 and y_0 values in the Baseline and Interest condition were either binned according to eccentricity and polar angle values in the Baseline (1st column in A.-C.) or Interest (2nd column in A.-C.) condition. The marginal histograms (bin width = 0.5 dva; y-axis: relative frequency) show the x_0 and y_0 distributions for each condition. Magenta histograms correspond to the Interest condition and gray histograms to the Baseline condition. Note that the range of the marginal y-axis is the same for all histograms. The large magenta dots (arrow tip) correspond to the means in the Interest condition and the tiny gray dots (arrow knock) to the mean in the Baseline condition. The gray line (arrow shaft) depicts the shift from the Baseline to the Interest condition. The magenta crosshair indicates the location of the overall x_0 and y_0 means for the Interest condition and the gray crosshair the location of the overall means for the Baseline condition. Please note that for subtle differences between the Baseline and Interest condition, the histograms and crosshairs almost coincide (see figure supplements). The light gray radar grid demarks the bin segments. Polar angle bins ranged from 0° to 360° with a constant bin width of 45° and eccentricity bins from 0 to 18 dva with a constant bin width of 2 dva. The maximal eccentricity of the stimulated visual field area subtended 8 dva. HCP = Human connectome project. Dva = Degrees of visual angle.

average out with an Independent condition (Figure 2-figure supplement 1-2, B., 3rd column as well as Figure 3-figure supplement 1-2, 2nd row). The same can evidently also happen with an anatomical criterion if the Baseline and Interest condition are subjected to cross-thresholding. Consequently, unless cross-thresholding can be omitted or demonstrated to be unbiased, an Independent condition might not be a safe option. Alternatively, we could use analyses without binning that control for circularity and regression artifacts or effects could be evaluated against appropriate null distributions that take into account all statistical dependencies (e.g., Holmes, 2009; Kriegeskorte et al., 2009). A combination of these approaches might be most fruitful. Regardless of the specific mitigation strategy, we believe that in light of the many layers of complexity in our analysis pipelines, we need to make it common practice to perform sanity checks using null simulations and empirical repeat data.

Uncontrolled post hoc binning analyses come in many flavours (e.g., centroids, shift vectors, eccentricity differences, x_0 and y_0 differences, and 1D or 2D bins) and are not restricted to pRF position estimates. For instance, biases should manifest equally when binning pRF size in a Baseline and Interest condition according to pRF positions from either of these conditions. Moreover, partial stimulation of pRFs likely results in heteroskedasticity and positively correlated errors for pRF size and position. This would, for instance, bias bin-wise pRF size vs eccentricity or pRF size vs pRF size comparisons where binning is based on non-independent eccentricity values. Likewise, fitting errors due to partial stimulation should be more pronounced whenever pRF size is larger, leading to stronger artifactual effects (for simulations using different levels of noise see Holmes, 2009). The same is to be expected based on a higher variability in pRF position estimates for wider pRFs. These factors might potentially explain why pRF position and size differences have been reported to be larger in higher-level areas where pRFs are wider. Moreover, the distribution of errors likely depends on the toolbox that was used for fitting (Lerma-Usabiaga, Benson, Winawer, & Wandell, 2020), making it hard to generalize across studies. Importantly, uncontrolled single bin (i.e., region of interest) analyses are equally affected by post-hoc binning (Kriegeskorte et al., 2009). And of course, delineations of visual areas in post hoc binning analyses should ideally also be based upon independent criteria as this is where selection starts.

The application of uncontrolled post hoc binning analyses in the pRF literature might have led to spurious claims about the plasticity of pRFs (see de Haas et al., 2014, 2020, for a possible example). Consequently, we urge researcher who engaged in post hoc binning to check for the severity of biases in their analyses by running adequate simulations and reanalyzing the original data wherever possible.

Without doubt, circularity and/or regression to the mean are thorny and omnipresent problems that can manifest subtly and diversely (e.g., Ball, Squeglia, Tapert, & Paulus, 2020; Barnett et al., 2005; Campbell & Kenny, 1999; Eriksson & Häggström, 2014; Gignac & Zajenkowski, 2020; Holmes, 2009; Kilner, 2013; Kriegeskorte et al., 2009; Preacher et al., 2005; Shanks, 2017; Vul, Harris, Winkielman, & Pashler, 2009). As such, we need to ensure that the validation of analysis procedures becomes part and parcel of the scientific process.

Materials and Methods

Post hoc binning using simulations

Stimuli and procedure

For the simulation analyses, we used data from a population receptive field (pRF) experiment involving a dynamic horizontal bar aperture (length of major axis: 17.15 degrees of visual angle, dva; length of minor axis: 1.27 dva). The bar aperture was centered and presented within the boundaries of a circular mapping area (diameter: 17.15 dva). It moved consecutively across the mapping area along cardinal (0/180° and 90/270°) and oblique axes (45/225° and 135/315°) and was superimposed onto a random dot kinematogram (RDK). The RDK comprised moving black dots (diameter: 0.13 dva) positioned within a square field (size: 17.03 × 17.03 dva). If a dot left the square field, it was moved back by 1 field width/height. The dots had a density of 6.89 dots/dva^2, a lifetime of 36 frames, were repositioned randomly once they had died, and oscillated according to a sine wave (A = 1.29 dva, f = 1 Hz, ω = 6.28 rad/s, ϕ = 0 rad). The sine wave was rotated with the current orientation of the bar aperture. The bar aperture and RDK were centered at the screen's midpoint.

A semi-transparent ($\alpha = 50\%$) array of 5 vertical ovals was superimposed onto the bar aperture. One of the ovals was centered at the screen's mid-point (length of major axis: 0.43 dva; length of minor axis: 0.28 dva) and the remaining ovals at an eccentricity of 4.29 dva (length of major axis: 0.86 dva; length of minor axis: 0.57 dva) and different polar angles (45°, 135°, 225°, and 315°). The ovals were presented as a rapid serial visual presentation (RSVP) task, where each trial started with 200 ms of oval presentation, followed by a blank (no ovals) of 600 ms. The ovals' orientation (45° left- or rightwards from vertical) and color (red, yellow, cyan, orange, brown, white, black, green, and blue) changed pseudorandomly in each trial with the exception that ovals of the same color were never presented simultaneously. Participants had to press a button whenever a rightwards oriented oval was presented in blue or green color. A black radar grid (line width: 0.02 dva) at low opacity ($\alpha = 20\%$) with 12 radial lines (at polar angles: 0 to 330° with a step size of 30°) and 18 circles (diameters: 0.95 to 51.42 dva with a step size of 2.97 dva) was superimposed onto the screen. The radial lines ran from the midpoint of the screen to the outermost circle.

The experiment comprised 4 attention conditions, in which participants were required to perform the RSVP task on different oval streams whilst ignoring other streams and the bar aperture. The condition of relevance here is the Center condition, where participants performed the task on the central oval stream. This condition resembles a standard pRF mapping experiment. Participants performed 2 sessions à 4 runs per condition on consecutive days. The order of conditions was pseudorandomized.

Within each run, the bar aperture moved along each axis twice, so that the starting point covered all chosen polar angles. Specifically, the sequence of starting points in each run was: 90°, 225°, 180°, 315°, 270°, 45°, 0°, and 135°. One bar sweep lasted 28 s (1 step/s). Consecutive bar apertures overlapped by 50%. After 4 bar sweeps, a blank interval of 28 s (without the bar apertures and RDK) was presented, during which participants had to refrain from doing the RSVP task (a brief tone cued the beginning and end of this interval). The position and lifetime of each dot in the RDK at the start of every 28s-interval was randomized. Experimental procedures were implemented in Matlab 2014a (8.3; https://uk.mathworks.com/) using Psychtoolbox-3 (3.0.11; Brainard, 1997; Kleiner et al., 2007) and approved by the University College London ethics committee. Written informed consent was obtained from all participants.

Apparatus

Functional and anatomical images were acquired at a field strength of 1.5 T on a Siemens Avanto magnetic resonance imaging (MRI) scanner. All stimuli were projected onto a screen (resolution: 1920 × 1080 pixels; refresh rate: 60 Hz; background color: gray) at the back of the MRI scanner. Participants viewed the experiment through a head-mounted mirror. The viewing distance was approximately 67 cm. To ensure that participants could view the screen without obstruction, the front visor of a 32 channel coil was removed, leaving 30 effective channels.

MRI acquisition

We collected anatomical images using a T1-weighted magnetization-prepared rapid acquisition with gradient echo sequence (repetition time, TR = 2.73 s; echo time, TE = 3.57 ms; voxel size = 1 mm isotropic; flip angle = 7°; field of view, FoV = 256 mm × 224 mm; matrix size = 256 × 224; 176 sagittal slices) and functional images using a T2*-weighted multiband 2D echo-planar imaging sequence (Breuer et al., 2005, TR = 1 s; TE = 55 ms; voxel size = 2.3 mm isotropic; flip angle = 75°; FoV = 224 mm × 224 mm, no gap, matrix size: 96 × 96, acceleration = 4, 36 transverse slices). The slice tab for the functional images was aligned to be roughly parallel to the calcarine sulcus so that the posterior third of the cortex was well covered.

Preprocessing

The initial 10 volumes of each run were discarded to allow for magnetisation to reach equilibrium. Using SPM8 (6313; `https://www.fil.ion.ucl.ac.uk/spm/software/spm8/`), functional images were then bias-corrected, realigned, unwarped, coregistered to the anatomical image, and finally projected onto an anatomical surface model constructed in FreeSurfer (5.3.0; Dale, Fischl, & Sereno, 1999; Fischl, Sereno, & Dale, 1999). We generated vertex-wise functional MRI (fMRI) time series per run by determining the functional voxel at half the distance between corresponding vertices in the pial surface and gray-white matter mesh. We then applied linear detrending to the time series of each run and z-standardized them. Surface projection, detrending, and z-standardization were performed in Matlab 2016b (9.1; `https://uk.mathworks.com/`) using SamSrf7 (7.05; `https://github.com/samsrf/samsrf/tree/3c7a0e25090e9097d5e2fd95696c00774acd26d6`).

PRF estimation and delineations

The vertex-wise preprocessed time series of the Center condition were averaged across the 2 sessions. We then fit a 2D isotropic Gaussian pRF model with 5 free parameters (x_0, y_0, σ, β_0, and β_1) to the vertex-wise average time series. To this end, we first predicted pRF responses by calculating the overlap between the pRF model and an indicator function of the bar aperture for each volume using a 100 × 100 pixel matrix. Specifically, we used a 3D search space of possible values for σ ($8.5\times2^{-5.6:0.2:1}$), x_0 and y_0, and generated pRF responses for each combination of these values. Values for x_0 and y_0 were first sampled from the polar angle system (polar angles: 0:10:350°; eccentricities: $8.5\times2^{-5:0.2:0.6}$) and then transformed to Cartesian coordinates. The pRF response per volume was expressed as mean percent overlap with the pRF model.

To obtain a predicted fMRI time series, we then convolved these pRF responses with a canonical hemodynamic response function (de Haas et al., 2014). Next, we calculated the Pearson correlation between the predicted and the observed fMRI time series and retained the combination of parameter values showing the largest R^2 with all R^2s $\geq .01$. These initial parameter estimates were then used as seeds for an optimization procedure aimed at further maximizing the Pearson correlation between the observed and predicted fMRI time series using a Nelder-Mead algorithm (Lagarias, Reeds, Wright, & Wright, 1998; Nelder & Mead, 1965). Lastly, we estimated β_0 and β_1 by performing linear regression between the observed and predicted time series. The final parameter maps were smoothed with a spherical Gaussian kernel (FWHM = 3mm). Vertices with a very poor R^2 ($< .01$) or artifacts ($\sigma \leq 0$, $\beta_1 \leq 0$ or $\beta_1 > 3$) were removed prior to smoothing. V1 hemifield maps were manually delineated based on smooth polar angle maps using polar angle reversals (Engel, Glover, & Wandell, 1997; Sereno et al., 1995; Wandell, Dumoulin, & Brewer, 2007). These delineations were used as a mask to extract V1 vertices. Fitting, smoothing, and manual delineations were performed in Matlab 2016b (9.1; `https://uk.mathworks.com/`) using SamSrf7 (7.05; `https://github.com/samsrf/samsrf/tree/3c7a0e25090e9097d5e2fd95696c00774acd26d6`).

