

Saccadic Biases

Alasdair D. F. Clarke, Matthew J. Stainer

October 13, 2015

Abstract

More bias modelling! Cause who can be bothered running actual experiments?

Much effort has been made to attempting to explain eye guidance during natural scene viewing [Tatler, 2009 VIS RES special issue]. However, underlying fixation placement appears to be a set of consistent biases in eye movement behaviour (e.g. see Clarke and Tatler, 2014). We present a model that parametrically accounts for where saccades are directed dependent on where in the bounds of a scene an observer is currently fixating. We find... [some very interesting stuff]. Given that much of our understanding in eye guidance is derived from how people look at pictures on a computer screen, it is important that we use these biases to form a frame upon which to build more sophisticated models of eye guidance that can account for the allocation of gaze above oculomotor behaviours that are independent of the image.

1 Introduction

Improve on last year's [Clarke and Tatler, 2014] effort. More sophisticated biases. And some examples of how to use biases for improved data analysis.

The human fovea provides a small window of high acuity vision to the world, and as such the locations that we select to view in the world can tell us about how we seek the information necessary to complete the task we are cur-

rently undertaking. Current understanding of eye guidance would suggest that fixation locations are selected based on a combination of low-level factors (such as visual salience [Itti and Koch, 2000] or orientation information [?]) and high-level factors [??]. However, there are also strong observable biases in eye movement including directional and amplitudinal [XXis that a word?XX] biases in saccades [??], and a strong tendency to fixate near to the centre of images [??]. Importantly, these biases are independent of the viewed content. If we are to gain a complete understanding of the factors that govern eye movement, we must therefore build models of eye guidance on the framework of these underlying biases.

1.1 The central bias

There is a strong tendency for people to look close to the centre of pictures [????] and movies [?] presented on computer screens. There have been a number of suggestions for why this might be. One possibility for this effect is that the muscles of the eye show a preference for the 'straight ahead' position, re-centring in the orbit of the eye socket for most comfortable contraction of the ocular muscles (an *orbital reserve* [?]). As most scene viewing experimental set-ups stabilise the head to increase the accuracy of the eye tracking, and most scenes are presented in the centre of computer displays, such a re-centring mechanism would mean that the centre of images would in-

deed be preferentially selected. However, when scenes are scrambled into four quadrants, fixations are located near to the centre of each quadrant, rather than the display centre, suggesting that the central tendency is responsive to the viewed content [?] rather than the frames of the computer monitor.

Another possibility for the central fixation bias is that it represents a *photographer bias* as photographers tend to frame their shots to include the most important content in the centre of the scene. However, when ? presented scenes where the image features were biased towards the edge of the scene, the central fixation bias persisted. The final possibility is that as a consequence of repeated exposure to photographer bias, the centre of scenes is simply where people are *trained* to look at images [?]. Such learning of spatial probabilities of targets can explain why, for example, people tend to look around the horizon when searching for people in natural scenes [??]. Expecting to find interesting content in the centre of scenes might be a consequence of this hypothesis typically being correct.

Clarke and Tatler [2014] revealed that the characteristics of the central bias is remarkably consistent across a series of eye movement databases.... [obviously you are in a better place than me to talk about this paper!]

1.2 Behavioural biases in saccades

Further to the observed bias towards the centre of images, it has been revealed that there are underlying biases in the characteristics of eye movement (in terms of the directions and amplitudes of saccades). It has been noted by several researchers that when viewing scenes, there is a higher proportion of eye movements in horizontal directions than vertical or oblique movements (Brandt, 1945; Crundall & Underwood, 1998; Gilchrist & Harvey, 2006; Foulsham, Kingstone & Underwood, 2008; Tatler

& Vincent, 2008). There are a number of possibilities as to why this tendency exists (as discussed in Foulsham, Kingstone, & Underwood, 2008). Firstly, there may be a muscular or neural dominance making oculomotor movements in the horizontal directions more likely. Secondly, the characteristics of photographic images may mean that content tends to be arranged horizontally by the photographer. In such situations, horizontal saccades may be the most efficient way to inspect scenes. Thirdly, using horizontal saccades in scene viewing might be a learned strategy. Observers may learn the natural characteristics of scenes based on previous experience, and therefore demonstrate an increased likelihood of moving in the horizontal direction. A final alternative explanation is that this tendency is a consequence of the aspect ratio of visual displays, which normally allow for larger amplitude saccades in the horizontal than vertical directions (Wartburg et al., 2007).

