

Perceptual scale for transparency: Common fate overrides geometrical and color cues

Zhehao Huang

Graduate Center for Vision Research,
State University of New York, College of Optometry,
New York, New York, USA



Qasim Zaidi

Graduate Center for Vision Research,
State University of New York, College of Optometry,
New York, New York, USA



Objects that pass light through are considered transparent, and we generally expect that the light coming out will match the color of the object. However, when the object is placed on a colored surface, the light coming back to our eyes becomes a composite of surface, illumination, and transparency properties. Despite that, we can often perceive separate overlaid and overlaying layers differing in colors. How neurons separate the information to extract the transparent layer remains unknown, but the physical characteristics of transparent filters generate geometrical and color features in retinal images, which could provide cues for separating layers. We estimated the relative importance of such cues in a perceptual scale for transparency, using stimuli in which X- or T-junctions, different relative motions, and consistent or inconsistent colors cooperated or competed in forced-preference psychophysics experiments. Maximum-likelihood Thurstone scaling revealed that motion increased transparency for X-junctions, but decreased transparency for T-junctions by creating the percept of an opaque patch. However, if the motion of a filter uncovered a dynamically changing but stationary pattern, sharing a common fate with the surround but forming T-junctions, the probability of seeing transparency was almost as high as for moving X-junctions, despite the stimulus being physically improbable. In addition, geometric cues overrode color inconsistency to a great degree. Finally, a linear model of transparency perception as a function of relative motions between filter, overlay, and surround layers, contour continuation, and color consistency, quantified a hierarchy of latent influences on when the filter is seen as a separate transparent layer.

reflected onto a second surface matches the perceived color and/or shape of the first object. Objects that pass light through are considered transparent, and the modification of light is more obvious. We expect that the modified light coming out will match the perceived color of the transparent object, but the situation is more complicated if the object is placed on a colored opaque surface so that the light coming back to an observer is the result of modifications by both the transparent object and the opaque surface. Perceptual scission (Heider & Koffka, 1933; Metelli, 1974) occurs if the shapes and colors of the underlying surface are seen as separate from the shape and color of the overlaying layer. The ability to disentangle the color of the surface from the color of the medium is essential to the success of vision. Among other functions, it enables us to judge the color of transparent objects (Ennis & Doerschner, 2021) and infer surface colors behind fog (D'Zmura et al., 2000). In conditions where a transparent layer is lying on the top of the background surface, the spectral distribution and intensity of light coming from each point of the overlaid image is a composite of illuminant spectrum passed through the spectrum of an intervening medium, reflected from the surface spectrum, and passed again through the spectrum of the intervening medium and does not in itself contain separable information about the characteristics of the components. However, the physical characteristics of transparency also create geometrical and color image features that evoke scission, and sometimes there is relative motion from movement of a transparent liquid or vapor, or from changes of viewpoint if the transparent layer has a volume or is in front of the surface. In this article, we quantify the relative importance of these features in giving the impression of transparency.

The continuation of contours from exposed to overlaid regions creates X-junctions in the image

Introduction

All objects modify the light that strikes them, but we only become perceptually aware of that when the light

Citation: Huang, Z., & Zaidi, Q. (2022). Perceptual scale for transparency: Common fate overrides geometrical and color cues. *Journal of Vision*, 22(6):6, 1–15, <https://doi.org/10.1167/jov.22.6.6>.



(Cavanagh, 1987; Kanizsa, 1979; Kersten, 1991; Metelli, 1974). X-junctions trigger transparency perception when there are multiplicative changes in contrast (Adelson & Anandan, 1990; Anderson, 1997; Beck et al., 1984; Metelli, 1974; Robilotto et al., 2002; Robilotto & Zaidi, 2004), and in achromatic stimuli, observers match the degree of transmittance to perceived contrast (Robilotto et al., 2002; Robilotto & Zaidi, 2004; Singh & Anderson, 2002, 2006). An opaque patch lying on a background creates T-junctions at the edge, which are generally considered to be cues for occlusion between opaque objects, but can elicit the impression of an illusory transparent layer under conditions of illusory modal contours (Kanizsa, 1979). In addition, X-junctions are not necessary for volumetric transparency (Fleming et al., 2011) or if there is a continuation of pattern from exposed to overlaid regions (Fuchs, 1923). We test the efficacy of T-junctions versus X-junctions when other cues are also present.

Relative motion has also been linked to transparency. Informal observations suggest that motion seems to enhance the impression of transparency in the presence of X-junctions, and enhances the perception of opacity in the presence of T-junctions (Khang & Zaidi, 2004), similar to its role in distinguishing reflections from paint (Doerschner et al., 2011), but this role has not been critically tested. In addition, color change and apparent motion can create an illusory transparent layer (Cicerone et al., 1995). Further, dynamic image deformation can lead observers to report transparent water or vapor flowing above the background (Kawabe & Nishida, 2018), and transparency is actually enhanced by the presence of T-junctions at the edge of the surface (Kawabe & Nishida, 2017). Motion is thus likely to be a complex cue for transparency, and we test its effect on the perceptual separation of layers.

When observers look at an overlaid surface through a transparent layer, the spectra of lights from an exposed background differ from lights from the portion overlaid by a filter by a double pass through the transmission spectrum of the filter, but, remarkably, when lights from exposed surfaces are absorbed in L, M, and S cone-photopigments, and plotted against lights from identical surfaces overlaid by a filter, the change can generally be defined by a multiplicative constant for each cone class. Hence, changes in spectra of lights transmitted through filters form a three-dimensional diagonal transform in cone space or an affine transform in cone-opponent space (Khang & Zaidi, 2002a, 2002b; Westland & Ripamonti, 2000; Zaidi, 1998, 2001). This transform provides a strong cue to the color of the filter, showing that there is sufficient information to estimate the color (not the spectrum) of the filter and suggesting that veridical perception of the filter color could be used as a measure of the degree of scission.

To test this suggestion, Khang and Zaidi (2004) estimated how accurately human observers separate the image into overlay and background components by placing moving transparent layers on chromatically different sets of background materials. To measure whether observers could tell whether the two filters are identical despite the local colors of the two overlaid regions being different, observers adjusted the color of the filter on a gray-level background to match the filter on a colored background. For six different colored filters, placed on six different colored backgrounds, the matched chromaticity was close to the actual chromaticity of the test filter against a gray background and differed significantly from the average chromaticity of the overlaid segment, providing ostensibly strong evidence for color scission. The same stimulus, with the surround changed to black, gave an impression of a spotlight on a dark surface, not a transparent filter; in this case, when observers matched the spotlight by a spotlight on a gray-level background, the light had the average chromaticity of the overlaid segment, which differed significantly from the actual chromaticity of the spotlight filter. These results suggest that observers can accurately scission colors of transparent layers in geometric configurations that support the perception of filters, but not if they support the perception of spotlights against black surrounds. However, the scission interpretation is limited by the fact that the color of the filter creates color contrast from exposed to overlaid areas, with an increase toward the filter color on the overlaid area, irrespective of the background colors, and observers could be matching the change in color with or without a transparency percept; thus, veridical filter color matching would not be a direct estimate of the degree of color scission.

To counter this limitation, we used the transparency cues we have described in cooperation or conflict to evoke gradations of perceived transparency, and asked observers to judge the probability of physical transparency in Likert-type paired comparisons (Spicker et al., 2017). From these comparisons, we estimated a perceptual transparency scale for human observers using a maximum likelihood variant of classical Thurstone scaling (1927).