Simulations

As outlined in the main text, we generated 6 simulation cases: a null effect, a null effect with condition cross-thresholding based on the Baseline condition, a null effect with condition cross-thresholding based on both the Baseline and Interest condition, a null effect with eccentricity-scaled noise, a true effect, and a null effect with equidistant binning (instead of decile binning which was applied to the other cases). These cases were chosen to illustrate a given issue in a clear fashion using an empirical pRF parameter distribution as a basis, not to mimic the exact properties of empirical data (which is unfeasible without explicit knowledge of the noise distribution).

For all simulation cases, x_0 and y_0 estimates from both cortical hemispheres were pooled and empty data points or obvious artifacts removed ($\sigma \leq 0$ and $\beta_1 \leq 0$). Moreover, all simulation cases followed the same general procedure of the null effect involving eccentricity as outlined in the main text (including parameters settings and the same seed for random number generation) with exceptions as follows.

1D post hoc binning analyses on eccentricity. For the simulation cases involving condition cross-thresholding, we removed simulated observations falling outside a certain eccentricity range (≥ 0 and ≤ 6 dva) in the Baseline or Baseline and Interest condition from all conditions (i.e., Baseline, Interest, and Independent). For the simulation case involving eccentricity-scaled noise, we used a small standard deviation ($sd = 0.25$ dva) of random Gaussian noise to disturb original observations with smaller eccentricities (≥ 0 and < 3 dva) and a larger standard deviation ($sd = 2$ dva) to disturb original observations with larger eccentricities (≥ 3 dva). For the simulation case involving a true effect, we induced a radial increase in eccentricity of 2 dva in the Interest condition. For the simulation case involving equidistant binning, we used a constant bin width of 1.75 dva and an overall binning range of 0 to 19.25 dva eccentricity. For all simulation cases, the Independent condition consisted of a second draw (resample) of the Baseline condition.

2D post hoc binning analyses on x_0 and y_0. Apart from a 1D binning analysis on eccentricity, we also conducted a 2D binning analysis on the simulated x_0 and y_0 values. To this end, we converted the x_0 and y_0 values to polar coordinates, that is, polar angle and eccentricity (Figure 1). We then binned the x_0 and y_0 values in the Baseline or Interest condition according to their polar coordinates in the Baseline, Interest, or Independent condition using equidistant bins and calculated the bin-wise x_0 and y_0 means for each condition. The condition-wise means were visualized as vector graphs. The polar angle bins ranged from 0° to 360° with a constant bin width of 45°. The eccentricity bins ranged from 0 to 22 dva (for the simulation case involving a true effect) or from 0 to 20 dva (for all other simulation cases) with a constant bin width of 2 dva. The 2D binning analysis was performed for all aforementioned simulation cases (apart from the case of equidistant binning of course).

Post hoc binning using repeat data

For the repeat data analysis, we used publicly available pRF estimates from the Human Connectome Project 7 T Retinotopy Dataset (Benson et al., 2018, 2020). These estimates stem from a split-half analysis where a 2D isotropic Gaussian with a subadditive exponent was fit to fMRI time series from the first and second half of 6 pRF mapping runs. For each half, 6 estimates were obtained for each grayordinate (vertex), that is, pRF polar angle, pRF eccentricity, pRF size, pRF gain, percentage of R^2, and mean signal intensity. The maximal eccentricity of the mapping area subtended 8 dva. For further details, see Benson et al. (2018).

Following Benson et al. (2018), we analysed complexes of visual areas across hemispheres for the 25th and 75th percentile participants of the R^2 distribution using delineations from Wang et al.'s (2015) atlas. Benson et al. (2018) generated the R^2 distribution by calculating the median R^2 for each participant across grayordinates from both cortical hemispheres within all areas of Wang et al.'s (2015) atlas. The posterior complex consisted of V1-V3, the ventral complex of VO-1/2 and PHC-1/2, the dorsal complex of V3A/B and IPS0–5, and the lateral complex of LO-1/2 and TO-1/2. For our purposes, we focused on the posterior and dorsal complexes, as those came with a larger number of available data points (which was particularly necessary to perform the 2D post hoc binning analysis and generate vector graphs).

To obtain x_0 and y_0 estimates, polar angle and eccentricity estimates were converted to Cartesian coordinates. The eccentricity, x_0, and y_0 estimates of the first half were used as a Baseline condition and those of the second half as an Interest condition. Grayordinates with unusual/implausible values ($R^2 \leq 0\%$ or $\sigma \leq 0$) in either condition were removed from both conditions.

Similar to the simulation-based analyses, binning was either based on the Interest or Baseline condition and bin-wise means were calculated. Moreover, binning was either performed with or without condition cross-thresholding. As for the latter case, we removed observations falling outside a certain eccentricity range (≥ 0 and ≤ 8 dva) or below a certain R^2 cut-off ($\leq 2.2\%$) in the Baseline or Baseline and Interest condition from both conditions. The R^2 cut-off of 2.2% was adopted from Benson et al. (2018).

The 1D binning analysis involving eccentricity and the 2D binning analysis involving x_0 and y_0 were conducted as for the simulated data, although, here, the eccentricity bins for the 2D analysis ranged from 0 to 18 dva with a constant bin width of 2 dva. All binning analyses (including those on simulated data) were implemented in Matlab 2016b (9.1; https://uk.mathworks.com/) using custom code.

Data and code availability

Preprocessed data, custom code, and figures are available at https://doi.org/10.17605/OSF.IO/WJADP.

Acknowledgements

This research was supported by European Research Council Starting Grants to DSS (WMOSPOTWU, 310829) and BdH (INDIVISUAL, 852885). BdH was further supported by the Deutsche Forschungsgemeinschaft (222641018–SFB/TRR 135 TP A8).

Declaration of competing interest

The authors declare no conflict of interest.

References

Alvarez, I., de Haas, B., Clark, C. A., Rees, G., & Samuel Schwarzkopf, D. (2015). Comparing different stimulus configurations for population receptive field mapping in human fMRI. *Front. Hum. Neurosci.*, *9*, 96. doi: 10.3389/fnhum.2015.00096

Ball, T. M., Squeglia, L. M., Tapert, S. F., & Paulus, M. P. (2020). Double dipping in machine learning: Problems and solutions. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging*, *5*(3), 261–263. doi: 10.1016/j.bpsc.2019.09.003

Barnett, A. G., van der Pols, J. C., & Dobson, A. J. (2005). Regression to the mean: What it is and how to deal with it. *Int. J. Epidemiol.*, *34*(1), 215–220. doi: 10.1093/ije/dyh299

Barton, B., & Brewer, A. A. (2015). fMRI of the rod scotoma elucidates cortical rod pathways and implications for lesion measurements. *Proc. Natl. Acad. Sci. U. S. A.*, *112*(16), 5201–5206. doi: 10.1073/pnas.1423673112

Benson, N. C., Jamison, K. W., Arcaro, M. J., Vu, A. T., Glasser, M. F., Coalson, T. S., … Kay, K. (2018). The Human Connectome Project 7 Tesla retinotopy dataset: Description and population receptive field analysis. *J. Vis.*, *18*(13), 1–22. doi: 10.1167/18.13.23

Benson, N. C., Jamison, K. W., Arcaro, M. J., Vu, A. T., Glasser, M. F., Coalson, T. S., … Kay, K. (2020). *The HCP 7T Retinotopy Dataset.* Retrieved from osf.io/bw9ec

Binda, P., Thomas, J. M., Boynton, G. M., & Fine, I. (2013). Minimizing biases in estimating the reorganization of human visual areas with bold retinotopic mapping. *J. Vis.*, *13*(7), 1–16. doi: 10.1167/13.7.14

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spat. Vis.*, *10*(4), 433–436. doi: 10.1163/156856897X00357

Breuer, F. A., Blaimer, M., Heidemann, R. M., Mueller, M. F., Griswold, M. A., & Jakob, P. M. (2005). Controlled aliasing in parallel imaging results in higher acceleration (CAIPIRINHA) for multi-slice imaging. *Magn. Reson. Med.*, *53*(3), 684–691. doi: 10.1002/mrm.20401

Campbell, D. T., & Kenny, D. A. (1999). *A primer on regression artifacts*. New York, NY: Guilford Press.

Carvalho, J., Invernizzi, A., Ahmadi, K., Hoffmann, M. B., Renken, R. J., & Cornelissen, F. W. (2020). Micro-probing enables fine-grained mapping of neuronal populations using fMRI. *Neuroimage*, *209*, 116423. doi: 10.1016/j.neuroimage.2019.116423

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*. Mahwah, NJ: Lawrence Erlbaum Associates.

Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis: I. Segmentation and surface reconstruction. *Neuroimage*, *9*(2), 179–194. doi: 10.1006/nimg.1998.0395

de Haas, B., Schwarzkopf, D. S., Anderson, E. J., & Rees, G. (2014). Perceptual load affects spatial tuning of neuronal populations in human early visual cortex. *Curr. Biol.*, *24*(2), R66–R67. doi: 10.1016/j.cub.2013.11.061

de Haas, B., Schwarzkopf, D. S., Anderson, E. J., & Rees, G. (2020). Retraction Notice to: Perceptual load affects spatial tuning of neuronal populations in human early visual cortex. *Curr. Biol.*, *30*(23), 4814. doi: 10.1016/j.cub.2020.11.015

Dumoulin, S. O., & Knapen, T. (2018). How visual cortical organization is altered by ophthalmologic and neurologic disorders. *Annu. Rev. Vis. Sci.*, *4*(1), 357–379. doi: 10.1146/annurev-vision-091517-033948

Dumoulin, S. O., & Wandell, B. A. (2008). Population receptive field estimates in human visual cortex. *Neuroimage*, *39*(2), 647–660. doi: 10.1016/j.neuroimage.2007.09.034

Engel, S. A., Glover, G. H., & Wandell, B. A. (1997). Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cereb. Cortex*, *7*(2), 181–192. doi: 10.1093/cercor/7.2.181

Eriksson, K., & Häggström, O. (2014). Lord's paradox in a continuous setting and a regression artifact in numerical cognition research. *PLoS One*, *9*(4), e95949. doi: 10.1371/journal.pone.0095949

Fischl, B., Sereno, M. I., & Dale, A. M. (1999). Cortical surface-based analysis: II. Inflation, flattening, and a surface-based coordinate system. *Neuroimage*, *9*(2), 195–207. doi: 10.1006/nimg.1998.0396

Galton, F. (1886). Regression towards mediocrity in hereditary stature. *J. Anthropol. Inst. Gt. Britain Irel.*, *1*, 246–263. doi: 10.1017/CBO9781107415324.004

Gignac, G. E., & Zajenkowski, M. (2020). The Dunning-Kruger effect is (mostly) a statistical artefact: Valid approaches to testing the hypothesis with individual differences data. *Intelligence*, *80*. doi: 10.1016/j.intell.2020.101449

Haak, K. V., Cornelissen, F. W., & Morland, A. B. (2012). Population receptive field dynamics in human visual cortex. *PLoS One*, *7*(5), e37686. doi: 10.1371/journal.pone.0037686

Holmes, N. P. (2009). The principle of inverse effectiveness in multisensory integration: Some statistical considerations. *Brain Topogr.*, *21*(3-4), 168–176. doi: 10.1007/s10548-009-0097-2

Infanti, E., & Schwarzkopf, D. S. (2020). Mapping sequences can bias population receptive field estimates. *Neuroimage*, *211*, 116636. doi: 10.1016/j.neuroimage.2020.116636

Kenny, D. A. (2005). Regression artifacts. In B. S. Everitt & D. C. Howell (Ed.), *Encycl. stat. behav. sci.* (pp. 1723–1725). Chichester, UK.

