Newly learned novel cues to location are combined with familiar cues but not always with each other

Stacey Aston[1], Ulrik Beierholm[1], & Marko Nardini[1]

[1] *Department of Psychology, Durham University, UK*

In Press, *Journal of Experimental Psychology: Human Perception and Performance*

Accepted 21st Feb 2022

Newly learned novel cues to location are combined with familiar cues but not always with each other

Abstract

Mature perceptual systems can learn new arbitrary sensory signals (novel cues) to properties of the environment, but little is known about the extent to which novel cues are integrated into normal perception. In normal perception, multiple uncertain familiar cues are combined, often near-optimally (reliability-weighted averaging), to increase perceptual precision. We trained observers to use abstract novel cues to estimate horizontal locations of hidden objects on a monitor. In Experiment 1, four groups of observers each learned to use a different novel cue. All groups benefitted from a suboptimal but significant gain in precision using novel and familiar cues together after short-term training (3 x ~1.5 hour sessions), extending previous reports of novel-familiar cue combination. In Experiment 2, we tested whether two novel cues may also be combined with each other. One pair of novel cues could be combined to improve precision but the other could not, at least not after three sessions of repeated training. Overall, our results provide extensive evidence that novel cues can be learned and combined with familiar cues to enhance perception, but mixed evidence for whether perceptual and decision-making systems can extend this ability to the combination of multiple novel cues with only short-term training.

Keywords

Cue combination, sensory integration, sensory augmentation

Public Significance Statement

- Human adults can learn novel relationships between arbitrary sensory signals and properties of the surrounding environment (novel cues).
- Newly learned novel cues are combined with familiar cues (natural relationships between sensory signals and properties of the surrounding environment) to enhance perception and decision-making.
- After repeated training, the enhancement from combining some novel cues with familiar cues is as good as it can. In other words, human adults make optimal use of the novel information.
- Whether or not this ability can be extended to the combination of two novel cues may depend on the two novel cues to be combined.

Introduction

A mature perceptual system can learn new mappings between arbitrary sensory signals and properties of the environment (novel cues), such as an artificial correlation between the brightness and stiffness of an object (Ernst, 2007) or an auditory cue to depth (Negen et al., 2018), among others (Di Luca et al., 2010; Haijiang et al., 2006; Harrison & Backus, 2012; Michel & Jacobs, 2008). However, little is known about the extent to which novel cues are integrated into the normal perceptual experience. In normal perception, there are often multiple uncertain familiar sensory cues (natural mappings between sensations and physical properties of the surrounding environment) providing similar information about the state of the surrounding world, such as disparity and texture cues to the slant of a surface (Knill & Saunders, 2003). An important feature of familiar cue use is that when multiple cues are available, rather than throwing one piece of information away and using only the most reliable cue, a mature perceptual system tends to combine the cues in line with reliability-weighted averaging - the Bayes-optimal solution to cue combination that maximises precision (Alais & Burr, 2004; Ernst & Banks, 2002; Hillis et al., 2004; Knill & Saunders, 2003).

A limited number of studies suggest newly learned novel cues are also combined with familiar cues (Ernst, 2007; Gibo et al., 2017; Michel & Jacobs, 2008; Negen et al., 2018). Importantly, although combination of novel and familiar cues is often suboptimal, with the gain in precision from combining the two cues less than that predicted by reliability-weighted averaging (Ernst, 2007; Gibo et al., 2017; Negen et al., 2018), it is “Bayes-like” in the sense that it shows some signatures of Bayes-optimal combination, such as weighting by reliability (Negen et al., 2018).

The ability to learn novel cues and combine them with familiar cues has vast applications for sensory substitution and augmentation. In the case of sensory substitution, it means that perceptual systems receiving disrupted familiar cues (for example, in partial vision loss) could not only learn to replace the disrupted input with a novel cue (Abboud et al., 2014; Auvray et al., 2007; Bach-y-Rita et al., 1969; Maidenbaum et al., 2014), but could combine the novel cue with disrupted familiar cues to make more precise judgements than using either cue alone would allow. Similarly, in the case of a healthy perceptual system, novel cues can be introduced to enhance the normal perceptual experience. New technologies offer a variety of options for providing perceptual systems with new sensory signals. To make the best use of these technologies, the design of new sensory signals should be grounded in research that explores which novel cues are most efficiently learned and combined with familiar or other novel cues, as well as the training conditions that best promote integration of new sensory signals into the normal perceptual experience.

Here, we asked whether observers combine novel and familiar cues to increase precision above what is possible using the most reliable single cue alone, and how any such gains in precision differ from the optimal or maximum gain predicted by reliability-weighted averaging. In Experiment 1, we trained observers to use abstract novel cues to estimate the horizontal location of hidden objects on a computer screen. The novel cues were the colour of a pair of lines (*colour* cue), the angle between two lines (the *angle* cue), the axis ratio of an oval (the *shape* cue), and the height of a bar (the *height* cue). We refer to our novel cues as abstract as they do not have a natural relationship to location. This contrasts with previous studies where observers learned to use an echolocation cue to make depth judgements (Negen et al., 2018) or made movements with the assistance of a force cue that guided movements in a particular direction (Gibo et al., 2017).

Observers completed a task that began with a short training period to teach (or reinforce) the mapping between the novel cue and location. After training, observers completed a series of trials where they were required to use either the novel cue, a familiar cue (e.g., a noisy dot-cloud), or the novel and familiar cues together to estimate the location of a hidden object. Forty observers were divided into equal groups so that each observer learned only one novel cue with each observer completing the same task on three different days (three sessions). This aspect of the design provided the observers with repeated training, allowing them not only to learn the mappings to location over time, but also to learn to discriminate finer differences in the novel cues (i.e. perceptual learning - an improvement in discrimination ability for a stimulus (cue) that was not previously well discriminated; Fahle & Poggio, 2002). We considered that it was important to allow for perceptual learning as single cue reliabilities may be changing as discrimination ability improves, and changing cue reliabilities could be a barrier to reliability-weighted averaging and Bayes-like combination (Alais & Burr, 2004; Ernst & Banks, 2002; Hillis et al., 2004; Knill & Saunders, 2003).