Methods

Stimuli

Materials

A $46^\circ \times 26^\circ$ (1920×1080 pixels) monitor screen was covered with a variegated background of randomly oriented elliptical patches centered on randomly chosen pixels on the monitor, with long axes randomly ranging from 0.85° to 1.56° , and short axes randomly

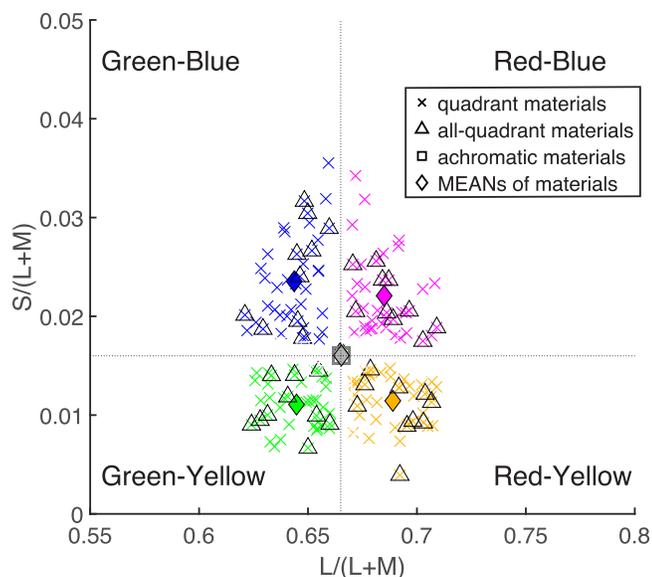


Figure 1. MacLeod–Boynton chromaticity of 280 reflectances of natural and man-made materials under equal energy light, used to simulate background surfaces. Background surfaces consisted of four sets of 40 materials each from the single quadrants of the color space designated with approximate color names for convenience, whereas the fifth set was equally balanced across all four quadrants, and the sixth set consisted of 40 achromatic materials with a similar luminance distribution.

ranging from 0.35° to 0.65° . The total number of ellipses per image was 8192, which we had found previously provided complete coverage in every case, so no iterations were needed. Ten different sets of ellipses were generated and one set was chosen randomly on each trial. To simulate the colors of background surfaces, we used 280 reflectance spectra materials chosen from measurements of natural and man-made objects that were previously used by [Khang and Zaidi \(2004\)](#) and [Smithson and Zaidi \(2004\)](#). From the MacLeod–Boynton chromaticity plot of all 280 materials under equal energy light ([Figure 1](#)), we selected 4 sets of 40 materials each from single quadrants of the color space, while the fifth set was equally balanced across all four quadrants, and the sixth set consisted of 40 achromatic materials. On each trial, reflectances from one set were randomly assigned to ellipses, and a central disk covered by one of six Kodak CC30 color filters ([Khang & Zaidi, 2004](#)), randomly chosen on each trial. The six filters are shown on the six background surfaces in [Figure 2](#).

Geometric configurations

We constructed geometric and motion configurations that required the transparent filter layer, the overlaid background surface layer, and the exposed surround surface layer to be oriented and moved independently, so each stimulus was constructed from the three simulated

layers (illustrated later in this article). By separately manipulating the three layers, we simulated seven geometric configurations that combined geometric, motion cues to transparency, while maintaining colors in the overlaid and exposed regions that were physically consistent with transparency (Videos in [Figure 3](#) top).

(i) **Static X-junctions:** stationary circular filter on the background. The construction in terms of the three layers is illustrated in [Figure 4A](#), where the overlaid surface is just cut out from the exposed surround, so replacing it yields a surface with continuous ellipses. (ii) **Moving X-junctions:** the simulated filter moving back and forth horizontally over the surface. Both configurations contain X-junctions on the boundaries of the disks as realistic cues for transparency. By asking observers to compare the first two configurations, we can quantify if motion enhances the perception of transparency. (iii) **Static T-junctions:** the circular overlaid region is the mirror-reversed version of the static X-junction case, thus creating T-junctions at the boundary with the surround ([Figure 4B](#)) and allowing a comparison of static X-junctions versus T-junctions in transparency perception. (iv) **Moving T-junctions:** the simulated filter and overlaid background move together as if one layer, thus creating moving T-junctions on the boundary. By asking observers to compare the static and moving T-junction conditions, we can quantify if motion also enhances the perception of opacity. [Figure 5](#) (left) presents a comparison of Moving X and Moving T configurations and their separation into the three layers. Note that the overlaid surface changes on each frame of the Moving X configuration as if cut out from the background surface on each frame, which is physically equivalent to the filter moving over a continuous stationary surface. The overlaid surface in the Moving T configuration moves with the filter, but is unchanging, which is physically equivalent to an opaque patch moving over a surface. (v) **Dynamic T-junctions:** the moving circular region on every frame is the mirror-reversed version of the moving X-junction condition, so the filtered region moves as if it is covering a stationary but changing overlaid surface on each frame, thus creating dynamically changing T-junctions on the moving boundary as opposed to X-junctions. In this critical condition, the moving filter seems to uncover new areas of the background on every frame, but these are different from what was on the exposed surround in the same location. The main motivation for this condition is that the overlaid and surround layers share a motion-defined common fate, not shared with the moving filter, thus this condition can reveal whether this common fate overrides T-junctions in transparency perception when compared with Moving T-junctions where the filter and overlaid layers share a common fate. [Figure 5](#) (right) presents a comparison of Moving X and Dynamic T configurations and their separation into the three layers. On each frame, the overlaid surface in the Dynamic T configuration is

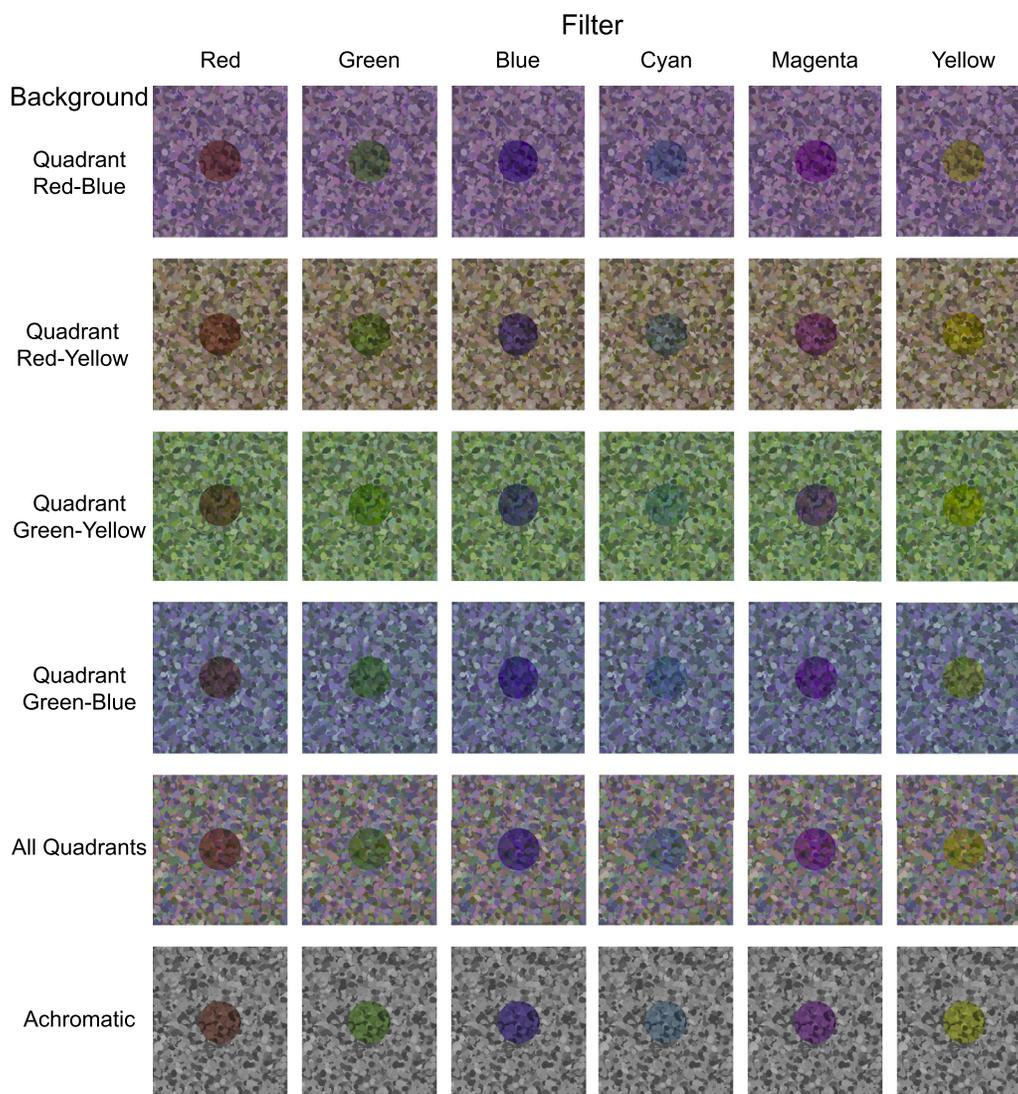


Figure 2. Six Kodak CC30 color filters (red, green, blue, cyan, magenta, and yellow) placed on six backgrounds: Quadrant Red–Blue, Quadrant Red–Yellow, Quadrant Green–Yellow, Quadrant Green–Blue, All Quadrants, and Achromatic.

the mirror-reversed version of the overlaid surface in the Moving X configuration, so it changes on each frame unlike the Moving T configuration. (vi) Relative Motion: the filter moves as in the Moving X-junction condition, but the overlaid background moves at one-half the speed in the same direction. This condition also creates Dynamic T-junctions, but no pair of layers share a common fate. (vii) Overlaid Motion: the filter and surround are stationary, but the overlaid background moves, as if the disk were an aperture. In this condition, T-junctions are created at the filter’s edge, but the overlaid layer does not share a common fate with the filter or surround layers, which however do so. This study aimed to estimate a scale for the degree of perceptual transparency of the filter layer across these seven combinations of geometric and motion cues.