Kilner, J. M. (2013). Bias in a common EEG and MEG statistical analysis and how to avoid it. *Clin. Neurophysiol.*, *124*(10), 2062–2063. doi: 10.1016/j.clinph.2013.03.024

Klein, B. P., Harvey, B. M., & Dumoulin, S. O. (2014). Attraction of position preference by spatial attention throughout human visual cortex. *Neuron*, *84*(1), 227–237. doi: 10.1016/j.neuron.2014.08.047

Kleiner, M., Brainard, D. H., Pelli, D. G., Broussard, C., Wolf, T., & Niehorster, D. (2007). What's new in Psychtoolbox-3? [ECVP abstract supplement]. *Perception*, *36*. doi: http://journals.sagepub.com/doi/10.1177/03010066070360S101

Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S., & Baker, C. I. (2009). Circular analysis in systems neuroscience: The dangers of double dipping. *Nat.*

Neurosci., *12*(5), 535–540. doi: 10.1038/nn.2303

Lagarias, J. C., Reeds, J. A., Wright, M. H., & Wright, P. E. (1998). Convergence properties of the Nelder-Mead simplex method in low dimensions. *SIAM J. Optim.*, *9*(1), 112–147. doi: 10.1137/S1052623496303470

Le, R., Witthoft, N., Ben-Shachar, M., & Wandell, B. (2017). The field of view available to the ventral occipito-temporal reading circuitry. *J. Vis.*, *17*(4), 1–19. doi: 10.1167/17.4.6

Lerma-Usabiaga, G., Benson, N., Winawer, J., & Wandell, B. A. (2020). A validation framework for neuroimaging software: The case of population receptive fields. *PLOS Comput. Biol.*, *16*(6), e1007924. doi: 10.1371/journal.pcbi.1007924

Makin, T. R., & De Xivry, J. J. O. (2019). Ten common statistical mistakes to watch out for when writing or reviewing a manuscript. *Elife*, *8*. doi: 10.7554/eLife.48175

Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *Comput. J.*, *7*(4), 308–313. doi: 10.1093/comjnl/7.4.308

Prabhakaran, G. T., Carvalho, J., Invernizzi, A., Kanowski, M., Renken, R. J., Cornelissen, F. W., & Hoffmann, M. B. (2020). Foveal pRF properties in the visual cortex depend on the extent of stimulated visual field. *Neuroimage*, *222*, 117250. doi: 10.1016/j.neuroimage.2020.117250

Preacher, K. J., MacCallum, R. C., Rucker, D. D., & Nicewander, W. A. (2005). Use of the extreme groups approach: A critical reexamination and new recommendations. *Psychol. Methods*, *10*(2), 178–192. doi: 10.1037/1082-989X.10.2.178

Schwarz, W., & Reike, D. (2018). Regression away from the mean: Theory and examples. *Br. J. Math. Stat. Psychol.*, *71*(1), 186–203. doi: 10.1111/bmsp.12106

Sereno, M. I., Dale, A. M., Reppas, J. B., Kwong, K. K., Belliveau, J. W., Brady, T. J., . . . Tootell, R. B. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science (80-.).*, *268*(5212), 889–893. doi: 10.1126/science.7754376

Shanks, D. R. (2017). Regressive research: The pitfalls of post hoc data selection in the study of unconscious mental processes. *Psychon. Bull. Rev.*, *24*(3), 752–775. doi: 10.3758/s13423-016-1170-y

van Dijk, J. A., de Haas, B., Moutsiana, C., & Schwarzkopf, D. S. (2016). Inter-session reliability of population receptive field estimates. *Neuroimage*, *143*, 293–303. doi: 10.1016/j.neuroimage.2016.09.013

van Es, D. M., Theeuwes, J., & Knapen, T. (2018). Spatial sampling in human visual cortex is modulated by both spatial and feature-based attention. *Elife*, *7*. doi: 10.7554/eLife.36928

Vo, V. A., Sprague, T. C., & Serences, J. T. (2017). Spatial tuning shifts increase the discriminability and fidelity of population codes in visual cortex. *J. Neurosci.*, *37*(12), 3386–3401. doi: 10.1523/JNEUROSCI.3484-16.2017

Vul, E., Harris, C., Winkielman, P., & Pashler, H. (2009). Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. *Perspect. Psychol. Sci.*, *4*(3), 319–324. doi: 10.1111/j.1745-6924.2009.01132.x

Wandell, B. A., Dumoulin, S. O., & Brewer, A. A. (2007). Visual field maps in human cortex. *Neuron*, *56*(2), 366–383. doi: 10.1016/j.neuron.2007.10.012

Wang, L., Mruczek, R. E., Arcaro, M. J., & Kastner, S. (2015). Probabilistic maps of visual topography in human cortex. *Cereb. Cortex*, *25*(10), 3911–3931. doi: 10.1093/cercor/bhu277

Yildirim, F., Carvalho, J., & Cornelissen, F. W. (2018). A second-order orientation-contrast stimulus for population-receptive-field-based retinotopic mapping. *Neuroimage*, *164*, 183–193. doi: 10.1016/j.neuroimage.2017.06.073

558 # Supplementary figures

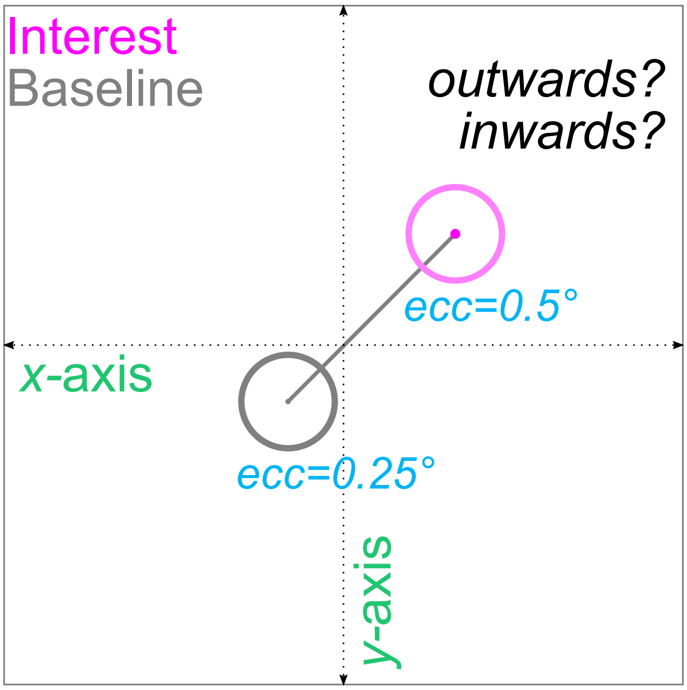


Figure 1-figure supplement 1. **Interpretation of changes in eccentricity.** The same as Figure 1, although here, the pRFs shifts from one visual field quadrant to another in the Interest compared to the Baseline condition. This can happen due to noise or when visual field maps partially cover the ipsilateral hemifield. In such cases, an increase or decrease in eccentricity does not necessarily correspond to an outwards or inwards shift in the traditional sense. For instance, imagine that a pRF sits at x_0 = -0.18 dva and y_0 = -0.18 dva in the Baseline condition (ecc = 0.25 dva) but at x_0 = 0.36 dva and y_0 = 0.36 dva in the Interest condition (ecc = 0.5 dva). This would result in an increase in eccentricity, which might be interpreted as an outwards shift, although the pRF shifts effectively radially inwards until it reaches the origin and then outwards. We can likewise imagine that the pRF shifts horizontally to x_0 = 0.36 dva and y_0 = -0.36 dva in the Interest condition. Importantly, removing such shifts would again bias noise components in non-predictable ways (see condition cross-thresholding in the main text and Figure 2-figure supplement 1-2) and therefore does not seem a valid option. Dva = Degrees of visual angle.

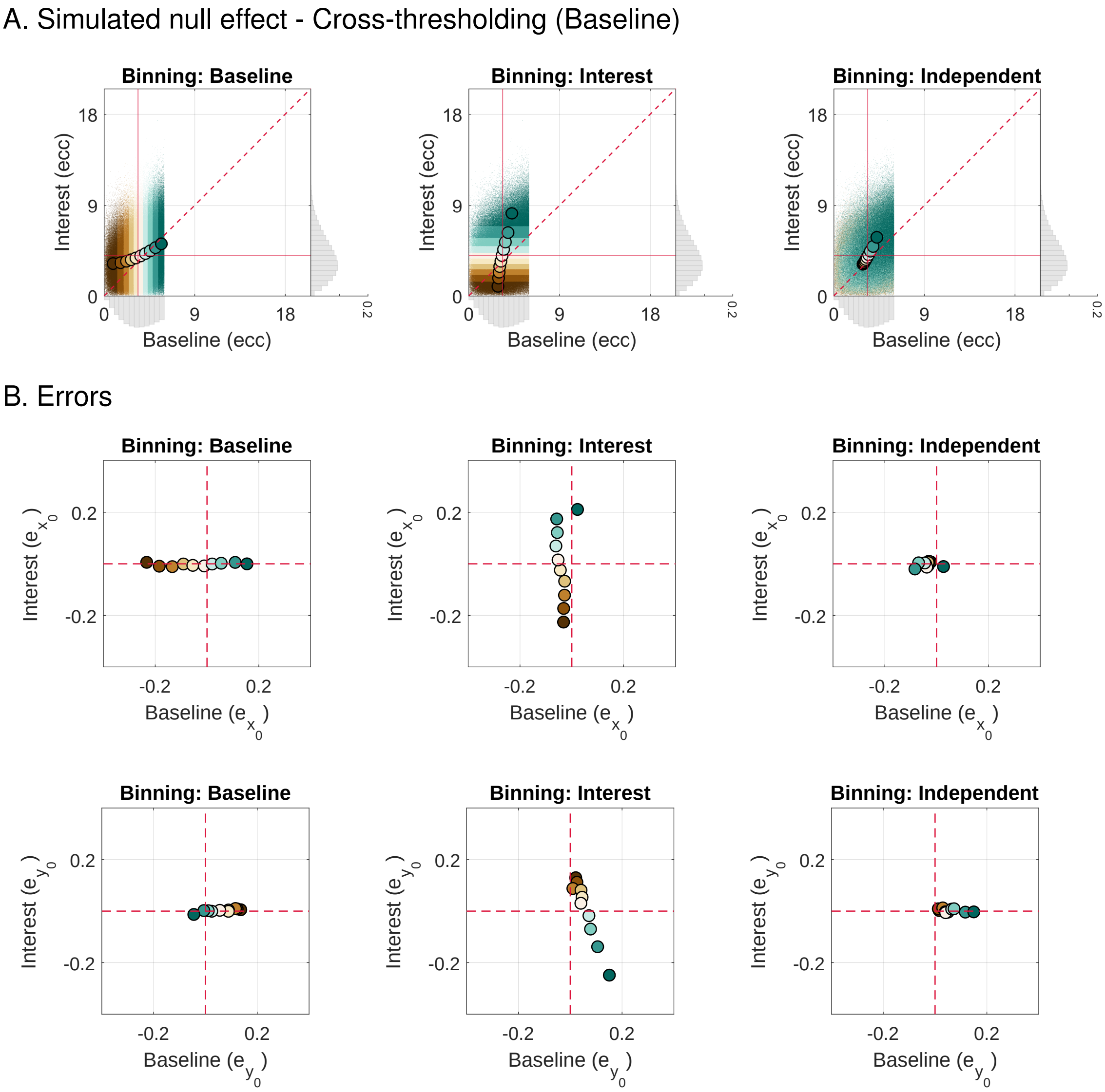


Figure 2-figure supplement 1. **Simulated 1D post hoc binning analysis on eccentricity | Null effect — Cross-thresholding (Baseline).** The same as in Figure 2, although here, simulated observations falling outside a certain eccentricity range (≥ 0 and ≤ 6 dva) in the Baseline condition were removed from all conditions — a simulation case we term *cross-thresholding (Baseline)*.

