The curse of three dimensions: Why your brain is lying to you

SUSAN VANDERPLAS, HEIKE HOFMANN and DI COOK, Iowa State University

General Terms: Design, Human factors, Experimentation

Additional Key Words and Phrases: Statistical graphics, optical illusion, perceptual theory, visualization

ACM Reference Format:

VanderPlas, S., Hofmann, H., and Cook, D. 2014. The Curse of Three Dimensions: Why Your Brain is Lying to You. ACM Trans. Appl. Percept. 1, 1, Article 1 (January 2014), 19 pages.
DOI = 10.1145/0000000.0000000 http://doi.acm.org/10.1145/0000000.0000000

Author's address: S. VanderPlas, Snedecor Hall, Iowa State University, Ames, IA 50010; email: skoons@iastate.edu

DOI 10.1145/0000000.0000000 http://doi.acm.org/10.1145/0000000.0000000

Contents

1. INTRODUCTION

motivate power of graphics first?

One of the basic principles of statistical graphics is that a chart should accurately reflect the data.

XXX what are statistical graphics, and how are they different from 'visualizations'/sketches?

The lie factor needs more details for a proper introduction, like the one you do a bit later. Take out some detail and just mention the lie factor as one measure to assess whether a picture is true to the data it's supposed to show.

Tufte's "Lie Factor", for instance, is calculated based on the ratio of the effect size shown in the graphic to the effect size in the data [Tufte 1991]. If there is a systematic difference, the lie factor will be notably different from 1 (values between .95 and 1.05 are typically acceptable). While Tufte's lie factor is an effective measurement of the mapping from data to graphics, it does not give us any insight about the mapping from the chart to the brain.

Ideally, charts not only represent the data accurately, but also allow readers to draw accurate conclusions from the graph. While the human visual system is quite good at accurately interpreting statistical charts [Cleveland and McGill 1984], we need to be aware of contextual misperceptions that lead us to the wrong conclusion.

let's not undermine the whole idea of graphics in the intro ... in most situations our visual system is pretty good in interpreting graphics ... maybe cite Cleveland ordering of visual tasks, that got replicated by Kosara et al. – **need to know more here, I can't find a similar paper on Google Scholar.**

In this paper, we examine situations in which low-level human perceptual processes interfere with making accurate judgements from displays and suggest an experimental methodology for estimating the psychological "lie factor" of a chart due to one of the identified conceptual misperceptions.

Are there any other attempts to ensure that a chart is representing the data? This section needs a bit more lit review.

There are relatively few examples of the effect of optical illusions and other misperceptions on statistical graphics, but Amer [2005] and Poulton [1985] have documented the effect of the poggendorff illusion on line graphs in different contexts.

I haven't found much, but there's at least this bit.

1.1 The Curse of Three Dimensions

The human visual system is largely optimized for perception of three dimensions. Biologically, binocular vision ensures that we have the necessary information to construct a functional mental representation of the three-dimensional world, but even in the absence of binocular information the brain uses numerous heuristics to parse otherwise ambiguous two-dimensional retinal images into meaningful three-dimensional information. Predictably, however, these heuristics are not without drawbacks; the same two-dimensional neural representation might correspond to multiple three-dimensional objects, as in the well-known Necker Cube (Gregory 1968; shown in figure 1). Additionally, the same three-dimensional object often has infinitely many two-dimensional representations, for instance, when viewed from different angles. Optical illusions that demonstrate this phenomenon are the so-called "impossible objects" such as the Penrose triangle [Gregory 1968].

In the case of the Necker cube, the two different figures are equally salient, and thus the brain does not prefer one interpretation over the other and instead continuously switches between interpretations. In many other instances, experience with the real world informs the choice between multiple possibilities of rendering a three-dimensional object from the same two-dimensional representation. This indicates that processing occurs "top down" in that our previous experience influences our current perceptions. Without this top-down influence, the brain would not be able

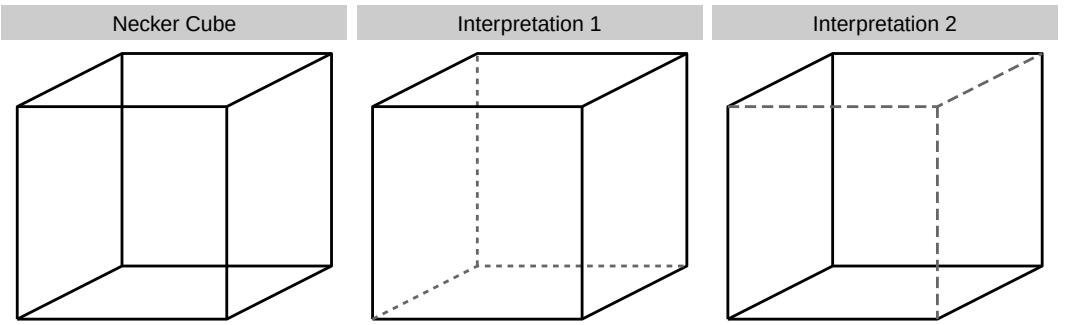


Fig. 1. The Necker Cube is a so-called "ambiguous object" because two different transparent objects produce the same retinal image (and thus the same perceptual experience). Commonly, the image seems to transition instantaneously from one possible mental representation to the other.

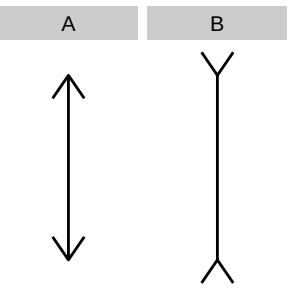


Fig. 2. The Müller-Lyer illusion. The central segment of figure A is perceived as shorter than the central segment of figure B, even though the two are actually the same length.

to map the picture back to a three-dimensional object. One of the most well studied examples of the influence of top-down processing is the Müller-Lyer illusion, shown in figure 2.

In the Müller-Lyer illusion, two vertical line segments are shown with arrows extending from both ends; one segment forms an acute angle with the arrows, the other segment forms obtuse angles with the arrows. The line segment adorned with arrows that form an acute angle appears to be shorter than the linen segment which forms an obtuse angle with the arrow segments.

One explanation for the Müller-Lyer illusion [Gregory 1968] is that the brain interprets the ambiguous lines as a common three-dimensional object common to everyday experience: corners of a room. Figure 2A occurs when viewing the outside corner of a rectangular prism, figure 2B occurs when viewing the prism from the inside. In regions which do not commonly have rectangular buildings, the illusion is significantly less pervasive [Ahluwalia 1978]. Figure 3 provides one possible context that would lead to the Müller-Lyer effect. This real-world experience carries with it an inferred perspective - when the arrows point inward, the object is typically closer than when the arrows point outward, which causes the brain to interpret the outward-pointing figure as larger when the retinal size of the two objects is identical. The perspective cues which contribute to the Müller-Lyer illusion allow for an accurate neural representation of the object in context; when misapplied to two-dimensional stimuli, these cues are responsible for the illusion's effect. This inferred "depth cue" [Gregory 1968] is reasonably consistent across individuals, suggesting that the phenomenon has a neurological basis.

A similar effect can also be found in the Necker Cube - whichever face appears to be furthest away also seems larger, even though any two parallel faces are equally sized in the image. This approach has proved to be very advantageous

Fig. 3. Real-world context that gives rise to the Müller-Lyer illusion. The highlighted areas correspond to the parts of the Muller Lyer illusion, and while the two arrows are obviously the same size in the real world, the black arrow takes up much more visual space than the white arrow.

for real world scenarios [Gregory 1968], as pictures of real objects are seldom ambiguous. This strategy also allows for high performance with limited neural bandwidth.

1.2 Three Dimensional Context of the Sine Illusion

While the classic Müller-Lyer illusion is seldom a factor in statistical charts, there are other illusions caused by the interpretation of a two-dimensional stimulus in the context of three-dimensional objects, leading to a distortion in the mental representation of the original stimulus. The sine illusion (also known as the line width illusion: VanderPlas and Hofmann 2014; Day and Stecher 1991) is one example of this phenomenon which occurs frequently in statistical graphics.

Give each of the two figures in the next sentence its own paragraph. First introduce the sine illusion. Then give an example outside the realm of vertical lines. Otherwise you're wasting good material.

