

# Performance Characterization of Watson Ahumada Motion Detector Using Random Dot Rotary Motion Stimuli

Siddharth Jain<sup>1\*</sup>

EECS Department, University of California, Berkeley, California, United States of America

## Abstract

The performance of Watson & Ahumada's model of human visual motion sensing is compared against human psychophysical performance. The stimulus consists of random dots undergoing rotary motion, displayed in a circular annulus. The model matches psychophysical observer performance with respect to most parameters. It is able to replicate some key psychophysical findings such as invariance of observer performance to dot density in the display, and decrease of observer performance with frame duration of the display. Associated with the concept of rotary motion is the notion of a center about which rotation occurs. One might think that for accurate estimation of rotary motion in the display, this center must be accurately known. A simple vector analysis reveals that this need not be the case. Numerical simulations confirm this result, and may explain the position invariance of MST(d) cells. Position invariance is the experimental finding that rotary motion sensitive cells are insensitive to where in their receptive field rotation occurs. When all the dots in the display are randomly drawn from a uniform distribution, illusory rotary motion is perceived. This case was investigated by Rose & Blake previously, who termed the illusory rotary motion the omega effect. Two important experimental findings are reported concerning this effect. First, although the display of random dots evokes perception of rotary motion, the direction of motion perceived does not depend on what dot pattern is shown. Second, the time interval between spontaneous flips in perceived direction is lognormally distributed (mode  $\approx 2$  s). These findings suggest the omega effect fits in the category of a typical bistable illusion, and therefore the processes that give rise to this illusion may be the same processes that underlie much of other bistable phenomenon.

**Citation:** Jain S (2009) Performance Characterization of Watson Ahumada Motion Detector Using Random Dot Rotary Motion Stimuli. PLoS ONE 4(2): e4536. doi:10.1371/journal.pone.0004536

**Editor:** Ernest Greene, University of Southern California, United States of America

**Received:** September 28, 2008; **Accepted:** December 30, 2008; **Published:** February 19, 2009

**Copyright:** © 2009 Jain. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** The equipment used in the research belonged to Donald A Glaser, who also provided funding to pay human subjects used in this research. The author was supported in part by a GSR (Graduate Student Researcher) funded by Donald A Glaser and in part by a GSI (Graduate Student Instructor) provided by EECS Department UC Berkeley.

**Competing Interests:** The author has declared that no competing interests exist.

\* E-mail: fd97207@yahoo.com

‡ Current address: Bellevue, Washington, United States of America

## Introduction

Many models of visual motion perception have been proposed [1,2,3,4,5]. Although much research has since been done on studies of human visual motion perception, little work has been done to psychophysically characterize the performance of these models. This is important for obvious reasons. A correct model of motion sensing should match human psychophysical performance on motion detection, and also agree with what is known currently about the neurophysiology of motion sensitive cells in the brain.

This paper presents a psychophysical performance characterization of Watson & Ahumada's model of visual motion sensing [3], the first one to do so in my knowledge. The ability of Watson Ahumada motion detector to detect motion in random dot kinematograms is compared against human psychophysical performance. The stimulus, termed the racetrack, consists of random dots displayed in a circular annulus. The dot pattern is refreshed periodically, and a certain fraction of dots are correlated to move either clockwise (CW) or counter-clockwise (CCW) in the next frame. By varying the fraction of dots to be correlated, the

amount of motion signal in the display can be controlled (see Movies S1, S2, S3 for illustration). There are many other parameters that can be varied, and performance of both the model and human observers is measured.

The model is able to match human performance with respect to most, but not all, stimulus parameters. For example, it is found that human observers are insensitive to the dot density in the display. The model shows similar behavior. The invariance of observer performance to dot density provides strong evidence against motion models based on matching dots to their nearest neighbors in the next frame [6,7]. Such models predict that observer performance should decrease with increase in dot density, according to the probability of mismatch formula [8]. This is because as the dot density increases, the chances that the nearest neighbor is not in fact the correlated partner from the previous frame increase. Another experimental finding is that a frame duration of about 30 ms is found to be optimal for motion perception. I explain this result in terms of the spatiotemporal receptive field (STRF) structure of motion sensitive cells. At any time instant  $t$ , the response of such cells is roughly based on the

value of the spatiotemporal stimulus from time  $t-T$  to  $t$ , with  $T$  of the order of 200 ms. When the frame duration is of the order of  $T$  or higher, the input is mostly constant within a window of  $T$  ms and therefore the cells fail to detect any motion. On the other hand if frame duration is very low the input may be changing at a rate that the cells cannot handle. This will again result in failure of cells to respond optimally.

The motion in racetrack is rotary as opposed to the more commonly encountered translational case. Associated with the concept of rotary motion is the notion of a center about which rotation occurs. One might think that for accurate estimation of rotary motion in the display, this center must be accurately known. A simple vector analysis presented in this paper reveals that this need not be the case. Numerical simulations confirm this result, and may explain the position invariance of MST(d) cells. Position invariance is the experimental finding that cells that are sensitive to rotary motion are insensitive to where in their receptive field rotation occurs [9].

A special case of the racetrack is when all dots are randomly drawn from a uniform distribution in each frame, i.e., there are no correlated dots. One would expect that in this case the perception would be that of random twinkling noise, since there is no motion embedded in the stimulus. However, about two-thirds of observers report perception of rotary motion. This illusory motion was investigated by Rose & Blake previously, who termed the phenomenon the omega effect [10]. The omega effect is a classic example of patternicity, the tendency of the brain to find meaningful patterns in meaningless noise [11]. Two important results concerning this effect are reported in this paper. First, although the display of random dots evokes perception of rotary motion, the direction of motion perceived does not depend on what dot pattern is shown. Second, the time interval between spontaneous flips in perceived direction is lognormally distributed (mode  $\approx 2$  s).

It may be worthwhile to mention some aspects of the ‘‘Materials & Methods’’ in this paper that are distinct from the traditional psychophysics paradigm. In the experiments described here, each trial has a 60 s duration. During this time, the direction of rotation changes randomly and the observer is faced with the task of continuously tracking the direction of rotation. Observer performance is calculated by cross correlating observer response with actual direction of rotation. The maximum value of the normalized cross correlation function denoted by  $\chi$  is taken to be a measure of observer performance. This method is distinct from traditional psychophysics paradigm, in which the display is shown to observer for fraction of a second, and the observer has to judge if motion was perceived CW or CCW. After many trials the confusion matrix and  $d'$  is calculated [12]. The reason for the new method is none, except that it naturally occurred to me. Also it is my opinion that sub-second trial duration may not provide enough time for visual system of observer to reach steady state. One would expect that trial duration should be such that the percent correct and  $d'$  should be independent of trial duration. This can only happen if the system is in steady state. A side-benefit of the new method is that it enables the calculation of reaction time of the observer. This is the time delay at which the normalized cross correlation function reaches its maximum value. It is found that for most observers, reaction time ranges from 0.5–2 s depending on how easy it is to detect motion in the display.

In summary, the paper can be said to have three main contributions:

1. It presents a psychophysical performance characterisation of Watson & Ahumada’s model of visual motion sensing. The model is found to provide a good fit to the experimental data for most, but not all, stimulus parameters.
2. It shows that for accurate estimation of rotary motion in a display, it is not necessary that the center of rotation be accurately known. This may explain the fact that rotary motion sensitive cells found in MST/MSTd areas of the brain are insensitive to where in their receptive field rotation occurs.
3. It presents two experimental findings concerning the omega effect. First, observer response is irreproducible. Second, the time interval between spontaneous flips in perceived direction is lognormally distributed (mode  $\approx 2$  s). These findings suggest the omega effect fits in the category of a typical bistable illusion, and therefore the processes that give rise to this illusion may be the same processes that underlie much of other bistable phenomenon.

## Previous Work

Visual motion perception has been a heavily researched topic and hence this paper will necessarily limit itself to a discussion of the most relevant work. Reviews reflecting the state-of-the-art in this area can be found in [13,14,15,16]. Three seminal models of visual motion perception were proposed by Adelson and Bergen (1985), van Santen and Sperling (1985), and Watson and Ahumada (1985) [1,2,3]. Central to the Adelson Bergen & Watson Ahumada models is the concept that the entire power spectrum of an image undergoing coherent translation lies on a plane in the spatiotemporal frequency domain [17]. Determining this plane is therefore equivalent to determining the motion of the image. In its original form the Adelson Bergen motion detector is limited to detecting motion in 1D. Its extension to 2D was provided by Heeger (1987), Simoncelli and Heeger (1998) [18,19]. The model has been refined further in Rust, Mante, Simoncelli, and Movshon (2006) where it is shown that it can capture the full range of pattern motion selectivity found in MT [20]. Emerson, Bergen, and Adelson (1992) did a study in which it was shown that the responses of V1 complex cells from cat’s striate cortex were well fitted by the Adelson Bergen model [21]. Moreover, cell responses were found to be inconsistent with the van Santen and Sperling model. Cells sensitive to rotary motion have been discovered in areas MST/MSTd of the brain [22,23,9]. These cells have large receptive fields compared to cells in V1/MT. Also, they are not sensitive as to where in their receptive field rotation occurs, a phenomenon termed position invariance [9].

Random dot kinematograms (RDKs) have been widely used in studies of visual motion perception [6,7,8,10,24,25,26,27,28]. Newsome & Pare (1998) have remarked that random dot displays are useful because they stimulate primary motion sensing mechanisms while minimizing familiar positional cues [24]. Newsome, Britten, & Movshon (1989) found that a dot correlation of at least six percent is required for monkeys to be able to detect motion in RDKs undergoing translational motion [25]. The present study gives a similar result for human observers. The effect of time-sampled displays on motion perception has been previously researched by Morgan (1980), Watson, Ahumada, & Farrell (1986) [29,30]. Williams & Sekuler (1984) had studied the effect of dot density on observer performance [8]. They formulated the probability of mismatch formula according to which observer performance should decrease with increase in dot density, a view challenged by the present paper.

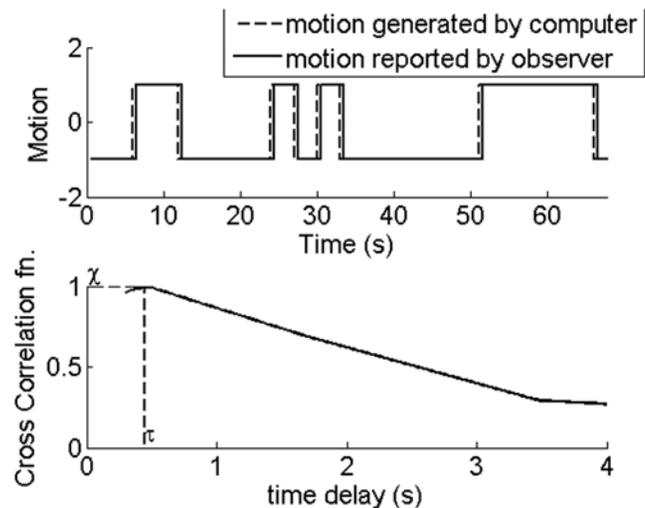
A special case of the racetrack is the omega effect, in which a display of dynamic uniformly distributed random dots in a circular annulus evokes perception of illusory rotary motion. This phenomenon was discovered by Rose & Blake (1998) [10] although they trace its origin to as far back as Mackay (1965) [31]. Recently several papers studying illusory motion from Glass

patterns have appeared in the literature [32,33,34,35]. Motion perception in such cases, where the spatial form of the stimulus is believed to guide motion perception, has generally been termed as implied motion in order to distinguish it from real motion, in which the display itself contains non-zero motion energy. Geisler (1999) had suggested motion streaks as providing a spatial cue that guides motion perception [36]. Barlow & Olshausen (2004) have explained the phenomenon of motion streaks and flow seen in Glass patterns by pointing out that the power spectrum of a motion blurred image or a Glass pattern exhibits strong anisotropy, which is a characteristic property of a moving image, and therefore excites the mechanisms that normally detect the distortions of local power spectrum caused by motion [37]. It is to be noted that the omega display does not display the anisotropy in power spectrum associated with Glass patterns, yet rotary motion is seen in it.