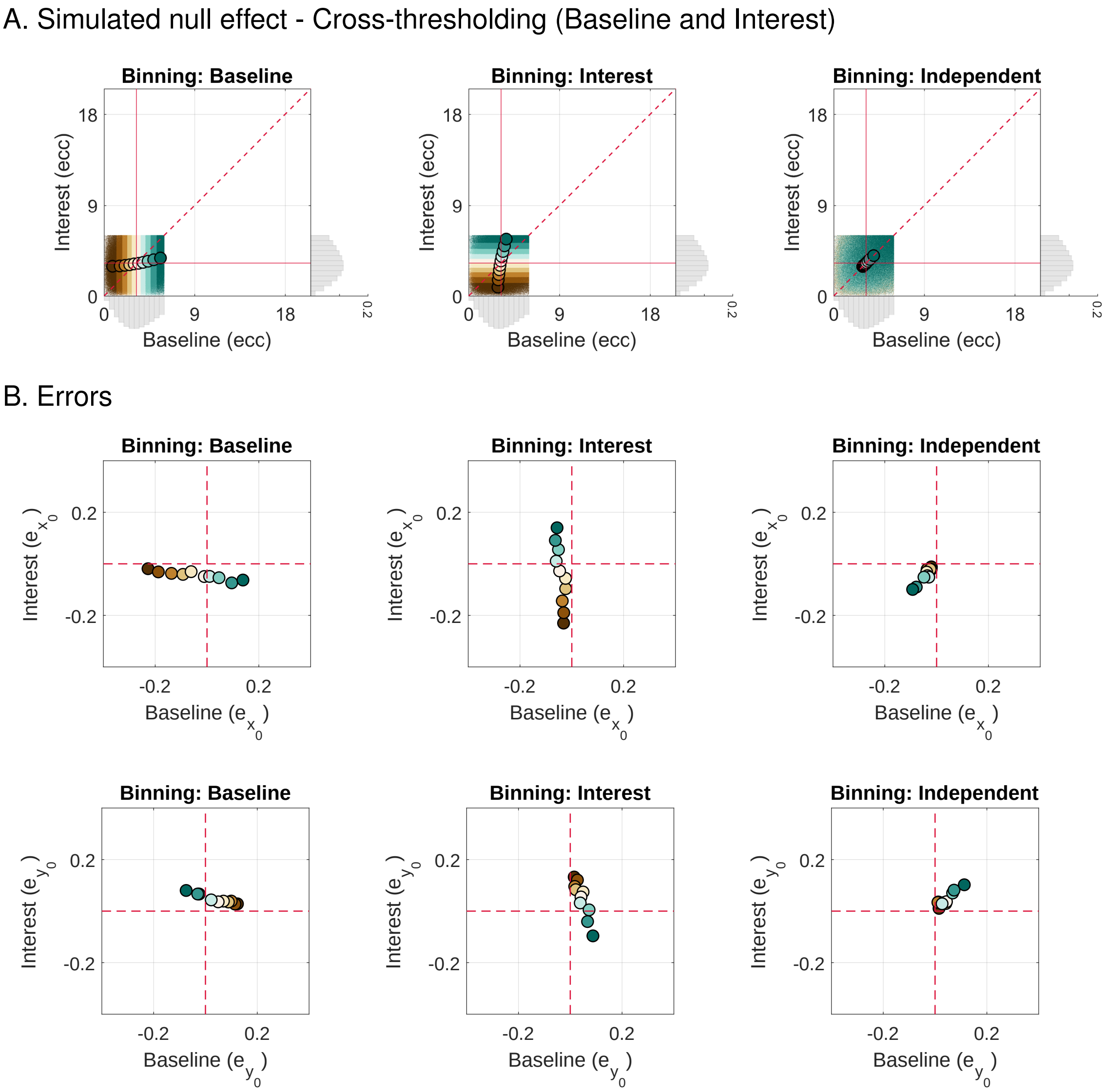


Figure 2-figure supplement 2. **Simulated 1D post hoc binning analysis on eccentricity | Null effect — Cross-thresholding (Baseline and Interest).** The same as in Figure 2-figure supplement 1, although here, condition cross-thresholding was based on both the Baseline and Interest condition — a simulation case we term *cross-thresholding (Baseline and Interest)*.

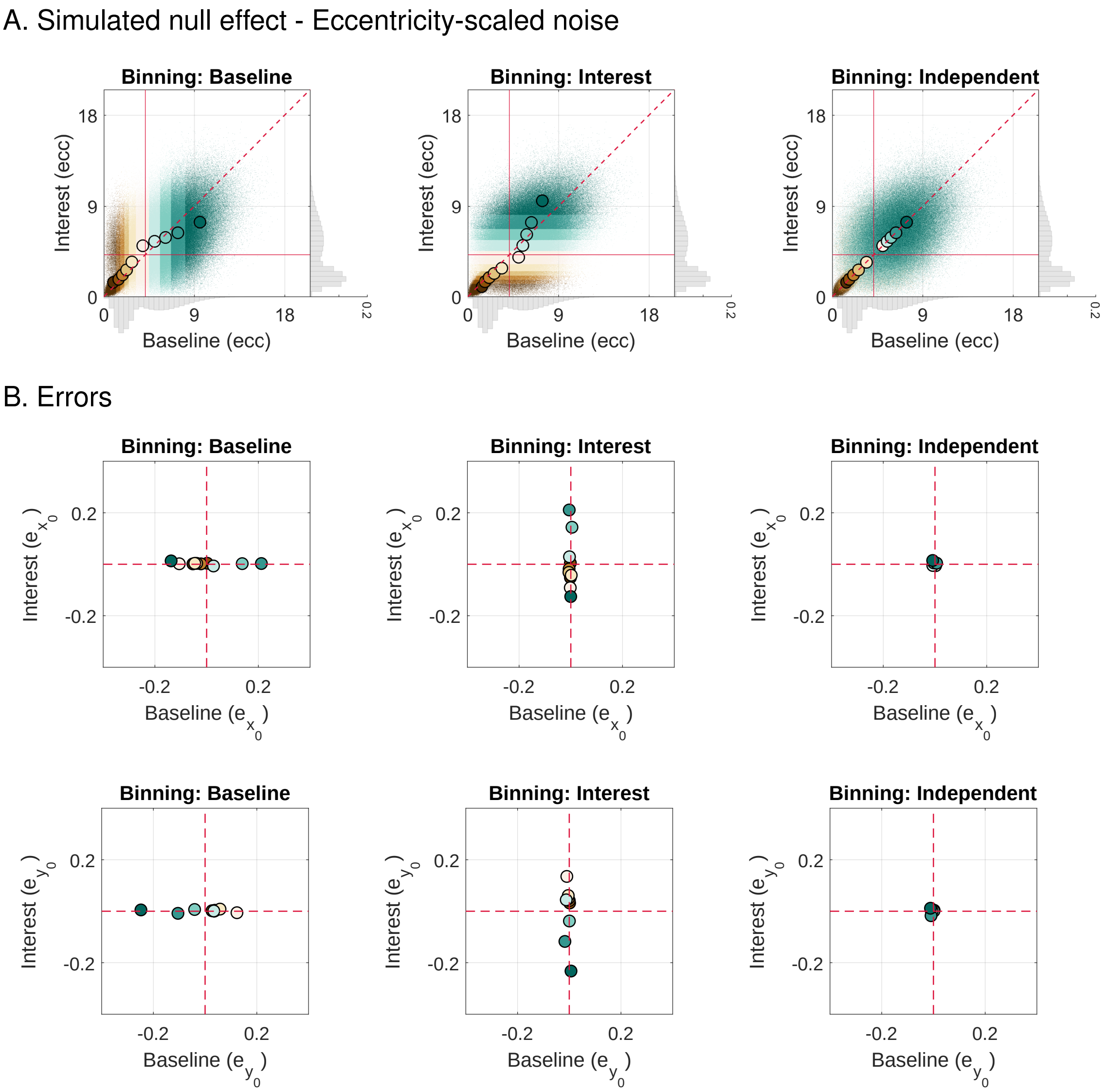


Figure 2-figure supplement 3. **Simulated 1D post hoc binning analysis on eccentricity | Null effect — Eccentricity-scaled noise.** The same as in Figure 2, although here, original observations having smaller eccentricities ($\geq$ 0 and $<$ 3 dva) were disturbed by random Gaussian noise with a smaller standard deviation ($sd =$ 0.25 dva) and those having larger eccentricities ($\geq$ 3 dva) by random Gaussian noise with a larger standard deviation ($sd =$ 2 dva) — a simulation case we term *eccentricity-scaled noise*.