Figure 4 shows the sine illusion in its original form [Day and Stecher 1991], as straight vertical lines of the same length with a nonlinear mean function. In this illusion, the vertical lines in the center of the figure appear much shorter than the vertical lines at the peak and trough of the sine curve. The illusion still persists when the image is rotated by 90^{deg}, and is incredibly difficult to mentally overcome (even when viewers are aware of the illusion's presence).

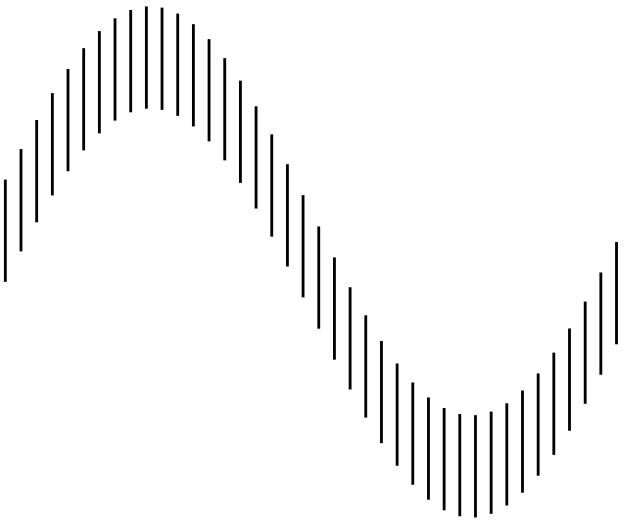


Fig. 4. The classic sine illusion. Each vertical line has the same length, though the lines at the peak and trough of the curve appear longer.

The sine illusion is also present in well-known graphics, such as the "Balance of Trade" chart from Playfair's Statistical Atlas [Playfair 1786], shown in figure 5. The balance of trade in 1765 seems to be approximately the same as the balance of trade in the years immediately preceding that year; this is in fact extremely misleading (using a straightedge along the chart vertically will demonstrate the issue). This problem has been well-documented [Cleveland and McGill 1984], but it is difficult to produce a (single) similar chart which conveys all of the information presented in figure 5.

In both figures, the illusion appears when the vertical length displayed in the chart does not match the perceived information. Like the Müller-Lyer illusion, the illusion is pervasive and very difficult to "un-see" or mentally correct. The sine illusion, which is also known as the line-width illusion, has also been documented in parallel sets plots [Hofmann and Vendettuoli 2013] and occurs when there is a nonlinear function with a large change in absolute slope; this change in slope can mask or exaggerate changes in variance. The illusion is also affected by the aspect ratio of the image and the aspect ratio of the chart's coordinate system. An interactive demonstration of the illusion is available at `http://bit.ly/1ldgujL`; manipulating the length of the lines and the amplitude of the underlying sine function also changes the chart's aspect ratio and the perceived strength of the illusion.

The illusion is not dependent on specifically identifying the vertical distance along a line. Figure 6(a) shows a scatterplot of data with a trend. A loess smooth is used to estimate the trendline. A visual assessment of variability along this trendline might result in a description such as 'homogeneous variance or slightly increasing variance for negative x, followed by a dramatic decrease in vertical variability for positive x'. Once the residuals are separated from the trendline as shown on the right hand side of the figure, it becomes apparent that this first assessment of conditional variability was not correct, and the steady decrease along the horizontal axis becomes visible.

Cleveland and McGill [1984] determined that comparison of the vertical distance between two curves is often inaccurate, as "the brain wants to judge minimum distance between the curves in different regions, and not vertical distance". While they do not explain a reason for this tendency, introspection does support their explanation: we judge the distance between two curves based on the shortest distance between them, which geometrically is the distance

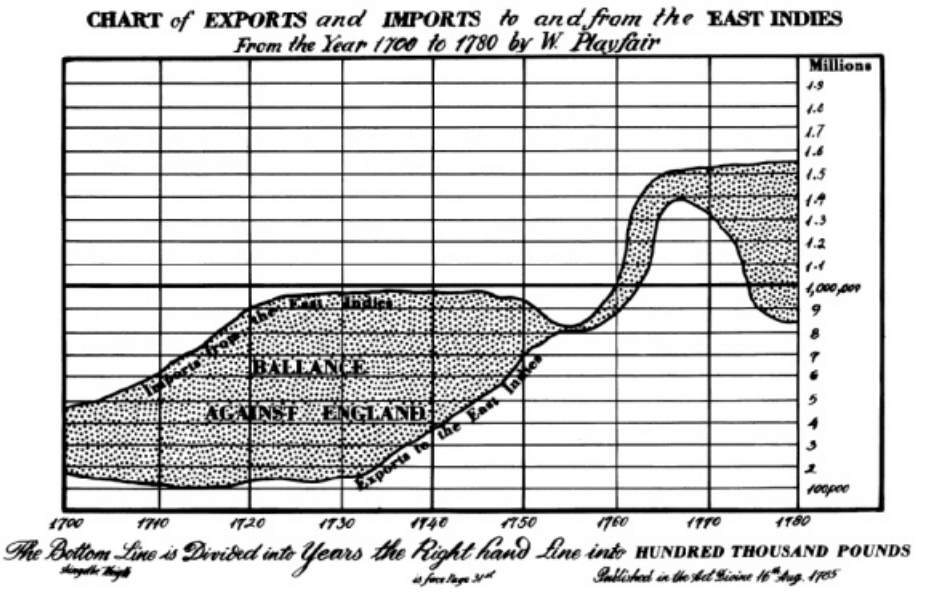
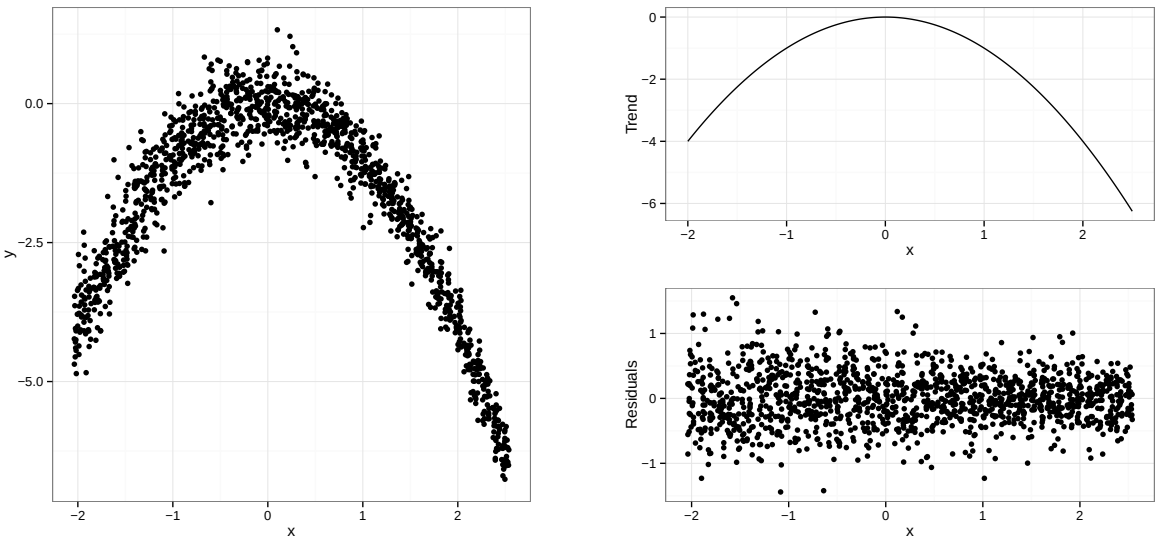


Fig. 5. Playfair's chart of trade between the East Indies and England, 1700-1780. The trade balance is influenced by the sine illusion: the difference between imports and exports in 1763 does not appear to be the same size as that in 1745, though the vertical distance is approximately the same.



(a) Data (b) Trend and Residuals

Fig. 6. Describe the conditional variability of the points along the x axis in (a). Is your description consistent with the residual plot in (b)?

along the line perpendicular to the tangent line of the curve. This comparison holds with scatterplots (such as figure 6) because when the points are dense, we examine variability by looking at the upper and lower contours of the data.

