How the Statistical Structure of the Environment Affects Perception of the Müller-Lyer Illusion

Stephen Blessing (sblessing@ut.edu)
Martina Svetlik (msvetlik@ut.edu)
Department of Psychology, University of Tampa
401 W. Kennedy Blvd., Box Q
Tampa, FL 33606 USA

Abstract

Past researchers have attempted to explain the Müller-Lyer illusion through such means as eye movements (Carr, 1935) and isolating parts from the whole (Pressey, 1971). A recent explanation is that observers are sensitive to statistical regularities within the environment (Howe & Purves, 2005a). Certain configurations of lines appear more often than others within the environment, and this accounts for the illusion. Experiment 1 investigated whether these observed environmental statistical regularities are matched by behavioral data. In Experiment 2 the probability of shorter or longer lines within the illusion was manipulated in order to influence people's perception of the illusion. The evidence from both experiments provides data that are consistent with the hypothesis that perception of the Müller-Lyer Illusion is affected by the statistical nature of the environment.

Keywords: perception; geometrical illusions; probability distributions

Introduction

The Müller-Lyer effect is a classic visual illusion. The standard presentation is shown in Figure 1a. The horizontal line contained within the figure on the bottom is generally perceived as longer than the horizontal line in the upper figure, though both are in fact the same length. Somehow the arrowhead adornments pointing out drive the observer to perceive that figure differently than when the arrowhead adornments are pointing in. Several hypotheses have been posited to explain this observation. Most concern themselves either with properties of the figure itself or properties of the visual system, though more recent attempts have examined the nature of the environment itself.

An older explanation by Gregory (1963) suggested that the illusion exists because the observer interprets the figure where the arrowheads point out as a concave corner (that is, the corner is farther away than the walls), whereas the figure with the arrowheads pointing in is interpreted as a convex corner (that is, the corner is closer). However, this account is called into question not only by the illusion persisting through variants of the original stimulus (see Figure 1) in which the corner explanation is not valid, and also through behavioral accounts as well (McGraw & Stanford, 2001).

Other explanations include an eye movement account, stating that because the eye moves more in analyzing the figure where the adornments point out than with the other figure, the observer perceives that line as longer (Carr, 1935). However, research has shown that even in the absence of eye movements, the illusion still exists (Bolles, 1969). Another account is referred to as assimilation theory, as it states that the illusion exists because the observer finds it hard to separate the single horizontal line from the rest of the figure (Pressey, 1971). However, this account has also shown not to be upheld by behavioral data (Bross, Blair, Longtin, 1978). More recently, Bulatov, Bertulis, and MicKiene analyzed the Müller-Lyer according to its spatial properties, and developed a neurophysiogical model. They obtained similar patterns between their model and behavioral data.

The research presented here takes the probabilistic explanation of the Müller-Lyer illusion by Howe and Purves (2005a) and attempts to match it to behavioral data. In contrast to the accounts mentioned above, this one bases itself on the statistical nature of the environment. In that regards, this account is consonant with theories that explain human behavior and cognition as adapting to the environment (e.g., Anderson, 1993; Gibson, 1966). Howe and Purves analyzed natural scenes for the standard version of the Müller-Lyer illusion and four of its variants. These natural scenes were gathered in a variety of settings using a laser device that converts 3-D space into numerical data representing the distance and direction of all points on visible surfaces that can then be related to retinal images (see Howe & Purves, 2005b, for a more detailed explanation). Using a selection of such scenes, they generated probability distributions of various line lengths and adornments. These distributions indicate that given a target adornment found in the scene, the probability of

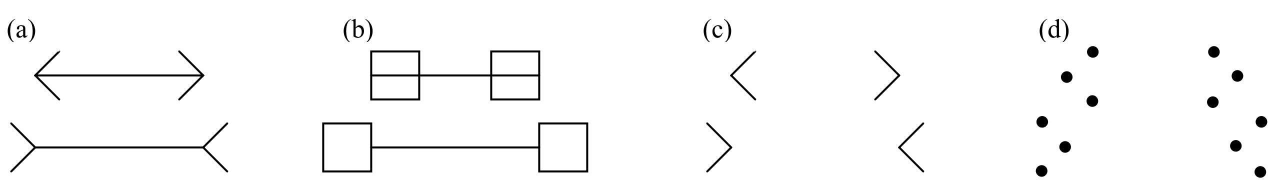

Figure 1: Standard Müller-Lyer illusion (a), box variant (b), shaftless variant (c), and dot variant (d).