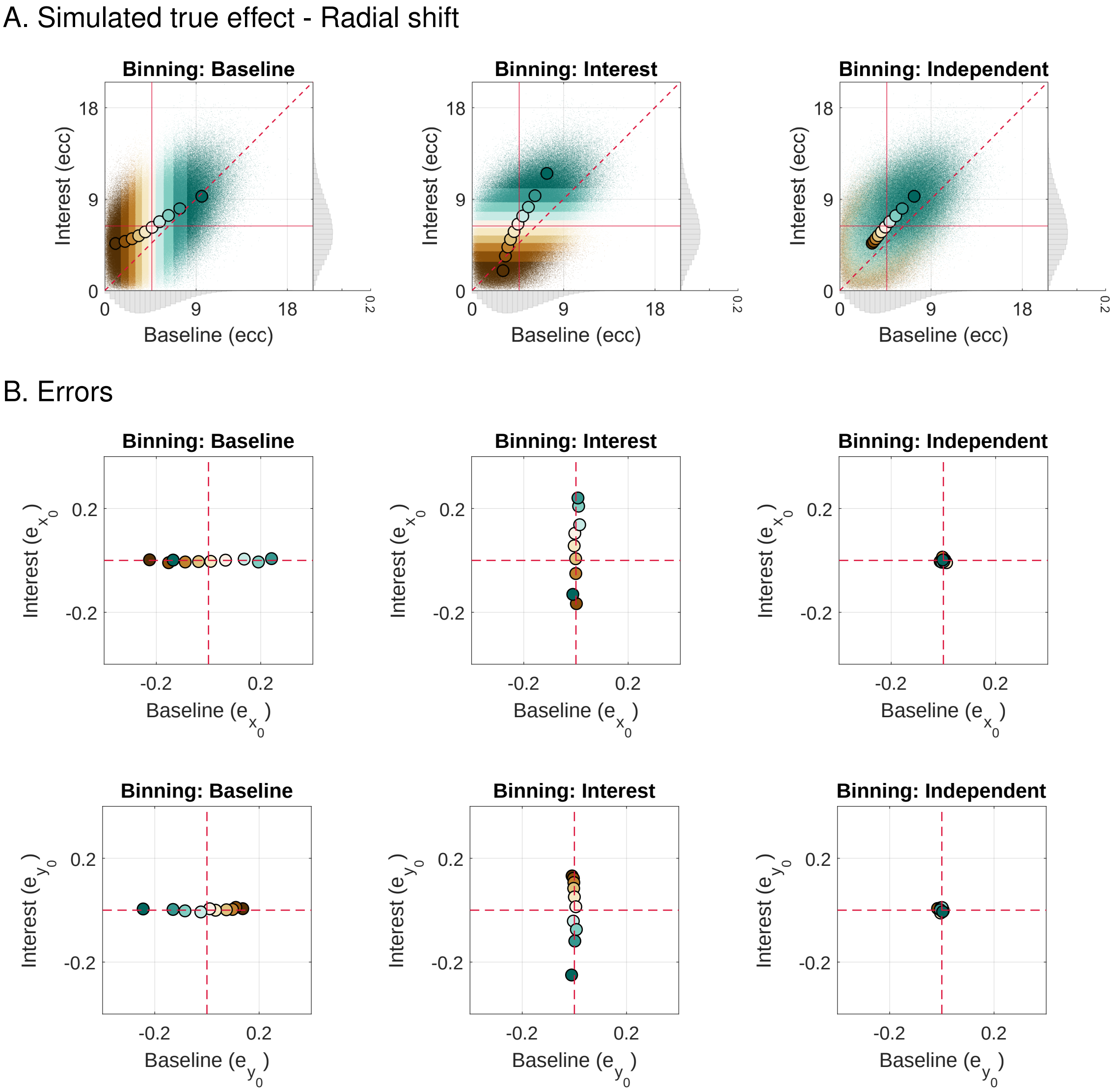


Figure 2-figure supplement 4. **Simulated 1D post hoc binning analysis on eccentricity | True effect — Radial shift.** The same as in Figure 2, although here, we simulated a true effect, that is, a radial increase in eccentricity of 2 dva in the Interest as compared to the Baseline condition.

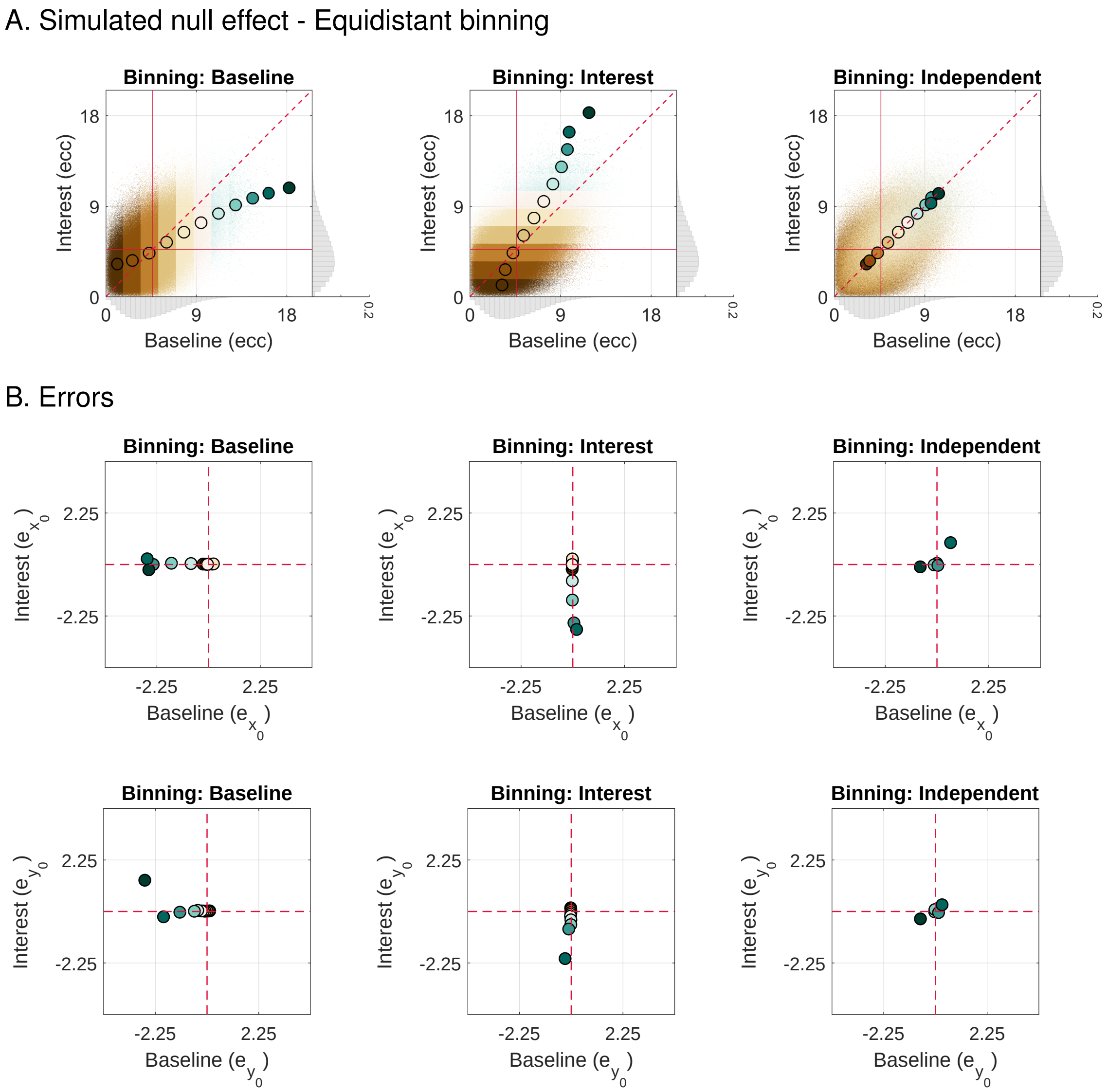


Figure 2-figure supplement 5. **Simulated 1D post hoc binning analysis on eccentricity | Null effect — Equidistant binning.** The same as in Figure 2, although here, equidistant binning was used. The equidistant bins ranged from an eccentricity of 0 dva to an eccentricity of 19.25 dva with a constant bin-width of 1.75 dva. Please note the different x- and y-axis ranges in B. relative to Figure 2 and other figure supplements (-4.5 to -4.5 vs -0.4 to 0.4, respectively) as well as the different number of bins (11 vs 10, respectively).

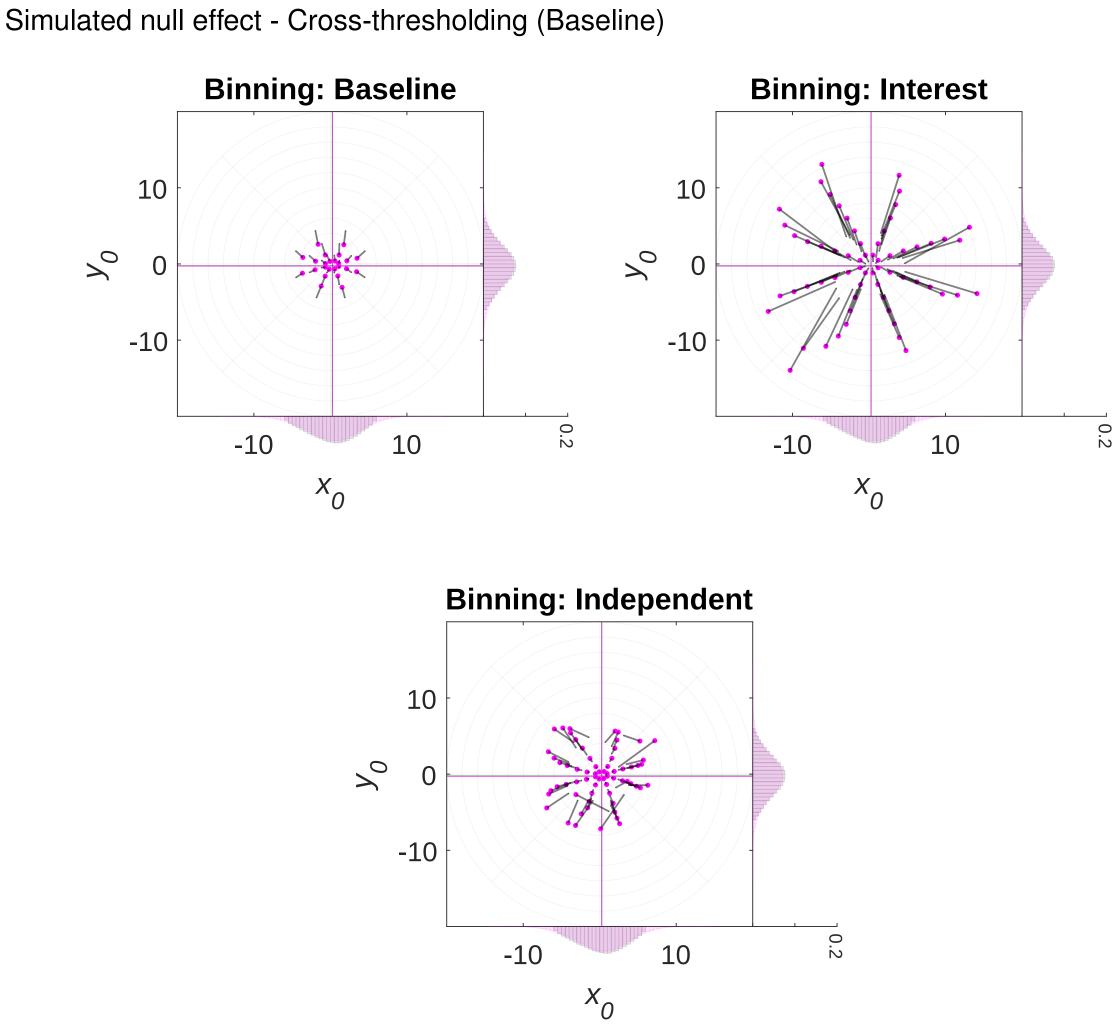


Figure 3-figure supplement 1. **Simulated 2D post hoc binning analysis on x_0 and y_0 | Null effect — Cross-thresholding (Baseline).** The same as in Figure 3, although here, simulated observations falling outside a certain eccentricity range (≥ 0 and ≤ 6 dva) in the Baseline condition were removed from all conditions — a simulation case we term *cross-thresholding (Baseline)*.