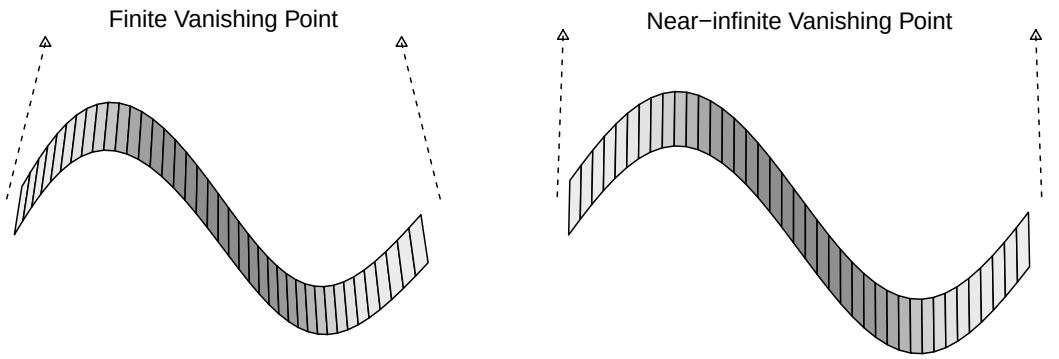


Fig. 7. Three dimensional context for the sine illusion. The second figure has a vanishing point closer to infinity, and very closely resembles the form of the classic sine illusion.

Day & Stecher [1991] suggest that the sine illusion is similar in principle to the Müller-Lyer illusion, attributing it to the perceptual compromise between the vertical extent and the overall dimensions of the figure. The sine illusion is similar to the Müller-Lyer illusion in another way, as well – there are three-dimensional analogues of the two-dimensional image that may influence the perceptual context. One of these contexts is shown in figure 7, generated from the same function shown in the two-dimensional analogue, figure 4, but with the length projected onto a third dimension. While the images do not match exactly, the similarities are striking. Additionally, the tendency to judge vertical distance using the extant width noted in Cleveland and McGill [1985] corresponds to the measurement of depth in the three-dimensional image. The main difference between the first three dimensional image shown in figure 7 and the original image is that the lines connecting the top and bottom sections of the curve are slightly angled in the three-dimensional version; this is due to the perspective projection used to create the image and the corresponding angles of rotation chosen such that the entire surface is visible.

As the vanishing point moves further away from the viewer and the 3d projection decreases in strength, the three-dimensional reconstruction of the image converges to figure 4. The second image in figure 7 shows a weaker 3-dimensional projection that is much closer to figure 4, however, the three-dimensional contextual information provided by the shading removes much of the illusion's distortions. This is similar to the Müller-Lyer illusion, as figure 3 is not at all ambiguous because the contextual depth information provided by the rest of the surface of the house is sufficient to remove the illusion that the closer corner is in fact larger due to the perspective.

Further evidence that places the sine illusion firmly into the area of a 3d contextual illusion is given by the non-response of individuals with depth-deficient vision. While it is difficult to provide experimental evidence suggesting that the sine illusion is due to depth perception directly, it is possible to examine whether the illusion occurs in people who do not have binocular depth perception. Conditions such as amblyopia (lazy eye) and strabismus (crossed eyes), when not corrected within a critical period during development [Hubel and Wiesel 1970], can result in weakened or absent depth perception [Henson and Williams 1980; Parker 2007; Holopigian et al. 1986]. In many cases, use of partial patching and early surgery can correct these problems before the critical period lapses, but before protocols were well established, this was not always completely successful.

We examined the effect of the sine illusion on DW, who has minimal depth perception due to strabismic amblyopia. DW was diagnosed as a young child, and prescribed complete patching to strengthen her intially non-dominant eye. As a result of the patching, DW developed near-independent control over both eyes (doctors now recommend partial patching as a result of this problem). She has 20/20 vision, and can wear glasses to correct the strabismus, but generally does not because they are not necessary for her to see well. DW is right-eye dominant in most contexts, but is left-eye dominant for driving, and can switch which eye is in focus at will.

We asked DW to view a subset of the sine illusion stimuli used in the experiment described in the next section, as well as the Müller-Lyer illusion, identifying the illusions as having lines that appeared the same length or different lengths (the stimuli are included in the appendix). DW identified both uncorrected sine illusion graphs as having lines of the same length, indicating that she did not appear susceptible to the sine illusion. In addition, DW identified the partially corrected images as having the same line length, indicating that the corrected image would produce similar conclusions as the uncorrected image (in this, she was not different from those with normal binocular vision). In fact, DW only identified the fully corrected $y = \exp(x)$ image as having lines of different length.

In addition to DW's resistance to the effects of the sine illusion, she also was not fooled by the Müller-Lyer illusion, instantaneously identifying the lines as the same length. This suggests that these two illusions are related to the presence of binocular depth perception, perhaps mitigated by experience.

One difference between the sine illusion and the Müller-Lyer illusion that may influence the tendency to see a three-dimensional "ribbon" instead of the two-dimensional sine curve is that the vertical lines in the sine illusion are ambiguously oriented - there is an entire plane of possible three-dimensional reconstructions for each line, and each possible rotation leads to a line of different length. It is this facet of the image that we believe partially contributes to the ambiguity of the image, though it is not a necessary feature for the illusion to persist, as the illusion also can be found in scatterplots and in "ribbon plots" such as figure 5.

2. MEASURING THE PSYCHOLOGICAL LIE FACTOR EXPERIMENTALLY

2.1 The Psychological Lie Factor

The psychological mechanisms which force three-dimensional context onto two-dimensional stimuli are useful adaptations to a three-dimensional world [Gregory 1968], but they do have disadvantages when applied to abstract two-dimensional stimuli, such as statistical charts. In order to assess the distortions due to the illusion, we need to quantify this distortion. For comparison, we will work from Tufte's lie factor [Tufte 1991, pg 57], which compares the size of an effect in the data with the size shown in a graphic, and is defined in equation 1.

$$\text{Lie Factor} = (\text{size of effect shown in chart})/(\text{size of effect shown in data}) \tag{1}$$

We will similarly define the psychological lie factor for this illusion as shown in equation 2.

$$\text{Psychological Lie Factor} = (\text{size of effect perceived})/(\text{size of the effect shown in the chart}) \tag{2}$$

In order to estimate the psychological lie factor that occurs due to this illusion, we assessed the strength of the illusion experimentally.

2.2 Study Setup

In order to determine the amount of the psychological distortion, participants were presented with different stimuli consisting of six subplots. One of these stimuli sets is shown in figure 8.

'similar to figure 8' is not enough information. Describe each one of the functions, and show tons of pictures.

These sets of charts were constructed such that line lengths along a curve were varied to different degrees to counteract the illusion. Figure 9 shows the amount of line correction used in each of the sub-plots in figure 8. Participants were asked to answer the question: "In which graph is the size of the curve most consistent?". The phrasing 'size of the curve' was chosen deliberately so as not to bias participants to explicitly measure line lengths.

Figure 10 shows another set of these stimuli using a different underlying mean function with the same underlying weight values.

