The saccadic flow baseline: Accounting for image-independent biases in fixation behaviour

Alasdair D. F. Clarke, Matthew J. Stainer, Ben Tatler & Amelia R. Hunt

January 22, 2016

Abstract

Much effort has been made to explain eye guidance during natural scene viewing. However, a substantial component of fixation placement appears to be a set of consistent biases in eye movement behaviour. We introduce the concept of *saccadic flow*, a generalisation of the central bias that describes the image-independent conditional probability of making a saccade to (x_{i+1}, y_{i+1}), given a fixation at (x_i, y_i). We suggest that saccadic flow can be used as a useful prior when carrying out analysis into fixation locations, and can be used as a submodule in models of eye movements during scene viewing. We demonstrate the utility of this idea by presenting bias-weighted gaze landscapes, and show that there is a link between the likelihood of a saccade under the flow model, and the salience of the following fixation. We also present a minor improvement to the central bias (based on using a multivariate truncated Gaussian), and investigate the leftwards and coarse-to-fine biases.

1 Introduction

The human fovea provides a small window of high acuity vision to the world, and the locations that we select to view through this window can tell us how we seek the information necessary to complete the task we are currently undertaking. Fixation locations are selected based on a combination of low-level factors such as visual salience [Borji and Itti, 2013] and high-level factors [Buswell, 1935, Yarbus, 1967, **?**]. However, there are also strong observable biases in eye movements that are independent of the content of the scene or the task being performed [Tatler and Vincent, 2009, **?**], such as a strong tendency to fixate near to the centre of images [Tatler, 2007, **?**, **?**]. If we are to gain a complete understanding of the factors that govern how we sample information, we must build models of eye guidance on the framework of these underlying biases, using them as a baseline against which to compare effects of the scene, task, image properties and individual differences.

1.1 Eye movement heuristics

One of the most influential models of eye movements of the last decade is the optimal search model [Najemnik and Geisler, 2008], which posits that human saccadic behaviour during visual search is consistent with predictions made by an ideal observer. The number of fixations human observers needed to make to find the target was closely matched by the ideal observer model, in which successive fixations were selected based on reducing uncertainty about the target's location, taking into account search history and target visibility across the visual field. The efficiency of human search (at least, in search for a Gabor patch hidden in $1/f$-noise) suggests this as a plausible mechanism for selecting fixations during search. Further evidence for this theory comes from Ma, Navalpakkam, Beck, Van Den Berg, and Pouget [2011] who find that human observers are near-optimal in a visual search task with line segments, and presented a neural network implementation of near-optimal search based on probabilistic population coding.

While this modelling framework is attractive, there are several issues. The computations driving each fixation are complex, and depend on a fairly precise representation of one's own acuity over the visual field for a wide range of possible target/background combinations. One might therefore question the assumption that these computations are undertaken to determine the location of each of the 3-4 fixations made on average every second during visual search. More importantly, Morvan and Maloney [2012] demonstrated that human observers are not able to use information about visual sensitivity in the periphery to rationally plan even a a single saccade to the optimal location in a target discrimination task. In their experiment, the observer simply has to select a location from which to detect a target that can appear with equal probability in one of two possible locations. If the locations are relatively close together, a location in between will maximise the probability of detecting a target appearing in either location. When the targets are too far apart to reliably detect the target from a point equidistant between them, the rational strategy is to look directly at one of the two possible target locations. Inconsistent with optimal viewing strategies, however, the observers did not systematically modify their choice about where to fixate according to the distance between the possible target locations. This striking failure of optimality has recently been replicated in a larger sample and generalised to other decisions in addition to eye movements [Clarke and Hunt, 2015]. To reconcile their results with those of Najemnik and Geisler [2008], Morvan and Maloney [2012] suggest *heuristics* guide saccade planning; that is, basic oculomotor biases such as a tendency to make saccades of particular amplitudes, and/or to particular regions of a display, or in particular sequences, depending on the current task.

This idea has recently been formalised in a model by Clarke, Green, Chantler, and Hunt [submitted], who demonstrate that a stochastic search model based on a memoryless random walk can find a target in noise in a similar number of fixations to human observers. The key component of this model was the use of the empirical distribution of saccades: for each saccade the model randomly samples a saccade from distributions estimating the likelihood a human observer made a saccade from (x_{i+1}, y_{i+1}) to (x_i, y_i). This stochastic model differs from the random baseline implemented by Najemnik and Geisler [2008], in which they randomly selected each fixation location from all possible points in the display, because it incorporates basic oculomotor heuristics that guide the eyes, without the need for complex computation of peripheral sensitivity or target location probability. In this paper, we re-implement and generalise this model, named *saccadic flow*, and examine the extent to which it is useful as a prior for analysing eye movements made with more natural (photographic) stimuli. This concept is illustrated in Figure 1.