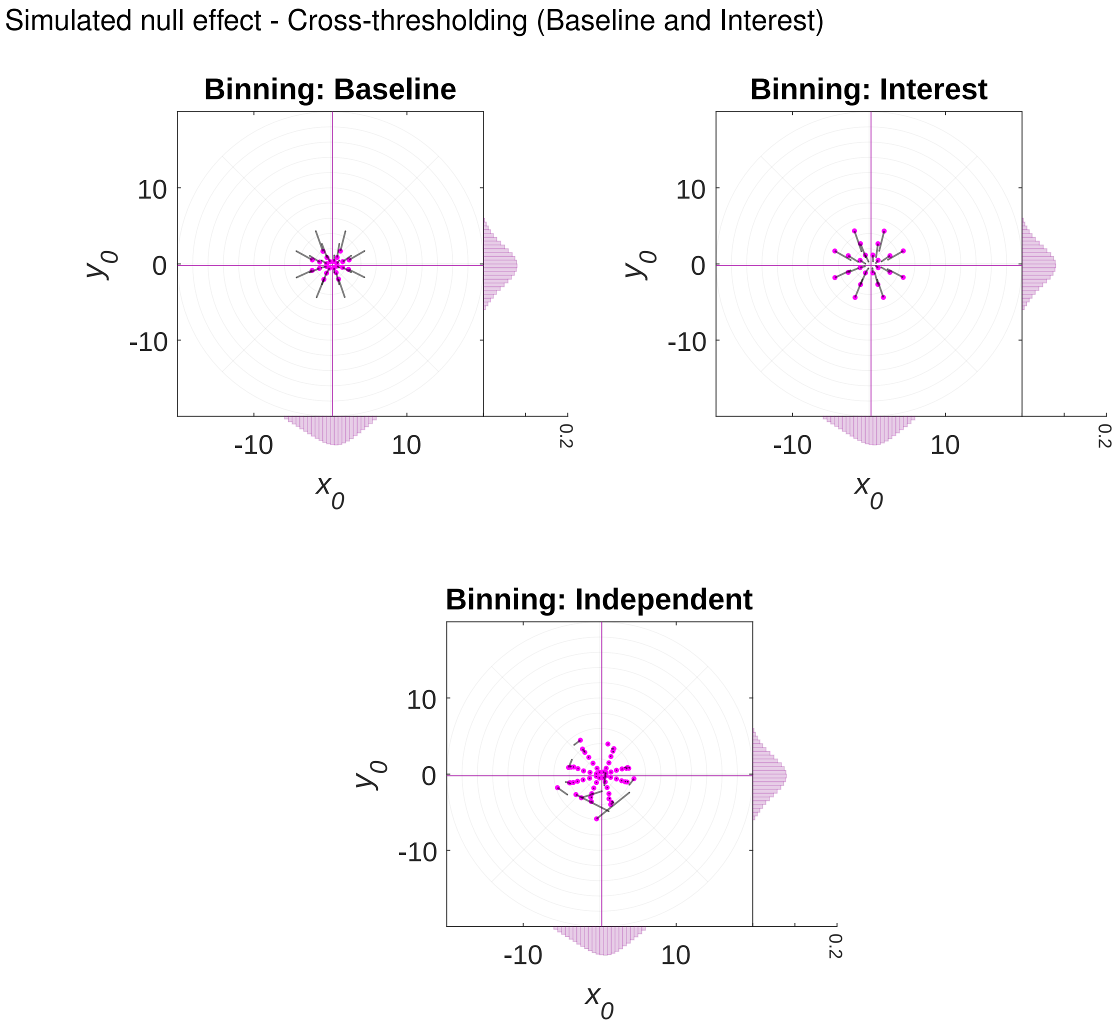


Figure 3-figure supplement 2. **Simulated 2D post hoc binning analysis on x_0 and y_0 | Null effect — Cross-thresholding (Baseline and Interest).** The same as in Figure 3-figure supplement 1, although here, condition cross-thresholding was based on both the Baseline and Interest condition — a simulation case we term *cross-thresholding (Baseline and Interest)*.

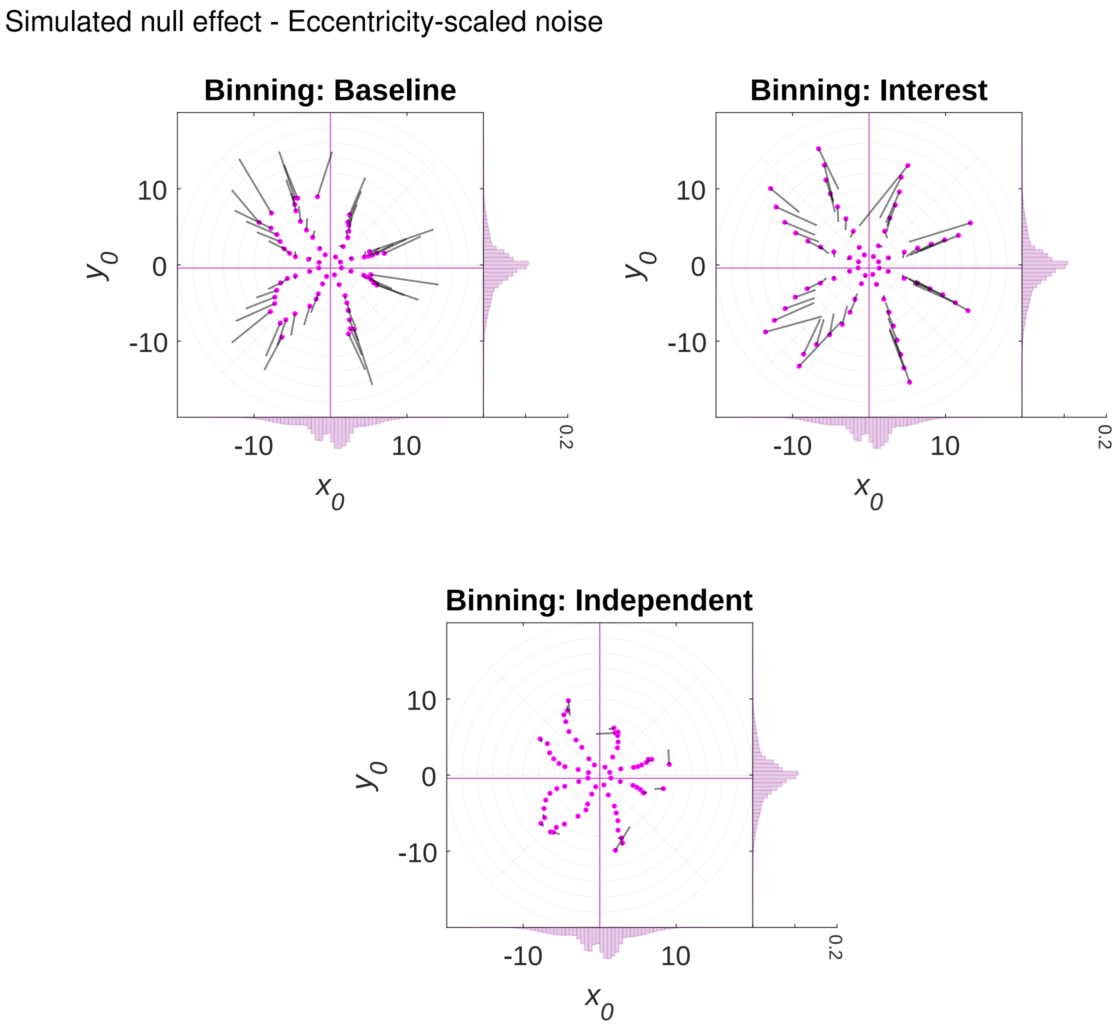


Figure 3-figure supplement 3. **Simulated 2D post hoc binning analysis on x_0 and y_0 | Null effect — Eccentricity-scaled noise.** The same as in Figure 3, although here, original observations having smaller eccentricities ($\geq$ 0 and $<$ 3 dva) were disturbed by random Gaussian noise with a smaller standard deviation ($sd =$ 0.25 dva) and those having larger eccentricities ($\geq$ 3 dva) by random Gaussian noise with a larger standard deviation ($sd = 2$ dva) — a simulation case we term *eccentricity-scaled noise.*

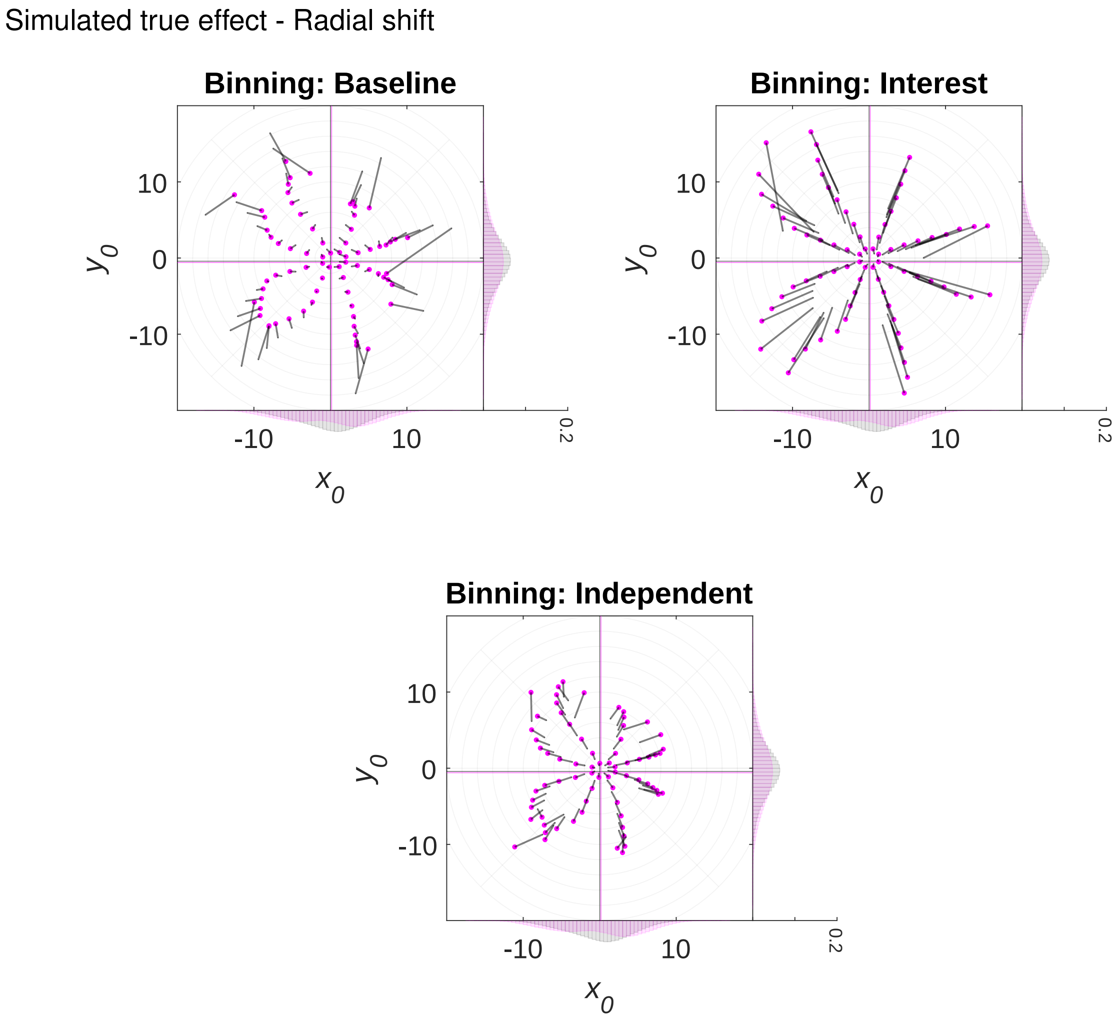


Figure 3-figure supplement 4. **Simulated 2D post hoc binning analysis on x_0 and y_0 | True effect — Radial shift.** The same as in Figure 3, although here, we simulated a true effect, that is, a radial increase in eccentricity of 2 dva in the Interest as compared to the Baseline condition. Note that the eccentricity bins ranged from 0 to 22 dva (instead of 0 to 20 dva) here.

