Further clarifying signal detection theoretic interpretations of the Müller–Lyer and sound-induced flash illusions

Jessica K. Witt Colorado State University, Fort Collins, CO, USA

J. Eric T. Taylor University of Toronto, Toronto, Ontario, Canada

Mila Sugovic EurekaFacts, LLC, Rockville, MD, USA

John T. Wixted University of California–San Diego, La Jolla, CA, USA

Introduction

Previously, we showed that an effect on *c*, even in the absence of any effect on d', can be reflective of a perceptual effect rather than a change in response bias (Witt, Taylor, Sugovic, & Wixted, 2015). Knotts and Shams (2016) agreed with this main point. In addition, they pointed out a potential source of confusion from our paper and defended their previous analyses of the sound-induced flash illusion.

First, Knotts and Shams (2016) questioned whether the Müller–Lyer illusion would only affect *c* without also affecting d'. Asserting that this issue is important seems to imply that effects on d' would change our main claim. However, the interpretation of a change in *c* as potentially reflecting a perceptual effect is completely orthogonal to whether or not there is any effect on d' as well. We elected to simulate a situation in which d' did not change in order to emphasize our point that *c* can reflect a perceptual effect even under that extreme scenario. But *c* can still capture a perceptual effect even if there is also a perceptually induced change in d'. This point was mentioned in our article when we said: "For discrimination experiments, d' can be interpreted as a perceptual effect related to changes in sensitivity, but *c* can be interpreted only as a bias without the ability to distinguish between perceptual bias and response-based bias" (p. 298). In this response, we take the opportunity to explain this key point in more detail.

Second, Knotts and Shams (2016) defended their comparison of the no-beep versus two-beep conditions. They argue that d' changed across those two conditions, thereby establishing that a perceptual effect occurred. Again, we did not dispute the interpretation of the observed change in d'. Instead, our focus was on the interpretation of a change in *c*, such as the change in *c* that occurred when comparing their one-beep versus two-beep conditions. We focused on that comparison both because it illustrates our point and because it seems more theoretically informative than a comparison between the no-beep versus two-beep conditions, which confounds multisensory cues that are consistent versus inconsistent. Regardless of which comparison one makes, we agree with Knotts and Shams (2016) that it is worth specifying the underlying signal detection models that can be used to interpret the results. They made an effort to do just that in their figure 3, but the models they presented are problematic. We present more viable models and emphasize that researchers should carefully consider both the information represented on the decision axis (the *x*-axis) and the theorized effect of an experimental manipulation. Knotts and Shams (2016) used "Perceived # of flashes" as their decision axis variable, and they selected a range of −6 to 6. However, perceiving a negative number of flashes is nonsensical. We illustrate better ways to conceptualize the perceptual effects they observed (and that showed up mostly as a change in *c*).

Müller–Lyer illusion simulations

Knotts and Shams (2016) took issue with our simulation of the Müller–Lyer illusion. First, they clarified a potential source of confusion by accurately noting that we did in fact model equal illusory effects on both the short and the long lines. We agree that our

Citation: Witt, J. K., Taylor, J. E. T., Sugovic, M., & Wixted, J. T. (2016). Further clarifying signal detection theoretic interpretations of the Müller–Lyer and sound-induced flash illusions. *Journal of Vision*, *16*(11):19, 1–7, doi:10.1167/16.11.19.

doi: 10.1167/16.11.19 Received February 10, 2016; published September 30, 2016 ISSN 1534-7362

description was unclear. As our main figure showed, we modeled equal shifts (13% of 6 cm) for both the 5-cm and 7-cm lines. Second, they argued that this way of simulating the illusion is not consistent with the literature. More specially, they claimed that the illusion should be bigger for the longer line than for the shorter line. Before addressing the validity of their claim, we first point out that their simulation does not challenge our main claim that a change in *c* could be due to either a perceptual bias or a response-based bias, and that other considerations are necessary to make this determination.