1.2 The central bias

There is a strong tendency for people to look close to the centre of pictures [Clarke and Tatler, 2014, Tatler, 2007, Tatler, Baddeley, and Gilchrist, 2005, **?**] and videos [Loschky, Larson, Magliano, and Smith, 2015, Tseng, Carmi, Cameron, Munoz, and Itti, 2009] presented on computer screens. There have been a number of suggestions for why this might be, the simplest being that the centre of the stimuli is the best place to look in terms of making best use of parafoveal vision. 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 indeed 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 Tatler [2007] 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 [Parkhurst, Law, and Niebur, 2002]. 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 [Ehinger, Hidalgo-Sotelo, Torralba, and Oliva, 2009, Torralba, Oliva, Castelhano, and Henderson, 2006, **?**]. 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 are remarkably consistent across a series of eye movement databases over tasks such as free-viewing, visual search and object

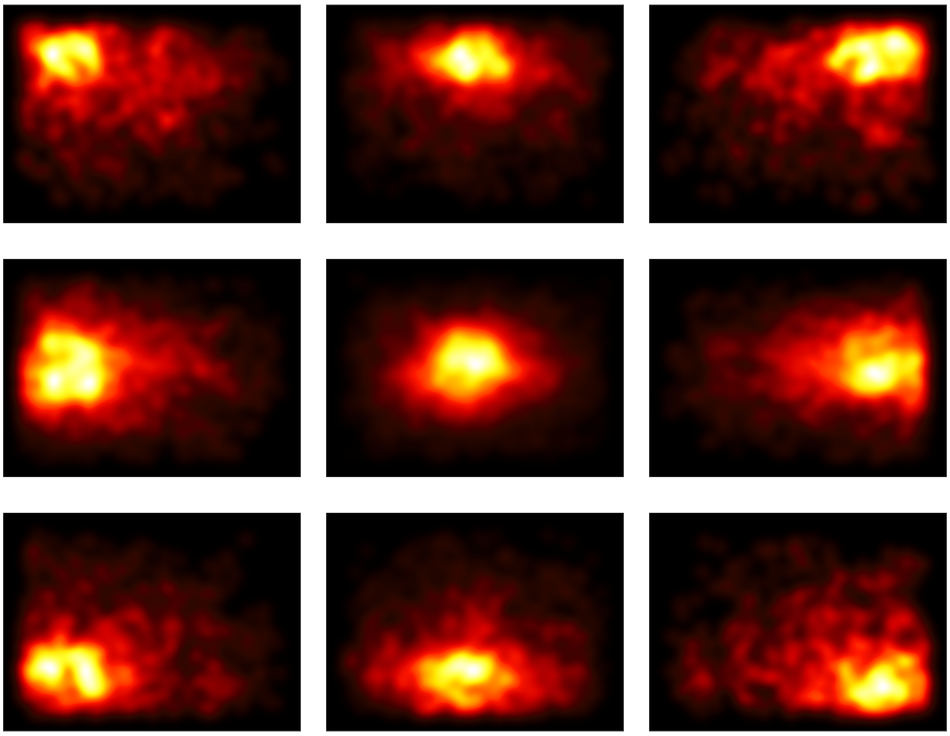

Figure 1: Saccade landing positions from fixations that were in different sections of the screen. Data from each plot has been separated into fixations in 9 spatial bins, with the screen being divided into thirds in both horizontal and vertical aspects.

naming. They proposed a simple, standardised central baseline based on a multivariate Gaussian, an demonstrated that it outperforms similar measures previously used in the literature.

1.3 Other Behavioural biases in saccades

While the central bias has attracted the most attention (at least in terms of models of visual attention), a number of other biases have been documented. These are discussed below.

1.3.1 Horizontal Saccades

Several researchers have noted that when viewing scenes there is a higher proportion of eye movements in horizontal directions than vertical or oblique movements [e.g. Foulsham, Kingstone, and Underwood, 2008, Gilchrist and Harvey, 2006, Tatler and Vincent, 2009]. There are a number of possibilities as to why this tendency exists. 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 [**?**].

Foulsham and colleagues have presented two interesting exceptions to the horizontal direction bias. Foulsham et al. [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 and 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.

1.3.2 Coarse-to-fine

Saccadic amplitudes get shorter and fixation durations get longer over time from scene onset [Over, Hooge, Vlaskamp, and Erkelens, 2007]. Replicated by Godwin, Reichle, and Menneer [2014] but they offer alternative reasons. And MacInnes, Hunt, Hilchey, and Klein [2014] among overs.