Physically inconsistent color configurations

In the configurations shown in the top row of Figure 3, the ellipses in the surround and overlaid circle were chosen from the same color set, so we call these conditions Color Consistent. For each configuration, we also created physically Color Inconsistent conditions by replacing the colors of the surrounding ellipses with achromatic shades (Figures 4C, D), because no filter with transmittance constant over time (no matter how complex spatially) could turn the achromatic ellipses into the dynamically varied colors seen under the filters (especially the Moving X-junctions and Dynamic T-junctions videos in Figure 3, bottom). It is true that the same achromatic color can be an addition of many different combinations of wavelengths, so different patches of the same achromatic color could possibly be seen as different colors when seen through

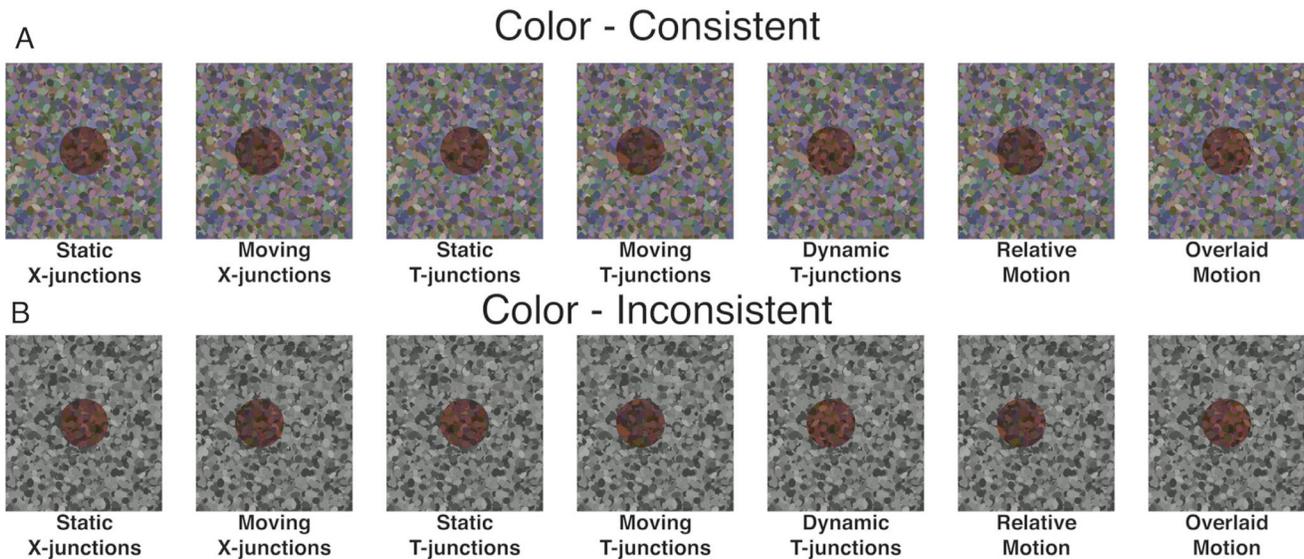


Figure 3. Video: (Top) red filter on the All Quadrants background set with seven geometric configurations. (Bottom) red filter on the All Quadrants set, but with an Achromatic surround, for physically inconsistent color configurations. The stimuli were presented on a 24-inch screen with the disks subtending 7° of visual angle, so the videos should be expanded to roughly the same size to replicate observers' percepts. Video is available on the journal website.

one color filter, but it would be physically impossible to obtain the pairs of almost complementary colors in the All Quadrants set by this process. The set of achromatic ellipses maintained the average luminance, but each individual color of the surround was replaced by a randomly selected luminance so that the luminance relations across the filter border were also not physically compatible with transparency. Consistent and inconsistent color configurations were compared across all geometric configurations to test whether color inconsistency vetoes perceived transparency or whether other cues can overcome it.

Experimental procedure

For each trial, an observer was presented with one stimulus each on the left and right halves of the screen with two different configurations randomly chosen from the 7 (geometric + motion) \times 2 (color consistency) = 14 configurations. Thus, there were a total of $(14 \times 13)/2 = 91$ distinct pairs. For each forced-preference paired comparison, one overlaid layer color set and one filter color were randomly chosen for both sides, that is, the same filter and overlaid layer colors on both sides, so each pair was repeated 72 times for each observer (6 filter colors \times 6 overlaid color sets \times two left–right permutations). In the color consistent conditions, the surround colors were from the same set as the overlaid layer, whereas in the color inconsistent conditions the surround was achromatic.

The videos in the top row of Figure 5 show two example trials. The left column shows Moving X-junctions on the left and Moving T-junctions on the

right, where the difference in perceived transparency is large, and the right column shows Moving X-junctions on the left and Dynamic T-junctions on the right, where the difference in perceived transparency is small despite there being T-junctions in the right stimulus and X-junctions in the left. For each session, the observer was given these instructions, “On each trial of the experiment, a scene of oriented ellipses will be displayed on the screen. There will be two disks, moving or not, on the left half and right half of the screen. Some of the disks will be transparent and some will be not. Sometimes it will be easy to tell and sometimes not. Your task is to look at both disks and decide which has a higher probability of being a transparent layer.” Observers were instructed to use buttons to report the judgment using a 5-point Likert scale: “Left disk has much higher probability,” “Left disk has slightly higher probability,” “Left and right disks have equal probability,” “Right disk has slightly higher probability,” and “Right disk has much higher probability.”

Observers

Five observers participated in the experiment. All of them had normal or corrected visual acuity and normal color vision. One of the observers is the first author of this article and was aware of the nature and the purpose of the experiment; the other observers were briefed about the experiment after the data had been collected. Observers gave informed consent, and the procedures were approved by the SUNY Optometry