## Materials and Methods

The experimental stimulus used in this study is termed racetrack. Three movies of the stimulus are included with this paper. A Java applet is also available online at <http://purl.oclc.org/NET/racetrack>. The racetrack stimulus consists of a random dot pattern displayed in a circular annulus. The dot pattern is refreshed periodically. A certain fraction  $c$  of the dots, referred to as correlated dots, are rotated by an angle  $\theta$  in the next frame. The remaining dots have their positions generated randomly and uniformly in Cartesian  $(x,y)$ , and are representative of noise. The algorithm for generating dots is such that if a dot is correlated in the present frame, it is guaranteed not to be correlated in the next frame. This eliminates the appearance of multiple dot trajectories, and thus the only motion cues in the stimulus are two dot apparent motion cues. Observers see a swarm of dots that appears to rotate clockwise (CW) or counter-clockwise (CCW). The direction of rotation changes randomly according to the polarity of a coin that flips every 3 s. They are instructed to click the left mouse button for CCW motion, and the right mouse button for CW motion. By cross correlating the observer response with the actual direction of rotation of the correlated dots, an estimate of observer performance and reaction time denoted by  $\chi$  and  $\tau$  respectively is obtained. This process is illustrated in Figure 1.  $\chi$  is defined as the maximum value of the normalized cross correlation function.  $\tau$  is the time delay at which  $\chi$  occurs. A  $\chi$  value of 1 indicates perfect detection of the embedded motion. At  $c = \theta$  the observer response can still be cross correlated with the input signal, which would have dictated the rotation of correlated dots if there were any in the stimulus. The  $\chi$  value obtained in this case reflects chance, or zero, detectability of embedded motion. Response reproducibility is quantified by cross correlating observer response curves in response to the same stimulus in multiple trials.

Definitions and default values of some parameters are as follows: dot correlation  $c$  = number of correlated dots/total number of dots; frame duration  $fd$  = length of time for which a frame stays on screen, default = 30 ms; dot density  $dd$  = dots per unit area, default = 5 dots/degrees<sup>2</sup>; angle of rotation  $\theta$  = angle by which correlated dots are rotated, default = 5°, the spatial hop size of a correlated dot =  $r\theta$  where  $r$  is distance of dot from center,  $\theta$  in radians; inner circle diameter  $ic$  = angle subtended by inner circle diameter at the eye, default = 7°; outer circle diameter  $oc$  = angle subtended by inner circle diameter at the eye, fixed at 10° in all experiments; dot diameter = 5', fixed in all experiments; duration of a trial = 60 s. Stimuli were displayed on a NEC MultiSync FP1370 22" (20" viewable image size) CRT monitor with display resolution = 640×480@100 Hz; black dots (luminance ≈ 0) against a background luminance of 10.8 cd/m<sup>2</sup> were displayed; viewing



**Figure 1.** The dotted curve is the motion generated by the computer, and the solid curve is the motion reported by the observer.  $\chi$  is the maximum value of the normalized cross correlation function and  $\tau$  is the time delay at which  $\chi$  occurs. doi:10.1371/journal.pone.0004536.g001

distance = 1.65 m. The range and default values of some parameters is summarised in Table 1.

The study was conducted over a period of several years, and new observers were recruited as old ones dropped out. In all experiments the number of observers is at least four, and number of trials  $\geq 20$  for each data point shown in the figures. Error bars in the figures equal one standard deviation (s.d.), unless otherwise stated. Custom software was written by the author in C# to generate the stimuli. The study was approved by Committee for Protection of Human Subjects (CPHS), UC Berkeley. Written informed consent was obtained from subjects.

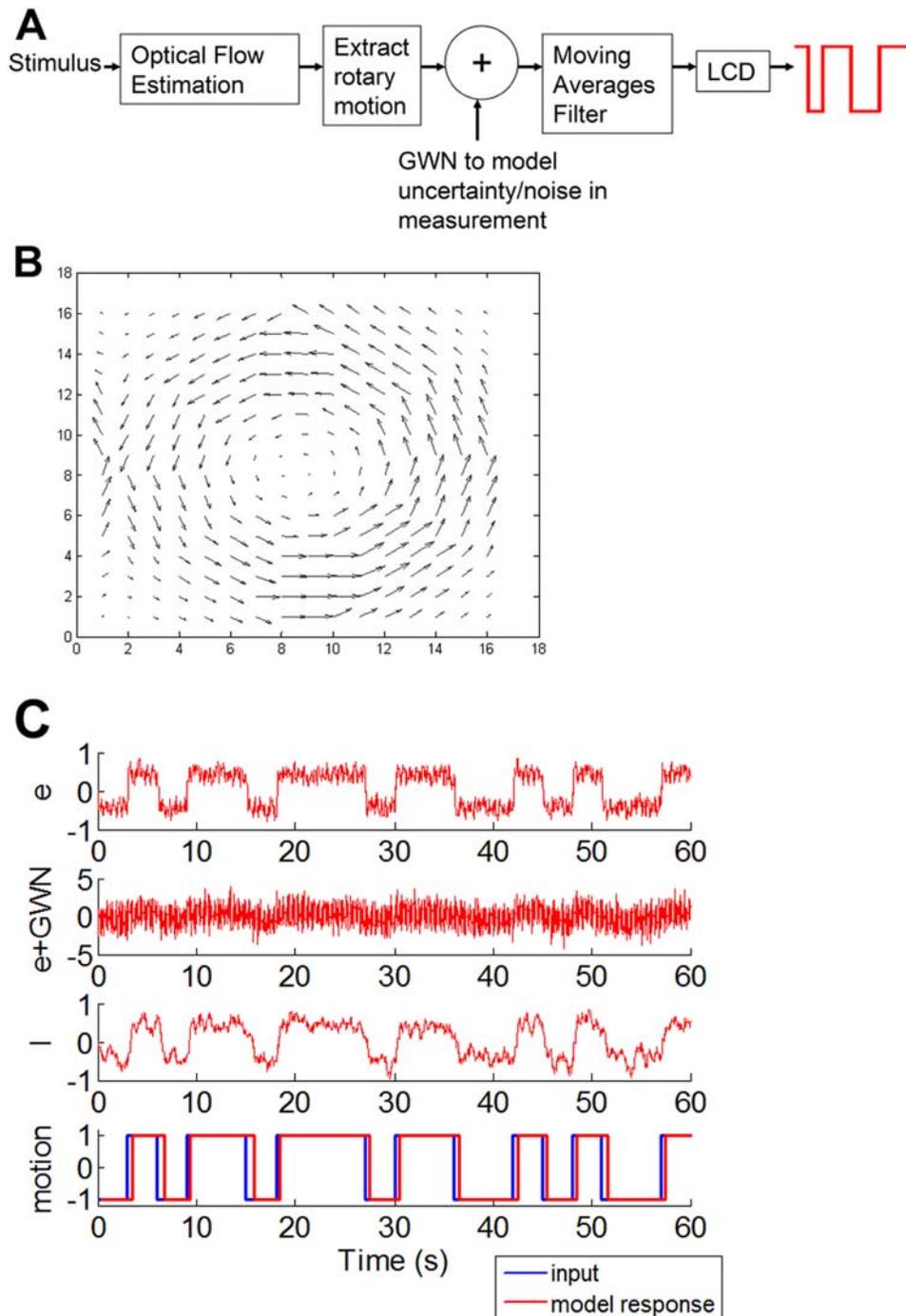
## Model Description

The following steps and Figure 2 describe the complete pipeline used to model observer responses to the racetrack:

**Table 1.** Default values and range of various parameters used in experiments.

Parameter	Range	Default value
Dot correlation $c$	0–0.5	-
Frame duration $fd$	10–100 ms	30 ms
Dot density $dd$	1–25 dots/degrees <sup>2</sup>	5 dots/degrees <sup>2</sup>
Angle of rotation $\theta$	1–20°	5°
Inner circle diameter $ic$	1–9.5°	7°
Outer circle diameter $oc$	-	10°
Dot diameter	-	5'
Duration of a trial	-	60 s
Dot luminance	-	≈ 0
Background luminance	-	10.8 cd/m <sup>2</sup>
Monitor resolution	-	640×480 pixels
Viewing distance	-	1.65 m

doi:10.1371/journal.pone.0004536.t001



**Figure 2. (a) Block schematic of the model (b) Optical flow output by Watson Ahumada motion detector (c) Model response at various other stages in the pipeline.**

doi:10.1371/journal.pone.0004536.g002

- Step 1: Stimulus is input to the Watson Ahumada motion detector, which at time  $t$  gives the instantaneous optical flow.
- Step 2: The optical flow is easily converted into a measure of rotary motion signal by taking cross products with radial vector, followed by weighted averaging. The weights are obtained in Step 1; for each velocity estimate the Watson Ahumada detector is able to provide a confidence/error

measure which is used as the weight. Area MST(d) in the brain is believed to carry out this type of processing, where the local motion signals from stage MT are pooled to estimate global patterns of rotation and expansion that guide in heading estimation [9,38,39]. The output of this step is denoted by  $e(t)$ .

- Step 3: The human visual system must integrate information over a certain interval of time to compute a reliable estimate of

motion. This is achieved by passing  $e(t)$  through a moving averages filter, with window size of half a second. The output of this step is denoted by  $I(t)$ .

- Step 4: While doing psychophysical experiments with human observers, the only information available is the direction in which the observer is perceiving motion. Therefore in order to compare model response with experimental psychophysics,  $I(t)$  is passed through a level crossing detector (LCD) with thresholds  $\pm B$ .  $B = 2\sigma(I)$  at  $c = 0$  under default parameters. This choice of  $B$  makes the events when  $I(t)$  may cross detection threshold, given there is no rotary motion in the stimulus, unlikely. The behavior of level crossing detector is as follows: when input crosses  $+/-B$  the detector signals CCW/CW motion respectively, and continues to do so until the input crosses threshold in the opposite direction. When that happens, the LCD flips to the opposite state.

$\chi$ ,  $\tau$  can now be computed for the model, and values compared to experimental psychophysics. In its present form the model is strictly deterministic. However, the human visual system necessarily exhibits some variability, characteristic of any real world physical system. In fact as shown in the results section, it is found that at  $c = 0$  observer responses are not reproducible. This variability in response is incorporated into the model by adding Gaussian White Noise (GWN)  $n(t)$  to  $e(t)$ , until the model response reproducibility also drops to zero at  $c = 0$ . This occurs for  $k = \sigma(n(t))/\sigma_\theta \approx$  approximately six, where  $\sigma_\theta$  stands for  $\sigma(e)$  at  $c = 0$  under default parameters. Accordingly  $k$  was fixed at six. Model simulations were run at a resolution of  $128 \times 128$  pixels, unless otherwise stated. Circular dots in psychophysical experiments were approximated as squares of equal area in model simulations. In all results, Watson Ahumada sensors are tuned to a center frequency of 0.64 cycles/degrees, unless otherwise stated. The reason for this setting is that it gave acceptable results. It may also be noted that most of the energy in power spectra of natural images is concentrated at relatively low spatial frequencies [40]. Default values of model parameters are summarised in Table 2. My source code for the Watson Ahumada component of the model is publicly available [41].