finding a matching adornment is greater at a shorter line length when the adornments are in the "pointing in" orientation. Put another way, it is more probable to find shorter Müller-Lyer-like configurations in the environment for when the tails point in than when the tails point out. Figure 2 shows such a distribution for the standard variation, based on the original data of Howe and Purves (2005a). For instance, consider the black arrowhead adornment at the top of the figure. The probability of finding a matching adornment to the left (thereby creating a figure where the adornments point in) of any specific length, such as the 25 units indicated in the figure, is greater than finding a matching adornment traveling to the right at that same length (which would create a figure where the adornments point out). Considering lines of length 25, one is 10% more likely to find the matching adornment to the left than to the right. The probability distribution essentially reverses if the initial adornment flipped, as the dark grey arrowhead and distribution in Figure 2 show.

If the probabilistic findings that Howe and Purves (2005a) show in their data are indeed a driver of our perceptions, then it should show up in behavioral data as well. The two experiments presented here attempt to check this claim. Experiment 1 examines some of the data they present regarding Müller-Lyer variants. As a stronger manipulation, Experiment 2 changes the distribution curves of what participants experience regarding figures with adornments pointing in versus pointing out, to see if that influences their perceptions, at least within the experimental session.

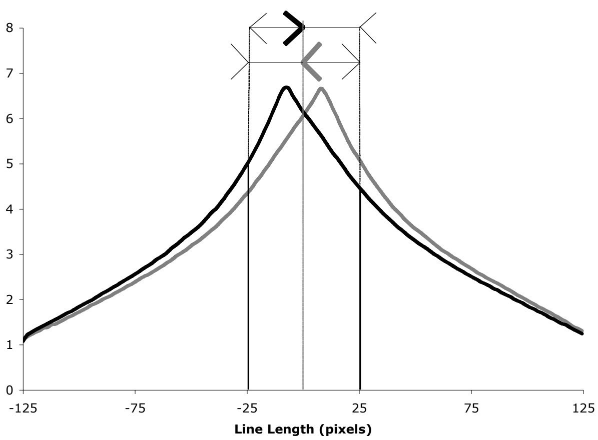

Figure 2. Probability distribution for the standard Müller-Lyer variant (adapted from Howe & Purves, 2005a).

Experiment 1

In addition to analyzing the natural scenes for the relative distributions of the standard version of the Müller-Lyer illusion, Howe and Purves also analyzed the scenes for several of its variants (see b-d in Figure 1). Perhaps not surprisingly, the distributions differ depending on the variant, but the general finding that shorter line lengths for when the adornment points in is more probable than when the adornment points out is true in all versions. (They also examined a fourth variant, one that was vertical in orientation; we did not use that variant in this experiment, due to a desire to have consistently horizontal stimuli.)

These different distributions suggest that the variants may differ in the magnitude of the illusion. The expectation would be that this difference should vary proportionately with the amount of overlap that one sees between the distributions for when the adornment is pointing in versus when the adornment is pointing out. That is, more overlap between the curves should indicate a weaker sense of the illusion, as that would mean the probability of shorter or longer lines becomes more equal to one another. Past researchers (Coren & Girgus, 1974) measured differences in the illusion magnitude of Müller-Lyer variants. However, they did not use one of the variants analyzed by Howe and Purves (2005a). This experiment serves as a partial replication of the prior study, with the addition of another variant. The results we obtain will be compared with the prior Coren and Girgus findings, in addition to all data being correlated with the environmental analysis of Howe and Purves.

Method

Participants The participants were 40 undergraduate students attending the University of Tampa recruited from various psychology classes. Participants received course credit for their participation.

Material Custom computer software was created to display the stimuli. The computers used were running the Windows XP operating system with 17" color CRT monitors. The software used the four variants displayed in Figure 1 (referred to as "standard," "boxes," "shaftless," and "dots"). Each of the variations could be displayed with the adornments either facing inward or outward. Furthermore, the length of each line was varied between 150 to 400 pixels, in 50 pixel increments, for a total of six possible lengths. This combination of six lengths, four variants, and two adornment types yields 48 possible stimuli. Each participant saw each stimuli exactly once. The order of stimuli presentation was determined randomly for each participant, with the only stipulation being that each variant had to be seen once before it could be repeated again. Each trial presented one of the 48 stimuli in the upper half of the screen, and below that was a line containing no adornments that whose size the participant could manipulate. The lower figure was placed off center to the upper one so that the participant could not line up the end points of the two lines.

