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

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

March 2, 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 of fixation locations, and can be used as a sub-module 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 our central bias model (based on using a multivariate truncated Gaussian), and investigate the leftwards and coarse-to-fine biases in scene viewing.

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, Land and Hayhoe, 2001, 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 [Foulsham and Kingstone, 2010b, Tatler and Vincent, 2009], such as a strong tendency to fixate near to the centre of images [Canosa, Pelz, Mennie, and Peak, 2003, Stainer, Scott-Brown, and Tatler, 2013, 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 influential model 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 found 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 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 [in press], 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). It is clear from Figure 1 that the distribution of saccade end points varies considerably depending on where the saccade is launched from. Thus a model that accounts for these launch-site dependent differences in exploration biases has the potential to offer a better account of viewing behaviour. 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.

This stochastic search model is related to the more general topic of saccadic biases. Recent work in this area by Le Meur and Coutrot [2016] independently arrived at a very similar model to Clarke et al. [in press] while investigating context-dependent and spatially-variant veiwing biases. Both their model and the Stochastic Search model partitioned the data into $k \times k$ subsets (Le Meur and Coutrot [2016] used $k = 3$ while Clarke et al. [in press] used $k = 5$) and then used non-parametric methods to model the distributions. In this paper, we re-implement and generalise this idea with a parametric 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 over a range of different tasks.

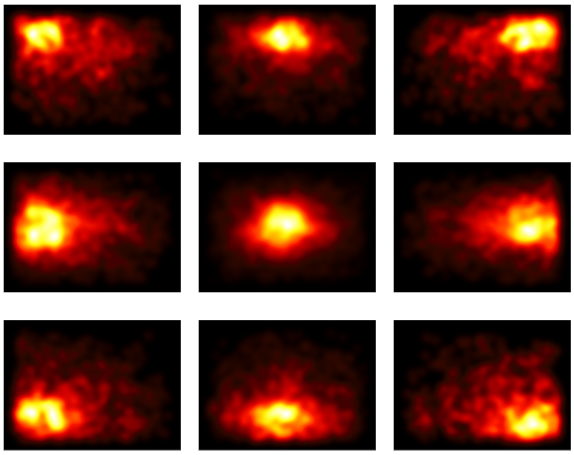

Figure 1: Saccade landing positions from fixations that were in different sections of the screen. Data from each plot has been separated into fixations in 9 spatial bins, with the screen being divided into thirds in both horizontal and vertical aspects.

1.2 The central bias

There is a strong tendency for people to look close to the centre of pictures [Canosa et al., 2003, 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 stimulus array 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, recentring in the orbit of the eye socket for most comfortable contraction of the ocular muscles (an *orbital reserve* [Fuller, 1996]). 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 [Stainer et al., 2013], suggesting that the central tendency is responsive to the scene itself rather than to the frame of the monitor upon which the scene is displayed.

Another possibility for the central fixation bias is that it represents a response to *photographer bias* in scenes, 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 [Birmingham, Bischof, and Kingstone, 2009, 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] showed 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 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.

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, Lappe, Pekel, and Hoffmann, 1998, Lee, Badler, and Badler, 2002, Tatler and Vincent, 2008]. 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 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 [von Wartburg, Wurtz, Pflugshaupt, Nyffeler, Luthi, and Muri, 2007].

Results from Foulsham and colleagues suggest that the outline of the displayed scene has a marked effect on saccade directions during viewing. Indeed, 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. Furthermore, 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 [Foulsham and Kingstone, 2010a]. 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. These findings suggest that directional biases in saccades are not only influenced by the shape of the displayed scene but also its content.

Coarse-to-fine: Another robust pattern in human saccadic behaviour is the tendency to make large eye movements after the initial scene onset, and smaller saccades as the trial unfolds [Antes, 1974, Over, Hooge, Vlaskamp, and Erkelens, 2007, Pannasch, Helmert, Roth, Herbold, and Walter, 2008]. This is often accompanied by an increase in fixation durations, and is framed as a move from ambient to focal processing [Follet, Le Meur, and Baccino, 2011, Unema, Pannasch, Joos, and Velichkovsky, 2005, Velichkovsky, Rothert, Kopf, Dornhöfer, and Joos, 2002]. Godwin, Reichle, and Menneer [2014] successfully replicated these findings, but they offered an alternative explanation, namely that this behaviours is driven stochastic factors that govern eye movements.