Foulsham and colleagues have presented two interesting exceptions to the horizontal direction bias. Foulsham, Kingstone and Underwood (2008) found that when the orientation of an image is rotated, the distribution of saccade directions follows the orientation of the scene. A second exception comes from using circular apertures (Foulsham & Kingstone, 2010). When a scene is presented in a circular aperture, the tendency to make horizontal saccades disappears, being replaced by a tendency to make vertical saccades relative to the image orientation. However, when using fractal images (where images do not have an obvious orientation), observers tend make horizontal saccades, regardless of the angle that the image is presented.

An example of saccadic flow is given in Figure 1. A model based on this idea has been developed by Clarke, Green, Chantler, and Hunt [2015]

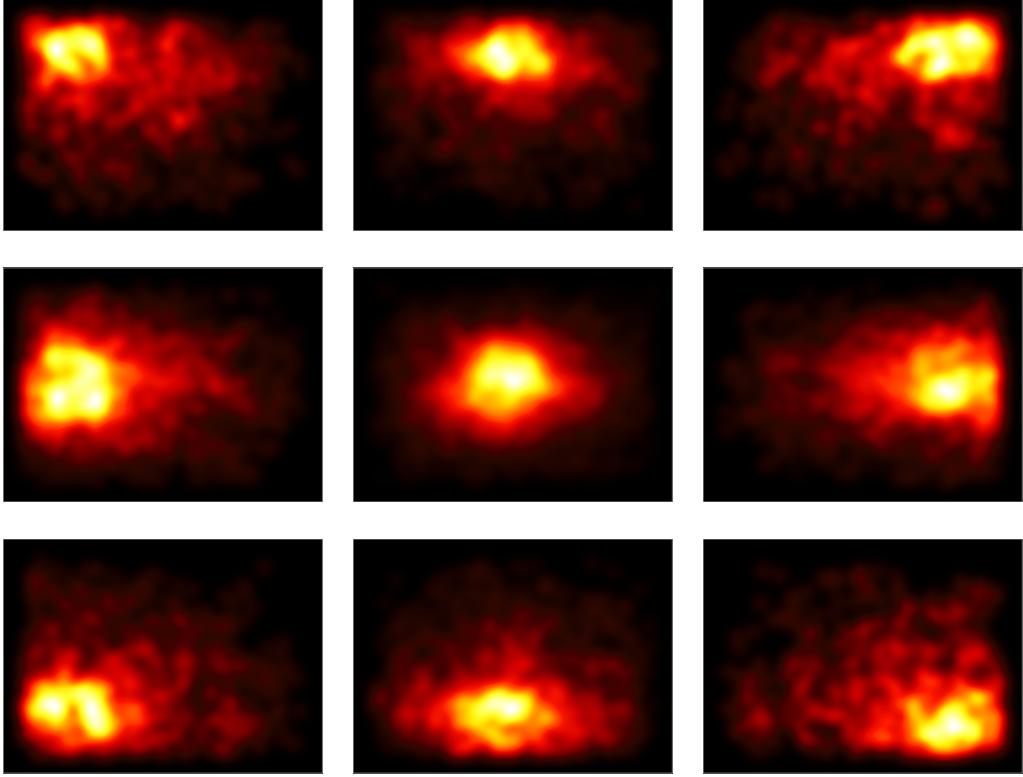


Figure 1: Saccade landing positions from fixations that were in different sections of the screen. Data from each plot has been separated into fixations in 9 spatial bins, with the screen being divided into thirds in both horizontal and vertical aspects.

1.3 The present study

The aim of the present study is to characterise the underlying biases of eye movement with which to understand fixation selection in natural scenes. We extend the previous work by examining [XXstuffXX] and

2 Using Biases

This section make use of the new central bias model (similar to Clarke and Tatler [2014] expect uses a truncated Gaussian distribution to take the image boundaries into account) and the *saccadic flow* model (described in 3.4). We present two examples of how these bias models

can be used as a prior in order to weight fixations. First of all, we will demonstrate how we can weight fixations in hotspot maps to reduce the noise and give an improved visualisation of the regions of the image that participants looked at more than expect. Secondly, we will demonstrate how these priors can be used in ROC analysis to improve results. We will re-analyse the effect of the scene context model developed by Ehinger, Hidalgo-Sotelo, Torralba, and Oliva [2009] as an example.