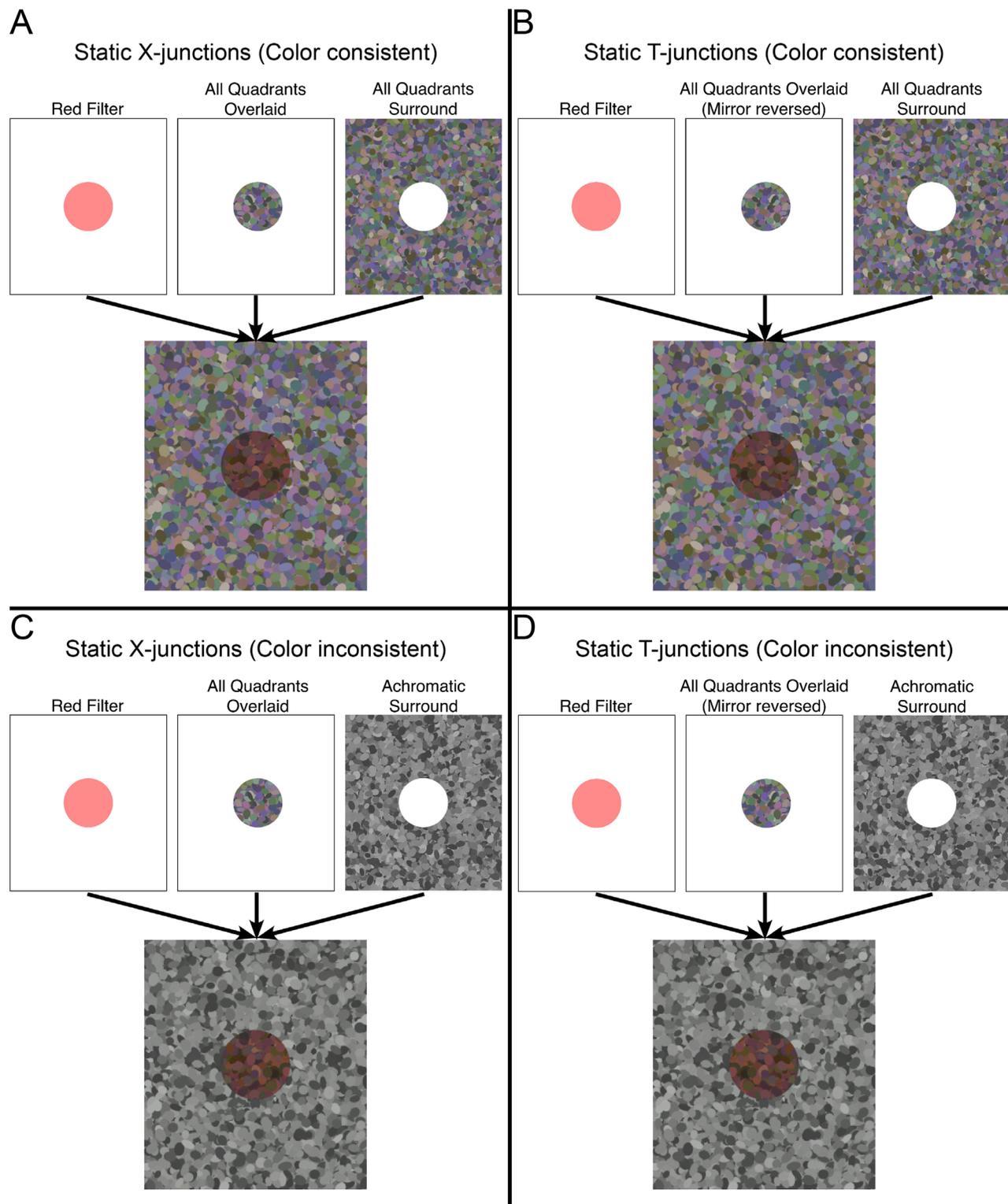


Figure 4. Stimuli were constructed by combining a transparent filter layer, an overlaid surface layer, and an exposed surround layer. (A) Construction of a Static X-junctions geometric configuration. The overlaid layer is cut out from the background, thus creating a continuous surface when replaced, and forming X-junctions on the boundary of the circular region when the filter is placed on the overlaid section. (B) Construction of Static T-junctions geometric configuration is similar to the Static X-junctions, except that the overlaid layer is mirror-reversed, so when it is placed inside the exposed background, and the filter laid on it, T-junctions form on the filter boundary. (C) When the exposed surface in the Static X-junctions is replaced by an achromatic background, while the overlaid layer remains the All Quadrants background, a color inconsistent condition is created. (D) A color inconsistent condition for the Static T-junctions configuration is constructed similarly.

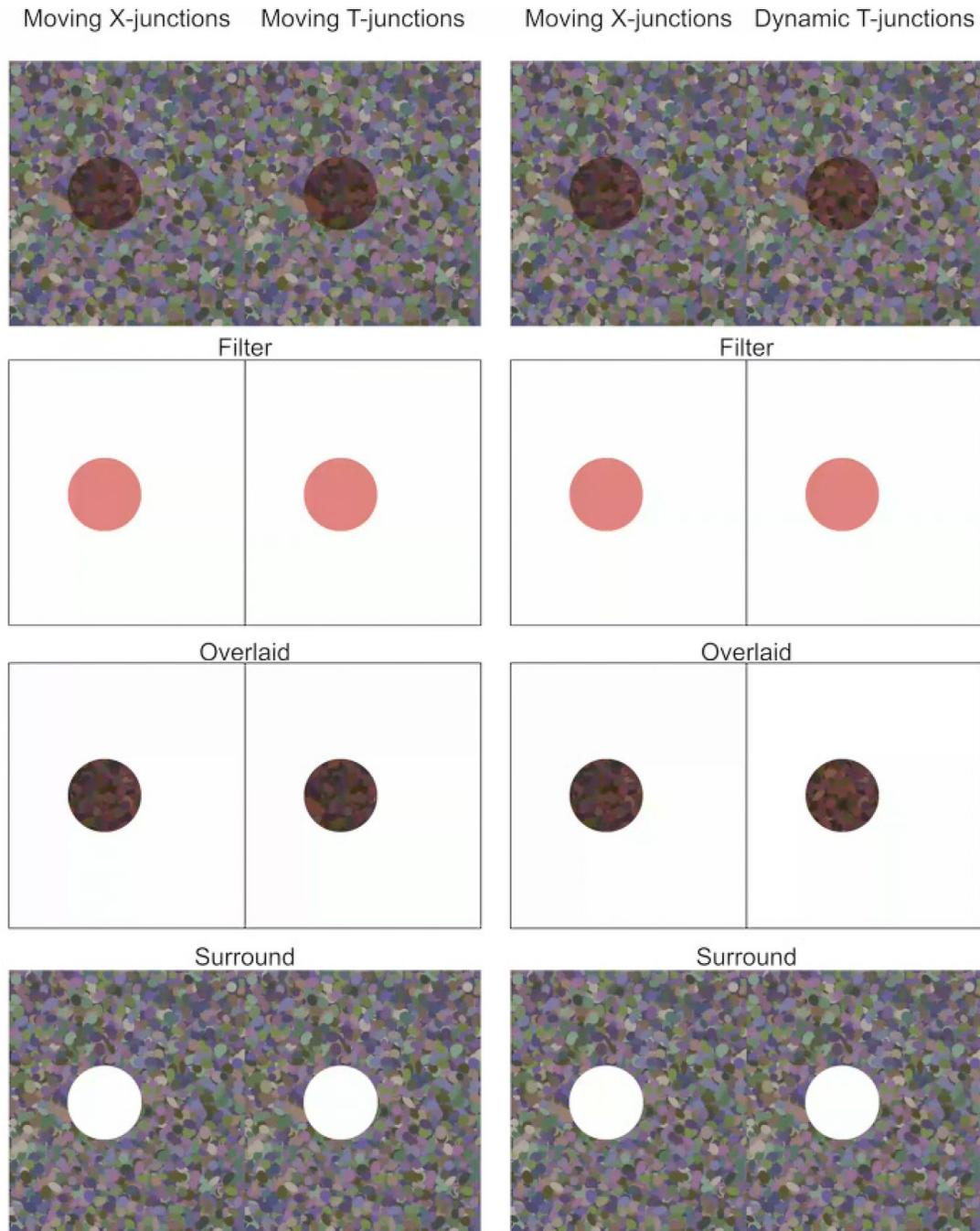


Figure 5. Video. Two example trials. (Left) Moving X-junctions on the left and Moving T-junctions on the right. (Right) Moving X-junctions on the left and Dynamic T-junctions on the right. The stimuli were presented on a 24-inch screen with the disks subtending 7° of visual angle, so the videos should be expanded to roughly the same size to replicate observers' percepts. Video is available on the journal website.

Institutional Review Board Committee in accordance with the Declaration of Helsinki.

Apparatus

Stimuli were shown on a VPixx LED monitor using MATLAB and Psychtoolbox (Brainard, 1997; Kleiner et al., 2007). The observers sat at a viewing distance of 63 cm. Stimuli were displayed at 16 bits per channel,

with a linear gamma. To compute screen RGB correctly for rendering light spectra, spectral distributions of the monitor primaries were measured with a SpectraScan PR650 photo-spectroradiometer.

Thurstone scaling

To estimate a perceptual scale for transparency we used a maximum likelihood variant of classical

Thurstone scaling to analyze the paired forced-preference results. In standard [Thurstone \(1927\)](#) scaling, the results from paired comparisons between m stimuli are stored in an $m \times m$ matrix C , where C_{ij} is the number of times that i is preferred over j , and C_{ji} is the number of times that j is preferred over i , so that $C_{ij} + C_{ji} = n$ the number of repeated choices. [Thurstone \(1927\)](#) assumed that the subjective value of each stimulus is a Gaussian random variable S_i with mean μ_i and variance σ_i^2 , so the task is to estimate μ . Forced preference between stimulus i and j , can be modeled as judging whether $S_i > S_j$, or equivalently $S_i - S_j > 0$. Based on the Gaussian assumption, $S_i - S_j$ is also a Gaussian with mean $\mu_i - \mu_j$ and variance $\sigma_i^2 + \sigma_j^2$, giving the probability of i being preferred over j :

$$P(S_i > S_j) = P(S_i - S_j > 0) = \Phi\left(\frac{\mu_i - \mu_j}{\sigma_i^2 + \sigma_j^2}\right), \quad (1)$$

where Φ is the Gaussian cumulative distribution. From the empirical results, $P(S_i > S_j)$ can be estimated by the proportion of times that i is preferred over j :

$$P(S_i > S_j) \approx \frac{C_{ij}}{C_{ij} + C_{ji}} \quad (2)$$

To simplify the model, all the subjective values are assumed to have variance equal to $\frac{1}{\sqrt{2}}$; therefore, $\sigma_i^2 + \sigma_j^2 = 1$, so $\mu_i - \mu_j$ can be estimated by:

$$\mu_i - \mu_j = \Phi^{-1}\left(\frac{C_{ij}}{C_{ij} + C_{ji}}\right), \quad (3)$$

If μ is the vector of all scale value $\mu = [\mu_1, \dots, \mu_m]$, then the log-likelihood of μ given C ([Tsukida & Gupta, 2011](#)):

$$L(\mu|C) \triangleq \log P(C|\mu) = \sum_{i,j} C_{ij} \log \Phi(\mu_i - \mu_j) \quad (4)$$