Each group of observers in Experiment 1 benefitted from a gain in precision using the novel and familiar cues together by the third session. The gain in precision was suboptimal but significant; location estimates were significantly less variable when both the novel and familiar cues were available than when observers used their best single cue alone. Our results show that observers can learn abstract novel cues to location and combine them with a familiar cue.

In Experiment 2, we tested if two novel cues may also be combined with each other. We tested this by teaching two different groups, each of ten observers, a different pairing of the abstract novel cues to location from Experiment 1 (the *colour* and *angle* cues or the *colour* and *shape* cues). In this experiment, observers received separate training with each novel cue. After training they completed a series of trials where they used either one of the novel cues, both novel cues, the familiar cue, or one of the novel cues and the familiar cue to estimate the location of the hidden object. As in Experiment 1, each observer completed the task three times on three different days. We found that one pair of novel cues could be combined to improve precision but the other could not, even after three sessions of repeated training.

Overall, our results provide extensive evidence that novel cues can be learned and combined with familiar cues to enhance perception, but mixed evidence for whether perceptual and decision-making systems can extend this ability to the combination of multiple novel cues with only short-term training.

Experiment 1: Methods

Overview

Forty observers completed the same task three times on three different days (three sessions). The task required the observers to use a novel cue, a familiar cue, or the novel and familiar cues simultaneously to estimate the location of a hidden target by using a computer mouse to adjust the horizontal position of a bar on a computer screen. The task began with a block of training trials that taught observers the mapping between a novel cue and horizontal location on the screen. The forty observers were split into four groups of ten with each group learning a different novel cue to location (Figure 1). The *colour* group learned to use the average colour of eight pairs of lines as a cue to location (the *colour* cue), the *angle* group learned to use the average size of the angle between eight pairs of lines as a cue to location (the *angle* cue), the *shape* group learned to use the average axis ratio of eight ovals as a cue to location (the *shape* cue), and the *height* group learned to use the average height of eight vertical bars as a cue to location (the *height* cue). All groups used the same familiar cue, that can be thought of as a dot cloud, though we will refer to it as the spread cue. The spread cue always consisted of eight stimuli (shapes that varied for each group to avoid giving information that conflicted with the novel cue) with varying position on the screen. The best way to utilise this cue was for observers to take the average horizontal location of the eight stimuli. We say the spread cue is a familiar cue as it naturally maps to horizontal location on the screen. This is unlike the novel cues, where the mapping must be learned.

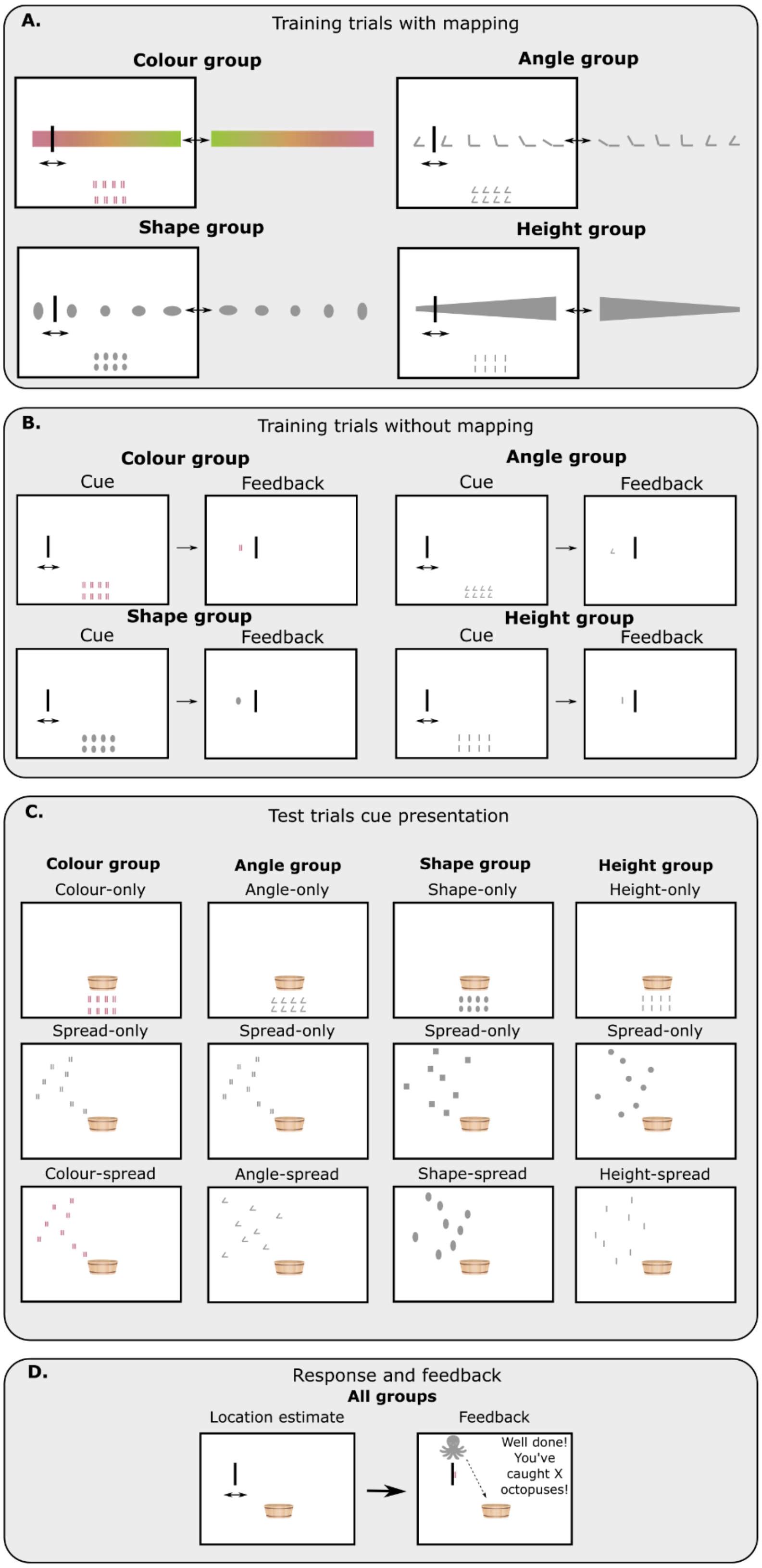

Figure 1: The task in Experiment 1. (A-B) The task began with a block of training trials where observers were taught a mapping between a novel cue (colour, angle size, the axis ratio of an oval, or