2.1 Gaze landscapes

One technique that is commonly used to visualise the spatial allocation of gaze is to create

'heatmap' plots where colour or luminance are used to indicate the density of fixation on those locations (Figure 2, column 2). Some argue that one problem with visualising data in this way is that they represent all fixations as equal. For example, a fixated location with a fixation of a second would be weighted equally with fixations that lasted half that time. If we want to make an assumption that fixation duration is intimately linked with the importance of that fixation (i.e. we will look longer at more informative information) then we can change our visualisation to weight fixations by their duration (Figure 2, column 3).

One advantage of the [Clarke and Tatler, 2014] model, and the saccadic flow model here is that we can represent fixations by the likelihood that they would occur based on the predictions of the models. As there is an image independent tendency to fixate in the centre of the scene (for example), then we might consider that saccades to locations less predicted by these behavioural and oculomotor biases might involve more high-level mechanisms. In Figure 2 (column 4 and 5) we present some overlaid heatmap data from the [Clarke, Coco, and Keller, 2013] dataset, where fixations are weighted by the inverse probability of them occurring based on the models of central bias and saccadic flow. These figures reveal that representing data in this manner can allow us to visualise information that was important enough to break the biases of looking at the scene centre, or making saccades in line with our saccadic flow model. We can therefore use this to remove some of the image-independent biases, and reveal the more important image *dependent* information.

Or do we call them hotspot maps?

2.2 Reanalysis of Clarke et al 2013

Do the flow-weights offer an improvement on predicting named objects over the fixations-

weights?

2.3 Reanalysis of Ehinger et al 2009

Re-analysis of Ehinger et al. [2009] comparing ROC methods for leave-one-out, uniform, central bias and flow. How well does their context model work in each test>

2.4 Comparison between synthetic and real data

To what extent does saccadic flow account for coarse-to-fine dynamics? Not that well. Not unexpected.

We can see from Figure ?? that both the central bias and the saccadic flow model do a good job of capturing the distribution of fixation locations in the x and y axes. However, the saccades generated by the flow model tend to be slightly larger than those made by human observers (Figure ??).

3 Biases

In this section, we will present an updated central bias distribution (using a truncated Gaussian); explore the degree of asymmetry in viewing positions, and describe the saccadic flow model.

3.1 Modelling Methods

In this section, we will give an overview of the methods and data used for the saccadic flow modelling.

3.1.1 Datasets

We will use a number of previously published datasets. The models will be trained on a subset of the 10 datasets used in Clarke and Tatler [2014]. These data are taken from Clarke et al. [2013], Einhäuser, Spain, and Perona [2008], Judd, Ehinger, Durand, and Torralba [2009],