And the scale values can be estimated as:

$$\arg \max_{\mu} \sum_{i,j} C_{ij} \log \Phi(\mu_i - \mu_j), \text{ subject to } \sum \mu_i = 0 \quad (5)$$

In pilot experiments, we realized that a two-alternative judgment was not providing sufficient nuance, so we used a 5-point Likert scale, adding strong and weak preferences and a neutral option. We used the [Spicker et al. \(2017\)](#) extension to Thurstone scaling. S and W matrices record the number of times that one stimulus is strongly or weakly preferred over

another, and N records the number of times that two stimuli are judged equal in value. If δ_0 is the subjective boundary between neutral and weakly preferred, and δ_1 be the subjective boundary between weak preferred and strongly preferred, the boundaries between adjacent options on the 5-point scale will be $[-\delta_1, -\delta_0, \delta_0, \delta_1]$. Then the log likelihood of scale values μ conditional on S , W and N is given by:

$$L(\mu|S, W, N) \triangleq \log P(S, W, N|\mu), \quad (6)$$

where

$$\begin{aligned} \log P(S, W, N|\mu) &= \sum_{i,j} S_{ij} \log(1 - \Phi(\mu_i - \mu_j - \delta_1)) \\ &\quad + \sum_{i,j} W_{ij} \log(\Phi(\mu_i - \mu_j - \delta_1) - \Phi(\mu_i - \mu_j - \delta_0)) \\ &\quad + \sum_{i,j} N_{ij} \log(\Phi(\mu_i - \mu_j - \delta_0) - \Phi(\mu_i - \mu_j + \delta_0)) \\ &\quad + \sum_{i,j} W_{ji} \log(\Phi(\mu_i - \mu_j + \delta_0) - \Phi(\mu_i - \mu_j + \delta_1)) \\ &\quad + \sum_{i,j} S_{ji} \log(\Phi(\mu_i - \mu_j + \delta_1)) \end{aligned} \quad (7)$$

The best estimate of μ is obtained by using convex optimization to solve:

$$\arg \max_{\mu, \delta_0, \delta_1} L(\mu|S, W, N), \text{ subject to } \sum \mu_i = 0, \delta_1 > \delta_0 > 0 \quad (8)$$

The results of the optimization form an interval scale, but not a ratio scale, so rank and differences are meaningful, but zero is arbitrary. In our case, it was set by assuming that $\sum \mu_i = 0$ in [Equation 8](#). The estimated mean of each configuration μ was taken as an estimate of the subjective scale value of transparency for each observer.

Results

In the scaling analysis, we pooled preferences from the 72 combinations (36 filter and overlaid color pairs times 2 left–right permutations) for each of the 14 configurations, thus obtaining a scale that was based on a wide range of similarities between the filter and overlaid layer colors (S , W , and N matrices for all observers are presented in [Appendix Figure A1](#)). The analysis estimated the subjective scale value of transparency of the 14 configurations for each of the 5 observers. The scale values for the Combined observer were estimated by pooling all 5 observer’s preferences before the analysis, thus giving 5 times the number of choices as each individual observer, hence more reliable probability estimates. As [Figure 6](#) shows there is remarkable concordance between observers.

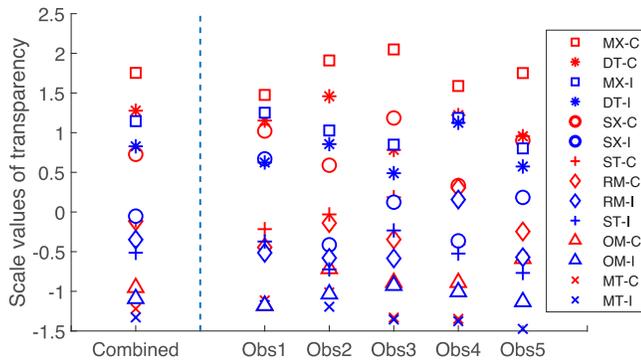


Figure 6. Transparency scale value for the combined observer and the five individual observers. Labels: MX = Moving X-junctions, DT = Dynamic T-junctions, SX = Static X-junctions, ST = Static T-junctions, RM = Relative Motion, OM = Overlaid Motion, MT = Moving T-junctions, C = Consistent (Red), I = Inconsistent (Blue).

Moving X-junctions, Static X-junctions, and Dynamic T-junctions values are positive; Relative Motion and Static T-junction values are approximately 0.0; Overlaid Motion and Moving T-junctions values are below negative, corresponding with impressions of transparency for positive values and opacity for negative values. Since the Thurstone scale is an interval scale, we used Pearson’s correlation coefficient to examine the correlation of observers’ scale value. The pairwise Pearson’s correlations between observers range from 0.87 to 0.95 (Table 1), and the correlations with the Combined observer range from 0.96 to 0.98; therefore, we discuss the results of the Combined observer in detail.

The results show that disk movement that continually generates X-junctions looks more transparent than its Static X-junction counterpart, in fact, Moving X-junctions even overcome physically impossible color inconsistency to have the third highest transparency value for the Combined observer. Dynamic T-junctions had the second highest scale value after Moving X-junctions for the Combined observer and was rated much more transparent than Relative Motion, even

though both configurations continuously generate new T-junctions. This result indicates that common fate could prevail over junctions in promoting a percept of transparency. Moving T-junctions configuration was consistently the least transparent configuration, showing that common fate between the filter and overlaid segment enhanced the effects of T-junctions for evoking opacity. It was interesting that the motion of just the overlaid layer inside an aperture did not evoke the impression of a colored transparent layer, possibly because observers added the color of the filter to the background. All Inconsistent Color conditions appeared to be a little less transparent than their color-consistent counterparts, but geometric configurations could overcome color inconsistency in evoking transparency, despite it being physically impossible for a transparent filter to create variegated colors from all four quadrants of color space when placed on an achromatic background. Although the effects of many of these cues have been examined in isolation, the cooperation and conflict of these cues compared together reveal a hierarchy of transparency enhancers and inhibitors, with effects that are remarkably similar across observers.

Latent factors

The perceptual scale is a function of stimulus configurations, and this would be useful in many applications, but we wanted to take advantage of the concordance across observers to identify the latent factors that influence transparency perception. Phenomenologically, in the simplest case, transparency perception requires seeing the transparent layer as separate from the overlaid surface, and that could be made easier by seeing surround and overlaid surfaces as connected. The connection between overlaid and surround surfaces is enhanced by contour continuation in X-junctions, so the presence or absence of contour continuation is likely to influence transparency perception. Transparency layer separation

	Obs 1	Obs 2	Obs 3	Obs 4	Obs 5	Combined
Obs 1	1					
Obs 2	0.9135*	1				
Obs 3	0.9464*	0.9313*	1			
Obs 4	0.8810*	0.9543*	0.8730*	1		
Obs 5	0.9339*	0.9525*	0.9521*	0.8725*	1	
Combined	0.9609*	0.9829*	0.9580*	0.9644*	0.9550*	1

Table 1. Pearson’s correlation coefficient between observers’ scale values. Obs = observer.

* $P < 0.001$.

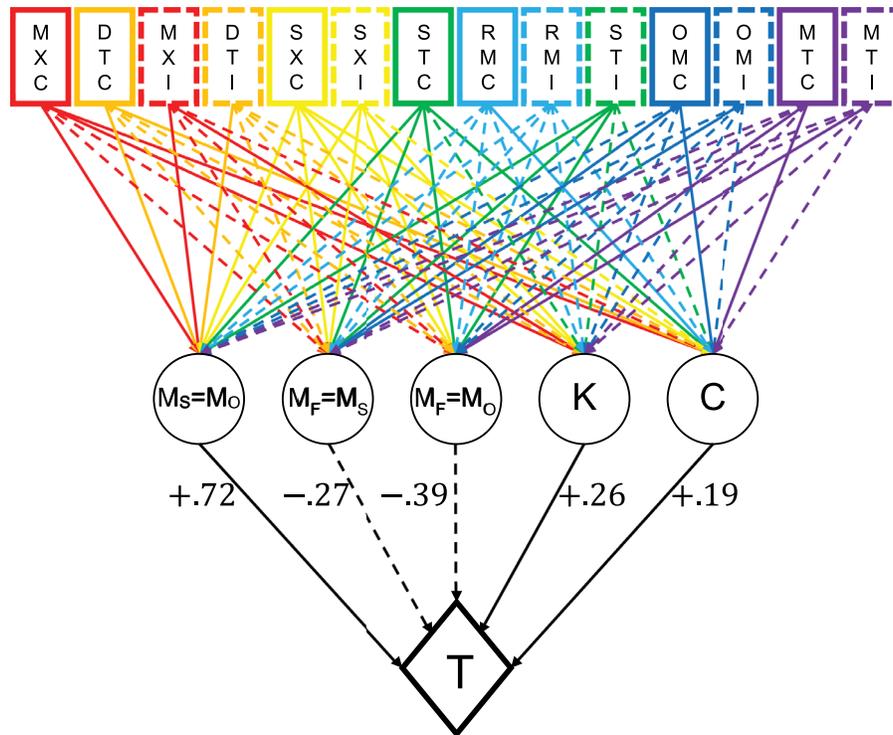


Figure 7. Latent factor perceptual transparency model. The top layer contains the experimental conditions in rectangles. The middle layer contains the latent factors in circles, and the color-coded solid and dashed arrows mean the latent variable is True (solid) or False (dashed) for that condition. The last layer is the transparency scale in the diamond, and the black solid (Positive correlation) and dashed (Negative correlation) arrows are labeled with the best fitting parameter values.