Leftwards bias: Several studies have shown that observers exhibit a bias to fixate the left half of a stimulus over the right [Brandt, 1945, Learmonth, Gallagher, Gibson, Thut, and Harvey, 2015, Nuthmann and Matthias, 2014, Ossandón, Onat, and König, 2014, Zelinsky, 1996]. This effects falls under the more general spatial attention bias of psuedoneglect [Bowers and Heilman, 1980], which also effects tasks such as line bisection tasks. The leftwards bias is typically short-lived, effecting only the first couple of saccades after scene onset, and while it is robust, it is comparatively weak compared to other biases in scene viewing. For example, Dickinson and Intraub [2009] found 62% of initial saccades were directed to the left half of the image during free viewing. There is some evidence that this bias is related to native reading direction [Friedrich and

Elias, 2014].

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. Krüger and Hunt [2013]), while saccadic momentum has been characterised as a basic tendency to repeat the same motor program [Wang, Satel, Trappenberg, and Klein, 2011]. Inhibition of return, unlike saccadic momentum, is task-dependent [Dodd, Van der Stigchel, and Hollingworth, 2009] and is disrupted by removing the scene or inhibited object [Klein and MacInnes, 1999, Takeda and Yagi, 2000]. MacInnes, Hunt, Hilchey, and Klein [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. The main contribution 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 3 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 demonstrate an interaction between the likelihood of a saccade under different bias models and bottom-up visual salience. Finally, we will investigate the short-comings of these generative models by comparing synthesised data to human eye movements. In Section 2 we will give full details of the saccadic flow model and an improved version of the central bias model.

2 Modelling Biases

In this section, we (i) update the central bias model of Clarke and Tatler [2014] to make use of a truncated Gaussian distribution that allows us to take the image boundaries into account. (ii) explore the strength of the leftwards bias in relation to the central bias, and (iii) describe the saccadic flow model.

2.1 Modelling Methods

Here, we give an overview of the methods and data used for the saccadic flow modelling.

2.1.1 Datasets

We will use a number of previously published datasets, covering a range of tasks, images, and experimental set-ups. This allows us to produce a model that will generalise well to other datasets. The models will be trained on eight of the ten datasets used in Clarke and Tatler [2014]. 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 is biased compared to the other datasets analysed. Both of these datasets are now used as test sets to evaluate how well our models generalise.

We also add four new datasets to the ten used by Clarke and Tatler [2014]. These will be used to test the model.

- 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] investigated 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 wide-screen 16:9 aspect ratio, the evaluations on this dataset are presented separately, and split by stimuli class.

  This gives us a relatively homogeneous training test, and a more heterogeneous test set. Hence, good performance on the test sets will likely be indicative of a generalisable result. An overview of the datasets used is given in Appendix Tables 1 and 2.

2.1.2 Pre-processing

As with Clarke and Tatler [2014], normalised all fixations to the image frame, keeping the aspect ratio constant. i.e., $(x, y) \in (-1.-1) \times (-a, a)$ with typically $a = 0.75$. The initial saccades after image onset (9.1% of the data) were excluded, giving us a total of 159,226 saccades. 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, by adding in horizontally and vertically reflected copies of the data. This has two advantages. (i) It is an easy way to make the saccadic flow bias symmetric in the horizontal or vertical directions. This is similar to how the central bias was defined Clarke and Tatler [2014]. (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]. However, as this acocunts for a relatively small proportion of the overall variance in the data (Section 2.3), we view this as an acceptable trade-off. Similarly, as we do not factor in the timecourse of the scanpath, we will not capture *coarse-to-fine* dynamics (saccadic amplitude tends to decrease with time from stimulus onset).

2.2 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 4.4%. Using 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.

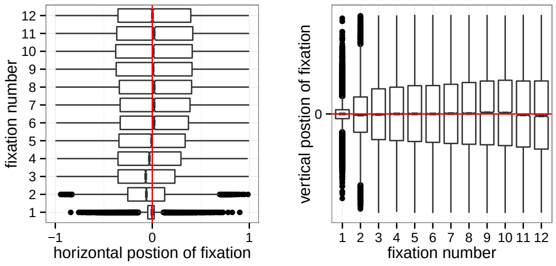

Figure 2: Boxplots showing the distribution of horizontal and vertical fixations by fixation number in the merged training set

2.3 Left v Right

As mentioned above, the downside of mirroring the saccades in our dataset is that our bias model will be symmetric and will be unable to exhibit the leftward bias observed in human fixation data. Here, we investigate the size of the leftwards bias by plotting how the distribution of horizontal fixation location varies with fixation number (Figure 2). We can see that while we do have a leftwards bias in our data, it is a small effect that only last for the first five fixations after scene onset. Furthermore, there is no sign of an asymmetry in the vertical direction. Fitting an ANOVA to predict the x-coordinates of the fixations given the fixation number gives adjusted $R^2 = 0.004$. If we limit our analysis to the first five fixations in each scanpath, this only increases to adjusted $R^2 = 0.01$. This suggests that by treating everything as symmetrical, we lose little explanatory power, while restricting the number of parameters, or increasing the amount of data available (by mirroring fixations).