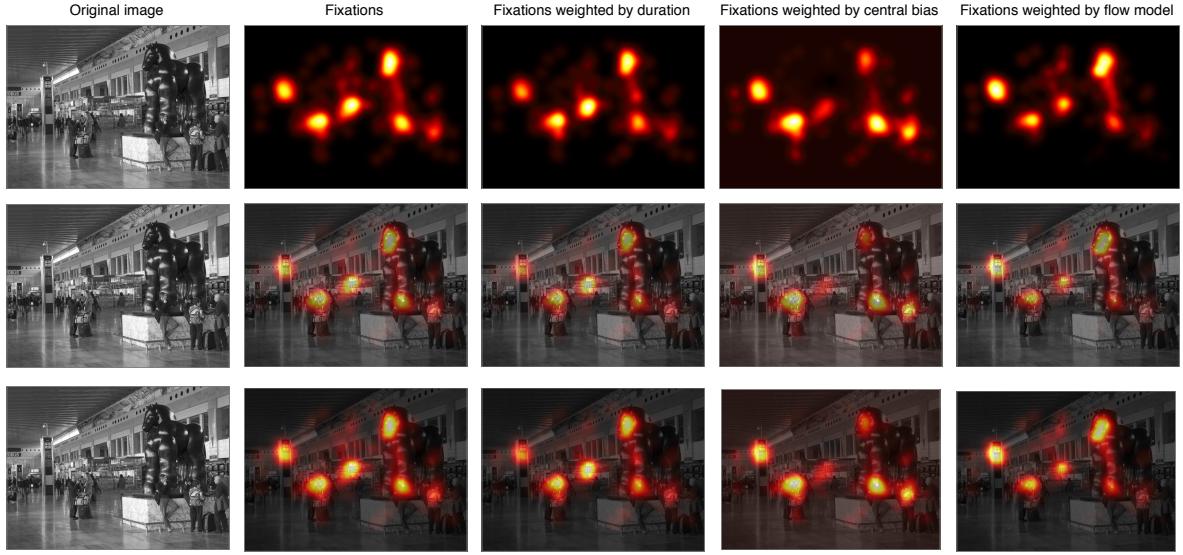


Figure 2: Traditional ‘heat map’ plots of fixations normalised by the central bias. This method allows us to characterise fixations that are less accountable for by image-independent central biases.

Tatler [2007], Tatler, Baddeley, and Gilchrist [2005], Yun, Peng, Samaras, Zelinsky, and Berg [2013]. The initial saccade after image onset (9.1% of the data) are excluded, giving us a total of 159,226 saccades. We chose to remove the data from Asher, Tolhurst, Troscianko, and Gilchrist [2013] from our training set as the images have an aspect ratio of 5:4, whereas the rest of the data in our training set has an aspect ratio of 4:3. The pedestrian search dataset [Ehinger et al., 2009] was removed from the training set as previous analysis [Clarke and Tatler, 2014] shows that it appears to be biased compared to the other datasets. Both these datasets are now used as test sets to evaluate how well our models generalise.

We also add a number of other datasets to our test suite collection.

- Clarke, Chantler, and Green [2009] has a dataset of fixations made during a visual search for a target on a homogeneous textured background (i.e. target in noise).

This dataset differs from the previous in that there is no semantic image content in the scene, and the stimuli had a 1:1 aspect ratio.

- Greene, Liu, and Wolfe [2012] released a dataset of observers viewing square greyscale photographs.
- Borji and Itti [2015] recently released a very large (≈ 0.625 million fixations, 2000 images) dataset collected over twenty different stimuli types. Given the size of this dataset, and the widescreen 16:9 aspect ratio, the evaluations on this dataset are presented separately, and split by stimuli class.

An overview of the datasets used is given in Appendix Tables 2 and 3.

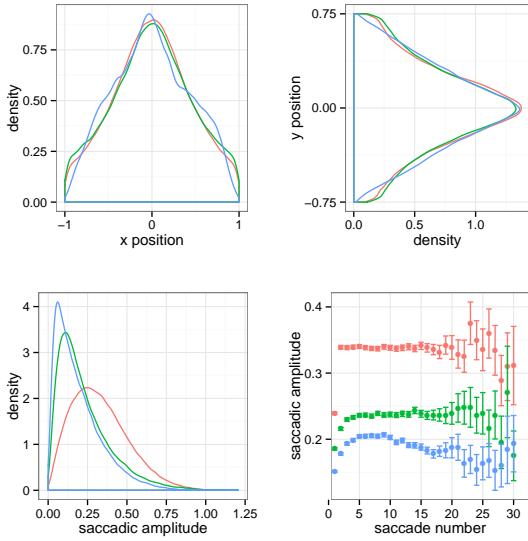


Figure 3: *blue*: human, *red*: central bias, *green*: saccadic flow. *top row*: Comparison of *x* and *y* fixation positions between human fixations and synthetic points generated from the central bias and flow model. *bottom row*: We can see that the flow model consistently makes saccades with a slightly larger amplitude than human observers. Distances are expressed relative to the width of the image.

3.1.2 Pre-processing

As with Clarke and Tatler [2014] we have normalised all fixations to the image frame, keeping the aspect ratio constant. ie, $(x, y) \in (-1, -1) \times (-a, a)$ with typically $a = 0.75$. The initial fixations and saccades were not included in the analysis. Saccades with a start or end point falling outside of the image frame were also removed.