1.3.3 Leftwards bias

This falls under the more general spatial attention bias of psuedoneglect [Bowers and Heilman, 1980], which also effects line bisection tasks, etc. Dickinson and Intraub [2009] found 62% of initial saccades were directed to the left half of the image. Half of the images where mirror reversed to avoid biases in the photographs.

[Brandt, 1945, Learmonth, Gallagher, Gibson, Thut, and Harvey, 2015, Nuthmann and Matthias, 2014, Ossandón, Onat, and König, 2014, Zelinsky, 1996].

Friedrich and Elias [2014] looked at the effect of native reading direction.

1.3.4 Saccadic Momentum and Inhibition of Return

Several studies have described sequential dependencies during free viewing that bias saccades to repeat the same vector and amplitude (known as saccadic momentum) and to bias saccades away from returning to previously-visited targets (known as inhibition of return). Although both of these phenomena bias fixations away from previously-fixated locations, they differ in that inhibition of return is bound to a location in the search array, i.e. it is coded in object-based or spatiotopic coordinates (e.g. **?**), while saccadic momentum has been characterised as a basic tendency to repeat the same motor program [**?**]. Inhibition of return, unlike saccadic momentum, is task-dependent [**?**] and is disrupted by removing the scene or inhibited object [**??**]. MacInnes et al. [2014] observed both of these mechanisms operating during free visual search of a complex scene, but presumably only saccadic momentum would be consistently observed for all tasks and images.

?

1.4 The present study

These biases, in particular, the central bias, are important to take into account when evaluating the performance of models of fixation location, and investigating relationships between eye movement data and other factors. **more details and examples**.

One of the main contributions of this manuscript is to introduce the *saccadic flow* model. This can be thought of as a generalisation of the central bias: instead of simply characterising the image-independent probability of fixating (x_i, y_i) we model the conditional probabilities $p(x_i, y_i | x_{i-1}, y_{i-1})$. i.e. the probability of making a saccade from to (x_i, y_i) given we are currently fixating (x_{i-1}, y_{i-1}).

In Section **??** we demonstrate how the central bias and saccadic flow can be used as priors and components of models to improve analysis and visualisation methods. In particular, we will present bias-weighted gaze landscapes, and reanalysis parts of two previously published papers [Clarke, Coco, and Keller, 2013, Ehinger et al., 2009]. Finally, we will investigate the short-comings of these generative models by comparing synthesised data to human eye movements.

In Section 3 we will give full details of the saccadic flow model. Furthermore, we will present an improved central bias distribution and discuss the importance of the left-wards bias (pseudo-neglect).

2 Using Biases

This section make use of an improved central bias model (similar to Clarke and Tatler [2014] except uses a truncated Gaussian distribution to take the image boundaries into account: see Section 3.2) and the *saccadic flow* model (described in Section 3.4). We present three 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 examine whether saccadic flow can be used to better understand the contribution of low-level features on fixation selection, and potentially lead to better evaluation of such computational saliency models. Finally, we use flow to generate a series of saccades and compare these to observed human saccades. Being able to generate realistic synthetic datasets is useful to create an image-independent baseline with which to examine spatial maps of prediction using signal detection [see Clarke and Tatler, 2014].

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 et al., 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.

The top row of Figure 2 demonstrates that weighting the fixations by the central bias and flow model both reduce the *importance* of some fixations. The central bias model punishes fixations near the center of the image, while the flow model punishes fixations that were well predicted by the oculomotor biases of the saccadic flow model. Conversely, the models reward unlikely fixations. The second row reveals an instance of where the car to the left received less fixations than the pub sign, but that these fixations are boosted in the central bias and saccadic flow models where 'unlikely' saccades were made to this location. In the third and fourth rows, there are examples of images with a photographer bias of content towards the centre of the photograph. This reveals an example of where down-weighting the central fixations might lose important content. where the central bias model reduces the influence fixations in the centre of pictures that have important content located there. Given the tendency for photographers to bias their photographs in the centre, reducing fixations to the castle in the painting (row 3) and the girl's face (row 4).

2.2 Removing biases when examining image-dependent information

By considering saccades by the probability that they were generated by the non-image biases, we can gain further insights into the image-dependent features that are important in attracting fixation. One feature that has been shown to correlate highly with fixation is visual salience (e.g. Parkhurst?). However, others have argued that this tendency is driven by the correlation between salient objects and the central bias (XXHenderson?XX), and that the oculomotor biases which favour a central tendency would give same fixation placement regardless of saliency (TatlerVincent2009). Here, we can examine this question by looking at the relationship between saccade probability and salience at fixated locations. If salience does draw fixation then we would expect there to be no relationship between our models of oculomotor bias and salience at fixation. Conversely if the effect of salience is a by-product of oculomotor biases then we should find that saccades that occur with a high probability from oculomotor biases would have higher salience scores at fixation than fixations from