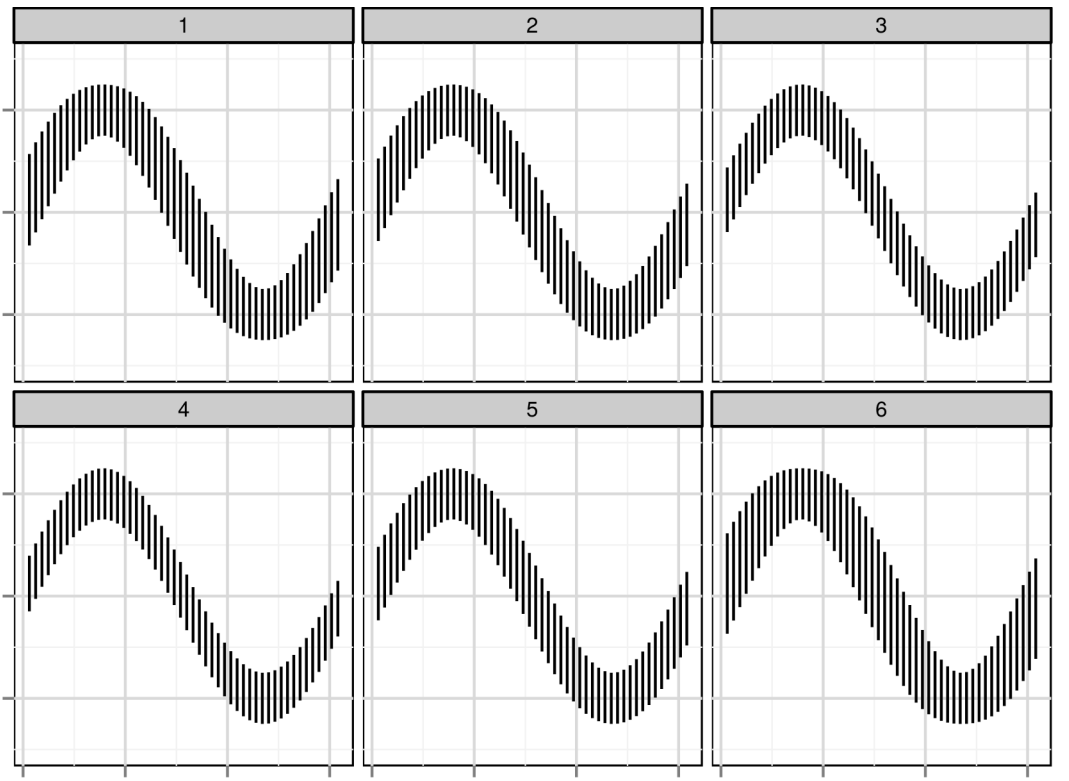


Fig. 8. One of the charts presented to participants through Amazon Mechanical Turk. Figure 9 shows the actual differences in line lengths. This chart corresponds to set #4 in table I. The plots are shown in random order, however; plot #1 corresponds to $w = 0.9$, plot #2 to $w = 0.7$, plot #3 to $w = 0.3$, plot #4 to $w = 0.1$, plot #5 to $w = 0.5$, and plot #6 to $w = 1.1$.

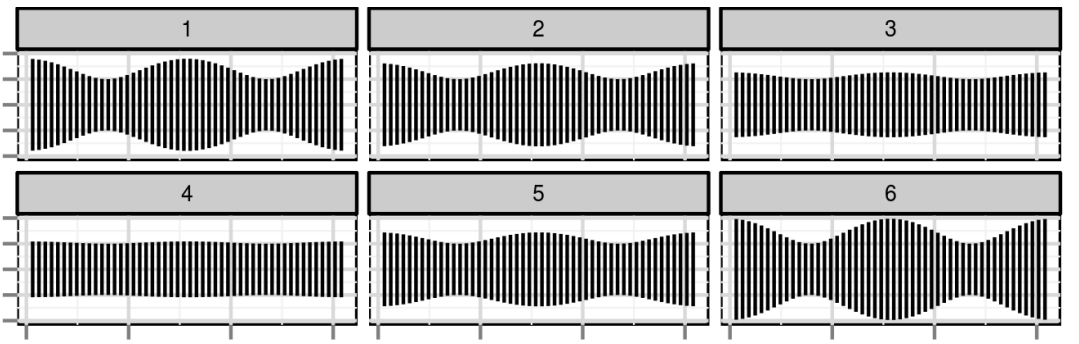


Fig. 9. De-trended line lengths for figure 8, demonstrating the distortion present due to the correction factor in each sub-plot. Comparing the distortion in the chosen sub-plot to the undistorted data produces an estimate of the psychological lie factor.

The underlying mean function change does change the effective lie factor (due to a different derivative, and thus, a different correction). The illusion, however, is still present; our goal is to determine whether the illusion's effects are consistent with regard to the strength of the preferred lie factor.

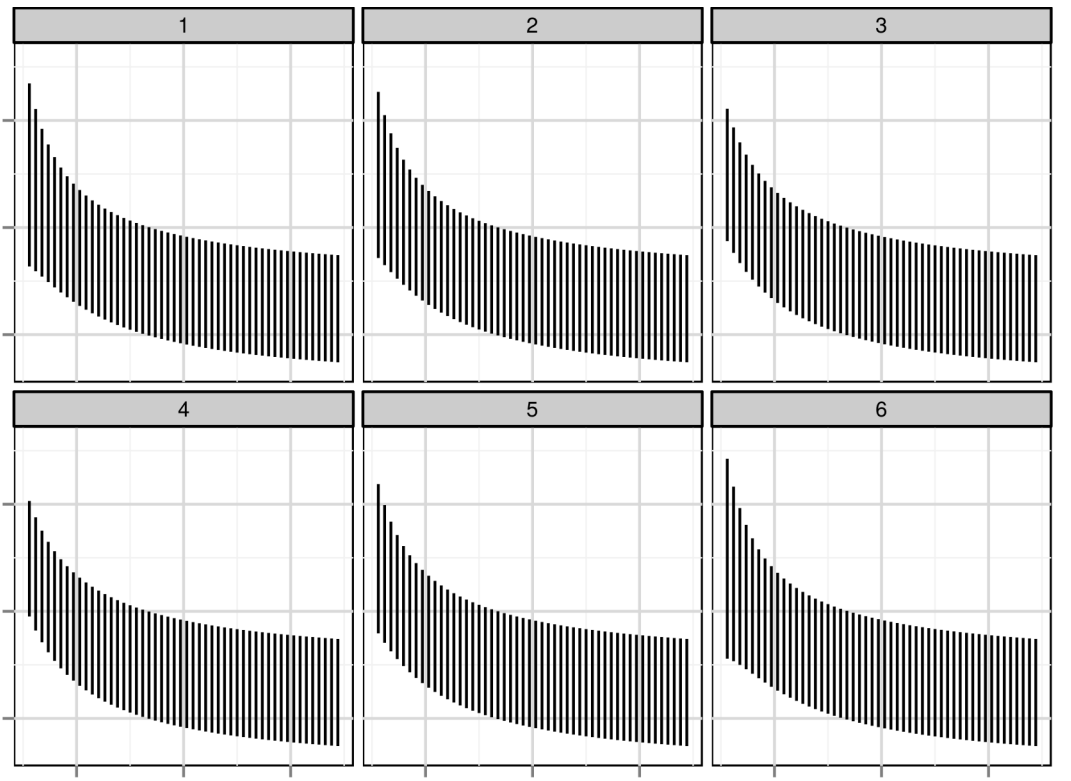


Fig. 10. Another chart presented to participants through Amazon Mechanical Turk. This chart corresponds to set #4 in table III. The plots are shown in random order, however; plot #1 corresponds to $w = 0.9$, plot #2 to $w = 0.7$, plot #3 to $w = 0.3$, plot #4 to $w = 0.1$, plot #5 to $w = 0.5$, and plot #6 to $w = 1.1$.

The amount of correction in each sub-plot was chosen such that each stimuli set contained a selection of curves corrected to various degrees using correction factors determined previously in an internal pilot study. The correction factor we use for extending the length of the line segment at location x is given as

$$(1-w)+w\cdot\left(1/\cos\left(\tan^{-1}(|f'(x)|)\right)\right) \tag{3}$$

where w represents a weight factor to allow variation of the strength of the correction. A value of $w = 0$ indicates that there is no correction, and a value of $w = 1$ indicates that the graph is fully corrected. Extending this approach, we can over-correct the graph, to test whether the geometric derivation of the correction is sufficient to remove the illusion. In this experiment, values of w were chosen from between 0 and 1.4.

In order to quantify the psychological lie factor D for each sub plot $k = 1,...,6$, we took the ratio of the maximum line length to the minimum line length shown in the plot. As participants were to choose the plot with lines that had the most even length, a large value of D indicates significant distortion.

$$D_k = (\text{maximum line length in sub-plot } k)/(\text{minimum line length in sub-plot } k) \tag{4}$$

The definition of D in equation 4 does deviate somewhat from our extension of Tufte's definition of the lie factor described in equation 2, but this is by necessity, as an uncorrected plot would show a difference of 0, which would

wreak havoc on any experimental measure of the lie factor, as that 0 would be in the denominator. Our modification preserves the interpretation of Tufte's lie factor, while adapting the computation for use in this experimental setting.