When fitting saccadic flow models, we *mirrored* the set of fixations, but adding in reflected copies of the data (reflected in the horizontal, vertical and both midlines). This has two advantages. (i) It is an easy way to make saccadic flow biases in the horizontal or vertical directions. This is similar to how the cen-

tral bias was defined Clarke and Tatler [2014], but by a different mechanism (with the central bias, the model fitting procedure is much simpler and so we just enforced zero mean and 0s in the covariance matrix). (ii) It increases the amount of data available for fitting by a factor of four. This is important as (due to the central bias) there are relatively few saccades that originate from the corners of the images. By equating all corners, we can pool the data and obtain more stable estimates for the underlying distribution.

The downside of mirroring saccades in this manner is that our model of saccadic flow will be insensitive to the *leftwards* bias in natural scene viewing [Nuthmann and Matthias, 2014]. This will be discussed in Section 3.3. Similary, as ignore the timecourse of saccades we will not capture *coarse-to-fine* dynamics (discussed previous in Section 2.4).

We will model and discuss saccadic flow, coarse-to-fine, and left v right.

3.2 Truncated Central Bias

First, we will update the central bias from Clarke and Tatler [2014] and use a truncated normal distribution. This is very straight forward. Re-fitting a multivariate gaussian to the data reduces the deviance in the central bias model by 0.06%. Usng a truncated Gaussian gives us an improvement of 7.28%. We can round the truncated Gaussian model to $\mu = (0, 0)$, $\sigma = (0.3, 0; 0, 0.12)$ with no loss of precision.

3.3 Left v Right

Initially more fixations to the left half of the image [Nuthmann and Matthias, 2014]. We replicate this here (Figure 4).

However, it has only a very small effect on explaining the variation over whole datasets: Figure 5.

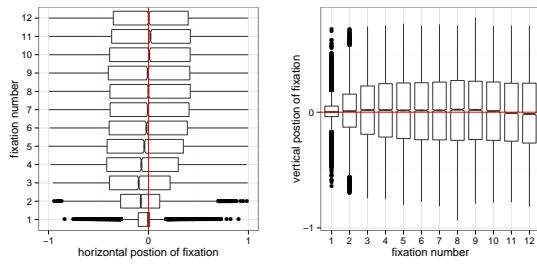


Figure 4: Distribution of horizontal and vertical fixations by fixation number.

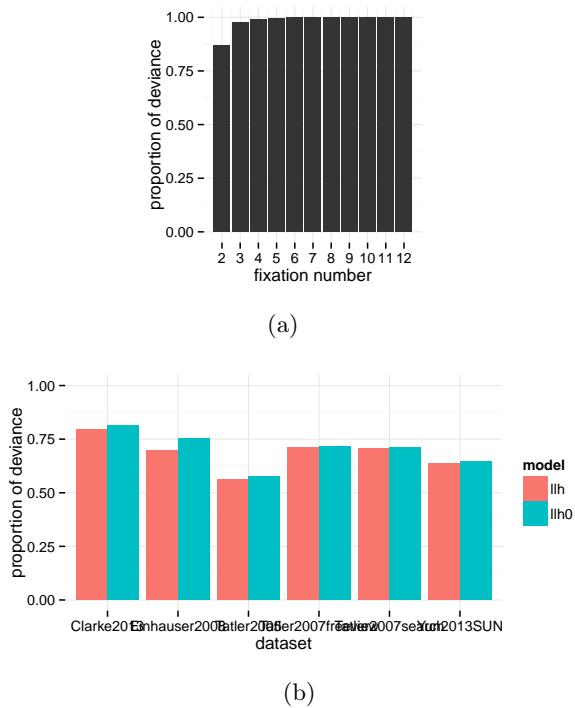


Figure 5: Modelling results

Hence we will ignore this effect from now on. By treating everything as symmetrical, we lose very little explanatory power, while restricting the number of parameters, or increasing the amount of data available (by mirroring fixations).

3.4 Saccadic Flow

Saccadic flow can be thought of as a generalisation of the central bias. Instead of computing the distribution of all saccadic endpoints in a dataset, we look at the distribution of saccade endpoints given the start points. So for a saccade from (x_0, y_0) to (x_1, y_1) we want to model $p(x_1, y_1 | x_0, y_0)$. This is illustrated in Figure 1.