## Results

We begin with a discussion of the omega effect ( $c = 0$  case of racetrack), and present two important results. First, although the display triggers perception of rotary motion, the direction of motion perceived is not dependent on what dot pattern is shown.

**Table 2.** Default values of model parameters used in simulations.

Parameter	Default value
Spatiotemporal filters	As in Watson & Ahumada (1985) [3]
Center frequency	0.64 cycles/°
Noise	$\sigma(n(t))/\sigma_\theta \approx 6$
LCD threshold $B$	$2\sigma(I)$ at $c = 0$
Moving averages filter	Impulse Response $h(t) = \begin{cases} 1/T & \text{if } 0 < t < T = 0.5 \text{ s} \\ 0 & \text{otherwise} \end{cases}$
Resolution	$128 \times 128$ pixels

doi:10.1371/journal.pone.0004536.t002

Second, the time interval between spontaneous flips in direction exhibits a lognormal distribution.

## Omega effect: Response Reproducibility and distribution of spontaneous flips in perceived direction

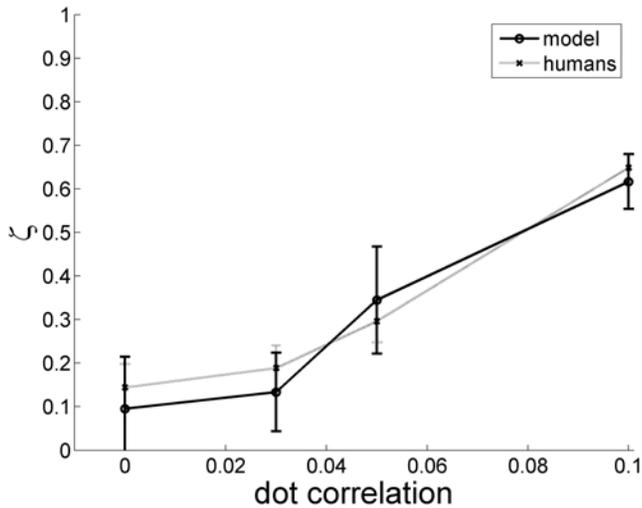
As mentioned earlier, the omega effect is the  $c = 0$  case of the racetrack. About two-thirds of observers report perception of rotary motion at  $c = 0$ , even though there is no motion embedded in the stimulus [42]. The perceived direction of motion changes randomly from time to time. After prolonged viewing most observers can usually choose the perceived direction of motion at will. For some observers the direction of motion switches when a sudden attention grabbing stimulus is given (such as a sudden tap on the back of the head). Some observers have even remarked that mere pressing of a mouse button causes the perceived direction of motion to reverse.

An important characteristic of the omega effect is that an observer gives different responses to the same stimulus in multiple trials. This is quantified in the following way. The observer response curves in response to the same stimulus in two separate trials are cross correlated. Let  $\zeta$  denote the maximum value of the normalized cross correlation function.  $\zeta$  is taken to be the measure of response reproducibility. It is found that the value of  $\zeta$  when the same stimulus is shown in multiple trials is no different than the value of  $\zeta$  when different stimuli are shown in multiple trials. Thus, the response reproducibility of the omega effect is zero. This may happen because the display is inherently ambiguous like most, if not all, bistable illusions. Both interpretations are equally likely and the brain randomly chooses a configuration at any time instant. It is found that  $\zeta = \mu_1 \pm \sigma_1 = 0.145 \pm 0.1048$  (mean  $\pm$  s.d.) based on 47 trials in which same stimulus is shown from trial to trial. Further,  $\zeta = \mu_2 \pm \sigma_2 = 0.118 \pm 0.1359$  based on 67 trials in which different stimuli are shown in multiple trials. One sided  $t$ -test to test the null hypothesis  $\mu_1 = \mu_2$  against the alternate hypothesis  $\mu_1 > \mu_2$  gives  $t = 1.196$ . At  $\alpha = 0.05$  level of significance the null hypothesis cannot be rejected ( $P$  value = 0.1158).

The foregoing discussion has shown that the reproducibility of response is zero for the omega effect ( $c = 0$ ). However, intuitively we expect if  $c$  is not zero, i.e., some dots are deliberately correlated to undergo rotary motion, then observers should start responding in direction of motion of correlated dots. Figure 3 shows the response reproducibility increases with  $c$  as expected ( $\zeta = 1$  reflects perfect reproducibility).

Figure 4(a) shows the histogram of the inter flip interval (IFI), which is the time interval between spontaneous reversals in perceived direction of motion, at  $c = 0$ . The mode of the histogram occurs at  $IFI \approx 2$  s. The histogram is well approximated by a lognormal distribution which is evident in Figure 4(b), where the pdf (probability density function) of  $\ln(IFI)$  is plotted together with a Gaussian fit. The IFI of many bistable illusions is lognormally distributed. Such distributions are common in biology, and one way to interpret them is in terms of the noise driven motion of a state point [43].

The mechanisms underlying omega effect are not clear. When dots are displayed in a circular annulus their freedom of movement is restricted. The dots at the boundary cannot move in all  $360^\circ$  directions. In the limit when the annulus width tends to zero, the dots can only move tangentially. This suggests an increase in the omega effect with decrease in annulus width which is experimentally true [10]. When the annulus has appreciable width the dots at boundary are more likely to bounce normal to the boundary. Some observers do report perception of a radial pulsating motion in the omega display [42]. Rose & Blake (1998) postulated that the omega effect arises because of interaction



**Figure 3. Response reproducibility  $\zeta$  vs.  $c$ .** Both model and humans show zero reproducibility at  $c=0$ , and the reproducibility steadily increases with  $c$ , because the motion signal gets stronger. doi:10.1371/journal.pone.0004536.g003

between cells that are sensitive to curvature in the display, and cells that are sensitive to motion [10]. For its part, the Watson Ahumada model outputs a zero mean white noise like signal in response to the omega display, since there are dots bouncing off in all directions randomly. This signal combined with the intrinsic noise  $n(t)$  (which at  $c=0$  is six times stronger than the Watson Ahumada signal) results in rapid zero mean fluctuations. Because of their stochastic nature, these fluctuations become large enough at times to cross the LCD thresholds. The IFI distribution resulting from such stochastic fluctuations is also shown in Figure 4(b) for comparison.

**Effect of dot correlation  $c$**

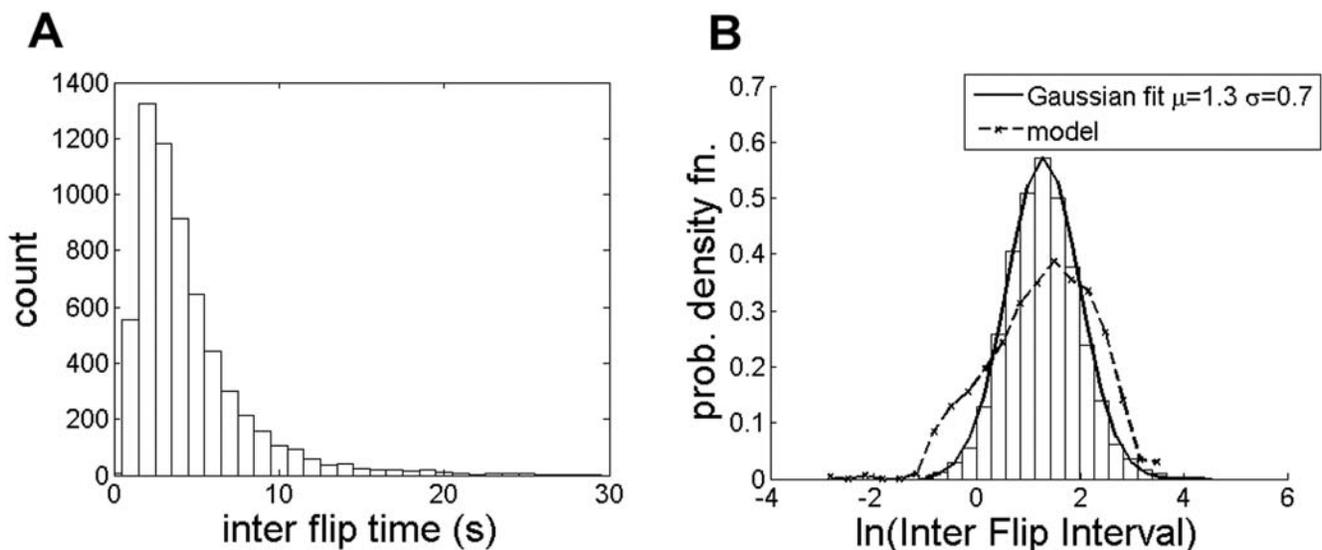
Figure 5(a) shows variation of signal detectability  $\chi$  vs. the dot correlation  $c$ . A  $\chi$  value of 1 means perfect detection, and  $\chi$  at  $c=0$

reflects the baseline zero level of  $\chi$  corresponding to chance detectability. The increase in  $\chi$  with  $c$  is easy to understand, as the value of  $c$  directly controls the amount of motion embedded in the stimulus. As can be seen from the figure, the model fits the experimental data very closely. If the threshold for motion perception is defined as the value of  $c$  for which  $\chi$  is one standard deviation higher than the  $\chi$  value at  $c=0$ , then this gives a threshold of  $c$  in the range of 0.03 to 0.06. This is comparable to thresholds reported elsewhere [24,25]. The experimental method described in this paper allows the measurement of the reaction time  $\tau$  of an observer. Figure 5(b) shows a graph of the reaction time  $\tau$  vs.  $c$ . For  $c \approx 0$ ,  $\tau$  is about 1.5 s, and decreases steadily with increase in  $c$ . It takes less time to recognize the motion signal as the signal gets stronger. At high values of  $c$ ,  $\tau$  is about half a second. The model is seen to fit the experimental data well. In general  $\chi$  and  $\tau$  are inversely correlated as shown in Figure 5(c). Parameters that tend to increase  $\chi$  tend to decrease  $\tau$  and vice-versa.