From a pilot study, we expected values around $w = 0.8$ to be sufficient to break the illusion, but did not know whether this would generalize to those outside of the statistical graphics community. In order to pinpoint the weight value necessary to correct the illusion more precisely, we chose twelve sets of 6 weight values each that were used to produce test plots similar to that shown in figure 8. These sets of weight values were chosen to allow greater precision estimates closer to $w = 0.8$, while still covering the range of w between 0 and 1.4. The sets of w used are shown in table I, along with corresponding lie factors D_k for stimuli with underlying function $sin(x)$ (other functions used will have different D_k due to the nature of the correction factor). Plots with weight values spread over the full range of w tested were considered "test" plots that could be used for verification purposes, while plots with weight values concentrated near $w = 0.8$ were considered to have higher difficulty (because the sub-plots were very similar). This allowed us to estimate w, and thus D, with higher precision while still exploring the entire parameter space.

Table I. Ordered weight factors and corresponding lie factors for the sine curve stimuli sets, as computed using equation 4.

set	difficulty	Weight (w)						Lie Factor (D) for $\sin(x)$ plots					
		sub-plot						sub-plot					
		1	2	3	4	5	6	1	2	3	4	5	6
1	test	0.00	0.20	0.40	0.80	1.25	1.40	1.00	1.18	1.35	1.71	2.11	2.24
2	test	0.00	0.15	0.35	0.80	1.20	1.40	1.00	1.13	1.31	1.71	2.06	2.24
3	1	0.00	0.20	0.40	0.60	0.80	1.00	1.00	1.18	1.35	1.53	1.71	1.88
4	1	0.10	0.30	0.50	0.70	0.90	1.10	1.09	1.27	1.44	1.62	1.80	1.97
5	2	0.05	0.30	0.50	0.65	0.80	1.00	1.04	1.27	1.44	1.57	1.71	1.88
6	2	0.10	0.30	0.55	0.70	0.85	1.00	1.09	1.27	1.49	1.62	1.75	1.88
7	3	0.40	0.60	0.70	0.80	0.90	1.05	1.35	1.53	1.62	1.71	1.80	1.93
8	3	0.35	0.65	0.75	0.85	0.95	1.05	1.31	1.57	1.66	1.75	1.84	1.93
9	4	0.35	0.50	0.60	0.70	0.80	0.95	1.31	1.44	1.53	1.62	1.71	1.84
10	4	0.40	0.55	0.65	0.75	0.85	1.00	1.35	1.49	1.57	1.66	1.75	1.88
11	5	0.50	0.65	0.75	0.80	0.90	1.00	1.44	1.57	1.66	1.71	1.80	1.88
12	5	0.50	0.60	0.70	0.75	0.85	1.00	1.44	1.53	1.62	1.66	1.75	1.88

Table II. Ordered weight factors and corresponding lie factors for the exponential curve stimuli sets, as computed using equation 4.

set	difficulty	Weight (w)						Lie Factor (D) for $\exp(x)$ plots					
		sub-plot						sub-plot					
		1	2	3	4	5	6	1	2	3	4	5	6
3	1	0.00	0.20	0.40	0.60	0.80	1.00	1.00	1.21	1.42	1.63	1.84	2.05
4	1	0.10	0.30	0.50	0.70	0.90	1.10	1.11	1.32	1.53	1.74	1.95	2.16
5	2	0.05	0.30	0.50	0.65	0.80	1.00	1.05	1.32	1.53	1.69	1.84	2.05
6	2	0.10	0.30	0.55	0.70	0.85	1.00	1.11	1.32	1.58	1.74	1.90	2.05
7	3	0.40	0.60	0.70	0.80	0.90	1.05	1.42	1.63	1.74	1.84	1.95	2.10
8	3	0.35	0.65	0.75	0.85	0.95	1.05	1.37	1.69	1.79	1.90	2.00	2.10
9	4	0.35	0.50	0.60	0.70	0.80	0.95	1.37	1.53	1.63	1.74	1.84	2.00
10	4	0.40	0.55	0.65	0.75	0.85	1.00	1.42	1.58	1.69	1.79	1.90	2.05
11	5	0.50	0.65	0.75	0.80	0.90	1.00	1.53	1.69	1.79	1.84	1.95	2.05
12	5	0.50	0.60	0.70	0.75	0.85	1.00	1.53	1.63	1.74	1.79	1.90	2.05

I know that all of these tables probably belong in the appendix, but I'm not entirely sure that's best. Thoughts?

Table III. Ordered weight factors and corresponding lie factors for the inverse curve stimuli sets, as computed using equation 4.

set	difficulty	Weight (w)						Lie Factor (D) for $1/x$ plots					
		sub-plot						sub-plot					
		1	2	3	4	5	6	1	2	3	4	5	6
3	1	0.00	0.20	0.40	0.60	0.80	1.00	1.00	1.14	1.28	1.43	1.57	1.71
4	1	0.10	0.30	0.50	0.70	0.90	1.10	1.07	1.21	1.36	1.50	1.64	1.78
5	2	0.05	0.30	0.50	0.65	0.80	1.00	1.04	1.21	1.36	1.46	1.57	1.71
6	2	0.10	0.30	0.55	0.70	0.85	1.00	1.07	1.21	1.39	1.50	1.60	1.71
7	3	0.40	0.60	0.70	0.80	0.90	1.05	1.28	1.43	1.50	1.57	1.64	1.75
8	3	0.35	0.65	0.75	0.85	0.95	1.05	1.25	1.46	1.53	1.60	1.67	1.75
9	4	0.35	0.50	0.60	0.70	0.80	0.95	1.25	1.36	1.43	1.50	1.57	1.67
10	4	0.40	0.55	0.65	0.75	0.85	1.00	1.28	1.39	1.46	1.53	1.60	1.71
11	5	0.50	0.65	0.75	0.80	0.90	1.00	1.36	1.46	1.53	1.57	1.64	1.71
12	5	0.50	0.60	0.70	0.75	0.85	1.00	1.36	1.43	1.50	1.53	1.60	1.71

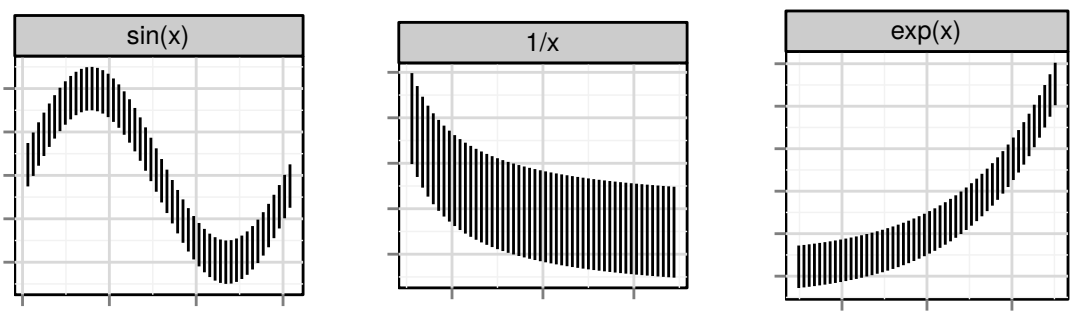


Fig. 11. Mean functions used during the experiment: $\sin(x)$, $\exp(x)$, and $1/x$. These functions are nonlinear, easily differentiable (for the correction factor), and are similar to trends commonly found in statistical graphics.

Three underlying mean function families were tested using these weight values: $\sin(x)$, $\exp(x)$, and $1/x$. These functions, shown in figure 11, were chosen because they are nonlinear functions which commonly occur in statistical graphics and are easily correctable using equation 3. The specific mean functions used in the experiment were $y = \sin(2x)$, $y = \exp(x/2)$, and $y = 5(6x)^{-1}$; the modifications allowed us to control the aspect ratio of the plot coordinate systems so that the uncorrected plots had aspect ratios between 0.75 and 0.85. As no x or y units were provided on the graph, these functional modifications served as experimental controls but did not change the information provided to the participants.