3.4.1 Modelling

To characterise how the distribution of saccadic endpoints varies with the start point, we used a sliding window approach. All saccades that originated in a $n \times n$ window were taken and used to fit a multivariate Gaussian distribution. This window was then moved over the stimuli in steps of $s = 0.01$. Parameter sets estimated from windows containing less than 250 data-points were removed. Multivariate polynomial regression was then used to fit 4-th order polynomials to each of the parameters. Robust estimation was used (`r1m` from the `textttMASS` library) to stop the model fits being overly influenced by outlier points from the image boundary. We experimented with varying the window size ($n \in \{0.05, 0.1, 0.2\}$). However, as this parameter was found to have a negligible result, we only report the results for $n = 0.05$.

3.4.2 Results

Figure ?? shows how the parameters for the multivariate Gaussian distribution vary over horizontal position for a selection of vertical positions. The regression coefficients given in Table 1 allow us to estimate the conditional

probability of a saccade to (x_1, y_1) given the starting fixation (x_0, y_0) .

How well does this model account for the fixations in our datasets? Figure 6 shows the deviance of the flow model expressed as a proportion of the deviance of the Clarke-Tatler central bias. For reference, we also show the results for re-fitting the central bias to each dataset. From this figure, we can see that the flow-normal model approximately halves the deviance.

As the flow:normal model is significantly more complex, requiring nine times as many parameters, it is important to test for robustness. We can test how well our model generalises on testing it on other datasets, for example, Borji and Itti [2015].

3.4.3 Discussion

We put the Flow:normal model forward as a robust prior for image-content independent saccadic behaviour. This model can be thought of as a partner of the Clarke-Tater central bias, and we expect that in some cases, the simpler central bias will be more appropriate, while in others, the more complex flow model is a better choice. We have demonstrated that although this model requires more parameters, it generalises well from one dataset to another and is a far better baseline for modelling a scan-path than the central bias.

There are two main simplifications to our modelling work. First of all, we are using an unbounded distribution (ie, $(x, y) \in \mathbb{R}^2$) to model bounded data. While it is possible to deal with this issue, by either applying a transform $(-1, 1) \rightarrow \mathbb{R}$ (such as $z = \log(\frac{x'}{1-x'})$, where $x' = \frac{x+1}{2}$), or fitting a truncated multivariate Gaussian, we decided that given the good performance of the model as is, it was not worth adding the additional complexities to our model at this time.

The second simplification is that we are treating the data as normal. From Figure 1

we can see that the data is clearly skewed, particularly in the corners. We will attempt to address these issues in the following section.

4 Discussion

4.1 Scenes and natural viewing behaviour

That observers organise their viewing behaviour on computer screens around the reference frames provided by the bounds of scenes (see also ?) causes problems for relating findings of eye guidance in scenes to eye guidance in natural behaviour, as the bounds of such reference frames are unclear in the real world. While it has been suggested that we tend to fixate near to the centre of our ‘straight ahead’ head position [FOULSHAM WALKING, CRISTINO AND BADDELEY?], there are no discrete edges as are typical in computer based scene viewing paradigms. If fixation locations are constrained by the bounds of the scene, this highlights the care we must take about the generalisations we make from findings in the lab to the real world (see [kingstonepaper 2010]).

Acknowledgements

A Dataset Details

Here are all the details on the datasets used in this paper. (Table 2 and 3).

References

Matthew F Asher, David J Tolhurst, Tom Troschianko, and Iain D Gilchrist. Regional effects of clutter on human target detection performance. *Journal of vision*, 13(5):25:1–15, 2013.

Ali Borji and Laurent Itti. Cat2000: A large scale

parameter	equation
$\Omega_{x,x}$	$= 0.33 + 0.38x^2 - 0.29y^2 + 0.02x^4 + 0.22y^4$
$\Omega_{x,y}$	$= x + y + x^2 + y^2 + x^3 + y^3 + x^4 + y^4$
$\Omega_{y,x}$	$= x + y + x^2 + y^2 + x^3 + y^3 + x^4 + y^4$
$\Omega_{y,y}$	$= x + y + x^2 + y^2 + x^3 + y^3 + x^4 + y^4$
α_{x^2}	$= x + y + x^2 + y^2 + x^3 + y^3 + x^4 + y^4$
α_{y^2}	$= x + y + x^2 + y^2 + x^3 + y^3 + x^4 + y^4$
ν	$= x + y + x^2 + y^2 + x^3 + y^3 + x^4 + y^4$

Table 1: Parameter model - clearly I still have to fill in all the coefficients!