Procedure The procedure used is similar to the one followed in Crawford, Huttenlocher, and Engebretson (2000). For each trial, the chosen stimulus was shown in the upper part of the screen. Beneath that figure was a horizontal line presented without adornment. The participant could manipulate this plain line by tapping the 's' key to make the line shorter or by tapping the 'l' key to make the line longer (both done in single pixel increments).

The participants were told to match this line's length as closely as possible to the horizontal line contained or implied within the stimulus above (in both the "shaftless" and "dots" variants, the horizontal line is not actually drawn). The lower line was offset 75 pixels either to the left or to the right of the stimulus, so that the participant could not adopt a strategy of visually lining up the endpoints (the offset direction was randomly chosen for each trial). No time restraints were placed on any trial, allowing the participant to work at their own pace. For each trial, the difference between the line that the participant could adjust and the length of the actual stimulus was recorded.

Results

The difference between the participant's length estimate and the actual line length was obtained for each trial. We divided this difference by the actual line length, to get the proportion error for each trial. For each participant on each variant, we took their average error across length on the stimuli with adornments pointing out (typically overestimates) and subtracted their average error across length on the stimuli that had the adornments pointing in (typically underestimates). This difference we consider to be the magnitude of the illusion for each variant.

Data from five participants were discarded, because they were over three standard deviations away from the average performance on one of the variants ("boxes"). It appears they were matching the subset of the full horizontal line.

Table 1 displays the data from both our experiment and the experiment by Coren and Girgus (1974) that also examined variants of the Müller-Lyer illusion (Coren & Girgus did not compare the boxes variant; they did examine two additional variants, but as Howe and Purves did not analyze them, those data are not included.) Our data are the computed magnitude illusion as described above, and the Coren and Girgus results are a similar measure, based on information in their paper (they tested 10 participants per variant, with 2 trials per participant).

Table 1: Differences in Illusion Magnitude of Müller-Lyer Variants.

Variant	Coren & Girgus Results	Current Results
Standard	0.27	0.25 (0.07)
Shaftless	0.20	0.23 (0.08)
Dots	0.08	0.09 (0.07)
Boxes		0.14 (0.10)

A repeated measures ANOVA on our results shows a significant difference of variant, $F(3,96)=38.08$, $p<.01$, $\eta^2 = .54$. Performing pairwise comparisons, it is found that all variants are significantly different from one another ($p<.05$), except for the standard and shaftless variants. One obvious result seen in Table 1 is that the ordering from the variant with the most illusion magnitude to the least (Standard, Shaftless, and Dots) is the same for both our results and the results from Coren and Girgus (1974). It should also be noted that in both sets of results the dots variant is the weakest.

Our hypothesis is that, if the perception of the illusion has a basis in the environment, there should be a relationship between our results and the data and graphs provided by Howe and Purves (2005a). We analyzed their data first with respect to the amount of overlap between the two distributions they plot on their graph (for an example, see Figure 2; Howe & Purves, 2005a, provided figures for all variants). One distribution plots the probability of finding a matching adornment based on length if the first adornment initially points to the right (the black distribution in Figure 2), and the other distribution plots the probability if the initial adornment points left (the grey distribution). The amount of non-overlap was found by computing the proportion of non-shared area between the two curves. All curves have much overlap, as can be seen in Table 2.

Table 2: Comparison of Current Results with Proportion Non-Overlap between Distributions.

Variant	Proportion Non-Overlap	Magnitude of Illusion
Standard	0.07	0.25
Shaftless	0.03	0.23
Dots	0.01	0.09
Boxes	0.04	0.14

What is perhaps more important than the proportions, is the relative rankings. The Standard variant has the least overlap, and it has the strongest illusion magnitude. The Dots variant has the least overlap (there is a difference in the two distributions, but obviously very little). Comparing the non-overlap of the shaftless and boxes variants in Table 2, one would have expected perhaps reversed results in Table 1 concerning these two variants. In looking at those two distributions in particular, the boxes variant does indeed have less overlap, but the modes on the shaftless variant distributions are much further apart, by over a 2 to 1 factor. The standard's variants modes are farthest apart.