the height of a bar) and horizontal location on a computer screen. In the first set of training trials (A), observers could see the novel mapping on the screen and had to select the location along the mapping that corresponded to the average novel cue value of eight stimuli shown at the bottom of the screen. The direction of the mapping was randomly chosen for each observer. In the second set of training trials (B), the mapping was not shown but observers could continue to learn the mapping through feedback. (C) In test trials, observers used either the newly learned novel cue, a familiar spread cue (e.g., a dot cloud), or both the novel and familiar cue together to estimate the position of a hidden object (an octopus hiding in the sea). (D) After issuing a response by positioning a vertical bar horizontally across the screen, observers received feedback and, if they "caught" the octopus, saw an animation of the octopus moving into their bucket.

In the training block, observers first completed a set of trials where the mapping between the novel cue and location was shown on the screen (Figure 1A). In these "with mapping" trials, the novel cue was presented at the bottom of the screen and observers were required to estimate the average colour, angle size, axis ratio, or height of the cue, indicating their response by moving a vertical bar to the correct location along the mapping. Observers then completed a set of "without mapping" trials (Figure 1B) that encouraged them to learn the relationship between the cues and location as the mapping was no longer shown. Learning of the mapping was reinforced through feedback in these trials, with observers shown the correct average colour, angle size, axis ratio, or height in the correct location as feedback. The direction of the mapping (left-to-right or right-to-left) on the screen was randomly determined for each observer.

After observers completed the training block, the test trials began (Figure 1C). At the start of the test block, observers were instructed that they would now begin to use the newly learnt novel cue, along with a familiar cue (i.e., a dot-cloud, or the spread cue) to estimate the location of a hidden object – an octopus hiding in the sea. On each trial, observers were presented with either the novel cue (*colour-only*, *angle-only*, *shape-only*, or *height-only* trials), the familiar cue (*spread-only* trials), or the novel and familiar cue together (*colour-spread, angle-spread, shape-spread*, or *height-spread* trials). In *colour-only* and *angle-only* trials, observers were presented with eight pairs of lines (in fixed positions) at the bottom of the screen. The average colour of the pair of lines or angle between them provided a novel estimate of location according to a trained mapping. In *shape-only* trials observers were presented with eight ovals (in fixed positions) at the bottom of the screen. The average vertical to horizontal axis ratio of the ovals provided a novel estimate of location according to a trained mapping. In *height-only* trials observers were presented with eight vertical bars (in fixed positions) at the bottom of the screen. The average height of the vertical bars provided a novel estimate of location according to a trained mapping. In *spread-only* trials, eight pairs of parallel and grey lines (*colour* and *shape* groups), grey squares (*shape* group), or grey circles (*height* group) were spread out across the screen. The position of each pair of lines, square, or circle was drawn from a Gaussian distribution, centred on the hidden location, such that the mean or centroid of the locations was the best estimate. In *colour-spread* or *angle-spread* trials, the eight pairs of lines were spread across the screen and had the property of the novel cue (either the relevant colours or angles between the lines). In *shape-spread* trials the eight ovals were spread across the screen and had the property of the novel cue (the relevant axis ratios). In *height-spread* trials the eight bars were spread across the screen and had the property of the novel cue (the relevant bar heights).

Trials of all types were interleaved for each group (e.g., *colour-only*, *spread-only*, and *colour-spread* for the *colour* group). After the cue(s) appeared on each trial, observers adjusted the horizontal position of a vertical line (width 10 pixels), using a mouse, to their best guess of the hidden location (Figure 1D). Feedback was given indicating if the observers had "caught" the octopus along with an indicator of the true hidden location that displayed the corresponding novel cue values (the correct average colour, angle size, axis ratio, or height). If the octopus was caught, an animation showed the octopus move across the screen from its hidden location to the bucket. The octopus was caught if any part of the vertical line overlapped with the feedback marker, meaning there was a tolerance of 26 pixels.

Observers

Forty observers were recruited using Durham Psychology Department's Participant Pool programme or through word of mouth. Each observer was assigned to either the *colour* group*, angle* group*, shape* group, or *height* group such that there were ten observers in each group (*colour* group: 7 female, age range 19-29 years; *angle* group: 8 female, age range 19-27 years; *shape* group: 9 female, age range 18-42 years; *height* group: 8 female, age range 18-21 years). All observers had normal or corrected to normal visual acuity (self-report) and no colour vision deficiencies (assessed using Ishihara Colour Plates). Each observer was given either £8 per hour or participant pool credits for their time.

Apparatus

Stimuli were shown on a 10-bit ASUS Proart LCD screen (ASUS, Fremont, CA) with observers seated so that their eyes were approximately 60 cm from the screen. The monitor was controlled using a 64-bit Windows machine, equipped with an NVIDIA Quadro K600 10-bit graphics card (NVIDIA, Santa Clara, CA), running MATLAB scripts that used Psychtoolbox routines (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997). The stimuli were colourimetrically calibrated using a linearized calibration table based on measurements of the monitor primaries made with a Konica Minolta CS2000 spectroradiometer (Konica Minolta, Nieuwegein, Netherlands). Conversions to CIELUV used the measured white point of the monitor: $(Y, x, y) = (205.24, .31, .34)$ in CIE 1931 Yxy colour space.