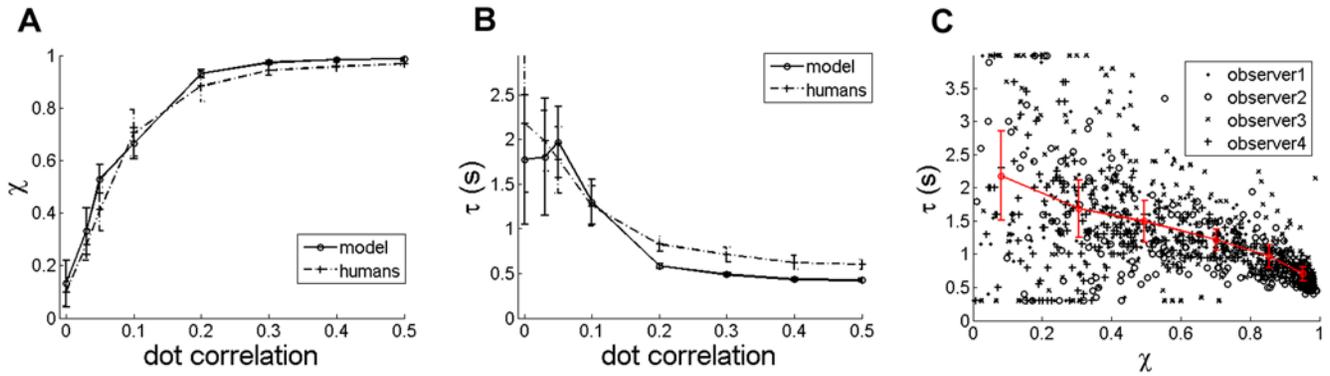
**Effect of frame duration  $fd$**

Figure 6 shows that  $fd \approx 30$  ms is optimal for motion perception. The same sequence of frames that evoke perception of vivid motion at  $fd \approx 30$  ms, fail to evoke any perception of motion at  $fd \gg 30$  ms. The explanation proposed for the  $fd$  effect seen here is as follows. The motion computed by local motion detectors at time  $t$  is based on the spatiotemporal signal from time  $t-T$  to time  $t$ , where  $T \approx 200$  ms is the temporal size of receptive fields of simple/complex cells found in the primary visual cortex [44]. When  $fd$  is too large the input is mostly constant within a window of 200 ms and so motion sensitive cells will fail to detect any motion. On the other hand, if  $fd$  is too small the input may be changing at a rate that the cells cannot handle. The bandwidth of the stimulus, viewed as a continuous-time signal, is directly proportional to the rate at which the individual racetrack frames are played. When  $fd$  is too low, the correlated dot will stay in the receptive field (RF) of a motion sensitive cell for a very brief interval of time, and will not excite the spatiotemporal RF profile of the cell.

The model results are close to that of experiment, except for the  $\chi$  values at  $fd = 10$  ms. This may be because of the high bandwidth of neurons used in the model. It is interesting to note that without



**Figure 4. (a) histogram of Inter Flip Interval (IFI) at  $c=0$  (mode  $\approx 2$  s) (b) normalised histogram of  $\ln(\text{IFI})$  together with a Gaussian fit.** The pdf of  $\ln(\text{IFI})$  given by the model is also shown. doi:10.1371/journal.pone.0004536.g004



**Figure 5. (a)  $\chi$  vs. (b)  $\tau$  vs. (fd = 30 ms, ic = 7°, dd = 5 dots/degrees<sup>2</sup>) (c)  $\tau$  vs.  $\chi$  scatter plot and piecewise linear fit for experimental data.**  
doi:10.1371/journal.pone.0004536.g005

noise,  $\chi$  at fd = 100 ms is at the baseline zero level. If noise is added,  $\chi$  rises above zero level, and matches value given by human observers. This is reminiscent of the beneficial effect noise may sometimes play in a system, by stochastically boosting a subthreshold signal in the manner of stochastic resonance [45,46,47].

**Effect of dot density dd**

Figure 7 shows the effect of varying the dot density dd in the display. Humans display a remarkable indifference to the dot density in the display. This shows that it is the relative proportion of the correlated dots that matters, not their absolute number. The experimentally observed independence of observer performance on dot density cannot be explained by models of motion perception based on matching dots or features to their nearest neighbors in the next frame [6,7]. Such models display a marked dependence on dot density in the display according to the probability of mismatch formula [8]. As the dot density increases there are more dots per unit area, and the chances that the nearest neighbor is not the correlated partner increase.

A derivation of the probability of mismatch formula follows. A correlated dot is displaced by a distance  $h$  in the next frame. A nearest neighbor model operates by matching dots to their nearest

neighbors in the next frame. The matching directly gives the local motion vectors, which is the output of the model. Therefore, for the correlated dot to be matched correctly to its partner, no dot should fall within a circle of radius  $h$  in the next frame. The probability of this happening, which is equal to the probability of no mismatch equals:

$$P(\text{no mismatch}) \approx \left(1 - \frac{\pi h^2}{A}\right)^N$$

where  $A$  is the area of display,  $N$  is total number of dots in the display,  $N = A \cdot dd$ , and we assume that dots are uniformly distributed.

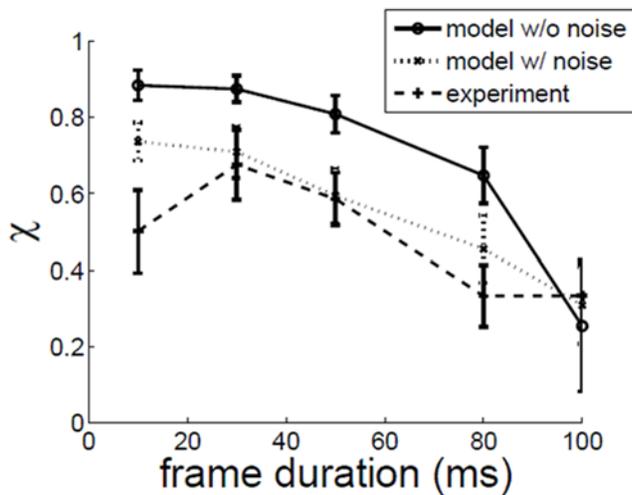
Approximating  $1-x$  as  $\exp(-x)$  for  $x$  sufficiently small, and substituting  $A \cdot dd$  for  $N$ ,

$$P(\text{no mismatch}) \approx \exp(-\pi h^2 dd)$$

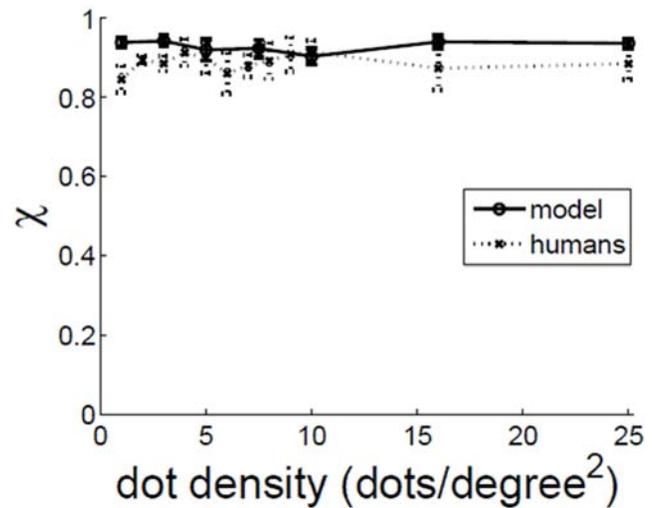
therefore,

$$P(\text{mismatch}) = 1 - P(\text{no mismatch}) = 1 - \exp(-\pi h^2 dd) \quad (1)$$

which the formula given in Williams & Sekuler (1984) [8]. The



**Figure 6.  $\chi$  vs. frame duration fd.** c = 0.1, ic = 7°, dd = 5 dots/degrees<sup>2</sup>.  
doi:10.1371/journal.pone.0004536.g006



**Figure 7.  $\chi$  vs. dot density dd.** c = 0.2, ic = 7°, fd = 30 ms. Model simulations done at 256×256 pixel resolution for dd > 10.  
doi:10.1371/journal.pone.0004536.g007

probability of mismatch values for the dot density range used in Figure 7, are tabulated in Table 3. This formula makes it explicit that as the dot density increases, there would be more and more mismatches, and therefore observer performance should decrease with increase in dot density. In reality, however, observer performance is independent of dot density in the display. The Watson Ahumada motion detector is able to capture this independence as shown in Figure 7.

It may be noted that if some of the assumptions leading to the formula in equation (1) do not hold, the analytic form of  $P(\text{mismatch})$  may no longer be given accurately by  $1 - \exp(-\pi h^2 dd)$ . However, the central thesis of the formula that observer performance should decrease with dot density will still remain true. This is because as the dot density increases, there are more dots per unit area, and therefore the expected separation between dots would decrease. When the expected dot separation becomes less than the hop size  $h$ , the matching would be dominated by mismatches, and performance would decline. It may be appropriate here to remark on the study of Grzywacz, Watamaniuk and McKee (1995) (figure 1 in their paper) [48]. It appears to me that the authors correctly simulated the Adelson Bergen model and found that it is insensitive to dot density. However, they concluded incorrectly, misguided by the probability of mismatch formula, that psychophysical results should depend on dot density.

**Effect of spatial hop size  $h$**

The hop size is the amount of displacement given to the correlated dots. By default the correlated dots are rotated by an angle of  $5^\circ$ . With  $ic = 7^\circ$ , and angle subtended by outer circle fixed at  $10^\circ$ , this translates to average displacement of  $\frac{7+10}{2} \times 5 \times \frac{\pi}{180} = 0.37^\circ$  visual angle on the eye. Figure 8 shows the effect of varying the hop size for the model and humans, at different dot densities. The correlated dots were rotated by angles of  $\{1,5,10,15,20\}$  degrees, corresponding to average displacements of  $\{0.074, 0.37, 0.74, 1.11, 1.48\}$  degrees visual angle on the eye. The curves for the model and humans are approximately similar. Note in particular that changing dot

**Table 3.** Probability of mismatch values for dot densities in Figure 7.

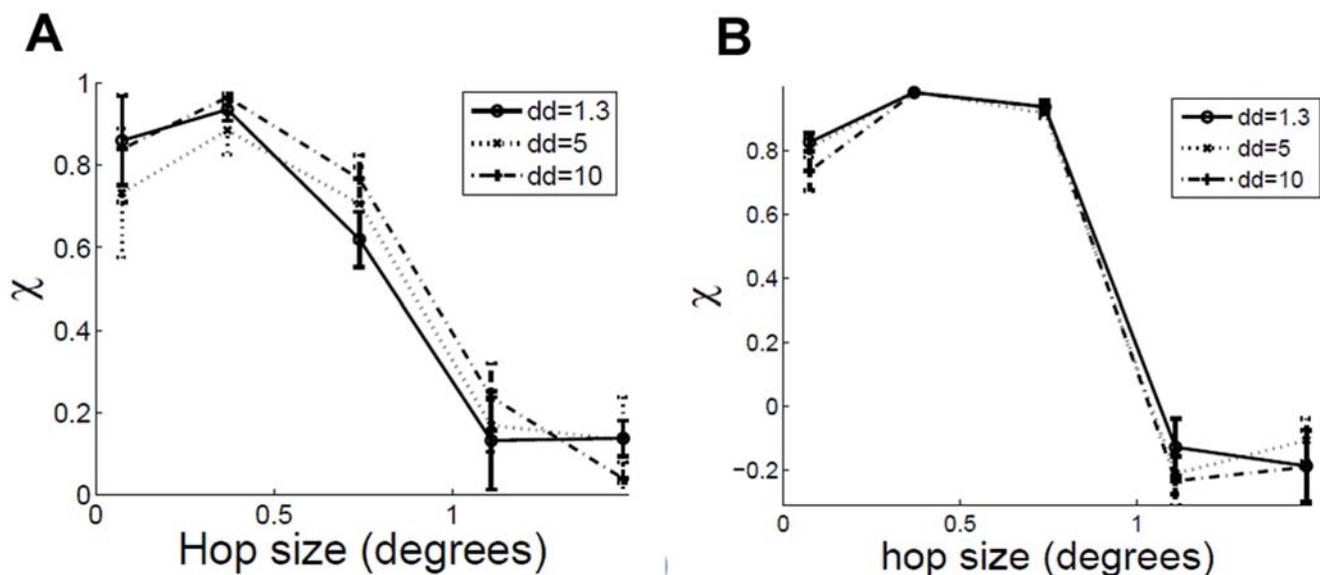
Dot density (dots/degrees <sup>2</sup> )	Probability of mismatch = $1 - \exp(-\pi h^2 dd)$ (up to 4 decimal places) $h = 0.37^\circ$
1	0.3495
2	0.5769
4	0.8210
8	0.9680
16	0.9990
25	1.0000