Each participant was presented with eleven sets of graphs (each "set" consisting of 6 separate plots), consisting of one "easy" test set, five stimuli sets of difficulty level 1 through 5 with the sine curve as the underlying function, and another five graph sets (also of difficulty levels 1 to 5) with either the exponential or the inverse curve as the underlying function. After the easy introductory chart, which was presented first, the order of the plots was randomized across difficulty level as well as function type. The test chart consisted of a set of six sine curves with a very low level difficulty level, and was used as an introduction to the testing procedure. Participants were asked to select a single plot out of the 6 plots presented as having lines which were the most "even".

2.3 Data Collection

Participants for the study were recruited through the Amazon Mechanical Turk web service, which connects workers with tasks that are not easily automated for a small fee. In exchange for completing at least 11 trials, participants were paid $1. Given the anonymity of web-based data collection, we informed participants that a unique IP address was required to participate in the experiment; responses from duplicate IP addresses with different Turk IDs were grouped and only the first response was paid. This procedure was used to lower the probability of a single user completing the task multiple times, in order to ensure that we could accurately estimate variation among individuals.

1598 responses from 115 users at 110 unique IP addresses were collected. We removed data from participants who did not complete at least 10 trials (allowing for one trial to be skipped or otherwise not completed), and of the collected responses, 30 trials from 4 participants were removed.

In addition to the requirement that participants complete at least 10 trials, the participant also was required to complete at least 4 trials of a specific underlying function for those trials to be included. This condition ensured that for any specific function, there were enough trials to estimate an initial effect; an additional 26 trials were excluded based on this condition.

There were 106 users who completed at least 10 trials each (providing enough data that we could accurately fit individual-level parameters), and those users completed 1542 trials which were used for this analysis. Though participants were asked to participate in 11 total trials, some participants continued to provide feedback beyond the eleven trials required to receive payment through Amazon. For any subsequent responses we randomly selected one of the 32 possible stimuli. This approach allowed us to collect some data in which a single participant provided responses to all three underlying functions. We did not exclude data based on user responses to avoid biasing our conclusions; the possibility of significant variability between individuals in the preference for a specific w value was too large to filter out even those who chose plots corresponding to $w = 1.4$.

Due to the experimental design, all participants completed trials with underlying function $y = \sin(x)$, for a total of 815 trials. As each participant who completed only the required 11 trials saw charts with either $y = 1/x$ or $y = \exp(x)$ as the underlying function, each of these functions had fewer trials; 316 and 411, respectively.

2.4 Analysis

Psychological "Lie Factor". As the strength of the correction varies across the horizontal range of the curve, we quantify the psychological distortion as the ratio of the maximum line length to the minimum line length for each sub-plot k: $D_k = l_{\max}/l_{\min}$. For a given set of 6 sub-plots, j, D_{jk} would denote the sub-plot distortion factor. Let D_{ijk} denote the distortion factor corresponding to participant i's choice of sub-plot k in stimulus set j during a single trial. In this experiment, there are 32 stimuli sets (2 test sets + 10 sets for each of 3 underlying functions), so $1 \leq j \leq 32$.

The correction for the sine illusion by default extends the line segments, so that if the initial line segments were all of length 1, the correction will produce corresponding line segments of length greater than or equal to 1. In addition, due to the underlying functions we have chosen, the minimum line length (assuming a starting line length of 1) after correction is approximately 1; this allows us to simplify D_k as

$$D_k = \text{maximum line length in plot } k$$

Without any correction for the sine illusion, this factor is, like Tufte's lie-factor, equal to one. Values above one indicate that at least in some areas of the curve line segments are extended.

We compute this quantity for each sub-plot in each stimulus presented to the participant. The participant's choice therefore provides us with an estimate of what value of D constitutes the most consistent line length (out of the set shown). As each set of 6 plots is not guaranteed to contain a plot with $w = 0$, corresponding to constant length, choosing a plot with $D = 1.4$ indicates more distortion if there is a sub-plot with $w = 0$ ($D = 1$) present than if least distorted sub-plot present has $w = 0.4$ ($D = 1.2$ for plots of $y = \sin(x)$). Correcting for this bias, the set of $\{D_{\cdot k}\} = \{D_{\cdot 1}, \ldots, D_{\cdot 6}\}$

that is available to choose from produces an estimate of the overall psychological 'lie factor' as

$$P_{ij} = D_{ijk} / \min_{1 \leq k \leq 6} D_{ijk} \tag{5}$$

for each plot and each participant. This normalization does conservatively bias the results, effectively shrinking the effect size we observe, but without the normalization we would be biased in favor of finding a significant result. Furthermore, while we could normalize relative to $\max_{1 \leq k \leq 6} D_{ijk}$, most stimuli sets contain $D > 1.85$ (for $y = \sin(x)$), where more than half of the stimuli sets do not contain $D \approx 1$. This is due to the range of w values chosen for the stimuli sets, as we had pilot data suggesting that $w = 0.8$ was a commonly preferred weight value.

That is, P is the ratio of the lie factor of the chosen plot to the smallest lie factor available in the set of available plots. When the set of available plots contains an uncorrected plot, $P = D$; when an uncorrected plot is not presented, $P < D$, since each $D_j >= 1$, $j = 1, ..., 6$. This transformation is a conservative approach to estimating the lie factor, but allows us to show a variety of scaled transformations and estimate the effect with more granularity around $w = 0.8$.

By considering each participant's answers for the plot with the most consistent line length, we can obtain an estimate of the psychological distortion from the sine illusion on an individual level. Estimating distortion factors for each participant facilitates comparison of these estimated values to determine whether the illusion is a product of an individual's perceptual experience or whether there is a possible underlying perceptual heuristic for the sine illusion common across the majority of participants. If the illusion is a learned misperception rather than an underlying perceptual "bug", we would expect there to be considerable variability in the estimated individual lie factor P_i for each unique participant i, $1 \leq i \leq 123$, as it is likely that personal experience varies more widely than perceptual heuristics and their underlying neural architecture.

Each set of w values as defined in table I corresponds to a value of P as defined in equation 5. We test for only a set of discrete values of w, which is reflected directly in the number of different values of P we can observe. This approach allows us to use a finite set of stimuli for testing, so that we can explicitly control the range of w displayed in each set of plots. To mathematically model a continuous quantity (the real domain of possible P values) using discrete data, we employ a Bayesian approach to model an overall psychological lie factor θ and individual participant lie factors θ_i.

Plots used in the experiment have factors P ranging between 1 and 2.5, so we can use a truncated normal data model for participant i viewing plot j, with $P = p_{ij} \sim N(\theta_i, \sigma)$ and independent flat priors $\pi(\theta) = 0.4$ and $\pi(\sigma) = 2.5$ for $\sigma \in [.1, .5]$. These prior distributions $\pi(.)$ represent our expectations of the values of θ and σ before the experiment; assigning them constant values indicates that we had little useable knowledge about the joint or marginal distributions of θ and σ before the experiment was conducted. Using Bayesian estimation, we can then obtain posterior distributions for θ_i and θ, the individual and overall mean lie factors. We are not particularly interested in the actual values of σ, but the additional parameter is a useful tool to better estimate possible values for θ.

2.5 Results

The posterior density of θ for each function is shown in figure 12, along with separate posterior densities for each individual θ_i. θ is reasonably similar for all three functions, suggesting that while function type may moderate the size of the effect, the illusion occurs regardless of function type. Individual curves have different variability due to the number of trials completed, and are necessarily more spread out as there is less data with which to estimate the individual posterior distributions.

On an individual level, figure 13 shows the posterior density for θ_i for four of the participants who completed at least 6 trials in each category. While in many cases, the most probable θ_i is similar across trials, individuals do seem to have been somewhat more affected by the illusion when the underlying function was sinusoidal, though this may reflect a discrepancy in the number of trials rather than a stronger illusion. Alternately, as the illusion depends on variable slope, it is possible that the monotonic exponential and inverse stimuli induced a weaker three-dimensional context. The posterior densities for these individuals appear extreme because they completed more trials (and thus estimates are much more precise); the individuals in question are not necessarily more prone to the illusion than other participants.