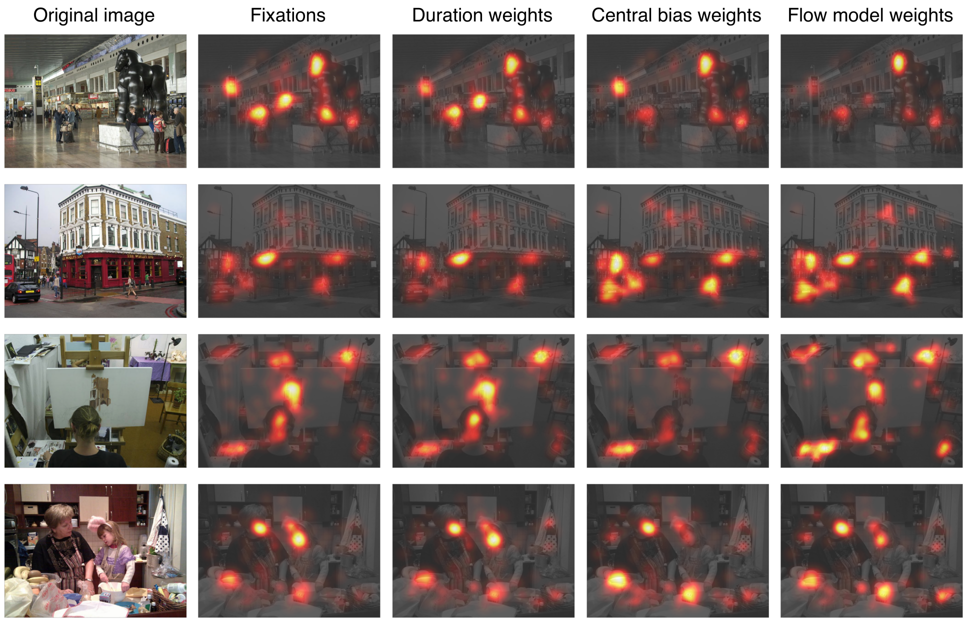

Figure 2: Examples of fixation heatmap plots from Clarke et al. [2013]. The same fixations are presented where the Gaussian at each fixation is weighted by the duration of the fixation, the centre bias model from [Clarke and Tatler, 2014] , and the saccadic flow model presented in this paper.

saccades that are less likely to occur.

We examined visual salience in the 1 degree area surrounding fixations from the XXCLARKE2013XX & Tatler 2007 free-viewing condition using the top 2 performing salience models in the MIT Saliency Benchmark that have open code - Artificial Whitening Saliency (AWS; XXREFXX) and RARE (REFS). We also examined salience as measured by the Graph Based Visual Salience (GBVS) algorithm (REFS), as this way of calculating visual salience includes a central bias. Examples of the maps can be seen in Figure 3 Maps were normalised to sum to 1, and data were analysed using linear mixed-effect models with the fixation weighting (duration, central bias or saccadic flow) as fixed effect factors, and image and participant as random effects.

2.2.1 Clarke 2013

There was no relationship between the duration of a fixation and salience using any of the salience algorithms (all p's > .05). The relationship between the probability of saccades occuring based on the central bias model and salience did not quite reach significance using AWS (p=.065), but was significant when using the RARE algorithm (p<.001), with a strong positive correlation confirming that saccades that landed close to the image centre were also highly salient, whereas fixations that were further from the centre were less salient. This was supported when using a centrally-biased salience algorithm (GBVS) in which a stronger relationship was observed. Finally, saccadic flow was significantly related to salience in all models, with saccades that were unlikely to be driven by behavioural bias having much lower salience scores than saccades that could be explained by the flow model.

2.3 Saccadic Flow as a Generative Model

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

We can see from Figure 4 that both the central bias and the saccadic flow model do an 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.

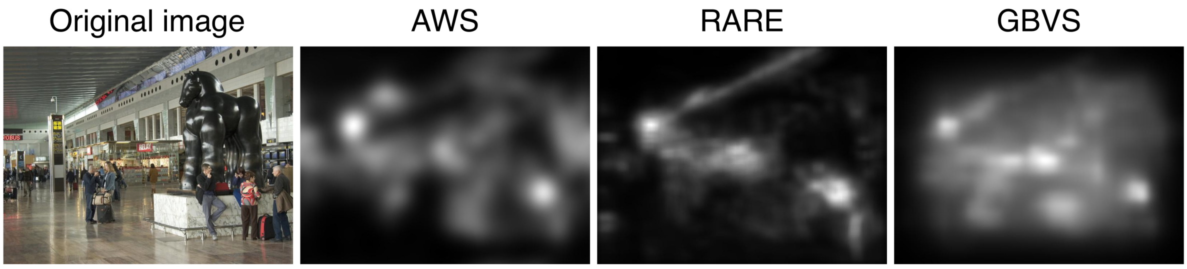

Figure 3: Example image with saliency maps made with the AWS, RARE and GBVS salience algorithms.