and connection between overlaid and surround surfaces are both enhanced by systematic luminance and color changes from exposed to overlaid regions that simulate physical reality, as in the color consistent conditions. However, achromatic surrounds in color inconsistent conditions simulate physical impossibility but do not veto transparency percepts, so we wanted to find out what weight observers give to the inconsistency relative to the other factors. Since all forced choices were between pairs with the same filter and overlaid layer colors, differences between stimulus color combinations do not affect any choice, so are not reflected in the perceptual scale. Since motion-created common fate is a strong grouping enhancer (Wagemans et al., 2012), the lack of relative motion between the surround and overlaid surfaces in conjunction with the motion of the transparent layer with respect to the two surfaces would enhance the separation of the transparent layer. We, thus, took contour continuation, color consistency, and relative motion as the latent factors underlying the perceptual transparency choices that generate the quantitative perceptual scale for our stimuli. We used a regression model to quantify the relative weights of these factors, without committing to a particular neural or behavioral decision process.

In the latent factor stage of the model (Figure 7), the pairwise relative motion was represented by three latent factors: $M_S = M_O$, when the overlaid surface moved with the same velocity (possibly zero) as the surround surface, $M_F = M_S$ when the filter layer moved with the same velocity as the surround surface, and $M_F = M_O$ when the filter layer moved with the same velocity as the overlaid surface. For each of the 14 experimental conditions, every latent factor was coded as a binary variable with +1 for True (solid arrows) and -1 for False (dashed arrows). The contour continuation factor K was +1 for X-junctions, and -1 for T-junctions, and the color consistency factor C was +1 for physically realistic consistency and -1 for physically impossible inconsistency. T represents the transparency scale in real numbers, and we fit T as an additive function of the latent factors (Equation 9) using the MATLAB function *fitlm*:

$$T = \alpha \mathbf{1} + L \cdot \mathbf{B} + \mathbf{E} \quad (9)$$

In Equation 9, T is a 14×1 array of empirical scale values, α is a constant parameter to be estimated, $\mathbf{1}$ is a 14×1 array of 1.0 s, and \mathbf{E} is a 14×1 array of errors ε . \mathbf{B} is a 5×1 vector of β parameters to be estimated, and L is the 14×5 matrix of the relation between

stimulus conditions and latent variables given by Equation 10:

$$L = \begin{matrix} & M_S = M_O & M_F = M_S & M_F = M_O & K & C \\ \begin{matrix} MX - C \\ DT - C \\ MX - I \\ DT - I \\ SX - C \\ SX - I \\ ST - C \\ RM - C \\ RM - I \\ ST - I \\ OM - C \\ OM - I \\ MT - C \\ MT - I \end{matrix} & \begin{bmatrix} +1 & -1 & -1 & +1 & +1 \\ +1 & -1 & -1 & -1 & +1 \\ +1 & -1 & -1 & +1 & -1 \\ +1 & -1 & -1 & -1 & -1 \\ +1 & +1 & +1 & +1 & +1 \\ +1 & +1 & +1 & +1 & -1 \\ +1 & +1 & +1 & -1 & +1 \\ -1 & -1 & -1 & -1 & +1 \\ -1 & -1 & -1 & -1 & -1 \\ +1 & +1 & +1 & -1 & -1 \\ -1 & +1 & -1 & -1 & +1 \\ -1 & +1 & -1 & -1 & -1 \\ -1 & +1 & +1 & -1 & +1 \\ -1 & -1 & +1 & -1 & +1 \\ -1 & -1 & +1 & -1 & -1 \end{bmatrix} \end{matrix} \quad (10)$$

The regression model fits so well with just additive binary latent factors ($R^2 = 0.976$, $F = 63.98$, $P < 0.001$), that making the factors continuous or adding interaction terms between them could not make the fit significantly better. The predicted transparent scale estimated from the best fitting model parameters is plotted against the Combined empirical scale in Figure 8, and the points lie close to the unit diagonal, consistent with the high correlation between the scales ($R^2 = 0.976$). Note that other sets of latent factors are unlikely to be sufficiently explanatory. For example, motion per se would not work because, while motion enhances the transparency effect of X-junctions, it also enhances the opacity effect of T-junctions. Similarly, even though pattern similarity between surround and overlaid surfaces works for Fuchs’s transparency (Fuchs, 1923; Masin, 1984, 1998, 1999), in our stimuli, the Relative Motion and Dynamic T-junctions conditions have the same pattern similarity, but evoke different probabilities of transparency.

The coefficients for the latent factors for the Combined observer are shown in Figure 7. The constant term in the regression equation is equal to

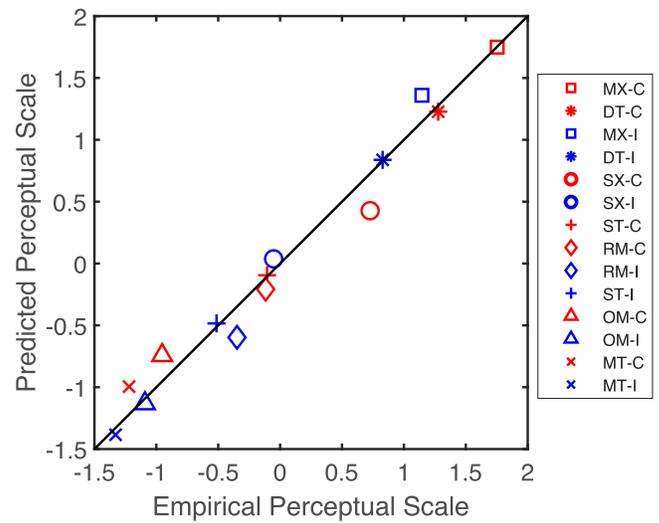


Figure 8. Perceptual scale predicted from the regression model versus the empirically derived perceptual scale for the Combined observer ($R^2 = 0.976$).

−0.09, so is much smaller than the other terms. The largest positive effect on transparency is from the lack of relative motion between surround and overlaid surfaces, while the next largest effects show that relative motion between the filter and surround or filter and overlaid surface enhances transparency, corresponding with the negative effect shown for a lack of relative motion. Together, the relative motion latent factors promote perceptually separating the filter layer from the two surfaces. All three motion effects are larger than the effect of contour continuation, while color consistency has the smallest effect, in fact, motion-defined common fate overcomes geometrical and color improbabilities and even impossibilities to create transparency percepts. Table 2 shows that the latent factors model predicts each observer’s perceptual scale extremely well (R^2 ranging from 0.939 to 0.978). It shows that different observers give different relative weights to the latent factors, but the signs are the same for all observers.

		Observer					
		Combined	Obs1	Obs2	Obs3	Obs4	Obs5
Factor	Constant	−0.085	−0.008	−0.120	0.006	−0.145	−0.004
	$M_S = M_O$	0.718	0.693	0.703	0.674	0.665	0.576
	$M_F = M_S$	−0.267	−0.240	−0.300	−0.094	−0.321	−0.189
	$M_F = M_O$	−0.394	−0.222	−0.415	−0.311	−0.528	−0.271
	K	0.261	0.404	0.194	0.373	0.166	0.337
	C	0.195	0.099	0.294	0.234	0.114	0.341
Fit	R^2	0.976	0.969	0.978	0.939	0.950	0.975

Table 2. Weights of the latent factors for all observers and R^2 measures of the fit of the model.