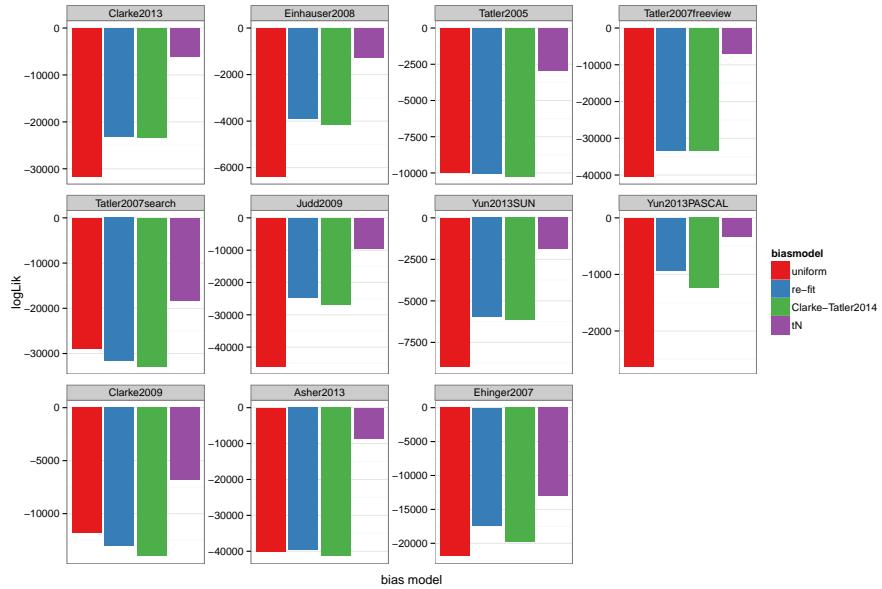


Figure 6: Flow:normal log likelihood results. We can see that re-fitting the central-bias to each specific dataset offers little improvement over using the Clarke-Tatler model, while the flow model offers a substantial improvement.

	Observers	Images	Task	Display duration
Clarke et al. [2013]	24	100	object naming	5000 ms
Yun et al. [2013] - SUN	8	104	image description	5000 ms
Tatler et al. [2005]	14	48	memory	variable
Einhäuser et al. [2008]	8	93	object naming	3000 ms
Tatler [2007] - free	22	120	free viewing	5000 ms
Judd et al. [2009]	15	1003	free viewing	3000 ms
Yun et al. [2013] - PASCAL	3	1000	free viewing	3000 ms
Ehinger et al. [2009]	14	912	visual search	variable
Tatler [2007] - search	30	120	visual search	5000 ms
Asher et al. [2013]	25	120	visual search	variable

Table 2: Summary of the 10 datasets used throughout this study.

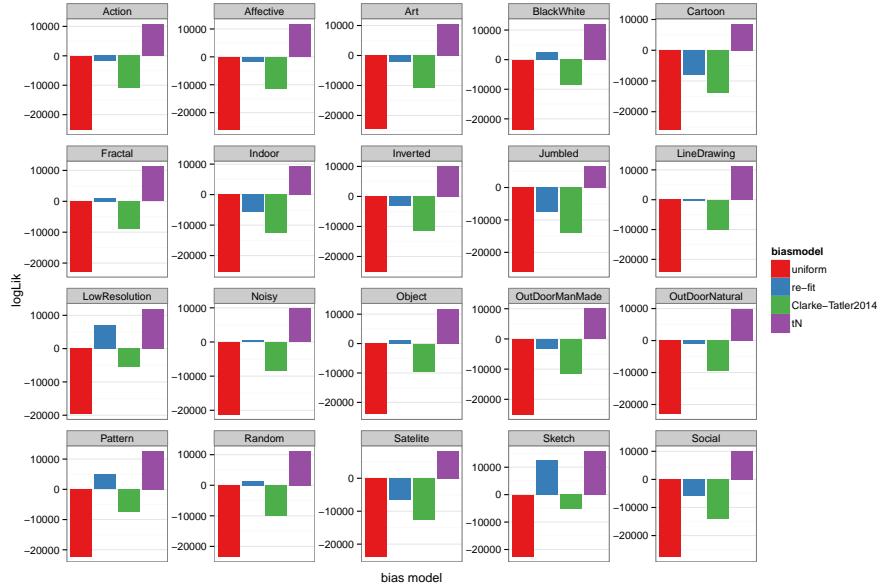


Figure 7: Flow:normal deviance results. We can see that re-fitting the central-bias to each specific dataset offers little improvement over using the Clarke-Tatler model, while the flow:normal model decreases the deviance by half.