Table 1: Linear Mixed-effect model outputs for relationship between duration, central bias and saccadic flow salience at fixation in the Clarke et al., (2013) dataset.

Salience model	Fixation weighting	β	SE	t	p
AWS					
	Duration	2.28e-09	2.95e-08	0.08	.996
	Central bias	5.5e-07	2.71e-07	2.03	.065 .
	Saccadic flow	1.82e-07	2.09e-08	8.724	<.001***
RARE					
	Duration	-1.52e-08	3.82e-08	-0.4	.89
	Central bias	1.39e-06	3.95e-07	3.51	<.001***
	Saccadic flow	2.31e-07	3.35e-08	6.89	<.001***
GBVS					
	Duration	-3.93e-09	1.94e-08	-0.2	.973
	Central bias	3.34e-06	1.69e-07	19.79	<.001***
	Saccadic flow	1.29e-07	1.83e-08	7.03	<.001***

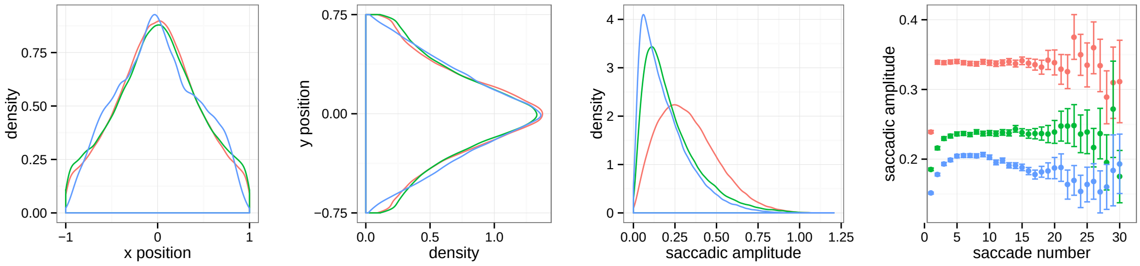

Figure 4: *blue*: human, *red*: central bias, *green*: saccadic flow. *top row*: Comparison of x and y fixation positions between human fixations and synthetic points generated from the central bias and flow model. *bottom row*: We can see that the flow model consistently makes saccades with a slightly larger amplitude than human observers. Distances are expressed relative to the width of the image.

2.4 Discussion

We have presented how biases such as saccadic flow and the central bias can be used in different ways. They can be used as a prior on the probability of making saccades to different regions of the image, allowing us to then more clearly visualise the image-dependant behaviour. These bias-weighted gaze landscapes can then be used in analysis as demonstrated in Section **??**.

We can also use the bias distributions to generate data

3 Modelling 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 uses 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], Tatler [2007], Tatler et al. [2005], Yun, Peng, Samaras, Zelinsky, and Berg [2013]. The inital 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.

- Jiang, Xu, and Zhao [2014] collected data from 16 observers viewing 500 natural scenes containing crowds of people (aspect ratio 4:3).
- 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 (≈ 0.625million 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 seperately, and split by stimuli class.

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

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 central 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 *corase-to-fine* dynamics (discussed previous in Section 2.3).

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 multivariante gaussian to the data reduces the deviance in the central bias model by 4.4%. Usng a truncated Gaussian gives us an improvement of 12%. We can round the truncated Gaussian model to $\mu = (0, 0)$, with a covariance matrix of $(0.32, 0; 0, 0.144)$ with no loss of precision. i.e. this is identical to Clarke and Tatler [2014] except with $\sigma = 0.32$ rather than 0.22

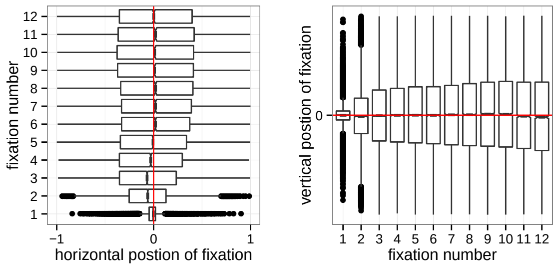

Figure 5: Distribution of horizontal and vertical fixations by fixation number.

3.3 Left v Right

Initially more fixations to the left half of the image [Nuthmann and Matthias, 2014]. We replicate this here (Figure 5).

However, it has only a very small effect on explaining the varition over whole datasets: fitting an ANOVA to predict the x coordinates of the fixations given the fixaiton number gives adjusted $R^2 = 0.004$. If we limit our analysis to the first 5 fixations in each scanpath, this only increases to adjusted $R^2 = 0.01$.