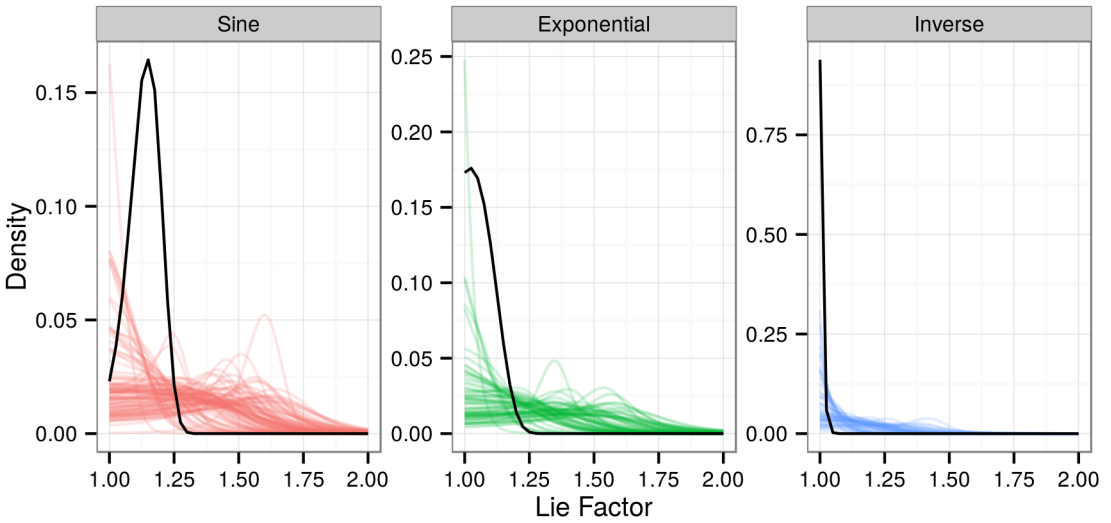


Fig. 12. Estimated densities for θ_i, shown in color, with the estimated overall density for θ shown in black. Individuals have extremely similar posterior distribution of θ_i, and even different functions have similar $\hat{\theta}$, suggesting a common underlying mental distortion. XXX shouldn't the posterior distribution be somewhat closer to the densities of individual θ_is?

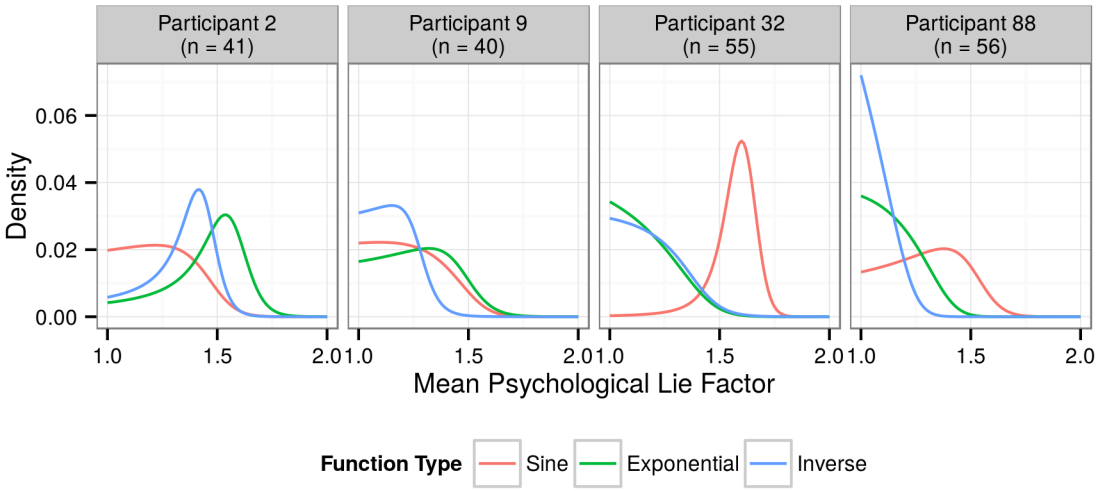


Fig. 13. Posterior distributions for θ_i for four of the participants who completed at least 6 trials of each of the three function types.

In order to appropriately compare intervals for each participant's θ_i (even though participants may have completed different numbers of trials), we simulated 11 new "data points" from our model (thus enforcing a uniform 11 trials per participant for each function type) to get a single new estimate of $\hat{\theta}_i$. For each participant, we generated 1000 of these $\hat{\theta}_i$ and used these simulated values to calculate the 95% credible intervals shown in figure 14. These intervals will allow us to consider the variability in θ_i due to participant preference rather than the number of trials a participant completed during the study. Removing this additional variability provides us with the opportunity to consider whether the sine illusion stems from an individual's perceptual experiences or from a lower-level perceptual heuristic.

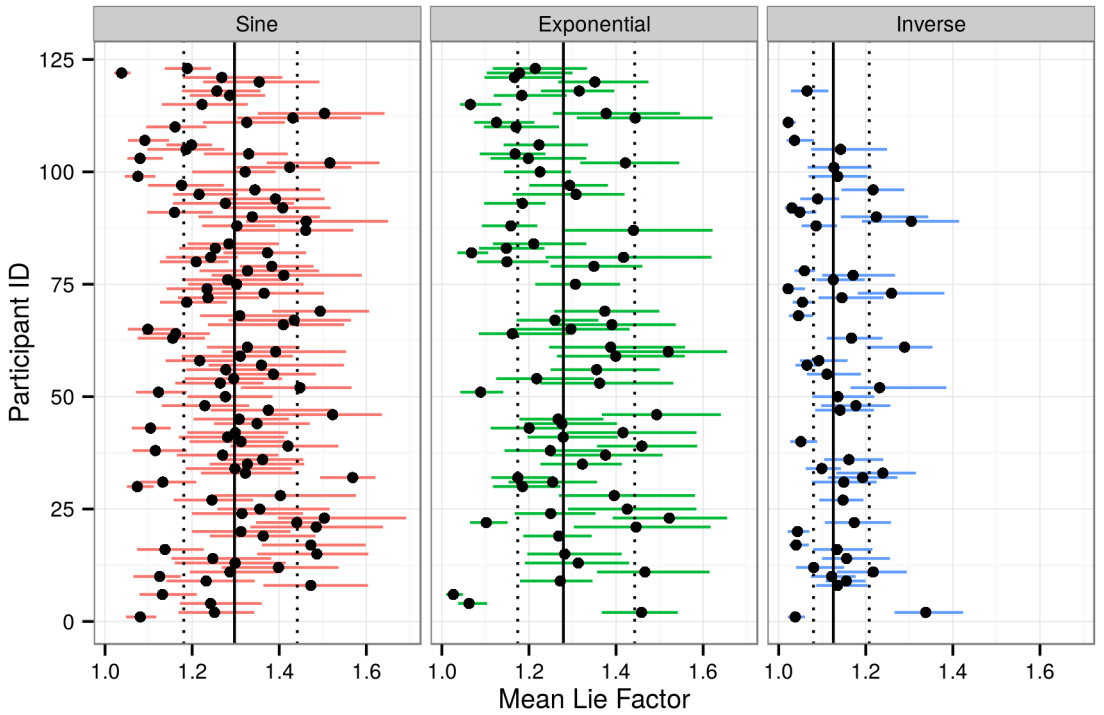


Fig. 14. 95% posterior predictive intervals for θ_i, calculated for each stimulus type. Vertical lines indicate the median estimate of the overall θ with a 95% credible interval.

Posterior predictive intervals for θ_i as shown in figure 14 suggest that overall, the θ_i are similar across individuals. Very few (5 each for $y = \sin(x)$ and $y = \exp(x)$, 16 for $y = 1/x$) of the intervals overlap the region (1, 1.05), which corresponds to an "acceptable" lie factor according to Tufte. This indicates significant distortion for most participants in our experiment, and the marked overlap of the intervals for each participant provides evidence consistent with a common magnitude of distortion. This suggests that there may be some common psychological strategy that is misapplied to the perception of these stimuli.

Comparison of the Preferred Stimuli. Estimates of $\hat{\theta} = E[\theta]$ for each function are 1.31, 1.29, and 1.14 respectively for exponential, inverse, and sine functions, suggesting a similar psychological distortion even for very different functions, though it seems as if the inverse function causes somewhat less distortion, possibly because the correction factor is not as proportionately large. Credible intervals can be found in Table IV. As all three of the credible intervals exclude 1.05, there is evidence that a psychological distortion is occurring; that is, there is evidence of a significant psychological lie factor. In addition, the method of adjusting the estimated lie factor we have used here is conservative; it is likely that because most stimuli contain a sub-plot with $w \geq 1$ our estimate of θ for the inverse function is low (as the lie factor for fully corrected inverse plots is greater than the corresponding lie factors for the other two functions used in this experiment).