Discussion

Previous researchers have shown how physical attributes (length and height of the adornments, the angle that the adornments make with the horizontal) of the Müller-Lyer illusion stimuli can vary the results (Bulatov, Bertuli, & Mickiene, 1997). Given that and with the experimental differences, the degree of similarity between our results and those of Coren and Girgus (1974) is quite high (considering that in their experiment participants manipulated a wooden version of the illusion; ours was done on a computer).

Comparing these behavioral results to the environmental data of Howe and Purves (2005a), there appears to be some degree of correspondence. The variant with the highest illusion magnitude is the variant one would expect based on the probability distributions (that is, the standard form of the illusion). Likewise, the variant with the least illusion magnitude, the dots variant, also corresponds. The middle two variants (boxes and shaftless) might have some interplay that would be difficult to account for between the overlap of the distribution curves and how far apart the two curves are, resulting in their lack of direct correspondence.

Experiment 2

Experiment 2 provides a stronger examination to the idea that people's perceptions of the environment are at least partly influenced by the statistical nature of the environment. If a person's perceptions are indeed influenced in such a way, then it might be possible, in a laboratory setting, to change their apparent perception of the illusion by manipulating in some way the statistical information they are receiving from the environment.

In this experiment we execute such a manipulation by having participants perform a judgment task in which they indicate which of two standard Müller-Lyer stimuli is longer. We manipulate the answers such that half of the participants see proportionately more stimuli in which the Müller-Lyer stimulus with the adornments pointing in is truly longer than a Müller-Lyer stimulus with the adornments pointing out. The other half of the participants encounters what would be considered the more likely scenario, given the environmental factors, where the longer stimulus is usually the figure with outward facing adornments. Our hypothesis is that the first group of participants, by observing more cases where what used to be statistically more likely the shorter stimulus is truly the longer stimulus, will become more accurate in their size adjustments of the illusion.

Method

Participants The participants were 65 undergraduate students attending the University of Tampa recruited from various psychology classes. Participants received either course or extra credit for their participation.

Material The same computers were used for Experiment 2, and the software was very similar as what was used in Experiment 1. For this experiment, only the standard variation of the Müller-Lyer illusion was used. The experiment was divided into three parts. The first and third sections were identical to the adjustment task used in Experiment 1, except with fewer stimuli. With only the one variant (the standard one), but the same six line lengths (150 – 400 pixels, at 50 pixel increments) and the two types of adornments, there were 12 possible stimuli. During each of the first and third parts of the experiment, each of these 12 stimuli was presented exactly once, as they were in Experiment 1. The middle part of the experiment was a judgment task where the participant had to decide on each trial which of two standard Müller-Lyer variants had the longer horizontal line. In each pair, one stimulus had the adornments pointing out, and the other had the adornments pointing in. These two stimuli were presented near the middle of the screen, with one variant on top, and the other beneath it (slightly offset to one side of the other, as in Experiment 1). The length of one line was between 150 – 400 pixels (50 pixel increments). The length of the other line was then set to be either 25% larger or smaller than that of the first line. This percentage was based on the illusion magnitude of the first experiment, so as to somewhat "erase" the effects of the illusion. For half of the participants, 70% of the time the stimuli that had the adornments pointing in would be longer (Group L), and for the other half of the participants, 70% of the time the stimuli with outward facing adornments would be longer (Group M). Participants in Group L, then, were receiving stimuli not consistent with their experiences outside of the experiment as shown in the Howe and Purves (2005a) data (that is, they were seeing a majority of cases where the variant with the adornments pointing in actually was longer).

Procedure First, participants were randomly placed into either Group L or Group M. The experiment consisted of three parts. The first and third parts were identical to Experiment 1, except that only one variant was used (the "standard" one), resulting in 12 stimuli being presented within each part. Participants in both groups saw the same stimuli in these two sections. Which stimulus was presented on each trial was randomly determined for each participant. As before, the difference between the length of the plain horizontal line that the participant could adjust and the horizontal line of the stimulus was recorded for both of these experimental sections.