In fact, intentionally or not, the authors further illustrated our main point by using a slightly less extreme case than we did, one in which almost all of the perceptual effect showed up as a change in *c* (with some of the effect also showing up as a change in d'). Although they only reported the results of their simulation in terms of d', if they had looked at their own estimated *c* parameter, they would see that *c* changed dramatically despite the fact that the criterion was fixed across conditions. Indeed, the change in *c* in their simulation (presented in their figure 2A, right column) is nearly identical to the change in *c* in our simulation (presented in their figure 2B, right column). The values of *c* for the tails-in versus tails-out conditions of their simulation (~ -0.65 and ~ 0.65, respectively) differ by approximately 1.30. In contrast, d' for those two conditions (~ 1.5 and ~ 2.0, respectively) changed comparatively little ($\Delta d' \approx 0.50$). Thus, even in their example, the measured parameters, interpreted as they commonly are, falsely suggest a large change in response bias despite the fact that the criterion did not shift at all. That aspect of their simulation (i.e., the large change in measured *c* despite the underlying criterion remaining constant across conditions) reinforces our main point that a change in the signal-detection parameter, *c*, can reflect a perceptual effect, not a response bias effect. Thus, the presence or absence of an effect on d' has no implications for the theoretical interpretation of an effect on *c*.

Nevertheless, it may still be useful to consider in some detail how a change in d' should be interpreted in the context of a perceptual bias. In cases for which it is known that there is a perceptual bias and not a response bias, the change in *c* can and should be interpreted as a perceptual bias. This is true regardless of whether there is also a change in d' (as shown by the similar effects in *c* across both simulations). In other words, the presence or absence of a change in d' has no bearing on the interpretation of the change in *c*. How, then, in the context of a perceptual shift, should the change in d' be interpreted? There are two, not mutually exclusive, possibilities: (a) that the perceptual shifts were unequal for the two types of stimuli, and (b) that the manipulation influenced the precision in one or both of the stimuli. If the perceptual shift was larger for one type of stimuli than another, d' should change. In their simulation, measured d' changed across conditions, correctly revealing some perceptual effect. The direction of change in d' reveals the direction of the effect. An increase in d' suggests a larger perceptual shift for the bigger stimulus (as in their simulation), and a decrease in d' suggests a larger perceptual shift for the smaller stimulus. A lack of effect on d' suggests equal perceptual shifts.

According to the second possibility, precision (i.e., the variance of perception across trials) may be affected. If the precision for either or both of the stimuli increases (i.e., reduced variance), d' or its unequal-variance counterpart, d_a, should also increase, but if precision decreases, d' or d_a should also decrease. One cannot tell by a change in d' alone whether the perceptual shifts were asymmetric or variance was impacted. However, receiver operating characteristics (ROC) analyses can reveal if there was an asymmetric change in the variances, and thus could be a useful tool.

How important is the difference between our simulation of equal effects and their simulation of unequal effects? The difference has no impact on the interpretation of *c*, as agreed by Knotts and Shams (2016). What, then, is the point of offering their version of the simulation? They argued that their version is grounded in the literature, and thus is a more accurate version. This would be a good reason if it were indeed true. However, we have not found literature supporting the claim that the Müller–Lyer illusion for a 7-cm line would be larger than for a 5-cm line. The article they cite as supporting this claim (Tudusciac & Nieder, 2010) did not even report the magnitude of the illusion across the various length lines, and thus cannot motivate this as a criticism of our simulation. Furthermore, our decision to simulate equal effects was intentional given pilot data (see Appendix) that showed similar effects for 5- and 7-cm lines and is also consistent with previous literature. For example, Daprati and Gentilucci (1997) used 5-, 6-, and 7-cm length lines while quantifying the magnitude of the Müller–Lyer illusion when grasping and manually estimating line length. Although they did not report analyses on this issue, their graphs clearly show no systematic increase in the magnitude of the illusion as line length increased. Thus, this criticism of our simulation is not founded in the literature, and our simulation is supported by prior results. So if a key point of Knotts and Shams (2016) is that we need to be consistent with the literature, our initial simulation already achieved that goal.

To summarize, in our simulation of a purely perceptual effect with no change in underlying criterion, measured d' did not change but measured *c* did. Thus, the *entirety* of the perceptual effect showed up in a change in *c*. In their simulation of a purely perceptual effect with no change in underlying criterion, measured d' changed a little and measured c changed a lot. Thus, *most* of the perceptual effect showed up in a change in c. And that fact reinforces our main point. Perceptual effects—even in the absence of an effect on measured d'—can show up as a change in c, which many researchers would mistakenly interpret as necessarily reflecting a change in response bias. In addition, a significant effect on d' does not mean that any effect on c can be interpreted as response bias. A perceptual effect can exert an influence on both d' and in c, as their simulation reveals.