Hence we will ignore this effect from now on. By treating everything as symmetrical, we lose very little explanitory power, while restricitng 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 datapoints were removed. Multivariate polynomial regression was then used to fit 4-th order polynomials to each of the parameters. Robust estimation was used (`rlm` from the texttttMASS library) to stop the model fits being overly infleunced 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 6 shows how the parameters for the multivariate Gaussian distribution vary over horizontal position for a selection of vertical positions. The regression coefficients (given in supplementary materials) allow us to estimate the conditional probability of a saccade to $(x_1, y1)$ given the starting fixation (x_0, y_0).

How well does this model account for the fixations in our datasets? Figure 7 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 as 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 $(-1, 1) \to \mathbb{R}$ (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

What isn't captured by our flow model? There will be some stuff in MacInnes et al. [2014]

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

And mention grants.

5 Author Contribution

All authors co-wrote the paper. The saccadic flow model was developed by ADFC. The gaze landscapes and saliency analysis were done by MJS.

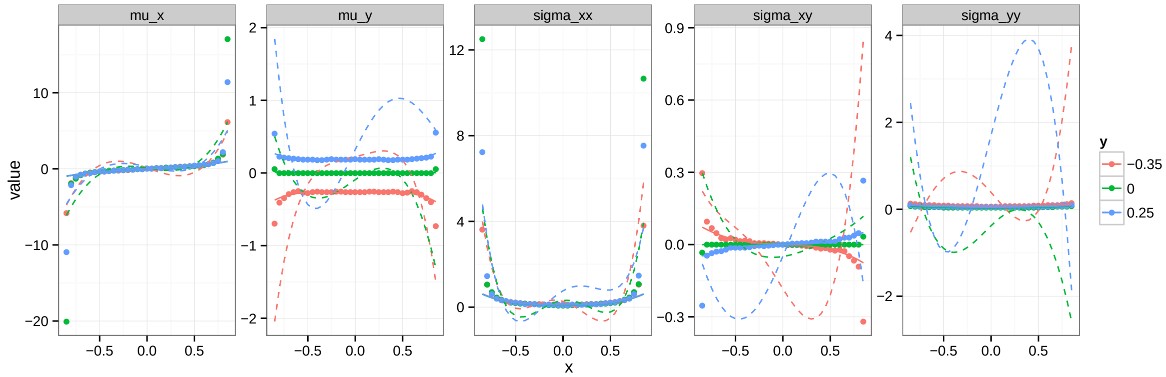

Figure 6: How the truncated Gaussian parameters vary with saccadic starting location. Dotted line show polynomial regression fits, solid line shows robust polynomial regression.

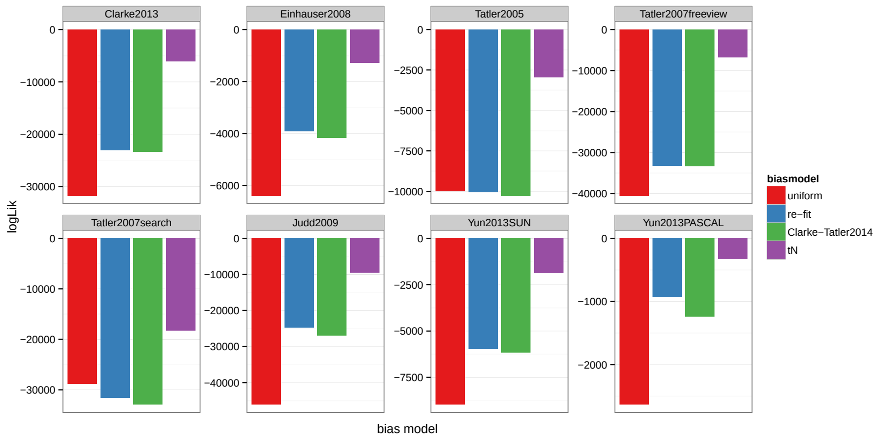

Figure 7: Flow:normal log likelihood results. We can see that re-fitting the central-bias to each specific dataset offers little improvement over using the Clarke-Tatler model, while the flow model offers a substantial improvement.

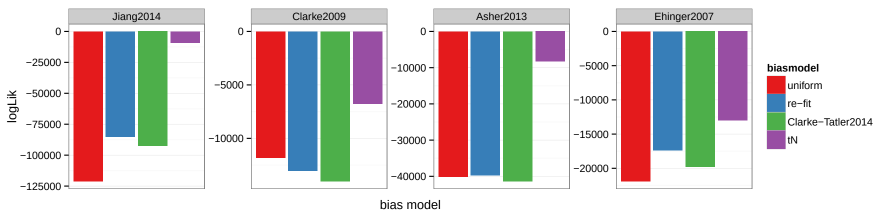

Figure 8: Doing the same but with some new testing datasets!