Discussion

The main contribution of this study is the perceptual scale for transparency estimated for cooperating or competing cues. The values in the perceptual scale in turn allowed us to infer weights of latent factors that control transparency perception. The neural locus and mechanisms of transparency perception are open questions (Qiu & von der Heydt, 2007). To understand the brain mechanisms that extract transparency from retinal images, one possible strategy would be to measure neuronal or voxel responses to the central disk for all 14 of our stimulus conditions and correlate these responses to the combined scale. A high correlation would indicate that the cell or brain area is segmenting the transparent layer from the background surface. One caveat is that the human results are combined over all combinations of filter and background colors, whereas for each cell, the stimuli may need to be restricted to the colors to which the cell responds reliably. Note that cells or brain areas that respond only to color contrast at the edge of the filter will respond just as strongly to the Moving T-junctions configuration as to the Moving X-junctions, so would not provide strong correlations with the perceptual scale where these conditions are at the opposite extremes. For a cell that shows evidence of responding well to perceptual transparency, a fit of the latent factors model to its responses will suggest how it combines geometric, motion, and color cues from earlier brain areas.

The perceptual scale shows that relative motion enhanced both transparency and opacity, depending on which layers were moving relative to each other, acting more like a potentiating agent than a cue, similar to its role in distinguishing reflections from paint (Doerschner et al., 2011). One possibility is that motion could be enhancing the effects of X-junctions and T-junctions by increasing the displayed number or their salience. Relative salience versus relative validity has been studied in associative learning and navigation (Kahnt et al., 2014; Leathers & Olson, 2012), but not yet in visual cue combination. Decreasing the sizes of ellipses in the displays would increase the number of X- or T-junctions, whereas increasing the sizes would increase the salience, but neither seems to have much effect on verbal ratings of transparency perception (Falkenberg & Faul, 2019). Instead, the Dynamic T-junctions condition, which is a novel contribution of this study, demonstrates that the main role of relative motion is increasing the likelihood of perceiving separate layers that do not share a common fate. In pointing out the role of certain shape changes in promoting impressions of transparency, particularly for shadows, Metzger (1936) mentions in a footnote that the law of common fate would add a contribution to transparency perception in a movie, but without presenting details

of the conditions or evidence, and it is highly unlikely that he was imagining a condition where common fate conflicts with other cues that are voting against transparency. The role we demonstrate for motion defined common fate explains motion enhancing effects in previous studies (D’Zmura et al., 2000; Falkenberg & Faul, 2019; Gerardin et al., 2006; Khang & Zaidi, 2002a), but goes beyond them in showing that common fate can override the information provided by junctions, which are otherwise quite powerful factors (Anderson, 1997). In fact, by replacing ellipses with rectangles in the overlaid surface (Figure 9), we demonstrate that, because of common fate, transparency can still be seen despite junction, color, and pattern information all being otherwise incompatible with transparency.

The high perceived transparency scale values for some color inconsistent stimuli, and the lowest weight for color inconsistency in the latent factors model for the Combined observer, point to another new result of this study, that color inconsistency did not veto transparency perception despite physical impossibility and violations of perceptually derived rules for luminance and color relations (Adelson & Anandan, 1990; Beck et al., 1984; D’Zmura et al., 2000; Metelli, 1974; Robiletto et al., 2002; Singh & Anderson, 2006).

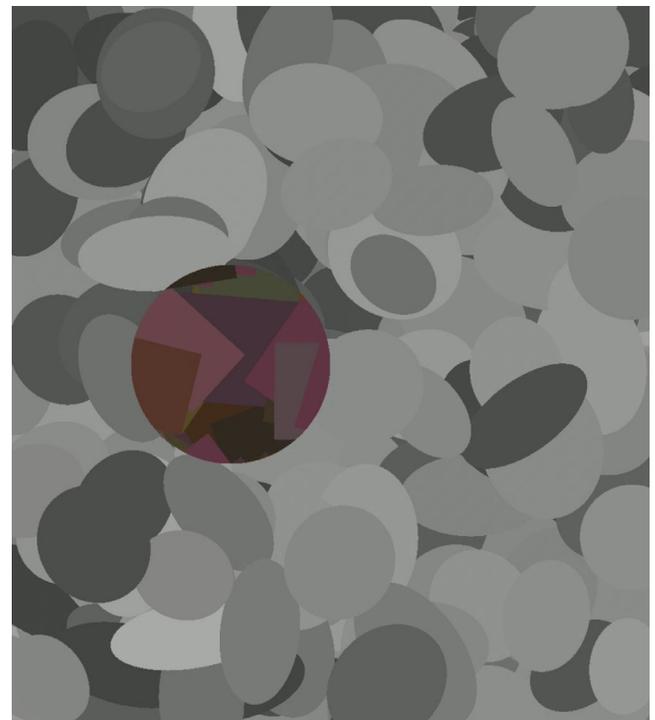


Figure 9. Motion-defined common fate overrides geometric, pattern and color incongruities in transparency perception. The overlaid pattern is composed of rectangles while the surround consists of ellipses in the Dynamic T-junction condition with color inconsistency. Video is available on the journal website.

The reasons why geometric and motion cues override color inconsistency would be interesting to explore.

Our latent factor model is extremely successful at explaining transparency-based choices with an additive regression model, but that does not mean that the decision process is restricted to a simple linear combination. Instead, as shown by Einhorn et al. (1979), the signs and weights in a regression equation could reflect the ambiguities that the organism faces regarding the substitutions and trade-offs between cues in a redundant environment, even if the choices arise from a much more complex process of cue search and attention which includes multiple hierarchical and conditional choice nodes. The regression weights could also reflect reliability of the cues, if the underlying process is optimal probabilistic cue combination that treats transparency estimates as noisy and uses a signal detection theoretic framework as in weak fusion models (Landy et al., 1995), but that would need to be investigated. The fact that certain latent factors can override physical impossibility in other factors is another new finding of this study. It may suggest temporal priority in processing that sometimes accumulates sufficient evidence for making a decision (Beck et al., 2008; Drugowitsch et al., 2012) before the lower ranking latent factors are considered, but that too remains to be tested.

Conclusion

A perceptual scale for transparency evoked by cooperation and competition between motion, geometry, and color cues, shows that relative motion-defined common fate leads to perceptual separation of transparency layers despite conflicting geometrical and color information, and that transparency can be seen despite color and luminance inconsistency if other cues dominate.

Keywords: transparency, perceptual scale, color scission, layers, relative motion, image junctions, latent factors

Acknowledgments

Supported by NIH grants EY007556 and EY013312 to QZ.

Commercial relationships: none.

Corresponding author. Qasim Zaidi.

Email: qz@sunyo.edu.

Address: State University of New York, 33 W 42nd St. New York, NY 10036, USA.

References

- Adelson, E. H., & Anandan, P. (1990, July 20). Ordinal characteristics of transparency. *AAAI-90 Workshop on Qualitative Vision*. Boston, Massachusetts.
- Anderson, B. L. (1997). A theory of illusory lightness and transparency in monocular and binocular images: The role of contour junctions. *Perception*, 26(4), 419–453.
- Beck, J. M., Ma, W. J., Kiani, R., Hanks, T., Churchland, A. K., Roitman, J., Shadlen, M. N., . . . Pouget, A. (2008). Probabilistic population codes for Bayesian decision making. *Neuron*, 60(6), 1142–1152.
- Beck, J., Prazdny, K., & Ivry, R. (1984). The perception of transparency with achromatic colors. *Perception & Psychophysics*, 35(5), 407–422.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10(4), 433–436.
- Cavanagh, P. (1987). Reconstructing the third dimension: Interactions between color, texture, motion, binocular disparity, and shape. *Computer Vision, Graphics, and Image Processing*, 37(2), 171–195.
- Cicerone, C. M., Hoffman, D. D., Gowdy, P. D., & Kim, J. S. (1995). The perception of color from motion. *Perception & Psychophysics*, 57(6), 761–777, <https://doi.org/10.3758/BF03206792>.
- Doerschner, K., Fleming, R. W., Yilmaz, O., Schrater, P. R., Hartung, B., & Kersten, D. (2011). Visual motion and the perception of surface material. *Current Biology*, 21(23), 2010–2016.
- Drugowitsch, J., Moreno-Bote, R., Churchland, A. K., Shadlen, M. N., & Pouget, A. (2012). The cost of accumulating evidence in perceptual decision making. *Journal of Neuroscience*, 32(11), 3612–3628.
- D’Zmura, M., Rinner, O., & Gegenfurtner, K. R. (2000). The colors seen behind transparent filters. *Perception*, 29(8), 911–926, <https://doi.org/10.1068/p2988>.
- Einhorn, H. J., Kleinmuntz, D. N., & Kleinmuntz, B. (1979). Linear regression and process-tracing models of judgment. *Psychological Review*, 86(5), 465.
- Ennis, R., & Doerschner, K. (2021). The color appearance of curved transparent objects. *Journal of Vision*, 21(5), 20–20.
- Falkenberg, C., & Faul, F. (2019). Transparent layer constancy is improved by motion, stereo disparity, highly regular background pattern, and successive presentation. *Journal of Vision*, 19(12), 16, <https://doi.org/10.1167/19.12.16>.