Research on the sound-induced flash illusion

The second point made by Knotts and Shams (2016) relates to our interpretation of previous research on the sound-induced flash illusion. The sound-induced flash illusion is the illusion that a single visual flash of light is perceived as being two flashes when accompanied by two auditory beeps (Shams et al., 2000). To determine if the auditory information produces a perceptual change or a decisional effect, the authors used signal detection measures to compare across conditions (Rosenthal, Shimojo, & Shams, 2009; Watkins, Shams, Tanaka, Haynes, & Rees, 2006). Their experiment involved three conditions: a baseline (no-beeps) condition, a one-beep condition, and a two-beep condition, and their analysis focused mainly on the no-beep versus two-beep conditions. Rosenthal et al. (2009) found that d' was lower when two beeps were present than in the baseline no-beeps condition and concluded that this significant effect on d' is due to perceptual processes, specifically multisensory integration (Rosenthal et al., 2009; Shams & Kim, 2010; Watkins et al., 2006).

Our purpose was to challenge their *stated logic* about what measured signal detection parameters can tell you about the effects of a manipulation on perception and response bias (logic that is widely accepted in the field). Rosenthal et al. (2009) said, “We used signal detection theory to differentiate between changes in participants’ general response bias (β) and changes in their perceptual sensitivity, d' (the ability to perceptually discriminate single and double flashes)” (p. 187). This statement clearly aligns with the idea that d' is the sole measure of perceptual processes, and this is the logic that our paper challenged because (we argued) perceptual effects also show up in measures of c. Although they did not analyze the one-beep versus two-beep conditions in terms of c and d', we did, and we showed what their stated logic would imply about the interpretation of the results.

In their experiment, d' for the one-beep condition was very similar to d' for the two-beep condition. Should this be interpreted to mean that perception in the one-beep condition did not differ appreciably from perception in the two-beep condition? According to commonly accepted logic (and their logic), the answer is yes, but we claim that the answer is no. Furthermore, there was a large difference in c between these two conditions. Should this be interpreted to mean that the sound-induced flash illusion (namely, that two beeps leads to the response of the visual perception of two flashes) is due to response bias when compared with a condition with one beep but due to perception when compared with a condition with no beeps? Using their logic, the answer would be yes, but the purpose of our paper was to make it clear that the answer is no. A change in c does not necessarily reflect a change in response bias. Perceptual effects show up in c too. Indeed, sometimes, a perceptual effect shows up exclusively in a change in c. This may be true of their one-beep versus two-beep conditions (just as it can be true of the Müller–Lyer illusion).

Although Knotts and Shams (2016) repeatedly stated that they agreed with our conclusions, their writing implies otherwise. For instance, they state “the purpose of comparing d' across the no beep and two beep conditions…was to track the magnitude of the perceptual component of the illusion that is *due to multisensory integration*” (Knotts & Shams, 2016, p. 7, emphasis in original). This statement implies that analyzing c would not tap into the perceptual component, or that comparison of the one-beep and two-beep conditions does not involve multisensory integration (discussed further below). They go on to say that although c can be reflective of perceptual processing, “this particular measure was not of primary interest to the goals of the study” (p. 7). Their stated goal of the study was to examine “how robust the illusion is by testing whether the frequency of the illusion can be reduced by providing feedback” (Rosenthal et al., 2009, p. 185). Our point is that the sound-induced flash illusion is a perceptual bias, and as such, should materialize in the measure of c, and so differences in c should be the primary focus of their study, not merely a way to “potentially provide a more complete capture of the perceptual effect” (Knotts & Shams, 2016, p. 7).