doi:10.1371/journal.pone.0004536.t003

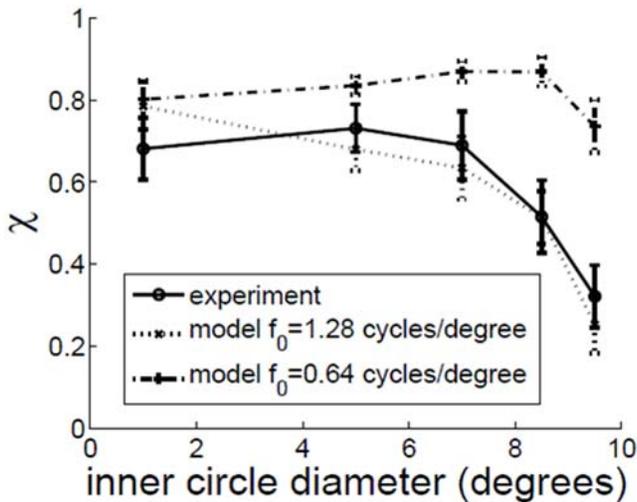
density does not produce any change in  $\chi$ . The figures show that as the hop size is increased, motion disappears in the display even though the dot correlation is very high ( $c = 0.4$ ). This is because if the hop size becomes greater than the RF size, motion sensitive neurons will fail to register motion. Also important is the decrease in  $\chi$  if the hop size becomes too small. In this case, the spatiotemporal profile of the stimulus will not cross-correlate well with the spatiotemporal RF of motion sensitive cells.

**Effect of inner circle diameter  $ic$**

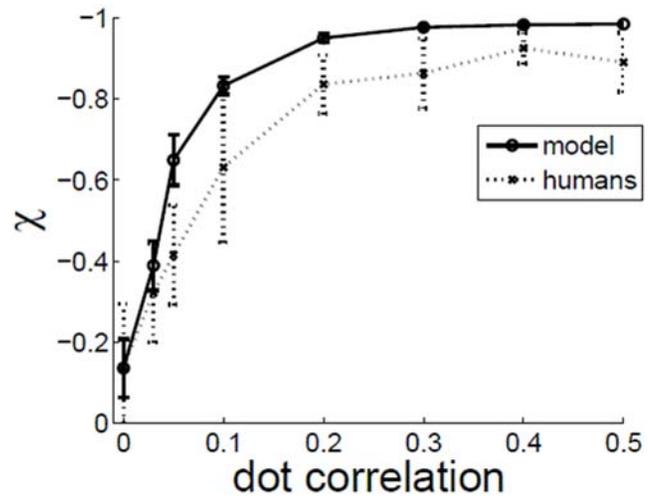
The angle subtended by the outer circle diameter is fixed at  $10^\circ$  in all the experiments. Figure 9 shows the effect of varying the angle subtended by the inner circle diameter  $ic$  ( $c = 0.1$ ). It is seen that observer performance falls off as the angle subtended by the inner circle diameter  $ic$  is changed from  $7^\circ$  to  $9.5^\circ$ . At  $ic = 9.5^\circ$  the annulus is very thin, and appears like a 1D ring rather than a 2D annulus. When  $ic$  is small, the noise in the display is uniformly distributed in the sense that if  $\theta$  is the angle made by the noise vector, then  $\theta$  is uniformly distributed from  $-\pi$  to  $+\pi$ . Denote the cross product of noise vector with the radial vector by  $x = \sin(\theta)$ . Then  $E[x] = 0$ , and the amount of noise is given by



**Figure 8.** (a)  $\chi$  vs. hop size  $h$  for human observers (b)  $\chi$  vs. hop size for model.  $c = 0.4$ ,  $fd = 30$  ms,  $ic = 7^\circ$ . doi:10.1371/journal.pone.0004536.g008



**Figure 9.  $\chi$  vs. angle subtended by inner circle diameter ic.** Angle subtended by outer circle diameter is fixed at  $10^\circ$ .  $c=0.1$ ,  $dd=5$  dots/degrees<sup>2</sup>,  $fd=30$  ms. Model simulations at  $256 \times 256$  pixel resolution.  $f_0$  denotes center frequency of Watson Ahumada sensors. doi:10.1371/journal.pone.0004536.g009



**Figure 10.  $\chi$  vs.  $c$  for contrast reversing dots.**  $fd=30$  ms,  $ic=7^\circ$ ,  $dd=2.5$  dots/degrees<sup>2</sup>. doi:10.1371/journal.pone.0004536.g010

$$E[x^2] = \int_{-\pi}^{+\pi} \sin^2\theta \frac{1}{2\pi} d\theta = \frac{1}{2}$$

On the other hand, when  $ic \rightarrow oc$ ,  $\theta$  is either  $-\pi/2$  or  $+\pi/2$  with equal probability.  $E[x]$  is still zero but

$$E[x^2] = (-1)^2 \frac{1}{2} + (+1)^2 \frac{1}{2} = 1$$

so the amount of noise has apparently doubled in this case. Model performance is seen to partially match psychophysical performance. The curve with center frequency equal to 1.28 cycles/° shows a better fit than the curve with center frequency equal to 0.64 cycles/°. Unfortunately I cannot say why the former curve shows a better fit.

### Effect of reverse contrast

If the stimulus is modified such that the correlated dots flip their polarity as they rotate, meaning black dots change to white and vice-versa, then the reverse-phi motion [49,50,1] takes place. It is found that the motion perceived by an observer is opposite to the physical displacement of the correlated dots. If the correlated dots move CCW(CW), observer perceives motion in CW(CCW) direction respectively. The Watson Ahumada model is able to capture this phenomenon as shown in Figure 10. If the observer perceives motion in a direction opposite to rotation of the correlated dots, the observer response is negatively correlated with the embedded motion. For this reason  $\chi$  in Figure 10 is defined as the minimum value of the normalized cross correlation function between the response and input function.

To understand why motion may be perceived in the opposite direction when dots reverse their contrast, consider the signal  $I(x,y,t)$ . It is well known [17] that the Fourier Transform of an image undergoing coherent translational motion lies on a plane, i.e., if  $I(x,y,t) = I(x - v_x t, y - v_y t, 0)$  then  $I(\omega_x, \omega_y, \omega_t) = J(\omega_x, \omega_y) \delta(\omega_x v_x + \omega_y v_y + \omega_t)$  where  $J(\omega_x, \omega_y)$

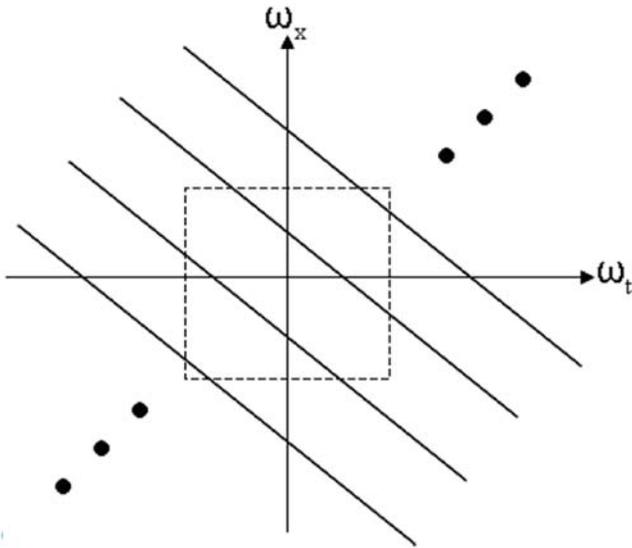
is the 2D Fourier Transform of  $I(x,y,0)$ , and  $(v_x, v_y)$  is velocity. The equation of plane is  $\omega_x v_x + \omega_y v_y + \omega_t = 0$ . This observation yields following algorithm to determine motion in a signal  $I(x,y,t)$ : find the best fitting plane to  $\|I(\omega_x, \omega_y, \omega_t)\|^2$  that passes through the origin. The velocity can be read off the equation of the plane. Now consider what happens when  $I(x,y,t)$  reverses its contrast every  $T$  s. The modified signal is given by  $L(x,y,t) = I(x,y,t)f(t)$ , where  $f(t)$  is a square wave alternating between  $+1$  and  $-1$  every  $T$  s. The

Fourier Series of  $f(t)$  is given by  $\sum_{n=-\infty}^{+\infty} a_n \exp(jn\omega_0 t)$ , with  $a_n = \begin{cases} 0 & n \text{ even} \\ c/n & n \text{ odd} \end{cases}$ ,  $c$  being a constant, and  $\omega_0 = 2\pi/T$ . This gives  $L(\omega_x, \omega_y, \omega_t) = \sum_{n=-\infty}^{+\infty} a_n J(\omega_x, \omega_y) \delta(\omega_x v_x + \omega_y v_y + \omega_t - n\omega_0)$ .

Note that  $a_0 = 0$ . Thus, the Fourier Transform of  $L(x,y,t)$  does not lie on a plane passing through the origin. Instead, the Fourier Transform of  $L(x,y,t)$  consists of infinitely many planes given by  $\omega_x v_x + \omega_y v_y + \omega_t = n\omega_0$  as illustrated in Figure 11. Assuming  $J(\omega_x, \omega_y)$  is mostly constant, the best fitting plane to  $\|L(\omega_x, \omega_y, \omega_t)\|^2$  (that also passes through the origin) is  $\perp$  to  $\omega_x v_x + \omega_y v_y + \omega_t = 0$ . If  $(a,b,1)$  is normal of the best fitting plane, then  $av_x + bv_y + 1 = 0$ .  $(a,b)$  is velocity of  $I(x,y,t)$  under reverse contrast. Letting  $\vec{v} = (v_x, v_y)$ , and  $\vec{w} = (a,b)$  we have

$$\vec{v} \bullet \vec{w} = -1$$

This equation can be satisfied by many  $\vec{w}$ . In particular  $\|\vec{w}\| = \frac{1}{\|\vec{v}\|}$ ,  $\angle \vec{w} = \angle \vec{v} + \pi$  is a solution, which is motion in opposite direction to  $\vec{v}$ . Note that  $\|\vec{w}\| = \frac{1}{\|\vec{v}\|}$  suggests that a faster moving particle should actually appear to move slower! This surprising prediction appears to be true within appropriate range. A display of alternating black and white stripes was made. The width of a stripe was  $0.25^\circ$ . The stripe pattern was translated to the right, and the stripes reversed their contrast after a time interval  $T$ . On viewing the display, motion was perceived in the leftward direction instead of right. With  $fd = T = 30$  ms and a hop size of  $0.125^\circ$ , the pattern appeared to be moving slower than with hop



**Figure 11. The Fourier Transform of an image undergoing coherent translational motion + periodic reverse contrast lies on infinitely many planes of the form  $\omega_x v_x + \omega_y v_y + \omega_t = n\omega_0$ , with  $n$  being an odd number.** The dashed lines denote the window of visibility [30].  
doi:10.1371/journal.pone.0004536.g011

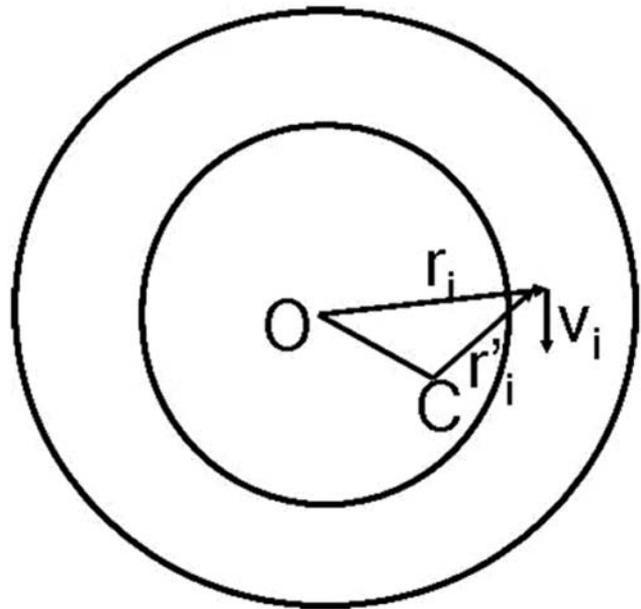
size of  $0.0625^\circ$ . Further experiments need to be done to gather numerical data to quantify the effect.