2.4 Saccadic Flow

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

2.4.1 Modelling

To characterise how the distribution of saccadic endpoints varies with the start point, we used a sliding window approach. All saccades that originated from a $n \times n$ window were taken and used to fit a truncated multivariate Gaussian distribution using the `tmvtnorm` library for `R`. This window was moved in steps of $s = 0.01$ from $[-1, -0.75]$ to $[1 - n, a - n]$. Windows containing less than 250 datapoints were discarded. 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 $n = 0.05$.

Multivariate polynomial regression was then used to fit 4-th order polynomials to each of the parameters. As polynomial regression performs poorly in the presence of outliers, we will also use robust estimation (`rlm` from the `MASS` library). This will stop the model fits being overly influenced by outlier points from the image boundary.

2.4.2 Results

Figure 3 shows how the parameters for the truncated multivariate Gaussian distributions 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, y_1) given the starting fixation (x_0, y_0). As the robust estimation methods give a far better fit to the data, we will use this version of the model and discard the polynomial regression version.

How well does this model account for the fixations in our datasets? Figures 4, 5 and 6 compared the log-likelihood of the flow model compared to the Clarke and Tatler [2014] central baseline and a uniform distribution. We can see

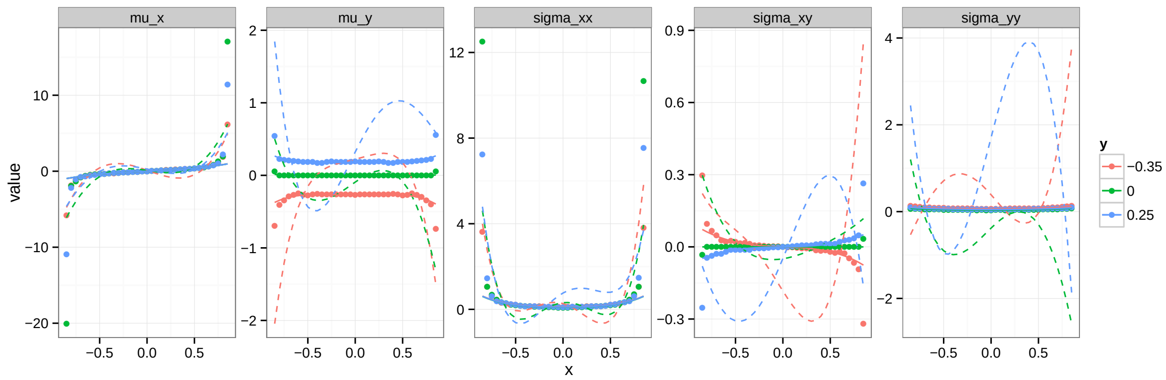

Figure 3: How the truncated Gaussian parameters vary with saccadic starting location. Dotted line show polynomial regression fits, solid line shows robust polynomial regression.

that in all cases, the log likelihood for the data under the flow model is higher than the central baseline and the uniform distribution. It is interesting to note that this holds even for datasets (those involving visual search) in which the central bias is outperformed (terms of log-likelihood) by the uniform distribution: chiefly the data from Asher et al. [2013], Clarke et al. [2009], Tatler [2007].

3 Using Biases

This section make use of an improved central bias model the *saccadic flow* model (described in Section 2.4). The new central bias model is similar to the model presented by Clarke and Tatler [2014] except for using a truncated Gaussian distribution to take the image boundaries into account. 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 gaze landscapes (also known as hotspot maps) to reduce noise and to give an improved visualisation of the image regions participants looked at more than expected. 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 theory [see Clarke and Tatler, 2014].

3.1 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 7, column 2). A potential problem with visualising data in this way is that such maps represent all fixations as being of equal importance to the inspection. 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 7, column 3).