When an effect produces large differences in c, other considerations must be taken into account to determine if the underlying effect reflects perceptual or decision processes. Shams and her colleagues have gone to great lengths to demonstrate that the sound-flash illusion is a perceptual effect. In fact, they have found little evidence for any role for response bias. Given this lack of response bias in the sound-flash illusion, how can they account for such a large difference in c? They cannot under the logic that c measures only response bias. Instead, the large difference in c across the one-

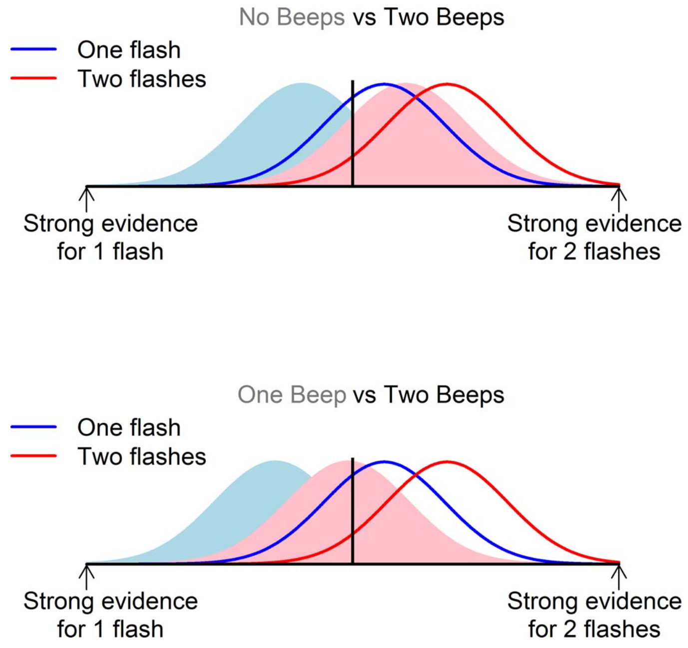

Figure 1. Hypothetical distributions of neural evidence for one flash versus two flashes as a function of number of visual flashes presented (one flash: blue curves; two flashes: red curves) and as a function of number of auditory beeps. The top panel shows the comparison for no beeps (pale and shaded curves) and two beeps (dark curves with no shading). The bottom panel shows the comparison for one beep (pale and shaded curves) and two beeps (dark curves with no shading). The vertical line represents the point of complete uncertainty as to the presence of one or two flashes. The dramatic shift towards perceiving two flashes (shift in curves to the right) is more clearly depicted in the bottom figure (especially by comparing the red curves across panels).

beep and two-beep conditions can and should be interpreted as a perceptual bias and not a response bias.

We agree with Knotts and Shams (2016), and in doing so will modify one of our early claims, with respect to whether d' can provide anything useful. We originally stated that the illusion "should present itself in the measure of c (or β), and not in d'" (Witt et al., 2015, p. 291). The first part of this statement is still true (the illusion should present itself in the measure of c). However, the illusion could present some of its magnitude in d' if, as we argued above, the illusion is asymmetrical. In focusing so much on the measure and interpretation of c, we neglected to discuss potential implications for any effects in d'. In the sections above, we have clarified these issues.

Selecting the conditions to compare

An issue tangentially debated across the two articles concerns the conditions that should be included in the critical comparison. We believe that the appropriate comparison is between the one-beep and two-beep conditions, but they defended their previous assertion that the correct comparison is between the no-beep and two-beep conditions. When interested in a multisensory effect, it can be sensible to compare unisensory to multisensory conditions, but one can also glean important insights by comparing two multisensory conditions as well. And, as we argue below, the two multisensory conditions are the more relevant for the sound-induced flash illusion.

When using signal detection theory (SDT) to analyze data, researchers should carefully consider the meaning of their decisional axis variable (the x-axis) and the changes that are theorized to occur as a result of the manipulation. Take their figure 3 as an example. Their x-axis is "Perceived # of flashes," but look at the range of this axis. It goes from -6 to 6. What could it possibly mean to perceive -2 flashes? In addition, they have centered the one-flash/one-beep curve at the location of -2 flashes. Do they think that one flash combined with one beep should result in the perception of -2 flashes? Determining the correct decisional axis variable requires careful theoretical consideration.

The participants' task in experiments testing the sound-induced flash illusion is to indicate if there was one flash or two flashes. Thus, a sensible x-axis could be the evidence in favor of one flash versus two flashes (see Figure 1). The axis would range from strong evidence that one flash occurred (on the far left) to strong evidence that two flashes occurred (on the far right).