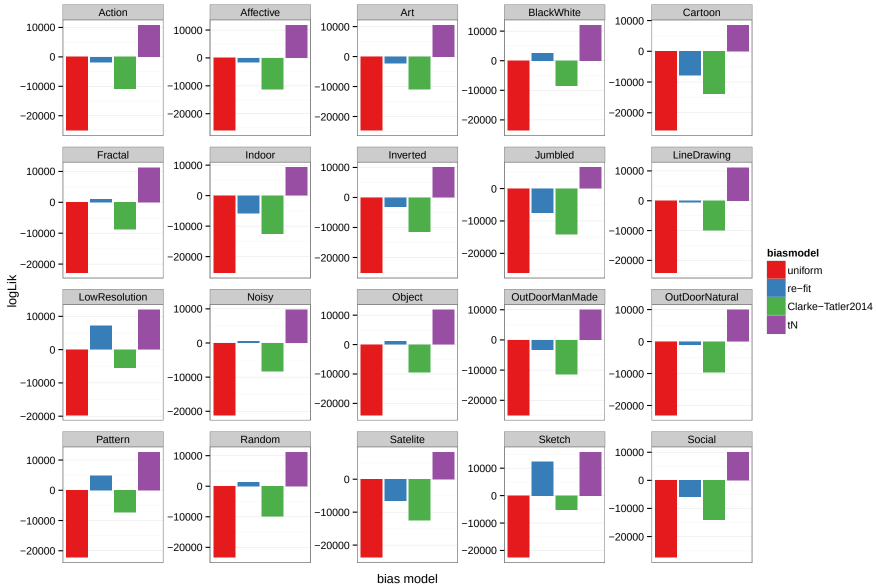

Figure 9: Flow:normal deviance results. We can see that re-fitting the central-bias to each specific dataset offers little improvement over using the Clarke-Tatler model, while the flow:normal model decreases the deviance by half.

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 Troscianko, and Iain D Gilchrist. Regional effects of clutter on human target detection performance. *Journal of vision*, 13(5):25:1–15, 2013.

A Borji and L Itti. State-of-the-Art in Visual Attention Modeling. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(1):185–207, 2013.

Ali Borji and Laurent Itti. Cat2000: A large scale fixation dataset for boosting saliency research. *arXiv preprint arXiv:1505.03581*, 2015.

Dawn Bowers and Kenneth M Heilman. Pseudoneglect: effects of hemispace on a tactile line bisection task. *Neuropsychologia*, 18(4):491–498, 1980.

Herman Francis Brandt. The psychology of seeing. 1945.

Guy Thomas Buswell. *How people look at pictures: a study of the psychology and perception in art.* Univ. Chicago Press, 1935.

Alasdair D. F. Clarke and Amelia R. Hunt. Failure of intuition when presented with a choice between investing in a single goal or splitting resources between two goals. *Psychological Science*, 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. Human search for a target on a textured background is consistent with a stochastic model. *submitted*, submitted.

Christopher A Dickinson and Helene Intraub. Spatial asymmetries in viewing and remembering scenes: Consequences of an attentional bias? *Attention, Perception, & Psychophysics*, 71(6):1251–1262, 2009.

Krista A Ehinger, Barbara Hidalgo-Sotelo, Antonio Torralba, and Aude Oliva. Modelling search for people in 900

	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
Tatler [2007] - search	30	120	visual search	5000 ms
Clarke et al. [2009]	7	360	visual search	variable
Ehinger et al. [2009]	14	912	visual search	variable
Asher et al. [2013]	25	120	visual search	variable
Jiang et al. [2014]	16	500	free viewing	5000 ms
Borji and Itti [2015]	120	4000	free viewing	5000 ms

Table 2: Summary of the 13 datasets used throughout this study.

	Eye tracker	Viewing distance	Screen size	Image size	Viewing angle	Chin / head rest
Tatler et al. [2005]	EyeLink I	60 cm	17"	800×600	$30 \times 22°$	no
Tatler [2007] - free	EyeLink II	60 cm	21"	1600×1200	$40 \times 30°$	no
Tatler [2007] - search	EyeLink II	60 cm	21"	1600×1200	$40 \times 30°$	no
Einhäuser et al. [2008]	EyeLink 1000	80 cm	20"	1024×768	$29 \times 22°$	yes
Judd et al. [2009]	?	2 feet	19"	$1024 \times 768*$	?	yes
Clarke et al. [2013]	EyeLink II	50 cm	21"	800×600	$31 \times 25°$	no
Yun et al. [2013] - PASCAL	EyeLink 1000	?	?	?	?	?
Yun et al. [2013] - SUN	EyeLink 1000	?	?	?	?	?
Clarke et al. [2009]	Tobii x50	87 cm	20"	1024×1024	$16.7 \times 16.7°$	yes
Ehinger et al. [2009]	ISCAN RK-464	75 cm	21"	800×600	$23.5 \times 17.7°$	yes
Asher et al. [2013]	EyeLink 1000	55 cm	?	1024×1280	$37.6 \times 30.5°$	yes
Jiang et al. [2014]	Eyelink 1000	57 cm	22"	1024×768	$38.8 \times 29.1°$	?
Borji and Itti [2015]	Eyelink 1000	106 cm	42"	1920×1080	$45.5 \times 31°$	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.