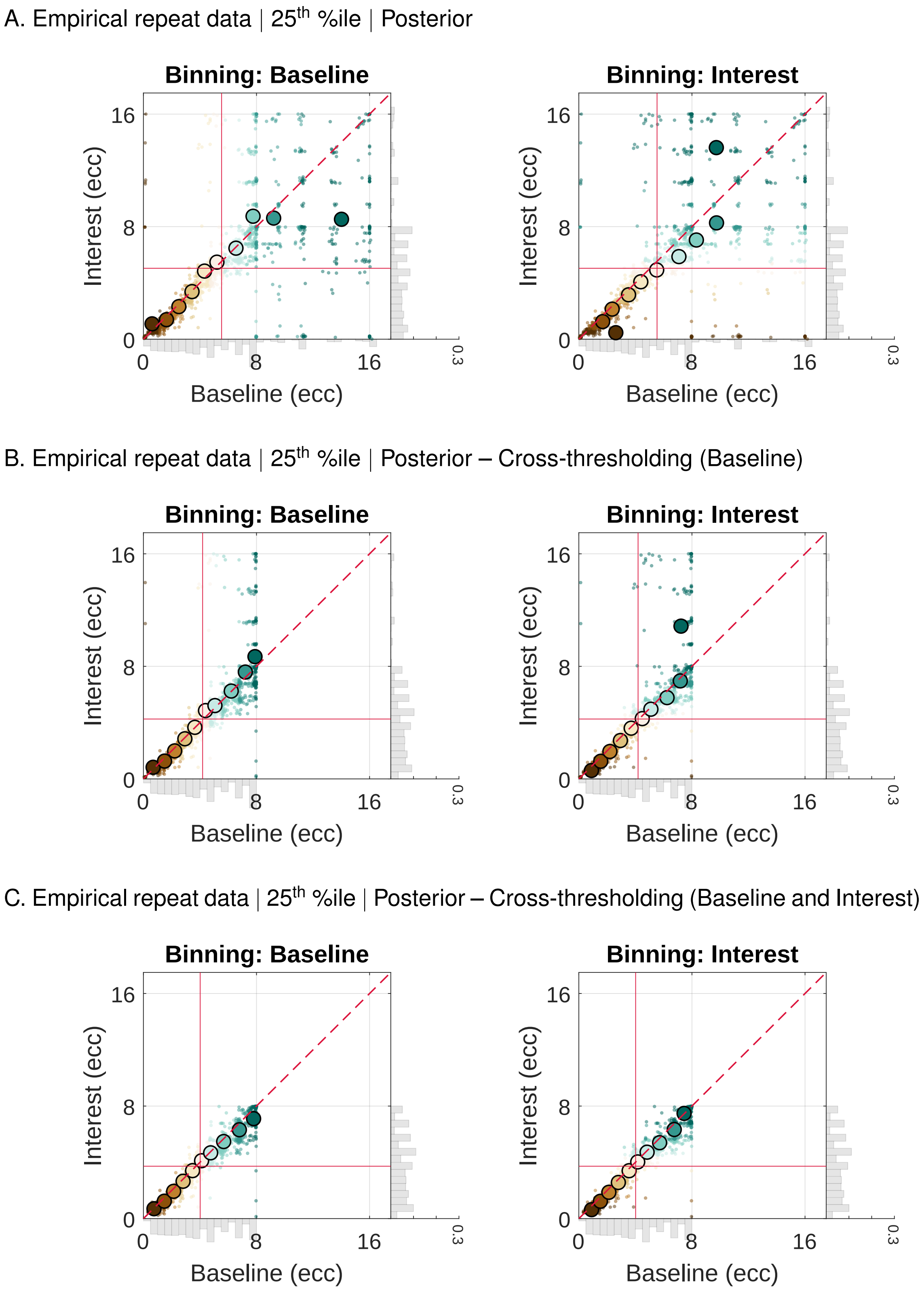
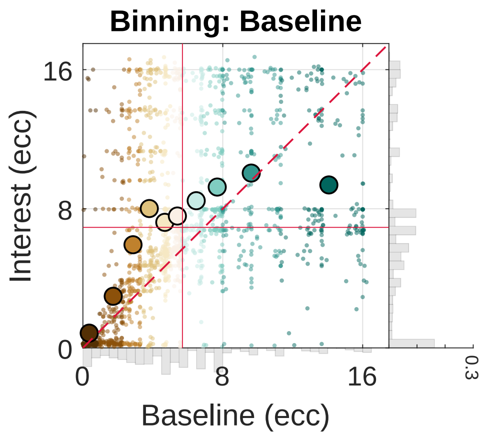
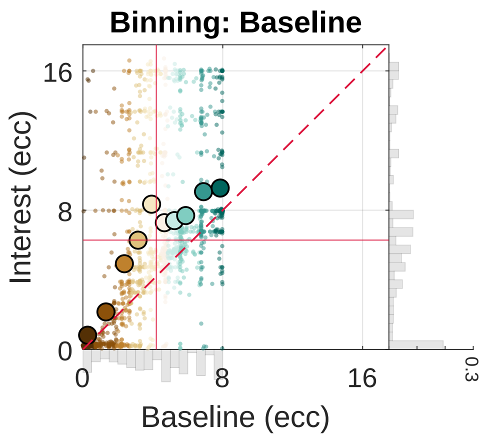


Figure 4-figure supplement 1. **Empirical 1D post hoc binning analysis on eccentricity | Repeat data | 25th %ile participant | Posterior.** The same as in Figure 4, although here, we used data from the posterior complex (V1-V3).

A. Empirical repeat data | 75th %ile | Dorsal



B. Empirical repeat data | 75th %ile | Dorsal – Cross-thresholding (Baseline)



C. Empirical repeat data | 75th %ile | Dorsal – Cross-thresholding (Baseline and Interest)

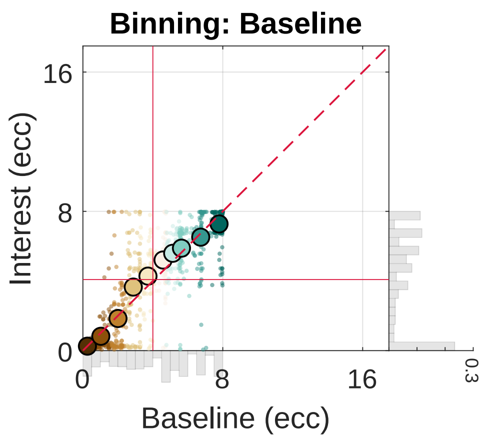


Figure 4-figure supplement 2. **Empirical 1D post hoc binning analysis on eccentricity | Repeat data | 75th %ile participant | Dorsal.** The same as in Figure 4, although here, we used the 75th %ile participant of the median R^2 distribution.

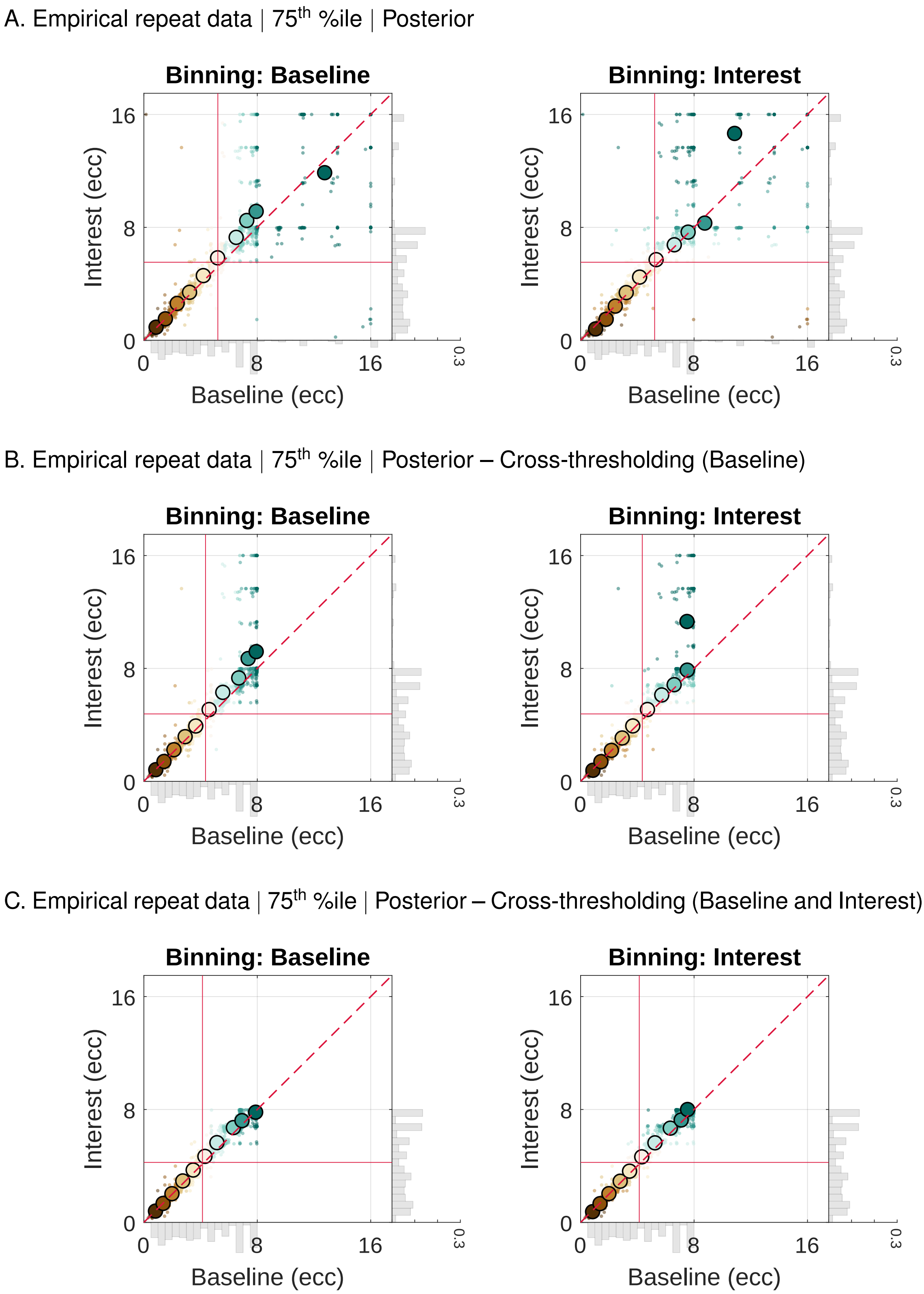


Figure 4-figure supplement 3. **Empirical 1D post hoc binning analysis on eccentricity | Repeat data | 75th %ile participant | Posterior.** The same as in Figure 4-figure supplement 1, although here, we used the 75th %ile participant of the median R^2 distribution.