Now, consider the theoretical claim that is made from the sound-induced flash illusion and how the effect of hearing two beeps should manifest in each of the distributions representing one and two flashes (compared to hearing one beep). There are several possibilities. One possibility is that, in response to two beeps, both the one-flash and two-flash distributions would be shifted to the right and to the same degree. This purely perceptual effect would cause a large change in measured c with no change in measured d' (as in our Müller–Lyer illusion simulation and as in the analyses of their one-beep vs. two-beep conditions). Another closely related possibility is that both distributions would shift to the right but with a larger effect on the one-flash distribution than the two-flash distribution. In that case, the purely perceptual effect would again cause a large change in measured c with a small change in measured d' (as in their Müller–Lyer illusion simulation). By comparing the one-beep to two-beep conditions, one can examine the change in d' as potentially indicative of asymmetrical perceptual shifts, whereas this information could not be determined by comparing the no-beep and two-beep conditions. Given comparable d' values, the data suggest that the number of beeps influenced one flash and two flashes similarly. In addition, the comparison between one-beep and two-beep conditions also reveals any asymmetries in the effect of one beep versus two beeps. Given that the absolute value of *c* is bigger for the two-beep condition than for the one-beep condition, this suggests that two beeps produced a larger perceptual shift compared with one beep. Again, this information could not be gleaned from a comparison of the no-beep and two-beep conditions.

A third possibility is that hearing two beeps in the two-flash condition would not only shift that distribution to the right but would also selectively reduce its variance, increasing the precision of perception in that condition and creating an unequal-variance model. Bayesian models of multisensory integration predict a reduction in variance, particularly when the auditory cues are consistent with the visual cues (e.g., Ernst, 2006). In the case of unequal variance, d' would not be the most appropriate dependent measure to use. Instead, a measure like d_a would make more sense. To compute d_a, one would need to perform confidence-based ROC analysis, as is commonly done in recognition memory experiments, which are almost invariably better modeled using an unequal-variance model than an equal-variance model (Macmillan & Creelman, 2008; Wixted, 2007). Given that Rosenthal et al. (2009) collected confidence ratings (participants expressed high or low confidence whether they claimed to see one flash or two), they have the data needed to perform these analyses, though the use of a more fine-grained confidence scale would be preferable.

Something worth noting is that the multisensory conditions include both a consistent component (e.g., two beeps paired with two flashes) and an inconsistent component (e.g., two beeps paired with one flash). If consistent and inconsistent components lead to different effects (e.g., bigger shifts or different variances), these would be confounded in the comparison of the two-beep condition to the no-beep condition. It would be difficult to understand which component of the multisensory condition (the consistent component or the inconsistent component) was responsible for any observed effects. While this confound also exists in the comparison of the two-beep to one-beep conditions, the confound would at least exist in both cases and thus could potentially cancel each other out.

Conclusion

In summary, the main point in Witt et al. (2015) had to do with the interpretation of a change in measured *c* across conditions. In their critique, the authors focused entirely on the interpretation of a change in d' across conditions. They claimed that a change in d' reflects a change in perception. Because this is not a point of contention, the purpose of their critique is not clear to us. Even in our abstract, we made it clear that "While d' can provide evidence for a perceptual effect, an effect solely on the criterion measure can also arise from a perceptual effect" (p. 289). And in our concluding summary, we said "…both perceptual biases and decision-based biases exert their influence on the criterion measure of *c*…An influence on *c* implies a bias, but the nature of this bias—be it a perceptual bias, a memory bias, a social bias, or a response-based bias—is not specified by current SDT techniques" (p. 299). That is our claim, and nothing in their critique addresses it.

Our main point is that *c* cannot, by itself, indicate whether an effect is due to a response bias. This point is best illustrated by cases where a universally-agreed upon perceptual effect (such as the Müller–Lyer illusion) produces an effect in *c* but not d'. In our efforts to illustrate our thesis, we presented evidence from these strongest cases, where perceptual effects occur in *c* and not in d'. However, our point is equally true in any SDT analysis with an effect on *c*: This measure cannot be construed strictly as an effect on response processes, regardless of independent effects on sensitivity (d'). The disagreement the authors have brought to bear is whether these strongest cases, with effects on *c* but not d', are legitimate examples of SDT analyses and interpretations. We defend our use of equal illusory effects for different line lengths in the Muller-Lyer illusion and the application of typically-flawed SDT logic (namely, that effects in *c* and not d' should be interpreted as effects on response strategy) in the sound-induced flash illusion.