In the second, middle part of the experiment the participants were still presented with two horizontal figures, but both had adornments (one pointing out, the other pointing in). They were instructed to indicate which of the two figures had the longer horizontal line. If the upper figure was longer, they were instructed to hit the 's' key, and if the lower figure was longer, they were instructed to hit the 'l' key. If the participants chose the incorrect answer they would receive feedback indicating so, the experiment would pause for 5 s, and then it would move on to the next trial. If the participant was correct, there was no pause. Participants in Group L saw one distribution of stimuli, and Group M participants saw a different distribution, as described above. There were a total 200 trials for part two, with the order of the stimuli being randomly chosen for each participant. How accurately participants performed on these trials were recorded. There were no pauses or breaks in the experiment between the three parts.

Results

The illusion magnitude as described in the results of Experiment 1 was computed for each participant for both the first and third parts of the experiment. In addition, an accuracy measure was computed for the judgment task that comprised the second part of this experiment. Being how it is critical for participants to pay attention and accurately process these stimuli, those participants with poor accuracy, defined as 80% or less, were not included in subsequent analyses. This affected eight participants in Group L (the harder of the two conditions, as the 25% difference between the two stimuli put the illusion magnitude right at the breakpoint for the average participant) and one participant in Group M. Thus, there were 25 participants in Group L (*M* = 0.908 for accuracy, *SD* = 0.050) and 29 participants in Group M (*M* = 0.906, *SD* = 0.051).

Table 2: Differences in Illusion Magnitude.

Condition	Pre-Judgment	Post-Judgment
Group L	0.236 (.066)	0.201 (.065)
Group M	0.225 (.073)	0.234 (.053)

The means and standard deviations of all conditions are shown in Table 2. Difference scores for each participant were obtained by subtracting post-judgment performance from pre-judgment performance. A between-subject *t*-test on these difference scores reveals a significant difference between the two groups ($t(52) = 2.17$, $p < .05$, $d = 0.55$; no difference existed before the judgment task, $t(52) = 0.41$, *n.s.*). Also of note, though not significant, is that whereas 76% of the participants in Group L improved their performance between the first and third parts, only 55% of the participants in Group M did so. This means that almost twice as many people in Group M actually did worse on their second attempt at the adjustment task.

Discussion

While the absolute differences in improvement between the two groups are not large, there is a difference (3.5% in Group L, -0.9% in Group M). If one examines relative changes, the Group L participants got 14.8% better at the task on average, whereas, the people in Group M were actually 4.0% worse, on average, at making their adjustments than they were initially.

We were surprised by the number of people in Group M who performed worse on the post-judgment task relative to their pre-judgment performance (45%). Past research has demonstrated an improvement across time in making such adjustments (this was one main finding of the Coren & Girgus, 1974, research discussed in Experiment 1). Prior to conducting the experiment, we expected both groups to improve, but that Group L, given their judgment experiences, would improve more. However, whereas Group L did improve, Group M did not. If anything, Group M's experience in the judgment task, where the stimuli they saw confirmed prior experience (that is, Müller-Lyer-like stimuli with adornments pointing out tend to be long) resulted in worse performance at the adjustment task.

General Discussion

Both experiments provide behavioral evidence for the findings and theory put forward by Howe and Purves (2005a). In Experiment 1, the differences that one sees in the magnitude of the illusion of the various Müller-Lyer variants in the behavioral data reflect to some degree those that are observed in the environmental probabilities found by Howe and Purves. Experiment 2 shows that a person's perception of the illusion can be manipulated in a laboratory setting by providing them a sequence of stimuli that is either consistent with or not consistent with their prior perceptions.

In future work we hope to further investigate the extent to which this probabilistic account of perception can be applied to behavioral findings. First, we are interested in how robust the finding is, and what may influence that robustness within individuals. In pilot testing for Experiment 2, we determined that 200 judgment trials was likely to be sufficient to detect if the effect existed or not, that the difference between the two stimuli, particularly for Group L participants, should be 25%, and that an adequate ratio to use was 70/30 between the longer/shorter stimuli depending on condition. One could manipulate any of these variables to determine its effect on perception of the stimuli, and which of these has the stronger influence. Also, it appears that there may be individual differences as to how strong the effect is. Some participants in Group M got considerably worse at the adjustment task after doing the judgments (5 participants had difference scores greater than 10%; none in Group L had such a large decrease in performance). Another interesting set of experiments to pursue would examine transfer issues and how long the effect lasts. Given the information we have about variants from Experiment 1, one could examine if seeing different probability distributions within one variant transferred to a different variant (Coren & Girgus, 1974, did see transfer across variants in terms of people making better size determinations after exposure to a different variant). Given the manipulations here, and the fact that we have a lifetime of exposure to such probability distributions as shown in Figure 2, one would not expect this effect to be long lasting, nor exist beyond the immediate experimental session. However, it would be interesting to see how durable the change might be.