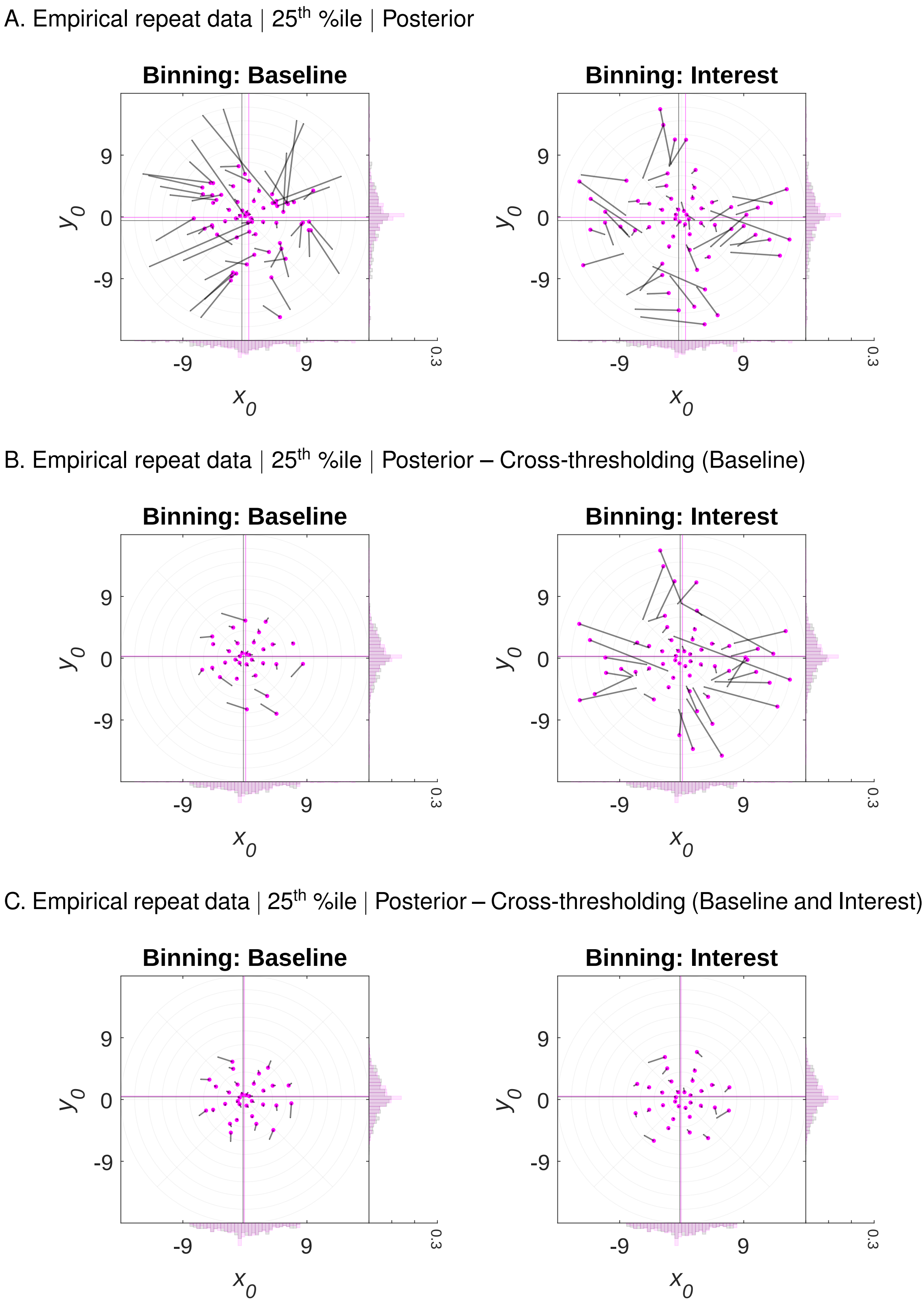
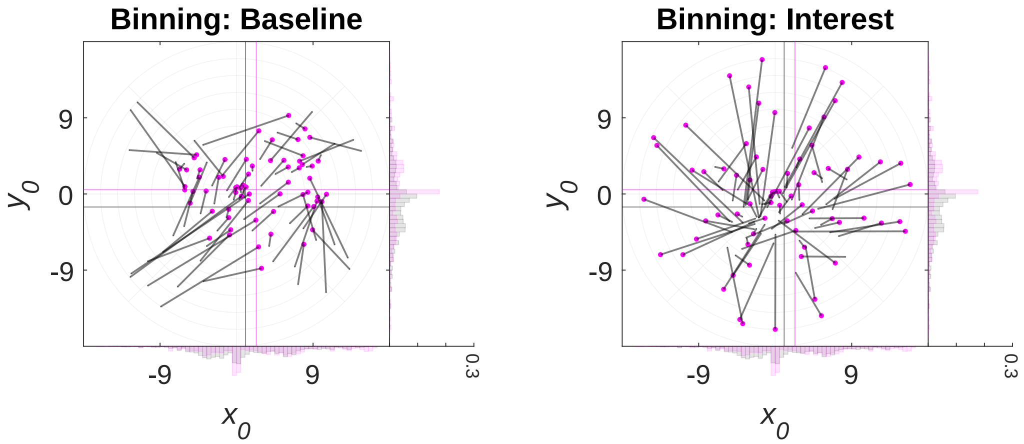
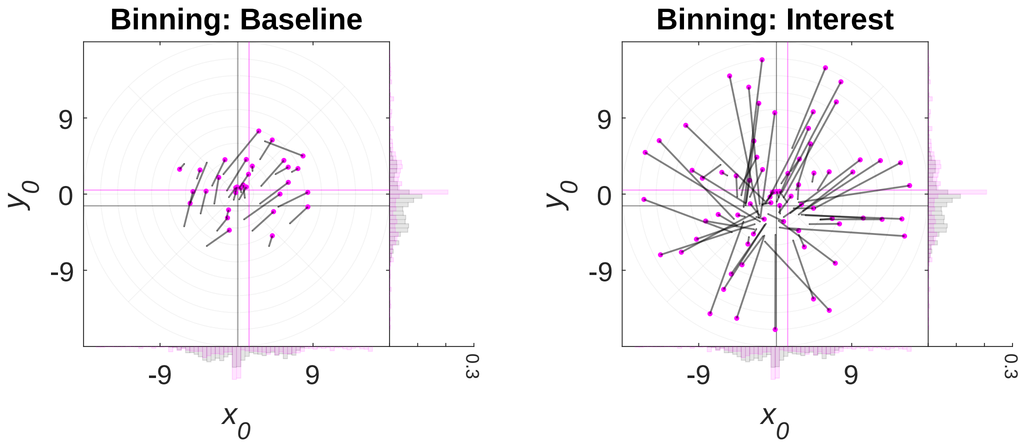


Figure 5-figure supplement 1. **Empirical 2D post hoc binning analysis on x_0 and y_0 | Repeat data | 25th %ile participant | Posterior.** The same as in Figure 5, although here, we used data from the posterior complex (V1-V3).

A. Empirical repeat data | 75th %ile | Dorsal



B. Empirical repeat data | 75th %ile | Dorsal – Cross-thresholding (Baseline)



C. Empirical repeat data | 75th %ile | Dorsal – Cross-thresholding (Baseline and Interest)

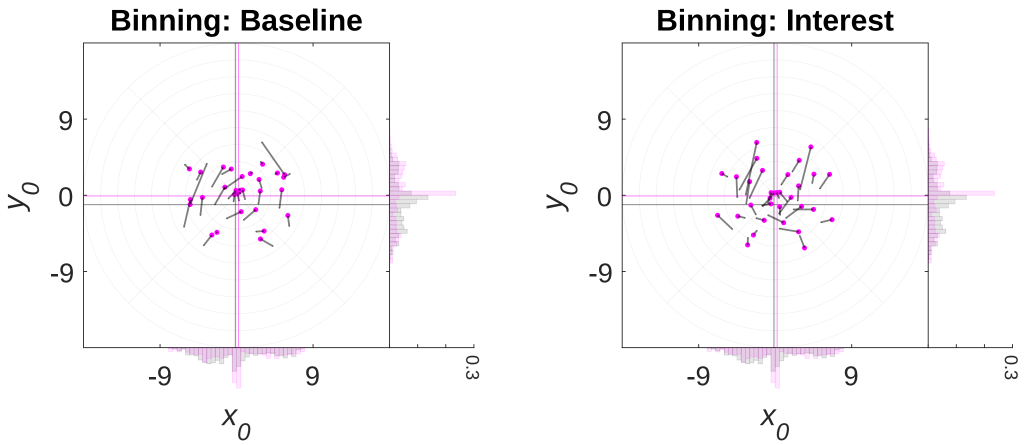


Figure 5-figure supplement 2. **Empirical 2D post hoc binning analysis on x_0 and y_0 | Repeat data | 75th %ile participant | Dorsal.** The same as in Figure 5, although here, we used the 75th %ile participant of the median R^2 distribution.

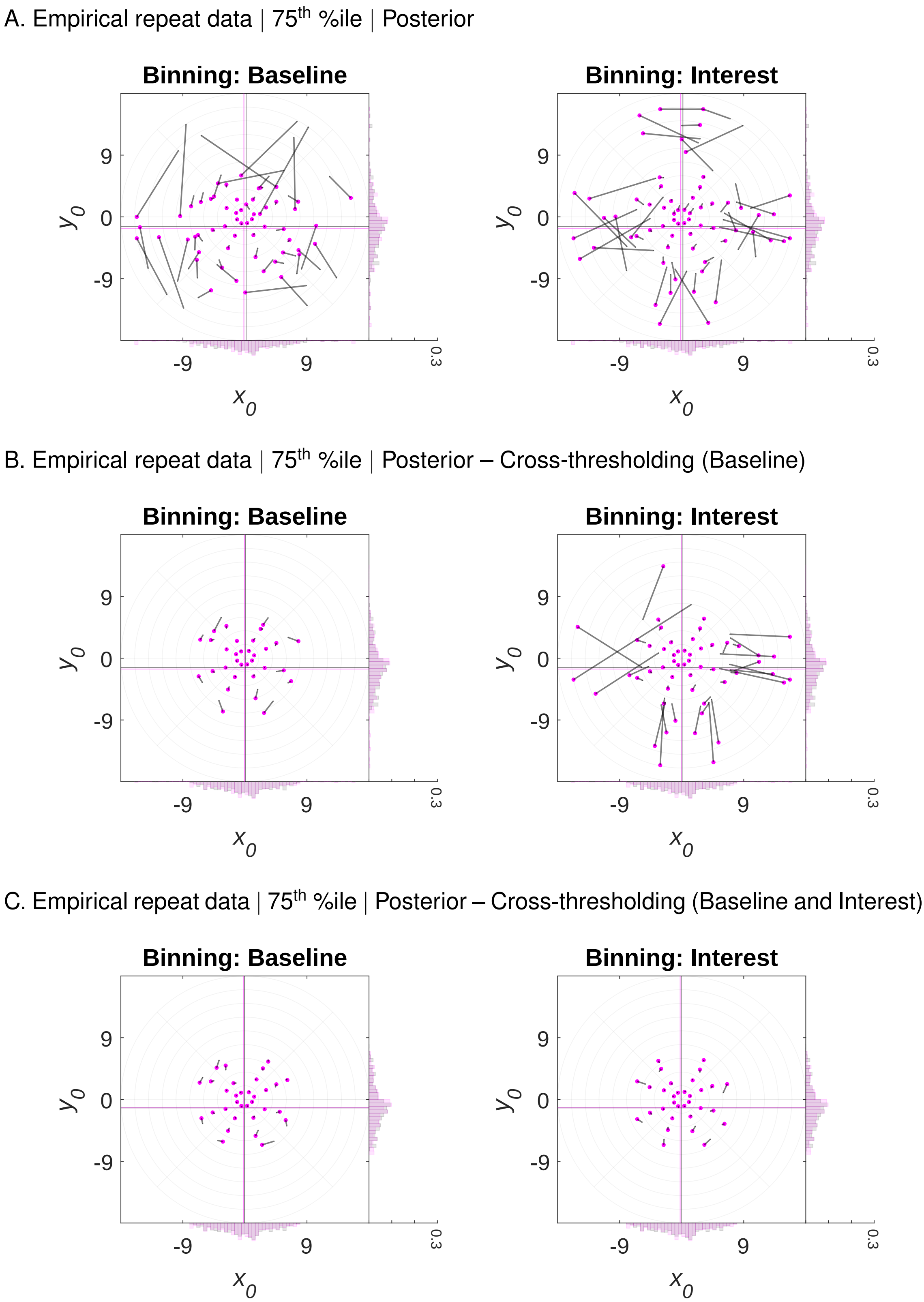


Figure 5-figure supplement 3. **Empirical 2D post hoc binning analysis on x_0 and y_0 | Repeat data | 75$^{\text{th}}$ %ile participant | Posterior.** The same as in Figure 5-figure supplement 1, although here, we used the 75$^{\text{th}}$ %ile participant of the median R^2 distribution.