Acknowledgments

This work was supported by a grant from the National Science Foundation (BCS-1348916) to JKW.

Commercial relationships: None.

Corresponding author: Jessica K. Witt.
Email: jessica.witt@colostate.edu.
Address: Colorado State University, Fort Collins, CO, USA.

References

Daprati, E., & Gentilucci, M. (1997). Grasping an illusion. *Neuropsychologia*, *35*(12), 1577–1582.

Ernst, M. O. (2006). A Bayesian view on multimodal cue integration. In G. Knoblich, I. M. Thornton, M. Grosjean, & M. Shiffrar (Eds.), *Human body* *perception from the inside out* (pp. 105–131). New York: Oxford University Press.

Knotts, & Shams, L. (2016). Clarifying signal detection theoretical interpretations of the Muller-Lyer and sound-induced flash illusions. *Journal of Vision*. *16*(11):18, 1–4, doi:10.1167/16.11.18.

Macmillan, N. A., & Creelman, C. D. (2008). *Detection theory: A user's guide* (2nd ed.). New York: Psychology Press.

Rosenthal, O., Shimojo, S., & Shams, L. (2009). Sound-induced flash illusion is resistant to feedback training. *Brian Topography*, *21*, 185–192.

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t-tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*, 225–237.

Shams, L., Kamitani, Y., & Shimojo, S. (2000). What you see is what you hear. *Nature*, *408*, 788.

Shams, L., & Kim, R. (2010). Crossmodal influences on visual perception. *Physics of Life Reviews*, *7*, 269–284.

Tudusciac, O., & Nieder, A. (2010). Comparison of length judgments and the Müller-Lyer illusion in monkeys and humans. *Experimental Brain Research*, *207*, 221–231.

Watkins, S., Shams, L., Tanaka, S., Haynes, J.-D., & Rees, G. (2006). Sound alters activity in human V1 in association with illusory visual perception. *Neuroimage*, *31*, 1247–1256.

Witt, J. K., Taylor, J. E. T., Sugovic, M., & Wixted, J. T. (2015). Signal detection measures cannot distinguish perceptual biases from response biases. *Perception*, *44*(3), 289–300.

Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review*, *114*, 152–176.

Appendix: Pilot experiment on Müller–Lyer illusion for various length lines

In a pilot study, we examined whether the magnitude of the Müller–Lyer illusion varied as a function of line length.

Method

Participants

Sixteen students participated in exchange for course credit.

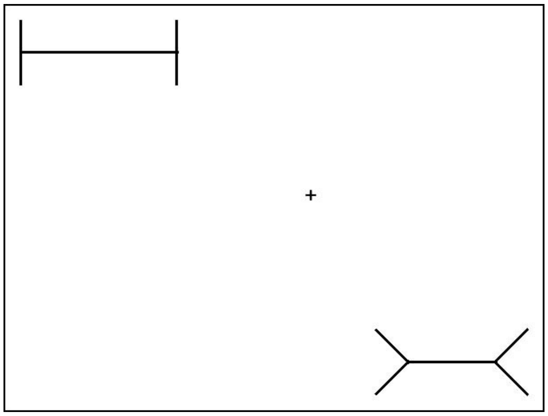

Figure A1. Sample display from the experiment.

Materials and apparatus

Stimuli were presented on a 19-in. monitor. On each trial, two lines were presented. The standard line was presented at the top-left corner of the screen and always had vertical lines at each end (see Figure A1). The standard line was set to one of six possible lengths ranging from 5.2 to 7.4 cm. The comparison line was presented at the bottom-right corner and had tails oriented in or tails oriented out. This line was initially set to be short (2.9 cm) or long (10.3 cm). Participants adjusted this line to be the same length as the standard line on each trial.