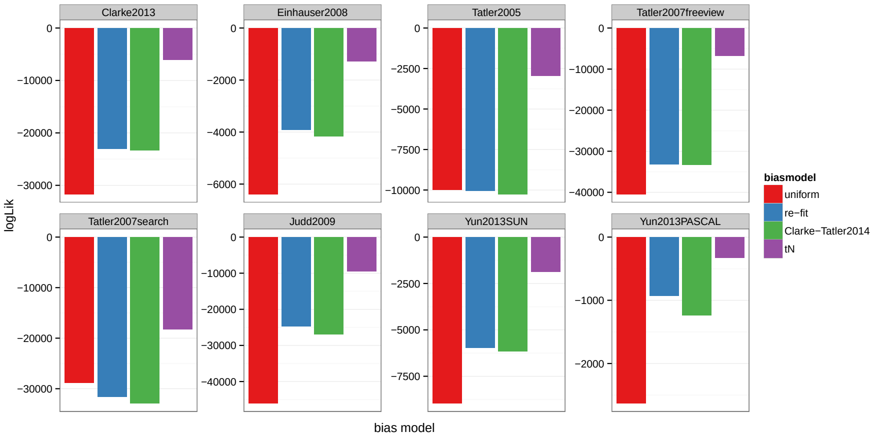

Figure 4: Flow:normal log likelihood results. We can see that re-fitting the central-bias to each specific dataset offers little improvement over using the Clarke-Tatler model, while the flow model offers a substantial improvement.

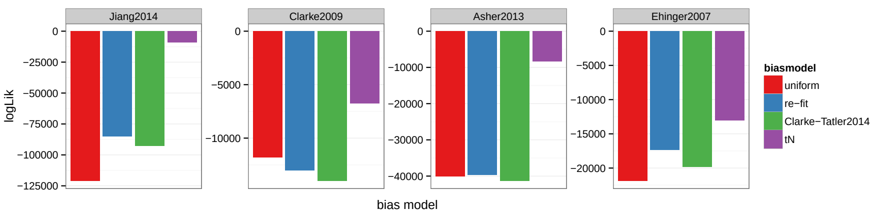

Figure 5: Doing the same but with some new testing datasets!

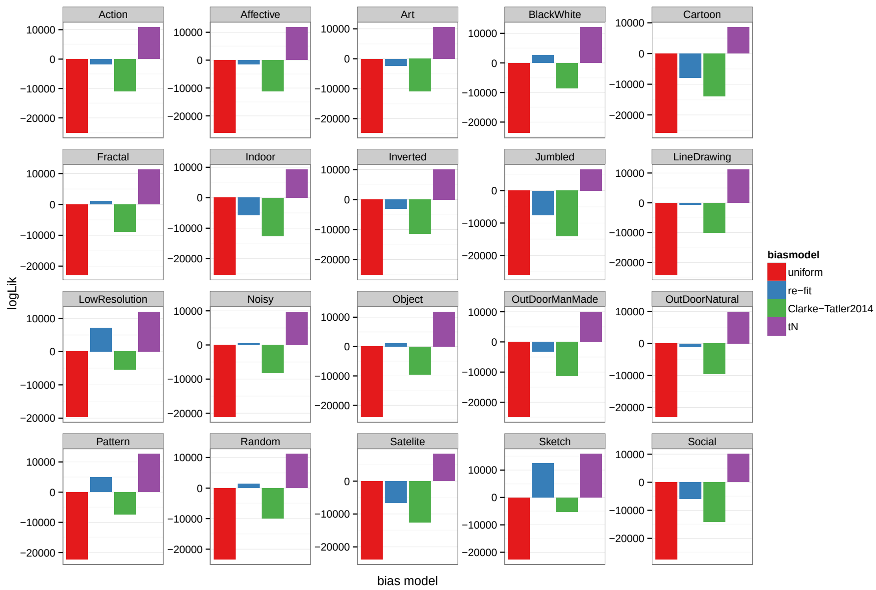

Figure 6: Flow:normal deviance results. We can see that re-fitting the central-bias to each specific dataset offers little improvement over using the Clarke-Tatler model, while the flow:normal model decreases the deviance by half.