**Model sensitivity to center position**

By definition of rotation, any measure of rotary motion has to be specified with respect to some center of rotation (more accurately the axis of rotation has to be specified). In all the results presented up till now, the center position used in the simulations was the true center of rotation of the dots. What happens if the true center of rotation is not accurately known, as must be the case in reality? Figure 12 shows a schematic in which point O is the true center relative to which the correlated dots are rotating, and point C is the center relative to which rotary motion is computed by the model.  $\vec{v}_i$  is a motion cue. The rotary motion relative to the true center O is given by  $\sum_i \vec{r}_i \times \vec{v}_i$ , whereas the rotary motion relative to C is given by  $\sum_i \vec{r}'_i \times \vec{v}_i$ . We have

$$\begin{aligned} S &= \sum_i \vec{r}'_i \times \vec{v}_i \\ &= \sum_i (\vec{CO} + \vec{r}_i) \times \vec{v}_i \\ &= \sum_i \vec{CO} \times \vec{v}_i + \sum_i \vec{r}_i \times \vec{v}_i \\ &= \vec{CO} \times \sum_i \vec{v}_i + \sum_i \vec{r}_i \times \vec{v}_i \\ &= \sum_i \vec{r}_i \times \vec{v}_i \text{ provided } \sum_i \vec{v}_i = 0 \end{aligned}$$

The condition  $\sum_i \vec{v}_i = 0$  is true in case of the racetrack. The uncorrelated dots are uniformly distributed and generate motion cues in all directions with equal probability. The correlated dots generate motion cues in tangential direction, which when summed over the entire  $360^\circ$  annulus add up to zero. The expected value of  $\sum_i \vec{v}_i$  is thus zero. Therefore it seems accurate knowledge of



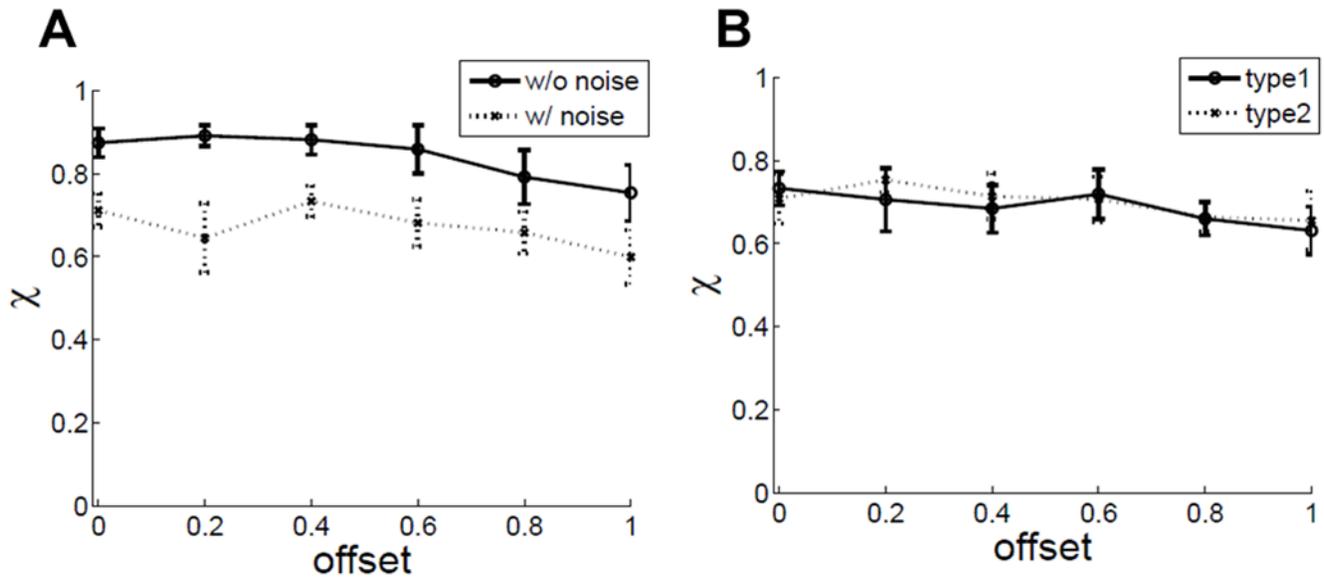
**Figure 12. Point O represents the true center of rotation, whereas point C is the center relative to which rotary motion is computed by the model.** The offset is given by  $\frac{\|\vec{OC}\|}{R_i}$  where  $R_i$  is radius of inner circle.  
doi:10.1371/journal.pone.0004536.g012

position of the true center relative to which rotation occurs is not needed. Figure 13(a) shows model sensitivity to knowledge of true center position. The rotary motion is computed by the model relative to a point C that is offset from the true center O. The offset is given by  $\frac{\|\vec{OC}\|}{R_i}$  where  $R_i$  is radius of inner circle. Two curves are shown: in one there is no noise added to the model, i.e.,  $n(t) = 0$ , and in the other the GWN equal to the default value of  $\sigma(\text{GWN})/\sigma_0 = 6$  is added to the model. It can be seen that the  $\chi$  values are not affected much by uncertainty in knowledge of true center position, and start to deteriorate only when the offset becomes very large. This may explain the experimentally observed position invariance of MST(d) cells, the fact that the cells are insensitive to where in their RF rotation occurs [9].

It seems that when only a sector of the racetrack is made visible, the condition  $\sum_i \vec{v}_i = 0$  may not hold true because of the correlated dots. However, if two sectors located diametrically opposite to each other are displayed then  $\sum_i \vec{v}_i = 0$ . Figure 13(b) shows  $\chi$  vs. offset for the two cases: type1 when only a single  $90^\circ$  sector is made visible, and type2 when two sectors located diametrically opposite to each other, and each  $45^\circ$  in size, are displayed. Interestingly the model is still robust enough to the offset even when only a sector of the racetrack is displayed, irrespective of whether it is type1 or type2.

**Effect of displaying only a sector**

Figure 14(a) shows the effect of displaying only a sector of the complete annulus on human observers. Two cases are considered. In type1, a single sector is shown that is randomly positioned. In type2, two sectors located diametrically opposite to each other, and each half the size of the sector in type1, are displayed. It is seen that  $\chi$  increases monotonically as the sector size increases. It is interesting to note that there is a significant difference in  $\chi$  for the two cases, even though the total area displayed is the same in the two cases. The



**Figure 13.  $\chi$  vs. center relative to which rotary motion is computed (a) full 360° annulus is visible (b) only 90° of annulus is visible.** Type1 – a single 90° sector is visible. Type 2 – two sectors located diametrically opposite to each other, and each 45° in size, are visible. Both curves are for the model.

doi:10.1371/journal.pone.0004536.g013

corresponding data for the model is shown in Figure 14(b). The model shows an increase in  $\chi$  with sector size. However, there is no difference between type1 and type2 for the model.

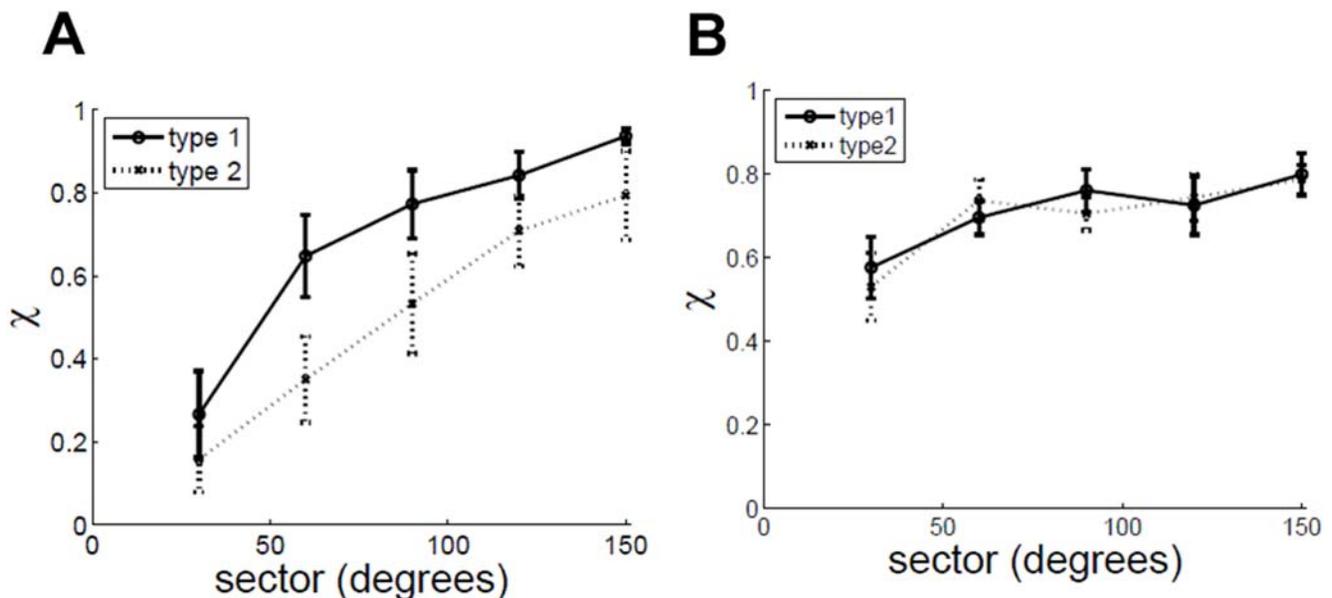
#### Effect of inserting random frames

Figure 15 shows effect of inserting  $K$  random frames between every pair of correlated frames in the stimulus. It is seen that observer performance does not fall to zero abruptly, but decreases in a graceful manner showing that the human visual system takes multiples frames into consideration when estimating motion. The

model performance also does not fall to zero abruptly, but degrades much more rapidly than human performance.