Stimuli

In *colour-only* trials, the novel colour cue appeared in a fixed location at the bottom of the screen. The novel colour cue was a set of eight pairs of parallel lines (length 24, width 5 pixels) where each pair of lines varied slightly in colour. The colour of the dots or pairs of lines was governed by a colour gradient from pink to green that mapped from 15% to 85% of the way across the screen from left to right or right to left (randomly flipped for each observer). The gradient was defined as a chord of a hue circle (chroma = 85) in CIELUV chromaticity space. The start and end values of the chord had CIE 1931 chromaticities of $(x, y) = (.3386, .2821)$ and $(x, y) = (.3476, .3960)$ and a luminance of $Y = 15$ cd/m^2. The colour gradient was defined in this way to ensure perceptual uniformity and defined a mapping from colour to location across the screen. The colours of the eight pairs of lines were defined by drawing eight horizontal positions from a Gaussian distribution centred on the hidden object's location with a standard deviation of 3 pixels. The colours of the eight pairs of lines were then taken to be the colours that corresponded to each of the sampled locations according to the mapping. In the training trials, the mapping was shown on the screen as a colour gradient.

In *angle-only* trials, the novel angle size cue appeared in a fixed location at the bottom of the screen. This cue was eight pairs of lines (length 24, width 5 pixels) where each pair formed an angle. Angles were always formed in either only the 1st or across both the 1st and 2nd quadrants such that one of the lines forming the angle was always the abscissa in the 1st quadrant. The size of the angle formed by each pair of lines was dictated by a pre-defined mapping of angle size to screen position. Angle sizes of 67.95° and 162.45° corresponded to 15% and 85% of the way across the screen, respectively, or vice versa (flipped at random for each observer). To set the angle sizes on each trial, eight horizontal positions were drawn from a Gaussian distribution centred on the hidden object's location with a standard deviation of 0.7 pixels. The angle sizes were then taken to be those that corresponded to each of the sampled locations according to the mapping. In the training trials, the angles corresponding to locations 17% to 85% of the way across the screen in steps of 4% were shown across the screen at their correct locations. On *angle-only* trials, the angles were always grey, as were the angles shown as part of the mapping. On *colour-angle* trials, each angle was also assigned a colour by the same method as the *colour-only* cue.

In *shape-only* trials, the novel shape cue appeared in a fixed location at the bottom of the screen. The novel shape cue was a set of eight ovals. The ratio of the vertical (a) to horizontal (b) axis varied for each oval, while maintaining the total area, and was defined based on a mapping of axis ratio to location across the screen. A location 15% of the way across the screen, from left to right, corresponded to a ratio of $a/b = 12.191/22.979$, while 85% of the way across the screen corresponded to $a/b = 22.979/12.191$ pixels, or vice versa (flipped randomly for each observer). To set the ratio for each oval, eight horizontal positions were drawn from a Gaussian distribution centred on the hidden object's location with a standard deviation of 0.7 pixels. The ratios were then taken to be those that corresponded to each of the sampled locations according to the mapping. In the training trials, only the shapes corresponding to locations 17% to 85% of the way across the screen in steps of 4% were shown. When the novel shape cue was paired with the familiar spread cue, the eight symbols representing the shape cue were spread across the screen.

In *height-only* trials, the novel bar height cue appeared in a fixed location at the bottom of the screen. The novel bar height cue was a set of eight vertical bars (width 5 pixels) whose heights varied. The heights were decided according to a linear mapping of bar height to screen position. A height of 8.69 pixels corresponded to 15% of the way across the screen, from left to right, and a length of 30.82 pixels to 85%, or vice versa (flipped randomly for each observer). To set the height of each bar, eight horizontal positions were drawn from a Gaussian distribution centred on the hidden object's location with a standard deviation of 0.2 pixels. The heights of the bars were then taken to be those that corresponded to each of the sampled locations according to the mapping. In the training trials, the mapping was shown on the screen as a truncated 2D cone with the height of the cone at each location corresponding to the bar height that mapped there. When the novel bar height cue was paired with the familiar spread cue, the eight symbols representing the bar height cue were spread across the screen.

In *spread-only* trials the familiar cue appeared on the screen. The familiar cue was effectively a "dot" cloud generated by drawing the position of each "dot" from a Gaussian distribution centred on the hidden object's location with a standard deviation of 237 pixels and were scaled so that the standard deviation of the eight sampled locations matched the population standard deviation. However, we only displayed a dot at each location for the *height* group. In *height-spread* trials, the *height* group saw eight bars of varying heights spread across the locations. For the *colour* group and *angle* group, in *spread-only* trials, we displayed a pair of parallel vertical lines at each location. In *spread-only* trials for the *colour* and *angle* groups, the pairs of lines were all grey. In *colour-spread* and *angle-spread* trials the pairs of lines spread across the screen were each assigned a colour by the same method as the *colour-only* cue or an angle size by the same method as the *angle-only* cue, respectively. In *spread-only* trials for the *shape* group, we displayed a grey square at each location. In *shape-spread* trials, eight ovals with varying axis ratios were shown at the different locations.

We used location estimation, with the spread of the stimuli being the familiar cue, as a framework to test for novel-familiar combination as this framework has been used multiple times to test the perceptual system's ability to learn novel stimulus distributions, or location priors (Bejjanki et al., 2016; Chambers et al., 2018; Kiryakova et al., 2020; Körding & Wolpert, 2004; Tassinari et al., 2006; Vilares et al., 2012). Those studies suggest that the spread of stimuli is an intuitive familiar cue to location that observers readily understand and can flexibly weight in relation to the mean of a novel location prior. We expect this to extend to combination with a novel cue.

The standard deviation of the Gaussian distribution from which the eight stimulus values were drawn varied for each novel cue. The variation was needed to account for the fact that the ability of participants to average the eight stimulus values varied with novel cue type. For example, in pilot testing participants produced more precise colour estimates from the eight pairs of lines than they did angle estimates from the eight angles. This led us to set a higher standard deviation for the Gaussian governing the colour cue than the Gaussian governing the angle cue so that variability using the two cues was better matched. The values that we used were determined in pilot testing and set such that, on average across pilot participants, variability using each novel cue and the familiar cue alone was roughly matched.

Task Parameters

In the training block there were two repeats of each of 36 possible hidden locations (15% to 85% of the way across the screen from left to right, sampled every 2%) for both the "with mapping" and "without mapping" trials (72 trials of each type). In the test block, the same 36 unique hidden locations were used, with each repeated five times for each trial type (e.g., *colour-only, spread-only,* and *colour-spread* for the *colour* group; 180 trials each). Trials of all types were interleaved and presented in a random order.