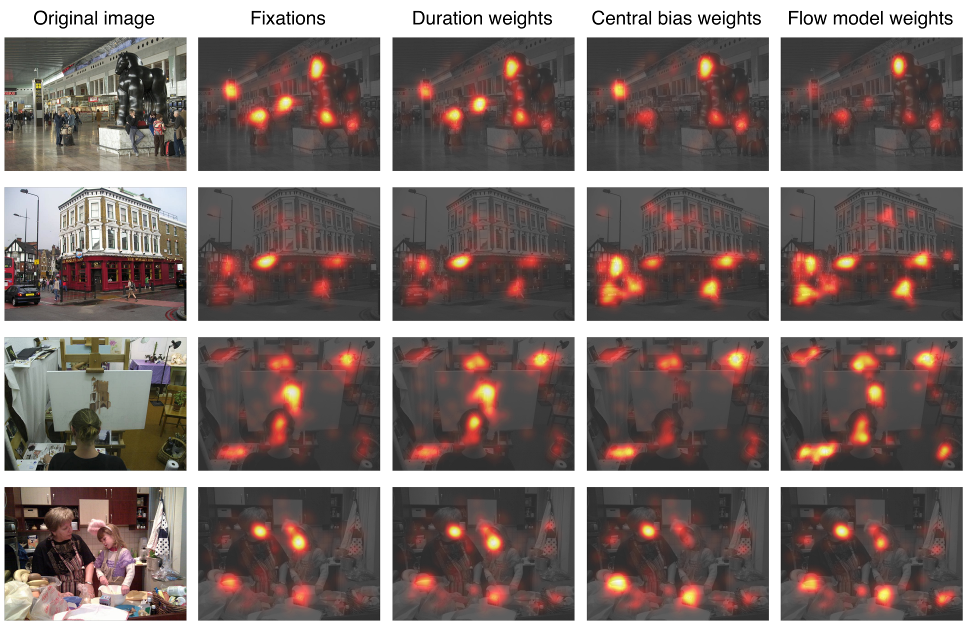

Figure 7: Examples of fixation heatmap plots from Clarke et al. [2013]. The same fixations are presented where the Gaussian at each fixation is weighted by the duration of the fixation, the centre bias model from [Clarke and Tatler, 2014] , and the saccadic flow model presented in this paper.

An 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 7 (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 provide visualisations that remove some of the image-independent biases, and reveal the more important image *dependent* information.

The top row of Figure 7 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) would perhaps punish centrally biased photographic composition. With the flow model, these areas are still represented as observers made saccades unlikely to be driven by behavioural biases to these regions.

3.2 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 with fixation is visual salience [Parkhurst et al., 2002]. However, others have argued that this tendency is driven by the correlation between salient objects and the central bias [**?**], and that the oculomotor biases which favour a central tendency would give same fixation placement regardless of saliency [Tatler and Vincent, 2009]. Here, we can examine this question by looking at the relationship between saccade probability and the ability of different conspicuity maps to predict fixation. We can therefore examine how the effect of visual salience observed in eye movement analysis is related to the behavioural biases of eye movement.

We examined the proportion of fixations that fell in the brightest 20% of pixels of salience maps by saccade likelihood from the flow and central bias models. Fixations were separated for each image into bins of 5% from the least-likely to the most-likely to be generated based on salience. We then examined what proportion of each of these bins were in the brightest 20% of salience maps using the Artificial Whitening Saliency (AWS; Garcia-Diaz, Leborán, Fdez-Vidal, and Pardo [2012]), RARE [Riche, Mancas, Duvinage, Mibulumukini, Gosselin, and Dutoit, 2013] and Graph-based visual saliency [GBVS;

Harel, Koch, and Perona, 2006] algorithms. We selected AWS and RARE as they are the two best performing salience models according to the MIT Saliency Benchmark [Bylinskii, Judd, Borji, Itti, Durand, Oliva, and Torralba, Judd, Durand, and Torralba, 2012] with publicly available code, and GBVS as it contains a bias towards the centre cause by summing neighbouring pixel values across the spatial prediction map.

Figure 8 reveals that the likelihood of making a saccade based on both the central bias and the flow model is highly related to salience in both AWS and RARE, with low-likelihood saccades being less likely to be to a salient region. Saccades that are very unlikely to be generated based on the oculomotor tendencies of eye movement (both flow and central bias) are therefore also less well explained by salience. Of the 5% fixations that were *most* likely from saccadic flow, 60% of fixations fell in the 20% thresholded region of the AWS map. However, of the 5% of fixations that were *least* likely from saccadic flow, only 40% of fixations fell in this region. This means that it may be important to consider, and potentially remove behavioural biases when attempting to predict fixation selection using feature-based models to ensure that any benefit in predictive power cannot be explained by behavioural biases correlating with salience. When examining a model that contains an inherent central bias (GBVS), we can see that weighting fixations by the Clarke and Tatler [2014] central bias model is highly related to the performance of GBVS in predicting fixation selection.