The estimated weight values corresponding to these θ are shown in figure 15. In all three cases, the experimentally-corrected plots appear less distorted than the uncorrected plots.

This experiment has demonstrated that the sine illusion results in mis-perception of graphically presented data. In particular, participants tend to see uneven line length when lines are even while missing uneven line length due to the illusion's effect.

Table IV. Credible intervals for the overall θ for exponential, inverse, and sine stimuli.

Function	95% Credible Interval for θ	Median
Sin	(1.0623636, 1.2310909)	1.1372727
Exp	(1.1984091, 1.4643182)	1.3004545
Inv	(1.1735455, 1.3952273)	1.2863636

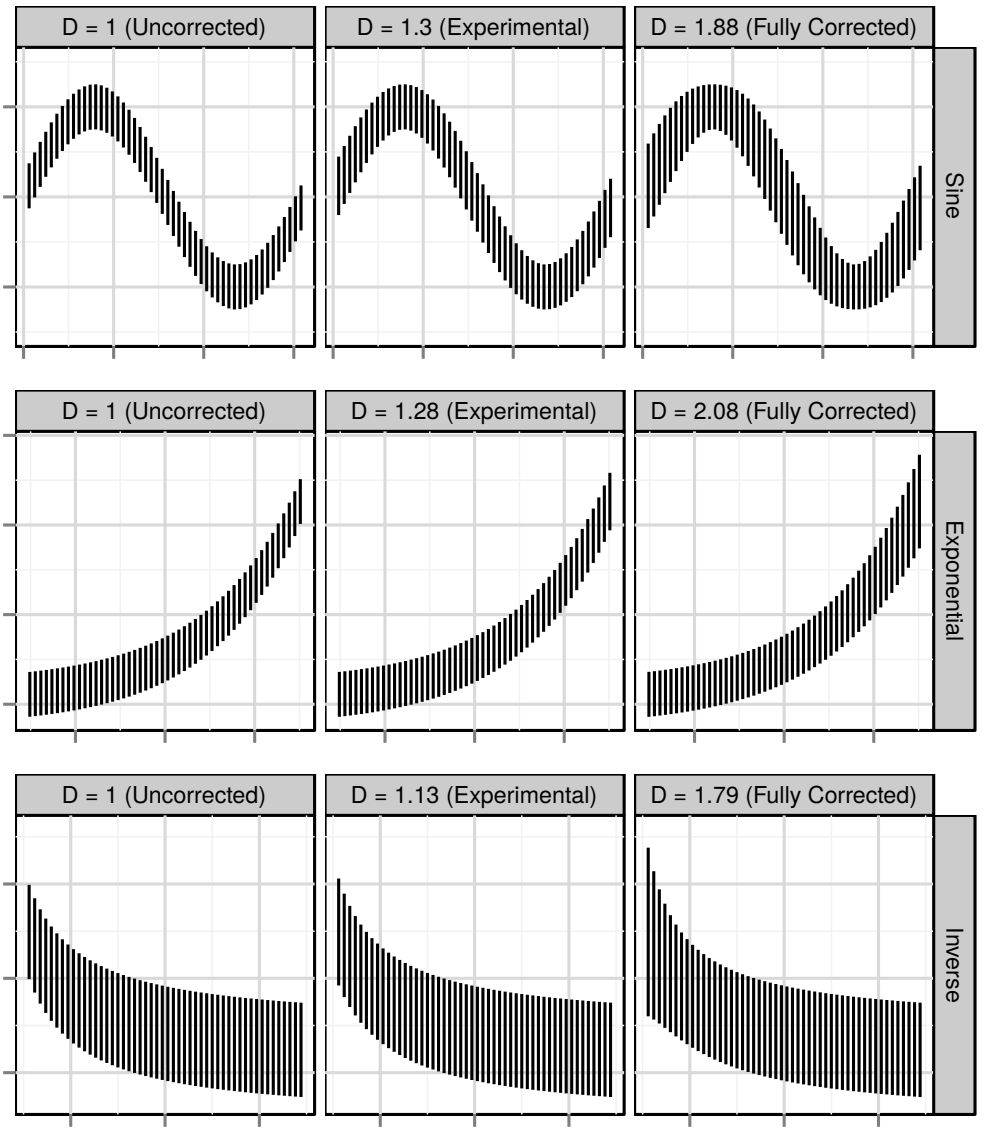


Fig. 15. Uncorrected, experimentally corrected (according to the median value of θ), and fully corrected stimuli for all three underlying functions used in the experiment. The corrected value shown here is equivalent to the distortion factor D defined as $D = \ell_{max}/\ell_{min}$.

3. CONCLUSIONS

The sine illusion arises from misapplication of three-dimensional context to a two-dimensional stimulus which results in nearly unavoidable perceptual distortions. These distortions impact the inferences made from charts which are similar to three-dimensional figures, even when we are not consciously aware that this context exists; the only immunity to the illusion we have found is for those who have never developed binocular depth perception. We have estimated that the illuson produces a distortion of about 135%. This distortion occurs entirely between the retinal image and the mental representation of the object; it is not due to the chart, rather, it is an artifact of our perceptual system.

As Tufte advocated for charts that showed the data without distortion, our goal is to raise awareness of perceptual distortions that occur within the brain itself due to misapplied heuristics. While applying corrections to the data to remove these distortions is somewhat radical, the persistence of the illusion despite awareness of its presence presents a challenge to those seeking to display data visually. In addition, many graph types can induce this illusion (scatterplots, ribbon plots, parallel sets plots), so avoiding a specific type of graph is not an effective solution. The best solution to this problem is to raise awareness: to demonstrate that optical illusions occur within statistical graphics, and to understand how these illusions arise.

REFERENCES

AHLUWALIA, A. 1978. An intra-cultural investigation of susceptibility to "perspective" and "non-perspective" spatial illusions. British Journal of Psychology 69, 233–241.

AMER, T. 2005. Bias due to visual illusion in the graphical presentation of accounting information. Journal of Information Systems 19, 1, 1–18.

CLEVELAND, W. S. AND MCGILL, R. 1984. Graphical perception: Theory, experimentation, and application to the development of graphical methods. Journal of the American Statistical Association 79, 387, pp. 531–554.

CLEVELAND, W. S. AND MCGILL, R. 1985. Graphical perception and graphical methods for analyzing scientific data. Science 229, 4716, 828–833.

DAY, R. H. AND STECHER, E. J. 1991. Sine of an illusion. Perception 20, 49–55.

GREGORY, R. 1968. Perceptual illusions and brain models. Proc. Roy. Soc. B 171, 279–296.

HENSON, D. B. AND WILLIAMS, D. E. 1980. Depth perception in strabismus. British Journal of Ophthalmology 64, 5, 349–353.

HOFMANN, H. AND VENDETTUOLI, M. 2013. Common angle plots as perception-true visualizations of categorical associations. Visualization and Computer Graphics, IEEE Transactions on 19, 12, 2297–2305.

HOLOPIGIAN, K., BLAKE, R., AND GREENWALD, M. J. 1986. Selective losses in binocular vision in anisometropic amblyopes. Vision research 26, 4, 621–630.

HUBEL, D. H. AND WIESEL, T. N. 1970. The period of susceptibility to the physiological effects of unilateral eye closure in kittens. The Journal of physiology 206, 2, 419–436.

PARKER, A. J. 2007. Binocular depth perception and the cerebral cortex. Nature Reviews Neuroscience 8, 5, 379–391.

PLAYFAIR, W. 1786. Commercial and Political Atlas. London.

POULTON, E. 1985. Geometric illusions in reading graphs. Perception & psychophysics 37, 6, 543–548.

TUFTE, E. 1991. The Visual Display of Quantitative Information 2 Ed. Graphics Press, USA.

VANDERPLAS, S. AND HOFMANN, H. 2014. Signs of the sine illusion - why we need to care. Journal of Computational and Graphical Statistics.