Data Analysis

Any response that was issued less than 500 ms after presentation of the cue(s) was considered a lapse and excluded from analysis. Detection of lapses was not performed online, but post-hoc in data analysis. Thus, participants were not informed when a response was classified as a lapse. To check that observers could use the cue(s), we calculated the correlation coefficient between the responses and the hidden location for each trial type (e.g., *colour-only, spread-only*, and *colour-spread* for the *colour* group) and for each observer within each session. Our *a priori* learning criteria were as follows. If $r \geq 0.7$ (Pearson's correlation) for all trial types within a session for a given observer, we conclude that the observer learned to use the cue(s) and they are included in all analyses including data from that session. However, if $r < 0.7$ for any trial type in a session, we conclude that the observer did not learn to use the cue(s) well enough, and they are excluded from analyses involving that session.

Our main research questions were: (1) do observers combine the novel and familiar cues to increase precision above what is possible using the most reliable single cue alone, and (2) if so, does the gain in precision using both cues compared to the best single cue differ from the optimal or maximum gain predicted by reliability-weighted averaging? Thus, our main measure of interest is precision or, equivalently, variability. We calculate measures of variability according to a method we recently described elsewhere (Aston et al., 2021). The method is designed to account for central biases in continuous responses that may reduce statistical power for detecting a gain in precision using multiple cues. To calculate measures of variability according to the method, we regress responses for each trial type on the true hidden object locations and calculate the standard deviation of the residuals. If the slope of the fitted regression line is significantly less than one, the standard deviation of the residuals is divided by the fitted slope of the regression line to correct for a central bias. Importantly, if there is no evidence of a central bias (the slope is not significantly less than one), no correction is performed. The mean strengths of the central bias for each trial type in the third session of each task (averaged across sessions and observers) were: *colour-only* $\beta = 0.04$, *angle-only* $\beta = 0.06$, *shape-only* $\beta = 0.05$, *height-only* $\beta = 0.1$, *spread-only* (*colour* group) $\beta = 0.07$, *spread-only* (*angle* group) $\beta = 0.07$, *spread-only* (*shape* group) $\beta = 0.08$, *spread-only* (*height* group) $\beta = 0.08$, *colour-spread* $\beta = 0.04$, *angle-spread* $\beta = 0.02$, *shape-spread* $\beta = 0.03$, and *height-spread* $\beta = 0.04$.

We will refer to our measures of variability as variable error. Our second main research question requires the comparison of variable error using both cues to the optimal prediction under the assumption of reliability-weighted averaging. Given variable errors for two single cues, σ_1 and σ_2, we can predict the optimal variable error using both cues, σ_b, using the equation below (Ernst & Banks, 2002).

$$\sigma_b^2 = \frac{\sigma_1^2 \sigma_2^2}{(\sigma_1^2 + \sigma_2^2)}$$

Pilot Experiment and Power Analysis

Five observers (4 female, age range 18-24 years) completed a pilot experiment using the novel colour cue to location. By the third session of the experiment, all five observers issued less variable (more precise) responses in the novel-familiar cue trials compared to trials where they used their most reliable cue alone. The mean reduction in variable error in the third session (in terms of screen proportion) was 0.013 with standard deviation 0.013. Based on this pilot data, we used G*Power (Faul et al., 2007) to calculate the statistical power that different sample sizes would allow for our most important research question: do observers issue less variable (more precise) responses using the novel and familiar cues together compared to the most reliable, or best, single cue. We planned to address this question by comparing variable error using the best single cue to variable error using the novel and familiar cues together using a one-tailed Wilcoxon signed-rank test. Based on the pilot data, we required 9 participants for 80% power. We chose to recruit ten observers for each novel cue type in the main experiment.

Open Practices Statement

This experiment was not pre-registered. The raw data files and analysis script are available online at https://osf.io/gj92a/.

Experiment 1: Results

Each row of plots in Figure 2 shows the data that pertains to a single group of observers. The top row shows data from the *colour* group, the second row is the *angle* group, the third is the *shape* group, and the bottom row is the *height* group. The left panel of plots shows variable error using the familiar and novel cues alone across sessions (Figure 2A-D). These plots show that variable error using the familiar cue is stable across sessions for all groups of observers but that some groups get better using the novel cue with increased training and exposure to the task. The right panel of plots shows variable error in each session using the worst of the two single cues (highest variable error), the best of the two single cues (lowest variable error), both cues together, and the optimal variable error using both cues together that would be achieved by taking a reliability-weighted average of estimates from the two single cues (Figure 2E-H). A visual inspection of Figure 2E-H shows lower median variable error using both cues together than the best single cue in all groups by the third session of the experiment, suggesting all groups of observers combined the newly learned novel cue with the familiar cue. However, the median variable errors using both cues are all higher than the optimal variable error from reliability-weighted averaging, suggesting that combination of novel and familiar cues was still suboptimal.