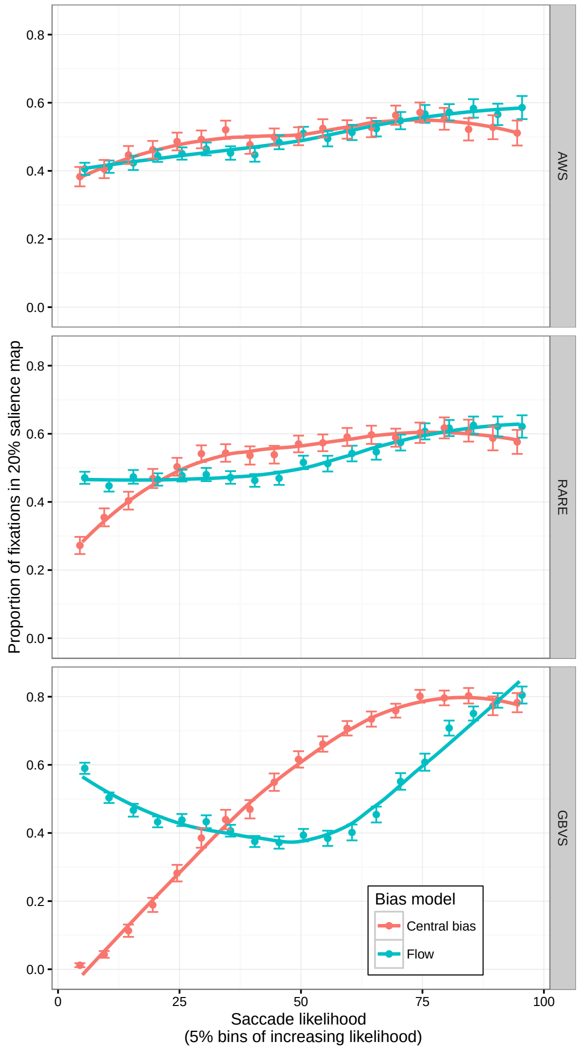

Figure 8: Saccades binned by probability of them occurring in 5% bins against the proportion of those fixations that fell in a 20% thresholded region of AWS, RARE and GBVS salience maps.

3.3 Saccadic Flow as a Generative Model

Another use of the saccadic flow model is to generate sequences of synthetic scan-paths. We can then compare the distributions with empirical scan-paths to determine which aspects of human saccadic behaviour are not captured by our model. To do this, we will create a merged dataset of fixations from the eight training datasets (175 000 fixations, including initial fixations, in total over 16 000 trials), and then generate a matched synthetic dataset such that the number of fixations in each trial is identical.

We can see from Figure 9(a) and (b) that both the central bias and the saccadic flow model do an good job of capturing the distribution of fixation locations over the x and y axes. While it is not surprising that the central bias closely matches the empirical distributions (as this is exactly what it has been fitted to), it is interesting that saccadic flow does just as good a job. Hence, the central bias can be thought of as a property of saccadic flow, and does not need to be accounted for separately.

When compared to the empirical distributions, both the central bias and saccadic flow appear to be slightly biased towards making fixations to the extreme edges of the image. This suggests that the truncated Gaussian distribution does not quite capture the effects of the image boundary on fixation selection and there is some additional aversion to fixating close to the screen edge.

Another discrepancy between the synthetic and empirical distributions can be seen with saccadic amplitudes. While the flow model is a better fit to the human data than the central bias, it still underestimates the proportion of very short saccades (Figure 9(c)). Interestingly, the flow model does manage to capture the initial increase in saccadic amplitudes after scene onset (Figure 9(d)), but it does not explain the subsequent coarse-to-fine dynamics that are seen in the empirical scan-paths.

3.4 Discussion

We have demonstrated different ways biases such as saccadic flow and the central bias can be used in eye movement research. 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. The likelihoods of fixations under the bias models are related to features such as salience. The interpretation of visual salience as a predictive model of fixation selection can be informed by considering how likely a saccade is to be generated by these models. Finally, we can also use the bias distributions to generate synthetic data that can be used as control points in ROC analysis, and to explore which aspects of human saccadic dynamics are not captured by the simple flow model.

4 Discussion

There has been much effort to generate a predictive model of human eye movements [Bylinskii et al., Judd, Ehinger, Durand, and Torralba, 2009]. We propose the saccadic flow model as a robust prior for the image-content independent saccadic behaviour that is evident when people look at pictures [Tatler and Vincent, 2009]. There are two ways in which models of eye movements may benfit from including such information. First, models may include saccadic flow in their calculation of spatial prediction. Understanding where someone is currently fixating in an image appears to dramatically influence where they will go next, meaning that as saccadic flow can be parametrically estimated from any point on an image, it may be that this can be used to weight models of low-level (i.e. visual conspicuity) and high-level (i.e. semantic interest) features. Thus, whether someone fixates one of two equally conspicuous, equally interesting objects may be simply determined by the way that the eyes *tend* to move.

The second potential utility of saccadic flow is to generate realistic control fixations with which to evaluate observed fixation data. In this way, saccadic flow can be thought of as a partner to the Clarke and Tatler [2014] central bias, and we expect that in some cases, the simpler central bias will be sufficient (for example, when examining the overall distribution of fixations rather than the sequence of saccades). However we have demonstrated that while the flow model requires more parameters (we use 16 coefficients to track how each of the five truncated Gaussian parameters vary as as a function of (x, y) - although many of them are ≈ 0), 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 limitations to our modelling work. Firstly, by using a truncated Gaussian we are unable to capture the skewed nature of the distribution of saccades originating from the corners (see Figure 1). We experimented with fitting a skew-normal distribution using the `sn` package for `R`, but met with limited success due to having to deal with the image boundaries. We expect this is one of the main reasons why our saccadic flow model generates saccades with on average greater amplitudes than those seen in empirical distributions. The second simplification is that by not taking the leftwards or coarse-to-fine biases into account, we are limiting the degree to which we can model the dynamics of human scan-paths.