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.

Tom Foulsham and Alan Kingstone. Asymmetries in the direction of saccades during perception of scenes and fractals: Effects of image type and image features. *Vision Research*, 50(8):779–795, 2010.

Tom Foulsham, Alan Kingstone, and Geoffrey Underwood. Turning the world around: Patterns in saccade direction vary with picture orientation. *Vision Research*, 48(17): 1777–1790, 2008.

Trista E Friedrich and Lorin J Elias. Behavioural asymmetries on the greyscales task: The influence of native reading direction. *Culture and brain*, 2(2):161–172, 2014.

Iain D Gilchrist and Monika Harvey. Evidence for a systematic component within scan paths in visual search. *Visual Cognition*, 14(4-8):704–715, 2006.

Hayward J Godwin, Erik D Reichle, and Tamaryn Menneer. Coarse-to-fine eye movement behavior during visual search. *Psychonomic bulletin & review*, 21(5):1244–1249, 2014.

Michelle R Greene, Tommy Liu, and Jeremy M Wolfe. Reconsidering yarbus: A failure to predict observersâĂŹ task from eye movement patterns. *Vision research*, 62: 1–8, 2012.

Ming Jiang, Juan Xu, and Qi Zhao. Saliency in crowd. In *ECCV*. IEEE, 2014.

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.

Gemma Learmonth, Aodhan Gallagher, Jamie Gibson, Gregor Thut, and Monika Harvey. Intra-and inter-task reliability of spatial attention measures in pseudoneglect. *PloS one*, 10(9):e0138379, 2015.

Lester C Loschky, Adam M Larson, Joseph P Magliano, and Tim J Smith. What would jaws do? the tyranny of film and the relationship between gaze and higher-level narrative film comprehension. *PloS one*, 10(11):e0142474, 2015.

Wei Ji Ma, Vidhya Navalpakkam, Jeffrey M Beck, Ronald Van Den Berg, and Alexandre Pouget. Behavior and neural basis of near-optimal visual search. *Nature neuroscience*, 14(6):783–790, 2011.

W Joseph MacInnes, Amelia R Hunt, Matthew D Hilchey, and Raymond M Klein. Driving forces in free visual search: An ethology. *Attention, Perception, & Psychophysics*, 76(2):280–295, 2014.

Camille Morvan and Laurence T. Maloney. Human visual search does not maximize the post-saccadic probability of identifying targets. *PLOS Computational Biology*, 8(2), 2012.

Jiri Najemnik and Wilson S Geisler. Eye movement statistics in humans are consistent with an optimal search strategy. *Journal of Vision*, 8(3), 2008.

Antje Nuthmann and Ellen Matthias. Time course of pseudoneglect in scene viewing. *Cortex*, 52:113–119, 2014.

José P Ossandón, Selim Onat, and Peter König. Spatial biases in viewing behavior. *Journal of vision*, 14(2):20, 2014.

EAB Over, ITC Hooge, BNS Vlaskamp, and CJ Erkelens. Coarse-to-fine eye movement strategy in visual search. *Vision Research*, 47(17):2272–2280, 2007.

Derrick J Parkhurst, K Law, and E Niebur. Modeling the role of salience in the allocation of overt visual attention. *Vision Research*, 42(1):107–123, 2002.

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 and Benjamin T Vincent. The prominence of behavioural biases in eye guidance. *Visual Cognition*, 17(6-7):1029–1054, 2009.

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.

Antonio Torralba, Aude Oliva, Monica S Castelhano, and John M Henderson. Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search. *Psychological review*, 113 (4):766–786, 2006.

Po-He Tseng, Ran Carmi, Ian GM Cameron, Douglas P Munoz, and Laurent Itti. Quantifying center bias of observers in free viewing of dynamic natural scenes. *Journal of vision*, 9(7):4: 1–16, 2009.

Alfred L Yarbus. *Eye movements and vision.* Plenum Press New York, 1967.

Kiwon Yun, Yifan Peng, D. Samaras, G.J. Zelinsky, 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.

Gregory J Zelinsky. Using eye saccades to assess the selectivity of search movements. *Vision research*, 36(14):2177–2187, 1996.