	Eye tracker	Viewing distance	Screen size	Image size	Viewing angle	Chin / head rest
Clarke et al. [2013]	EyeLink II	50 cm	21"	800 x 600	31 x 25°	no
Yun et al. [2013] - SUN	EyeLink 1000	?	?	?	?	?
Tatler et al. [2005]	EyeLink I	60 cm	17"	800 x 600	30 x 22°	no
Einhäuser et al. [2008]	EyeLink 1000	80 cm	20"	1024 x 768	29 x 22°	yes
Tatler [2007] - free	EyeLink II	60 cm	21"	1600 x 1200	40 x 30°	no
Judd et al. [2009]	?	2 feet	19"	1024 x 768*	?	yes
Yun et al. [2013] - PASCAL	EyeLink 1000	?	?	?	?	?
Ehinger et al. [2009]	ISCAN RK-464	75 cm	21"	800 x 600	23.5 x 17.7°	yes
Tatler [2007] - search	EyeLink II	60 cm	21"	1600 x 1200	40 x 30°	no
Asher et al. [2013]	EyeLink 1000	55 cm	?	1024 x 1280	37.6 x 30.5°	yes

Table 3: Details of the experimental setups in each of the 10 datasets analysed in the present study. We provide only information reported in the original articles. Question marks indicate information not reported in the original article. *For the Judd et al dataset images varied in pixel dimensions but the majority were at 1024 x 768.

- fixation dataset for boosting saliency research. *arXiv preprint arXiv:1505.03581*, 2015.
- Alasdair D. F. Clarke, Moreno I. Coco, and Frank Keller. The impact of attentional, linguistic, and visual features during object naming. *Frontiers in psychology*, 4:927, 2013.
- Alasdair DF Clarke and Benjamin W Tatler. Deriving an appropriate baseline for describing fixation behaviour. *Vision research*, 102:41–51, 2014.
- Alasdair DF Clarke, Mike J Chantler, and Patrick R Green. Modeling visual search on a rough surface. *Journal of Vision*, 9(4), 2009.
- Alasdair DF Clarke, Patrick R Green, Mike J Chantler, and Amelia R Hunt. Stochastic search. *submitted*, 2015.
- Krista A Ehinger, Barbara Hidalgo-Sotelo, Antonio Torralba, and Aude Oliva. Modelling search for people in 900 scenes: A combined source model of eye guidance. *Visual Cognition*, 17(6-7):945–978, 2009.
- Wolfgang Einhäuser, Merrielle Spain, and Pietro Perona. Objects predict fixations better than early saliency. *Journal of Vision*, 8(14):18: 1–26, 2008.
- Michelle R Greene, Tommy Liu, and Jeremy M Wolfe. Reconsidering yarbus: A failure to predict observersâŽ task from eye movement patterns. *Vision research*, 62:1–8, 2012.
- Laurent Itti and Christof Koch. A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision research*, 40(10):1489–1506, 2000.
- Tilke Judd, Krista Ehinger, Frédo Durand, and Antonio Torralba. Learning to predict where humans look. In *Computer Vision, 2009 IEEE 12th international conference on*, pages 2106–2113. IEEE, 2009.
- Antje Nuthmann and Ellen Matthias. Time course of pseudoneglect in scene viewing. *Cortex*, 52: 113–119, 2014.
- Benjamin W Tatler. The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. *Journal of Vision*, 7(14):4: 1–17, 2007.
- Benjamin W Tatler, Roland J Baddeley, and Iain D Gilchrist. Visual correlates of fixation selection: Effects of scale and time. *Vision research*, 45(5): 643–659, 2005.
- Kiwon Yun, Yifan Peng, D. Samaras, G.J. Zelniksky, and T.L. Berg. Studying relationships between human gaze, description, and computer vision. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 739–746, 2013. doi: 10.1109/CVPR.2013.101.