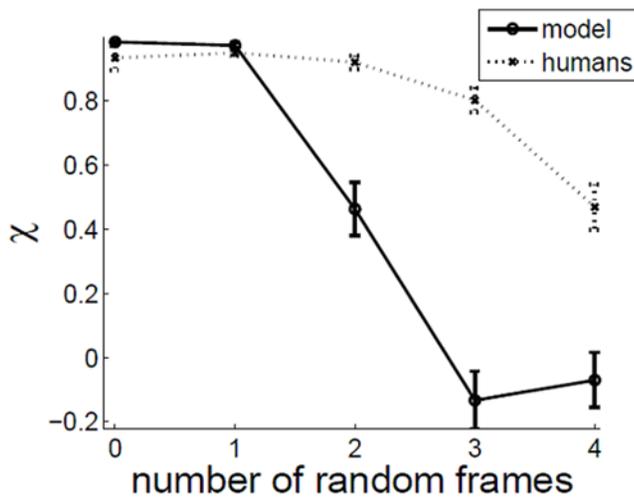
#### Dipoles

Instead of displaying dots in an annulus, each dot can be split into two dots – one black and one white forming a dipole. This results in what has been termed as the anti-Glass pattern [51]. The  $c = 0$  case creates a powerful motion illusion that has been previously investigated [35]. The addition of dipoles introduces several new parameters:



**Figure 14.  $\chi$  vs. sector.** In case of type 1 only one sector is displayed, whereas in case of type 2 two sectors located diametrically opposite to each other, and each half the size of sector in type 1, are displayed. (a) human performance (b) model performance.

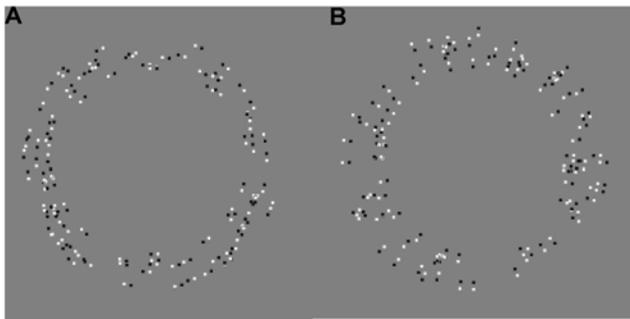
doi:10.1371/journal.pone.0004536.g014



**Figure 15. Effect of inserting K random frames between correlated frames.**  $c = 0.5$ ,  $fd = 30$  ms,  $dd = 5$  dots/degrees<sup>2</sup>,  $ic = 7^\circ$ . doi:10.1371/journal.pone.0004536.g015

- 1) the dipole spacing  $s$ ,
- 2) the black to white intensity ratio  $bwir$  defined as  $\frac{I_b - I_w}{I_w - I_0}$  where  $I_0$ ,  $I_b$ ,  $I_w$  are luminance of background (fixed at  $10.8$  cd/m<sup>2</sup>), black and white dots respectively,
- 3) the dipole orientation: tangential or radial as in Figure 16

Complex patterns of motion are perceived with dipoles in the display, e.g., if dipoles are oriented radially there is tendency to observe radial pulsating motion, even if dipoles are actually rotating with significant rotary motion. Depending upon the parameter settings, motion in opposite directions is also seen. It can become difficult to assign a single motion direction to the whole display, although there is no doubt that there is motion in it. Let RC (reverse contrast) ON denote the setting that if a dipole is correlated, then black changes to white and vice-versa in the next frame. With RC ON, the perception of motion can switch from normal  $\phi$  to reverse  $\phi$  depending upon dipole spacing. This section reports results of an experiment investigating  $\chi$  vs.  $bwir$  with center-to-center spacing equal to six minutes,  $c = \theta$ , and RC ON. The results are summarised in Figure 17. As can be seen the model is able to capture some aspects of psychophysical behavior, but not all of it. In Figure 17, the definition of  $\chi$  is modified as follows. Let  $\chi_+$  denote maximum value of normalized cross correlation function, and  $\chi_-$  denote minimum value of normalized cross correlation function. If  $|\chi_+| > |\chi_-|$ ,  $\chi = \chi_+$ , otherwise



**Figure 16. (a) tangential dipoles (b) radial dipoles.** Center-to-center spacing =  $17'$  in both cases. doi:10.1371/journal.pone.0004536.g016

$\chi = \chi_-$ . When  $|\chi_+| > |\chi_-|$ , observer perceives motion in the direction of displacement of correlated dots and therefore  $\chi$  is defined to be equal to  $\chi_+$  in this case. By similar reasoning,  $\chi$  is defined to be  $\chi_-$  for the other case.

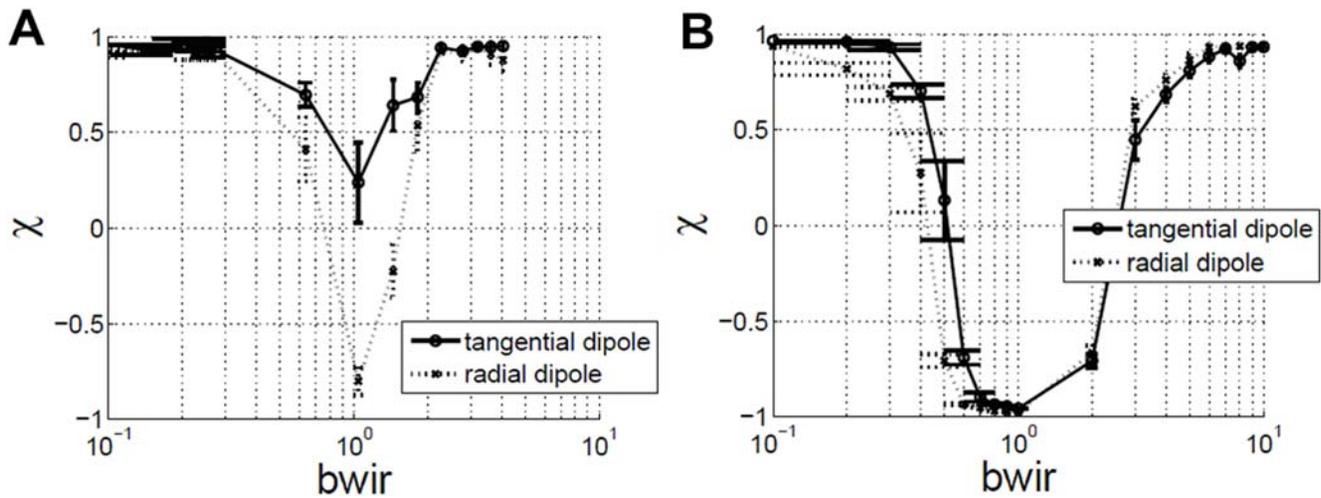
## Discussion

Although this paper shows that the Watson Ahumada motion detector does a good job at detecting motion in random dot kinematograms (RDKs) consonant with human psychophysical performance, it remains to be seen how well it would perform on real world imagery. The challenge here is that although it is straightforward to run the model on real world test cases, how do we accurately measure the optical flow perceived by humans on these test cases? Computer vision papers characterise optical flow performance of a model by either using synthetic image sequences designed to mimic the real world, or using real world image sequences in which the motion of the camera is carefully calibrated [52]. However, as we have already seen in this paper: 1. the same sequence of image frames can produce different perception depending on frame rate, 2. the human visual system takes multiple frames into consideration when determining motion. In the light of these remarks, it is not immediately obvious what the ground truth optical flow (ground truth being defined as the flow perceived by a human) would be for the test cases mentioned above. These caveats should be borne in mind while attempting a performance characterisation of the Watson Ahumada model using the computer vision paradigm. I have placed some preliminary work running the Watson Ahumada model on real world imagery online as a proof of concept [41].

The neurophysiological plausibility of a model is likely to attract heavy debate. Krekelberg (2008) has provided a comprehensive discussion on the biological plausibility of the Reichardt, Adelson Bergen, and gradient based motion detectors [13]. With respect to the Watson Ahumada model, DeAngelis et. al. (1995) and others have found that the Watson Ahumada filters provide an accurate model of simple cell receptive fields (RFs) [44]. Quoting DeAngelis et. al. [53]:

“Rather, simple cell RFs in the joint space-time domain appear to be fit well by a model first proposed by Watson and Ahumada ... Based on the Watson-Ahumada formulation, we have modelled space-time RFs of simple cells, as the weighted sum of two space-time separable subunits in a quadrature relationship. This model formulation provides a remarkably good fit to the data from most cells, regardless of their degree of space-time inseparability ... In conclusion, to account for space-time RFs of simple cells that differ widely in the degree of space-time inseparability, at least two separable subunits appear necessary as modelled by Watson and Ahumada.”

Although there are similarities between the Watson Ahumada motion detector and the Adelson Bergen motion detector, which is usually the *de facto* motion detection mechanism used in studies of visual motion perception, there are also some differences. The Adelson Bergen motion detector measures how much power the stimulus has within a spatiotemporal frequency band. Thus a detector tuned to  $(\omega_{x0}, \omega_{y0}, \omega_{t0})$  effectively samples the power spectrum of the stimulus within the vicinity of  $(\omega_{x0}, \omega_{y0}, \omega_{t0})$ . Such detectors have been proposed as models of V1 complex cells [1,21]. The responses of multiple such detectors tuned to different spatiotemporal frequencies are pooled to determine the best fitting plane in the frequency domain [18,19]. The best fitting plane defines the motion of the stimulus [17]. This processing, although still debatable, is believed to occur in MT. In contrast, with respect to the Watson Ahumada motion detector, information about the



**Figure 17.  $\chi$  vs. bwir (black to white intensity ratio).** Dipole spacing =  $6'$ ,  $c = 0.5$ ,  $dd = 2.5$  dots/degrees<sup>2</sup>, RC ON, (a) human observers (b) Watson Ahumada model with simulations at  $256 \times 256$  pixel resolution. doi:10.1371/journal.pone.0004536.g017

motion of the stimulus is encoded in the (most dominant) temporal frequency of oscillation of detector response as per the equation:

$$\omega_t = - (v_x \omega_x + v_y \omega_y)$$

The temporal frequencies of oscillation of different detectors tuned to different  $(\omega_x, \omega_y)$  are measured, and then above relationship is used to determine the motion of the stimulus. The neural locus of the stages that perform this computation is unclear. Also unclear is the relationship of the model to what we do know about motion processing in the brain beyond the first stage of spatiotemporal filtering. For example, the model does not state how simple V1 neuron outputs could be combined to generate speed tuned V1 complex and MT cells [54,55]. Perrone (2005) has put forward a model that explains how the magnitude of the Fourier transform of simple V1 neuron responses can be combined to generate the magnitude of the Fourier transform of a speed tuned neuron [56]. The input V1 neurons that Perrone's model uses are based on the Watson Ahumada filters.

It may be worthwhile to mention that the Watson Ahumada model has been proposed as a model of primary motion sensing mechanisms, what Cavanagh (1991) called passive motion detectors in his paper [57]. The human visual system is a complex parallel distributed system in which modules interact with each other and do not function in isolation, e.g., it is widely accepted now that motion perception interacts with form perception, a view that was not always held in this field. The interactions between modules can give rise to phenomenon that cannot be explained by either module alone. Benton, O'Brien, & Curran (2007) have recently provided example of a fractal rotation stimulus in which rotation is perceived within any arbitrary window applied to the stimulus [58]. The authors assert that the fact that observers can readily perceive fractal rotation is a clear example of a stimulus in which motion extraction is dependent upon the prior analysis of some spatial property (which happens to be the orientation in case of fractal rotation). The omega effect itself is believed to occur because of interactions between form and motion processing circuits in the brain. Although there are growing examples of such stimuli that point to interactions between form and motion, little is known about how these interactions occur. To my knowledge no

quantitative model has been put forward to explain these interactions.