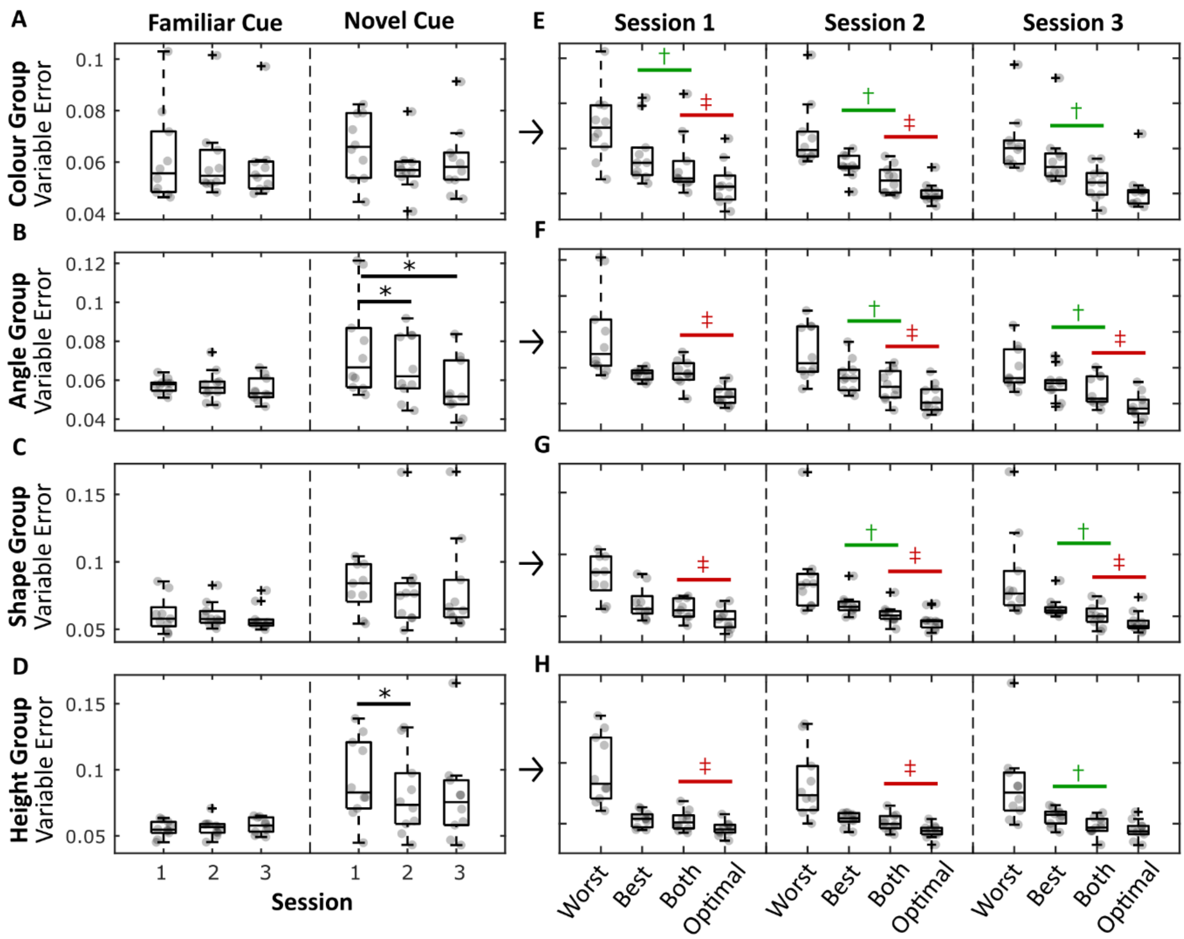

*Figure 2: Results of Experiment 1. (A-D) Variable errors using the familiar and novel cues alone for each group of observers across sessions. (E-H) Variable errors for each group of observers in each session using the worst single cue (novel or familiar), the best single cue (familiar or novel), both cues together, and the optimal variable error that could be achieved using both together by taking a reliability-weighted average of estimates from each single cue. The whiskers of the boxplots extend to adjacent values (the most extreme data points that are not more than 1.5 times the interquartile range above or below the upper and lower quartiles or that are not outliers). Outliers are indicated by black crosses and the black line across the box is the median value. Grey circles show individual variable errors for each observer. * indicates significant difference at the 5% significance level when testing for a difference in variable error across sessions. † indicates significant difference at the 5% significance level when testing for evidence of combination (best > both). ‡ indicates significant difference at the 5% significance level when testing for a difference from optimal (both ≠ optimal).*

Observers quickly learned to use the novel cues, and variability using the cues decreased with repeated training and exposure to the task

Thirty-eight of thirty-nine observers passed the *a priori* learning criteria in the first session of the experiment and each following session. To pass the learning criteria, an observer was required to show a correlation coefficient greater than 0.7 between their responses and the hidden target locations for each trial type. One observer's data from the first session (in the *shape* group) was lost as the computer crashed while the data was saving. That observer passed the learning criteria in both subsequent sessions. The remaining observer (in the *angle* group) also passed the learning criteria in the second and third sessions. Thus, observers quickly learned the mappings between the novel cues and location and could use the novel cues to complete the task.

We were interested in whether the observers' performance changed over the sessions as they gained more practice with the novel cues. To address this question, we performed a Friedman's Test to compare variable errors over time (session number was the independent variable) for each group separately. We used a Friedman's Test as variable errors were not normally distributed and, as the test relies on ranking the data rather than absolute values, does not depend on the measure of variable error that we use (we chose to use standard deviation, but could have used variance instead, leading to increased absolute differences between conditions). Both the *angle* group and *height* group significantly reduced their variable error over time using the novel cues (*angle* group: $\chi^2(2) = 10.4, p = .006$, Figure 2B; *height* group: $\chi^2(2) = 8.6, p = .014$, Figure 2D). Variable error using the angle size cue significantly decreased from sessions one to three ($W = 54, p = .004$) and two to three ($W = 53, p = .006$) in the *angle* group. Variable error using the bar height cue significantly decreased from sessions one to two ($W = 51\ p = .014$) for the *height* group. There was no change in variable error using the novel cue over time for the *colour* or *shape* groups (*colour* group: $\chi^2(2) = 1.4, p = .497$, Figure 2A; *shape* group: $\chi^2(2) = 2.89, p = .236$, Figure 2C); although we note that the median variable error reduces from 0.084 in session one to 0.064 in session three for the *shape* group with the lack of significance likely caused by the outlier values in sessions two and three (Figure 2C).

Variable error using the familiar spread cue did not change over time for any group of participants (*colour* group: $\chi^2(2) = 1.4, p = .497$, Figure 2A; *angle* group: $\chi^2(2) = 1.4, p = .497$, Figure 2B; *shape* group: $\chi^2(2) = 4.67, p = .0.97$, Figure 2C; *height* group: $\chi^2(2) = 2.4, p = .301$, Figure 2D).