- Fleming, R. W., Jäkel, F., & Maloney, L. T. (2011). Visual perception of thick transparent materials. *Psychological Science*, 22(6), 812–820. JSTOR.
- Fuchs, W. (1923). Experimentelle Untersuchungen über die Änderung von Farben unter dem Einfluss von Gestalten. *Zeitschrift Für Psychologie*, 92, 249–325.
- Gerardin, P., Roud, P., Süsstrunk, S., & Knoblauch, K. (2006). Effects of motion and configural complexity on color transparency perception. *Visual Neuroscience*, 23(3–4), 591–596, <https://doi.org/10.1017/S0952523806233352>.
- Heider, G. M., & Koffka, K. (1933). New studies in transparency, form and color. *Psychologische Forschung*, 17, 13–55.
- Kahnt, T., Park, S. Q., Haynes, J.-D., & Tobler, P. N. (2014). Disentangling neural representations of value and salience in the human brain. *Proceedings of the National Academy of Sciences of the United States of American*, 111(13), 5000–5005.
- Kanizsa, G. (1979). *Organization in vision: Essays on Gestalt perception*. Westpoer, CT: Praeger Publishers.
- Kawabe, T., & Nishida, S. (2017). Contour junctions defined by dynamic image deformations enhance perceptual transparency. *Journal of Vision*, 17(13), 15–15.
- Kawabe, T., & Nishida, S. (2018). Deformation-induced transparency resolves color scission. *Journal of Vision*, 18(8), 3, <https://doi.org/10.1167/18.8.3>.
- Kersten, D. (1991). Transparency and the cooperative computation of scene attributes. In M. S. Landy, & J. A. Movshon (Eds.), *Computational models of visual processing* (pp. 209–228). Cambridge, MA: MIT Press.
- Khang, B.-G., & Zaidi, Q. (2002a). Cues and strategies for color constancy: Perceptual scission, image junctions and transformational color matching. *Vision Research*, 42(2), 211–226, [https://doi.org/10.1016/S0042-6989\(01\)00252-8](https://doi.org/10.1016/S0042-6989(01)00252-8).
- Khang, B.-G., & Zaidi, Q. (2002b). Accuracy of color scission for spectral transparencies. *Journal of Vision*, 2(6), 3, <https://doi.org/10.1167/2.6.3>.
- Khang, B.-G., & Zaidi, Q. (2004). Illuminant color perception of spectrally filtered spotlights. *Journal of Vision*, 4 (9), 680–692.
- Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3? *Perception*, 36, ECVF Abstract Supplement.
- Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modeling of depth cue combination: In defense of weak fusion. *Vision Research*, 35(3), 389–412.
- Leathers, M. L., & Olson, C. R. (2012). In monkeys making value-based decisions, LIP neurons encode cue salience and not action value. *Science*, 338(6103), 132–135.
- Masin, S. C. (1984). An experimental comparison of three- versus four-surface phenomenal transparency. *Perception & Psychophysics*, 35(4), 325–332, <https://doi.org/10.3758/BF03206336>.
- Masin, S. C. (1998). The luminance conditions of Fuchs's transparency in two-dimensional patterns. *Perception*, 27(7), 851–859, <https://doi.org/10.1068/p270851>.
- Masin, S. C. (1999). Color scission and phenomenal transparency. *Perceptual and Motor Skills*, 89(3), 815–823, <https://doi.org/10.2466/pms.1999.89.3.815>.
- Metelli, F. (1974). The perception of transparency. *Scientific American*, 230(4), 90–99.
- Metzger, W. (1936). *Gesetze des Sehens [Laws of seeing]*. Frankfurt am Main, Germany: Kramer.
- Qiu, F. T., & von der Heydt, R. (2007). Neural representation of transparent overlay. *Nature Neuroscience*, 10(3), 283–284, <https://doi.org/10.1038/nn1853>.
- Robilotto, R., Khang, B.-G., & Zaidi, Q. (2002). Sensory and physical determinants of perceived achromatic transparency. *Journal of Vision*, 2(5), 3, <https://doi.org/10.1167/2.5.3>.
- Robilotto, R., & Zaidi, Q. (2004). Perceived transparency of neutral density filters across dissimilar backgrounds. *Journal of Vision*, 4(3), 5–5, <https://doi.org/10.1167/4.3.5>.
- Singh, M., & Anderson, B. L. (2002). Toward a perceptual theory of transparency. *Psychological Review*, 109(3), 492–519, <https://doi.org/10.1037/0033-295X.109.3.492>.
- Singh, M., & Anderson, B. L. (2006). Photometric determinants of perceived transparency. *Vision Research*, 46(6), 879–894, <https://doi.org/10.1016/j.visres.2005.10.022>.
- Smithson, H., & Zaidi, Q. (2004). Colour constancy in context: Roles for local adaptation and levels of reference. *Journal of Vision*, 4(9), 3, <https://doi.org/10.1167/4.9.3>.
- Spicker, M., Hahn, F., Lindemeier, T., Saupe, D., & Deussen, O. (2017). Quantifying visual abstraction quality for stipple drawings. *Proceedings of the Symposium on Non-Photorealistic Animation and Rendering - NPAR '17*, 1–10, <https://doi.org/10.1145/3092919.3092923>.
- Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review*, 34(4), 273–286, <https://doi.org/10.1037/h0070288>.

Tsukida, K., & Gupta, M. R. (2011). *How to analyze paired comparison data* (Technical Report UWEETR-2011-0004). Seattle, WA: University of Washington.

Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., . . . von der Heydt, R. (2012). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure–ground organization. *Psychological Bulletin*, 138(6), 1172.

Westland, S., & Ripamonti, C. (2000). Invariant cone-excitation ratios may predict transparency.

Journal of the Optical Society of America A, 17(2), 255, <https://doi.org/10.1364/JOSAA.17.000255>.

Zaidi, Q. (1998). Identification of illuminant and object colors: Heuristic-based algorithms. *Journal of the Optical Society of America A*, 15(7), 1767, <https://doi.org/10.1364/JOSAA.15.001767>.

Zaidi, Q. (2001). Color constancy in a rough world. *Color Research & Application*, 26, 192–200.

Appendix

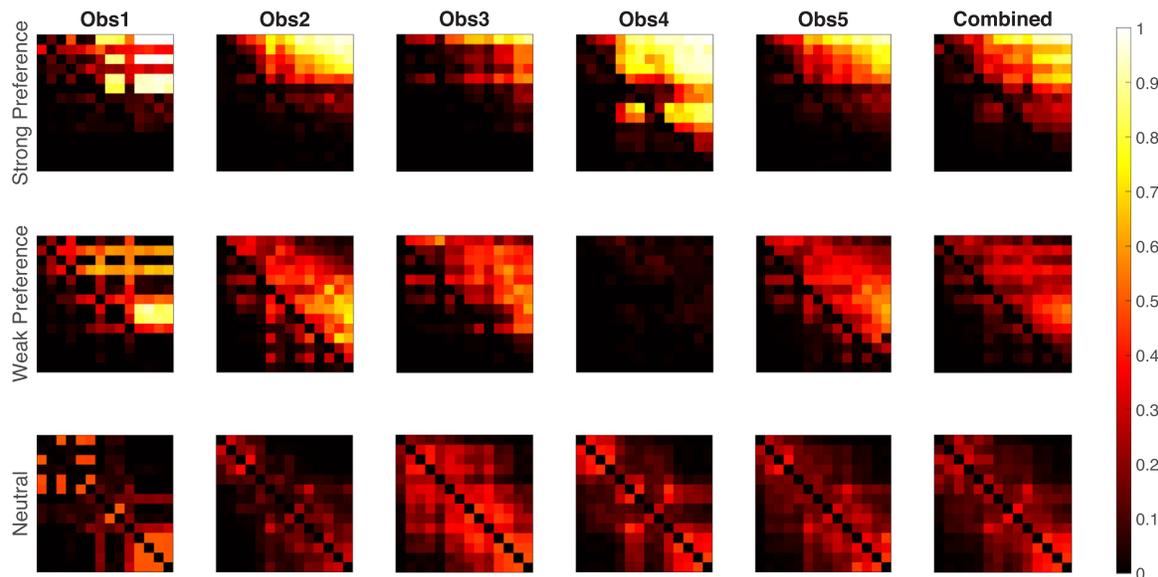


Figure A1. Matrices showing the frequency that one stimulus is strongly or weakly preferred over another or that the two stimuli are judged equal in value, for each observer and the Combined observer, used in constructing the Thurstone scale. The color of the pixel represents the rate of preference for the row condition over the column condition. The sequence of the conditions is the same as the transparency scale order.