Procedure

Participants were instructed to adjust the length of the bottom line to be the same as the length of the top line. To make adjustments, participants pressed keys 1, 2, 4, and 5 on a keypad. Keys 1 and 4 made the line shorter by small and large amounts, respectively, and keys 2 and 5 made the line bigger by small and large amounts (small = 0.6-mm change, large = 3.0-mm change). Participants were not limited in the number of adjustments or the time to make these adjustments. Once done with adjustments, participants hit Enter to begin a new trial. Each trial began with a fixation screen presented for 250 ms before both lines appeared simultaneously. Both lines were visible throughout the duration of the trial. Participants completed four blocks of trials. Each block contained 24 trials with each combination of tail orientation (tails in, tails out), initial comparison width (short, long), and standard line length (5.2–7.4 cm), and order within block was randomized.

Results and discussion

The dependent measure was perceived line length, which was estimated as the final adjusted width of the

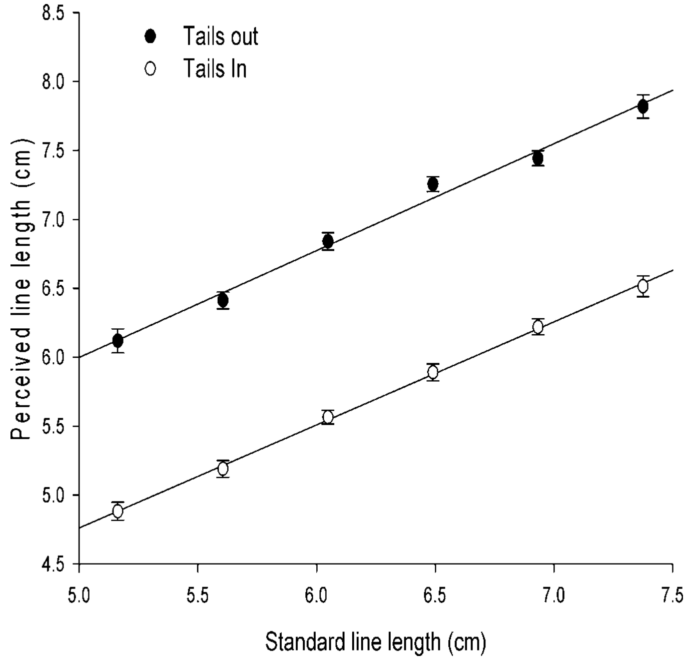

Figure A2. Mean perceived line length as a function of standard line length and tail orientation. Error bars are 1 *SEM* calculated within-subjects.

comparison line at the end of each trial. Mean perceived length was calculated for each participant for each line length and tail orientation. These were submitted to a repeated-measures analysis of variance with tail orientation and standard line length as within-subjects factors. Tail orientation significantly influenced perceived line length, $F(1, 15) = 199.46$, $p < 0.001$, $\eta_p^2 = 0.93$. Lines with tails oriented out looked longer than lines with tails oriented in. Standard line length significantly influenced perceived line length, $F(5, 75) = 196.01$, $p < 0.001$, $\eta_p^2 = 0.93$. Critically, the interaction between tail orientation and standard line length was not significant, $F(5, 75) = 1.08$, $p = 0.38$, $\eta_p^2 = 0.07$. As shown in Figure A2, the strength of the illusion based on tail orientation was consistent for all lengths of the standard line.

We also computed the difference score between tails out minus tails in for the shortest standard line (5.2 cm) and the longest standard line (7.4 cm). The difference score for these two lines were submitted to a paired-samples *t* test, which revealed a nonsignificant effect, $t(15) = 0.74$, $p = 0.47$. Running this through the Bayesian calculator provided by Rouder and colleagues (Rouder, Speckman, Sun, Morey, & Iverson, 2009) resulted in a scaled JZS Bayes factor of 4.10 in favor of the null hypothesis. This means that the data collected are over four times more likely to occur under the null hypothesis than under the alternative hypothesis that illusion magnitude increases with an increase in line length from 5 to 7 cm. In our pilot study, the second longest line was actually closest to the modeled 7-cm line, so we repeatedPosition 50378 the analysis comparing the shorted line to the second longest line. Again, the difference was not significant, $t(15) = 0.16$, $p = 0.88$, and the Bayes factor showed that the null hypothesis was 5.23 than the alternative hypothesis. The results from this pilot study support our decision to model equal effects of the Müller–Lyer illusion on the 5- and 7-cm lines (Witt et al., 2015).