Novel cues were combined with the familiar cue by, at most, the third session, but combination was often suboptimal

Recall that our main research questions were: (1) do observers combine the novel and familiar cues to increase precision above what is possible using the most reliable single cue alone, and (2) if so, does the gain in precision using both cues compared to the best single cue differ from the optimal or maximum gain predicted by reliability-weighted averaging? To answer (1), we performed a one-tailed Wilcoxon Signed-Rank test comparing variable error with the best of the novel and familiar cues to performance with both cues together for each group in each session of the experiment. If variable error using both cues was significantly less than variable error using the best single cue, we conclude that the observers in that group and session showed evidence of combination (green dagger and lines in Figure 2). To answer (2), we performed a two-tailed Wilcoxon Signed-Rank test comparing variable error using both cues to the optimal prediction (calculated from measured variable error using each single cue alone). If variable error using both cues differed significantly from the optimal prediction, we concluded that the observers in that group and session were, on the hole, sub-optimal (red double dagger and lines in Figure 2). If not, we conclude that they optimally combined the novel and familiar cues.

In the first session, only the *colour* group showed evidence of combination and all groups were suboptimal (rows 1-4 of Table 1; third column of plots in Figure 2). In the second session, all except the *height* group showed evidence of combination, but all groups remained suboptimal (rows 5-8 of Table 1; forth column of plots in Figure 2). In the third session, all groups showed evidence of combination, with only the *angle* and *shape* groups remaining suboptimal (rows 9-12 of Table 1; fifth column of plots in Figure 2).

Table 1: Statistical tests for evidence of combination and a difference from optimal for each group in each session of Experiment 1. A one-tailed Wilcoxon Signed-Rank test was used to test for evidence of combination and a two-tailed test was used to test for a difference from optimal. The columns "Best > Both" and "Both > Optimal" show the number of participants whose individual data satisfy the inequality out of the total number of participants included in the analysis of that session for that group.

Row No.	**Group**	**Session**	**Best > Both**	W	p	**Combine?**	**Both > Optimal**	W	p	**Suboptimal?**
1	*Colour*	1	8/10	51	.007	**Yes**	9/10	53	.006	**Yes**
2	*Angle*	1	4/10	20	.784	**No**	10/10	55	.002	**Yes**
3	*Shape*	1	7/9	36	.064	**No**	9/9	45	.004	**Yes**
4	*Height*	1	5/10	31	.385	**No**	10/10	55	.002	**Yes**
5	*Colour*	2	10/10	55	.001	**Yes**	10/10	55	.002	**Yes**
6	*Angle*	2	8/10	49	.014	**Yes**	9/10	54	.004	**Yes**

7	*Shape*	2	10/10	55	.001	Yes	8/10	50	.02	Yes
8	*Height*	2	6/10	38	.161	**No**	9/10	53	.006	Yes
9	*Colour*	3	10/10	55	.001	Yes	7/10	43	.131	**No**
10	*Angle*	3	7/10	47	.024	Yes	10/10	55	.002	Yes
11	*Shape*	3	9/10	49	.014	Yes	7/10	49	.027	Yes
12	*Height*	3	9/10	54	.002	Yes	7/10	38	.322	**No**

Experiment 1: Summary

In Experiment 1, we showed that observers can combine newly learned novel cues (colour, angle size, shape, and the height of a bar) to horizontal location with a familiar cue (a dot cloud) to improve location estimate precision. Variable error using the novel cues alone decreased across sessions, likely due to extra training and increased exposure to the task. Importantly, by the third session of the experiment, all four groups of observers had significantly lower variable error using the novel and familiar cues together compared to their best single cue (35/40 observers were better with both cues than their best single cue in total across the groups in the third session), a feature of integration of familiar cues. For two groups of observers, those who learned the colour and height cues, variable error using the novel and familiar cues together in the third session was not significantly different to the optimal variable error of an ideal observer who takes a reliability-weighted average of estimates from the two single cues.

These findings complement the limited number of previous studies showing that the human perceptual system can combine newly learned novel cues with familiar cues to improve precision. They extend the previous results to instances where observers must learn to use abstract novel cues to aid estimates of horizontal position on a computer screen.

In Experiment 2, we tested whether observers would also combine two newly learned novel cues (colour and angle size or colour and shape) to location with each other, as well as with a familiar cue (dot cloud).

Experiment 2: Methods

Overview

Two separate groups, each of ten observers, completed a task three times in three separate sessions. The task required the observers to use one of two novel cues, a familiar cue, or two of the cues simultaneously to estimate the location of a hidden target by using a computer mouse to adjust the horizontal position of a bar on a computer screen. As in Experiment 1, the task began a training period. However, there were now two blocks of training trials that taught observers the mapping between each novel cue and location separately. Observers completed the two novel cue training blocks in a random order. They were identical to the training blocks in Experiment 1 (Figure 1).

After observers completed both novel cue training blocks, the test trials began (Figure 3). At the start of the test block, observers were instructed that they would now begin to use the newly learnt novel cues, along with a familiar cue (a dot-cloud, or the spread cue) to estimate the location of a hidden object – an octopus hiding in the sea. The two different groups of ten observers (the *colour-angle-spread* group and the *colour-shape-spread* group) saw different combinations of trials.