Howe and Purves use their probabilistic-based theory to explain a much wider spectrum of perceptual phenomena, such as perception of line length and angle in general, and also other illusions such as the Poggenndorff Illusion (2005b). In findings similar to that discussed here for the Müller-Lyer illusion, they show how the probability distributions of these phenomena exist in the natural world. The implication is that our perceptions for all these phenomena are tuned to those statistical probabilities that exist within the environment.

Howe and Purves' theory could be placed with other theories that attempt to explain human cognition as adapting to the environment, such as Gibson's ecological approach (1966) and Marr's computational approach (1982). Anderson constructed his latest iteration of his ACT theory around such a rational analysis, as he termed it (1993; Anderson & Lebiere, 1998). In his work, Anderson and his colleagues have shown how various memory, categorization, and problem solving phenomena can be explained by understanding how the mind adapts itself to the statistical regularities of the environment (1991). Developing a perceptual model within such a system would be an interesting undertaking (the ACT theory has been augmented with perceptual capabilities, Lebiere et al., 2004). Perhaps our perception of the Müller-Lyer illusion, as well as that of visual space in general, is one more phenomenon that can at least partially be explained via such a mechanism.

Acknowledgments

The second author received an Undergraduate Honors Research Fellowship from the University of Tampa to assist in this project.

References

Anderson, J.R., (1991). Is human cognition adaptive? *Behavioral and Brain Sciences*, *14*, 471-517.

Anderson, J.R., (1993). *Rules of the Mind*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Anderson, J.R., Lebiere, C. (1998). *Atomic Components of Thought*. Mahway, NJ: Lawrence Erlbaum Associates.

Bolles, R. C. (1969). The role of eye movements in the Müller-Lyer illusion. *Perception & Psychophysics*, *6*, 175-176.

Bross, M., Blair, R., & Longtin, P. (1978). Assimilation theory, attentive fields, and the Müller-Lyer illusion. *Perception*, *7*, 297-304.

Bulatov, A., Bertulis, A., & Mickiene, L. (1997). Geometrical illusions: Study and modeling. *Biological Cybernetics*, *77*, 395-406.

Carr, H. A. (1935). *An introduction to space perception*. New York: Longmans, Gree, & Co.

Coren, S., & Girgus, J. S. (1974). Transfer of illusion decrement as a function of perceived similarity. *Journal of Experimental Psychology*, 102, 881-887.

Crawford, L. E., Huttenlocher, J., & Engebretson, P.H. (2000). Category effects on estimates of stimuli: Perception or reconstruction? *Psychological Science*, *11*, 280-284.

Gibson, J.J. (1966). *The senses considered as perceptual systems*. Boston: Houghton Mifflin.

Gregory, R.L. (1963). Distortion of visual space as inappropriate constancy scaling. *Nature*, *199*, 678-680.

Howe, C. Q., & Purves, D. (2005a). The Müller-Lyer illusion explained by the statistics of image-source relationships. *Proceedings of the National Academy of Sciences*, *102*, 1234-1239.

Howe, C. Q., & Purves, D. (2005b). *Perceiving geometry: Geometrical illusions explained by natural scene statistics*. New York: Springer.

Lebiere, C., Byrne, M.D., Anderson, J.R., Qin, Y., Bothell, D., & Douglass, S.A. (2004). An integrated theory of the mind. *Psychological Review*, *111*, 1036-1060.

Marr, D., (1982). *Vision*, San Francisco: W.H. Freeman.

McGraw, K.O., & Stanford, J. (2001). The apparent distance of interior and exterior corners: A test of Gregory's misapplied size constancy explanation for the Müller-Lyer illusion. *The Journal of General Psychology*, *121*, 19-26.

Pressey, A. W. (1971). An extension of assimilation theory to illusions of size, area and direction. *Perception & Psychophysics*, *9*, 172-176.