In conclusion, the contribution of this paper is to present a performance characterisation of the Watson Ahumada model of human visual motion sensing. The model performance is seen to match human performance with respect to most parameters. It is able to explain some key and important parts of the psychophysical data such as independence of observer performance to dot density in the display, and decrease of observer performance with frame duration of the display. The model insensitivity to the center position relative to which rotary motion is computed, together with the vector analysis presented in the paper, may explain the experimentally observed position invariance of MST(d) cells. In addition, this paper shows that the omega effect of Rose & Blake (1998) is a truly bistable illusion. Although the display of random dots triggers perception of rotary motion, the direction of motion perceived is independent of what dot pattern is shown. The time interval between spontaneous reversals in perceived direction is lognormally distributed as is the case for most bistable illusions. Therefore the processes that give rise to this illusion may be the same processes that underlie much of other bistable phenomenon.

## Supporting Information

**Movie S1** A movie of the racetrack for  $c = 0$ .  
Found at: doi:10.1371/journal.pone.0004536.s001 (8.63 MB AVI)

**Movie S2** A movie of the racetrack for  $c = 0.3$ .  
Found at: doi:10.1371/journal.pone.0004536.s002 (8.02 MB AVI)

**Movie S3** A movie of the racetrack for  $c = 0.5$ .  
Found at: doi:10.1371/journal.pone.0004536.s003 (8.94 MB AVI)

## Acknowledgments

This work was done when I was a graduate student in Donald A. Glaser's lab at UC Berkeley. I thank him and T. Kumar for the financial support and help which made this work possible. In particular Donald A. Glaser is credited with naming the stimulus described in this paper as the racetrack, and for introducing me to the  $c = 0$  case, which he invented to model the illusory motion perceived in Leviant's Enigma [59]. T. Kumar provided guidance in various aspects of this research, especially the planning of the psychophysical experiments and recruitment of subjects. I thank my friend David Gelbart for proof-reading this paper.

This paper is dedicated to Pink Floyd → The Division Bell → High Hopes.

## References

- Adelson EH, Bergen JR (1985) Spatiotemporal energy models for the perception of motion. *J Opt Soc Am A* 2(2): 284–299.
- van Santen JPH, Sperling G (1985) Elaborated Reichardt detectors. *J Opt Soc Am A* 2(2): 300–321.
- Watson AB, Ahumada A (1985) Model of human visual-motion sensing. *J Opt Soc Am A* 2(2): 322–342.
- Johnston A, McOwan PW, Buxton H (1992) A computational model of the analysis of some first-order and second-order motion patterns by simple and complex cells. *Proc Roy Soc B* 250: 297–306.
- Perrone JA, Thiele A (2002) A model of speed tuning in MT neurons. *Vis Res* 42(8): 1035–51.
- Ullman S (1979) *The Interpretation of Visual Motion*. MIT Press.
- Dawson MRW (1991) The How and Why of What Went Where in Apparent Motion: Modeling Solutions to the Motion Correspondence Problem. *Psychological Review* 98(4): 569–603.
- Williams DW, Sekuler R (1984) Coherent Global Motion Percepts from Stochastic Local Motions. *Vis Res* 24(1): 55–62.
- Graziano MSA, Andersen RA, Snowden RJ (1994) Tuning of MST Neurons to Spiral Motions. *J Neurosci* 14(1): 54–67.
- Rose D, Blake R (1998) Motion Perception: From Phi to Omega. *Proc Roy Soc B* 353(1371): 967–980.
- Shermer M (2008) Patternicity. *Sci Am* 299(6): 48.
- Green DM, Swets JA (1966) *Signal detection theory and psychophysics*. New York: Wiley.
- Krekelberg B (2008) Motion detection mechanisms. In Allan Basbaum, ed. *The Senses: A Comprehensive Reference*. Oxford: Elsevier Inc.
- Born RT, Bradley DC (2005) Structure and function of visual area MT. *Ann Rev Neurosci* 28: 157–189.
- Derrington AM, Allen HA, Delicato DS (2004) Visual mechanisms of motion analysis and motion perception. *Ann Rev Psych* 55: 181–205.
- Grzywacz NM, Merwine DK (2003) Neural basis of motion perception. *Encyclopedia of Cognitive Science* 3: 86–98.
- Watson AB, Ahumada A (1983) A look at motion in the frequency domain. Technical Report 84352, NASA Technical Memorandum.
- Heeger DJ (1987) Model for the extraction of image flow. *J Opt Soc Am A* 4(8): 1455–1471.
- Simoncelli EP, Heeger DJ (1998) A model of neuronal responses in visual area MT. *Vis Res* 38(5): 743–761.
- Rust NC, Mante V, Simoncelli EP, Movshon JA (2006) How MT cells analyze the motion of visual patterns. *Nat Neurosci* 9(11): 1421–1431.
- Emerson RC, Bergen JR, Adelson EH (1992) Directionally selective complex cells and the computation of motion energy in cat visual cortex. *Vis Res* 32(2): 203–218.
- Tanaka K, Saito H (1989) Analysis of Motion of the Visual Field by Direction, Expansion/Contraction, and Rotation cells clustered in the Dorsal Part of the Medial Superior Temporal Area of the Macaque Monkey. *J Neurophysiol* 62(3): 626–641.
- Sakata H, Shibutani H, Ito Y, Tsurugai K, Mine S, Kusunoki M (1994) Functional properties of rotation-sensitive neurons in the posterior parietal association cortex of the monkey. *Exp Brain Res* 101(2): 183–202.
- Newsome WT, Pare EB (1988) A Selective Impairment of Motion Perception following lesions of the Middle Temporal Visual Area MT. *J Neurosci* 8(6): 2201–2211.
- Newsome WT, Britten KH, Movshon JA (1989) Neuronal Correlates of a perceptual decision. *Nature* 341(6237): 52–54.
- Barlow H, Tripathy SP (1997) Correspondence Noise and Signal Pooling in the Detection of Coherent Visual Motion. *J Neurosci* 17(20): 7954–7966.
- Scase MO, Braddick OJ, Raymond JE (1996) What is noise for the motion system? *Vis Res* 36(16): 2579–2586.
- Watamaniuk SNJ, McKee SP, Grzywacz NM (1995) Detecting a Trajectory Embedded in Random-direction Motion Noise. *Vis Res* 35(1): 65–77.
- Morgan MJ (1980) Analogue models of motion perception. *Proc Roy Soc B* 290(1038): 117–135.
- Watson AB, Ahumada A, Farrell JE (1986) Window of Visibility: a psychophysical theory of feliity in time-sampled visual motion displays. *J Opt Soc Am A* 3(3): 300–307.
- MacKay DM (1965) Visual Noise as a tool of Research. *J Gen Psychol* 72: 181–197.
- Ross J, Badcock DR, Hayes A (2000) Coherent global motion in the absence of coherent velocity signals. *Curr Biol* 10: 679–682.
- Krekelberg B, Dannenberg S, Hoffmann KP, Bremmer F, Ross J (2003) Neural correlates of implied motion. *Nature* 424: 674–677.
- Krekelberg B, Vatakis A, Kourtzi Z (2005) Implied motion from form in the Human Visual Cortex. *J Neurophysiol* 94: 4373–4386.
- Viva MMD, Gori M, Burr DC (2006) Powerful Motion Illusion Caused by Temporal Asymmetries in ON and OFF Visual Pathways. *J Neurophysiol* 95: 3928–3932.
- Geisler WS (1999) Motion streaks provide a spatial code for motion direction. *Nature* 400: 65–69.
- Barlow H, Olshausen B (2004) Convergent evidence for the visual analysis of optic flow through anisotropic attenuation of high spatial frequencies. *J Vis* 4: 415–426.
- Perrone JA, Stone LS (1994) A model of self-motion estimation within primate extrastriate visual cortex. *Vis Res* 34(21): 2917–2938.
- Duffy CJ, Wurtz RH (1995) Response of Monkey MST Neurons to Optic Flow Stimuli with Shifted Centers of Motion. *J Neurosci* 15(7): 5192–5208.
- van der Schaaf A, van Hateren JH (1996) Modeling the Power Spectra of Natural Images: Statistics and Information. *Vis Res* 36(17): 2759–2770.
- Jain S (2007) Optical Flow or Motion Estimation using the Watson Ahumada (WA) algorithm. [http://www.codeproject.com/KB/audio-video/Optical\\_Flow\\_Estimation.aspx](http://www.codeproject.com/KB/audio-video/Optical_Flow_Estimation.aspx).
- Anonymous survey available at <http://purl.oclc.org/NET/survey>.
- Riani M, Simonotto E (1994) Stochastic Resonance in the Perceptual Interpretation of Ambiguous figures: A Neural Network Model. *Phys Rev Lett* 72(19): 3120–3123.
- DeAngelis DC, Ohzawa I, Freeman RD (1995) Receptive-field dynamics in the central visual pathways. *Trends Neurosci* 18: 451–458.
- Benzi R, Sutera A, Vulpani A (1981) The mechanism of stochastic resonance. *J Phys A* 14: 453–457.
- Gammaitoni L, Hanggi P, Jung P, Marchesoni F (1998) Stochastic Resonance. *Rev Mod Phys* 70(1): 223–287.
- Moss F, Ward LM, Sannita WG (2004) Stochastic Resonance and sensory information processing: a tutorial and review of application. *Clinical Neurophysiology* 115: 267–281.
- Grzywacz NM, Watamaniuk SNJ, McKee SP (1995) Temporal Coherence Theory for the detection and measurement of Visual Motion. *Vis Res* 35(22): 3183–3203.
- Anstis SM (1970) Phi movement as a subtractive process. *Vis Res* 10: 1411–1430.
- Anstis SM, Rogers BJ (1975) Illusory reversal of visual depth and movement during changes of contrast. *Vis Res* 15: 957–961.
- Burr D, Ross J (2006) The effects of opposite-polarity dipoles on the detection of Glass patterns. *Vis Res* 46: 1139–1144.
- McCane B, Novins K, Crannitch D, Galvin B (2001) On Benchmarking Optical Flow. *Computer Vision and Image Understanding* 84: 126–143.
- DeAngelis GC, Ohzawa I, Freeman RD (1996) Reply to TINS letter to the Editor by Wang et. al. *Trends in Neuroscience* 19: 386.
- Perrone JA, Thiele A (2001) Speed skills: measuring the visual speed analyzing properties of primate MT neurons. *Nat Neurosci* 4(5): 526–532.
- Priebe NJ, Lisberger SG, Movshon JA (2006) Tuning for Spatiotemporal Frequency and Speed in Directionally Selective Neurons of Macaque Striate Cortex. *J Neurosci* 26(11): 2941–50.
- Perrone JA (2005) Economy of scale: A motion sensor with variable speed tuning. *J Vis* 5(1): 28–33.
- Cavanagh P (1991) Short-range vs. long-range motion: Not a valid distinction. *Spatial Vision* 5: 303–309.
- Benton CP, O'Brien JMD, Curran W (2007) Fractal rotation isolates mechanisms for form-dependent motion in human vision. *Biol Lett* 3: 306–308.
- Kumar T, Glaser DA (2006) Illusory motion in Enigma: A psychophysical investigation. *PNAS* 103(6): 1947–1952.

## Author Contributions

Conceived and designed the experiments: SJ. Performed the experiments: SJ. Analyzed the data: SJ. Wrote the paper: SJ.