The work presented here improves on recent models by Clarke et al. [in press], Le Meur and Coutrot [2016] by offering a parametric model that avoids coarsely partioning the data into large bins. We have demonstrated that the Saccadic Flow model generalises well to unseen test datasets, although this is only likely to hold for stimuli that are broadly similar to the images used to train the model (photohraphs of natural and manmade scenes). As we move away from photographic images to stimuli such as computer interfaces, we would expect the Flow model to offer a poor account of the data [Le Meur and Coutrot, 2016]. While we have shown that the model performs well over a small range of tasks (freeviewing, scene description, object naming and visual search), we do observer differences in the log likelihood when different tasks are carried out while viewing the same images.

4.1 Conclusions

Behavioural biases in eye movement are prevelant during scene viewing. Our saccadic flow model allows calculation of saccade likelihood across an image based on empirical data of how the eye tends to move in many different scene viewing conditions, with flow providing a strong fit to several datasets. There are a number of ways that flow can be developed, and we propose that gaining a better understanding of the saccadic biases underlying fixation behaviour can only be a positive for our search to understand why people look where they look.

Acknowledgements

And mention grants.

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.

A Dataset Details

Here are all the details on the datasets used in this paper. (Table 1 and 2).

References

James R Antes. The time course of picture viewing. *Journal of experimental psychology*, 103(1):62, 1974.

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.

	Observers	Images	Task	Display duration
Clarke et al. [2013]	24	100	object naming	5000 ms
Yun, Peng, Samaras, Zelinsky, and Berg [2013] - SUN	8	104	image description	5000 ms
Tatler et al. [2005]	14	48	memory	variable
Einhäuser, Spain, and Perona [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 1: Summary of the 13 datasets used throughout this study. The top eight datasets were used to train the model, while the bottom five were only used for evaluation.

	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 2: 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.

Elina Birmingham, Walter F Bischof, and Alan A Kingstone. Get real! Resolving the debate about equivalent social stimuli. *Visual Cognition*, 17(6-7):904–924, 2009.

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.

Zoya Bylinskii, Tilke Judd, Ali Borji, Laurent Itti, Frédo Durand, Aude Oliva, and Antonio Torralba. Mit saliency benchmark. http://saliency.mit.edu/.

R.L. Canosa, Jeff B Pelz, N.R. Mennie, and J. Peak. High-level aspects of oculomotor control during viewing of natural-task images. *Proc. SPIE Human Vision and Electronic Imaging VIII*, 5007:240–251, 2003.

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. *Journal of Vision*, in press.

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.

Michael D Dodd, Stefan Van der Stigchel, and Andrew Hollingworth. Novelty is not always the best policy inhibition of return and facilitation of return as a function of visual task. *Psychological Science*, 20(3): 333–339, 2009.

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.

Brice Follet, Olivier Le Meur, and Thierry Baccino. New insights into ambient and focal visual fixations using an automatic classification algorithm. *i-Perception*, 2(6):592, 2011.

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, 2010a.

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, 2010b.

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.

James H Fuller. Eye position and target amplitude effects on human visual saccadic latencies. *Experimental brain research*, 109(3):457–466, 1996.

Antón A Garcia-Diaz, Víctor V Leborán, Xosé R XR Fdez-Vidal, and Xosé M XM Pardo. On the relationship between optical variability, visual saliency, and eye fixations: a computational approach. *Journal of Vision*, 12(6):17–17, 2012.

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.

J Harel, C Koch, and P Perona. Graph-based visual saliency. *Advances in Neural Information Processing Systems*, pages 545–552, 2006.

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

T Judd, F Durand, and A Torralba. A benchmark of computational models of saliency to predict human fixations. Technical Report MIT-CSAIL-TR-2012-001, MIT Computer Science and Artificial Intelligence Laboratory, Cambridge, MA, 2012.

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.

Raymond M Klein and W Joseph MacInnes. Inhibition of return is a foraging facilitator in visual search. *Psychological science*, 10(4):346–352, 1999.

Hannah M Krüger and Amelia R Hunt. Inhibition of return across eye and object movements: The role of prediction. *Journal of Experimental Psychology: Human Perception and Performance*, 39 (3):735, 2013.

Michael F Land and Mary Hayhoe. In what ways do eye movements contribute to everyday activities? *Vision research*, 41(25):3559–3565, 2001.

Markus Lappe, Martin Pekel, and Klaus-Peter Hoffmann. Optokinetic eye movements elicited by radial optic flow in the macaque monkey. *Journal of neurophysiology*, 79(3):1461–1480, 1998.

Olivier Le Meur and Antoine Coutrot. Introducing context-dependent and spatially-variant viewing biases in saccadic models. *Vision Research*, 121:72–84, 2016.

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.

Sooha Park Lee, Jeremy B Badler, and Norman I Badler. Eyes alive. In *ACM Transactions on Graphics (TOG)*, volume 21, pages 637–644. ACM, 2002.

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.

Sebastian Pannasch, Jens R Helmert, Katharina Roth, Ann-Katrin Herbold, and Henrik Walter. Visual fixation durations and saccade amplitudes: Shifting relationship in a variety of conditions. *Journal of Eye Movement Research*, 2(2):4, 2008.

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.

Nicolas Riche, Matei Mancas, Matthieu Duvinage, Makiese Mibulumukini, Bernard Gosselin, and Thierry Dutoit. RARE2012: A multi-scale rarity-based saliency detection with its comparative statistical analysis. *Signal Processing: Image Communication*, 28(6):642–658, 2013.

Matthew J Stainer, Kenneth C Scott-Brown, and Benjamin W Tatler. Behavioral biases when viewing multiplexed scenes: scene structure and frames of reference for inspection. *Frontiers in Psychology*, 4: 624, 2013.

Yuji Takeda and Akihiro Yagi. Inhibitory tagging in visual search can be found if search stimuli remain visible. *Perception & Psychophysics*, 62(5):927–934, 2000.

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 B T Vincent. Systematic tendencies in scene viewing. *Journal of Eye Movement Research*, 2008.

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.

Pieter J A Unema, Sebastian Pannasch, Markus Joos, and Boris M Velichkovsky. Time course of information processing during scene perception: The relationship between saccade amplitude and fixation duration. *Visual Cognition*, 12(3):473–494, 2005.

Boris M Velichkovsky, Alexandra Rothert, Mathias Kopf, Sascha M Dornhöfer, and Markus Joos. Towards an express-diagnostics for level of processing and hazard perception. *Transportation Research Part F: Traffic Psychology and Behaviour*, 5 (2):145–156, 2002.

Roman von Wartburg, Pascal Wurtz, Tobias Pflugshaupt, Thomas Nyffeler, Mathias Luthi, and RM Muri. Size matters: Saccades during scene perception. *Perception*, 36(3):355, 2007.

Zhiguo Wang, Jason Satel, Thomas P Trappenberg, and Raymond M Klein. Aftereffects of saccades explored in a dynamic neural field model of the superior colliculus. *Journal of Eye Movement Research*, 4 (2):1–16, 2011.

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.

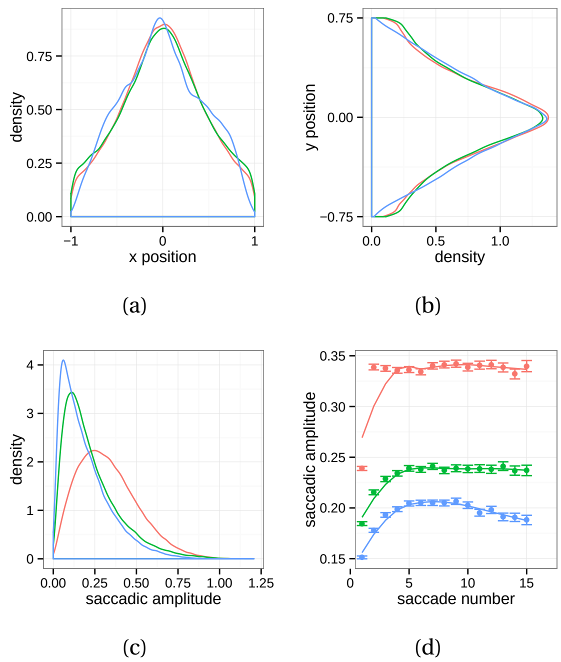

Figure 9: *blue*: human, *red*: central bias, *green*: saccadic flow. *top row*: Comparison of x and y fixation positions between human fixations and synthetic points generated from the central bias and flow model. *bottom row*: We can see that the flow model consistently makes saccades with a slightly larger amplitude than human observers. Distances are expressed relative to the width of the image. Best fit line in (d) fitted with loess regression. All distances are given in normalised units in which the width of an image is 2 (see Section 2.1).