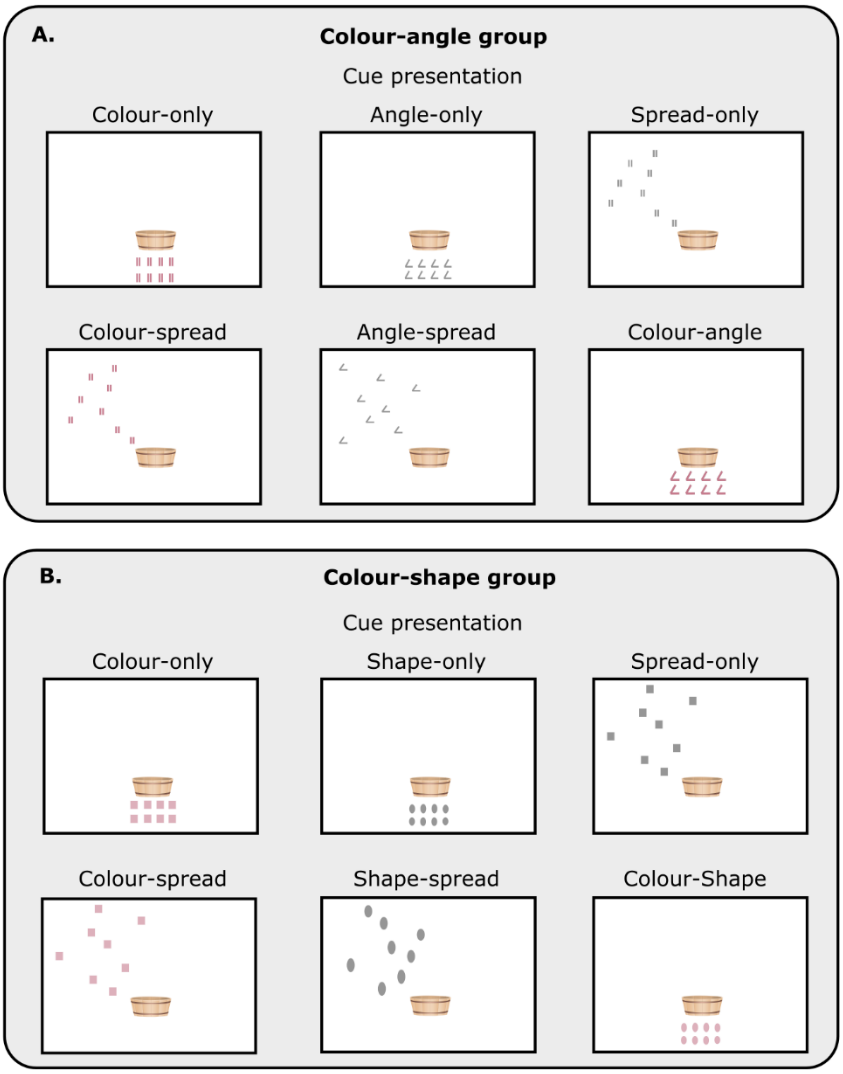

Figure 3: The test trials in Experiment 2. (A-B) In test trials, observers used either one of the newly learned novel cues, a familiar spread cue, both the novel cues together, or one of the novel cues and the familiar cue together to estimate the position of a hidden object (an octopus hiding in the sea).

On each trial, the *colour-angle-spread* group of observers were presented with either the colour cue, angle cue, or spread cue alone (*colour-only*, *angle-only*, or *spread-only* trials), or with a pairing of two cues (*colour-spread*, *angle-spread*, or *colour-angle* trials). In *colour-only* and *angle-only* trials, observers were presented with eight pairs of lines (in fixed positions) at the bottom of the screen. The average colour of the pair of lines or angle between them provided a novel estimate of location according to the trained mappings. In *spread-only* trials, eight pairs of parallel and grey lines (no novel cue information) were spread out across the screen. The position of each pair of lines was drawn from a Gaussian distribution, centred on the hidden location, such that the mean or centroid of the locations was the best estimate. In *colour-spread* or *angle-spread* trials, the eight pairs of lines were spread across the screen and had the property of the novel cue (either the relevant colours or angles between the lines). In *colour-angle* trials, the eight pairs of lines appeared in their fixed positions at the bottom of the screen and had the property of both novel cues (both the relevant colours and angles between the lines).

The *colour-shape-spread* group of observers also experienced the *colour-only, spread-only*, and *colour-spread* trials, with the small difference that cues were no longer presented as pairs of lines but as grey or coloured squares. This group of observers also experienced *shape-only, shape-spread*, or *colour-shape* trials. In *shape-only* and *colour-shape* trials, observers were presented with eight ovals (in fixed positions) at the bottom of the screen. Either the average axis ratio of the ovals alone (*shape-only* trials) or both the average axis ratio and colour of the ovals (*colour-shape* trials) provided a novel estimate of location according to the trained mappings. In *shape-spread* trials, the eight ovals were spread across the screen and had the property of the novel cue (the relevant axis ratios).

For both groups of observers, trials of all types were interleaved. After the cue(s) appeared on each trial, observers adjusted the horizontal position of a vertical line, using a mouse, to their best guess of the hidden location. Feedback was given indicating if the observers had “caught” the octopus along with an indicator of the true hidden location that displayed the corresponding novel cue values (the colour or angle size, or the colour and shape). If the octopus was caught, an animation showed the octopus move across the screen from its hidden location to the bucket.

Observers

Ten observers were recruited for the *colour-angle-spread* group (6 female, age range 22-28 years) and ten for the *colour-shape-spread* group (9 female, age range 19-36 years) using Durham Psychology Department's Participant Pool programme or through word of mouth. All observers had normal or corrected to normal visual acuity (self-report) and no colour vision deficiencies (assessed using Ishihara Colour Plates). Each observer was given either £8 per hour or participant pool credits for their time. All observers gave written, informed consent prior to taking part in the study. Ethical approval was received from the Durham University Psychology Department Ethics Board (reference number: 17/07).

Apparatus and Stimuli

The apparatus and stimuli were the same we have already described for Experiment 1.

Task Parameters

In the colour, angle, and shape cue training blocks there were two repeats of each of 36 possible hidden locations (15% to 85% of the way across the screen from left to right, sampled every 2%) for both the "with mapping" and "without mapping" trials (72 trials of each type). In the test block, the same 36 unique hidden locations were used, with each repeated three times for each trial type (*colour-angle-spread* group: *colour-only, angle-only, spread-only, colour-spread, angle-spread, colour-angle*; *colour-shape-spread* group: *colour-only, shape -only, spread-only, colour-spread, shape-spread, colour- shape*; 108 trials each). Trials of all types were interleaved and presented in a random order.

Data Analysis

The analysis procedure was identical to Experiment 1. The mean strengths of the central bias for each trial type in the third session for the *colour-angle-spread* group (averaged across sessions and observers), where zero would indicate no bias and larger numbers indicate increasing bias, were: *colour-only* $\beta = 0.1$, *angle-only* $\beta = 0.05$, *spread-only* $\beta = 0.09$, *colour-spread* $\beta = 0.06$, *angle-spread* $\beta = 0.02$, and *colour-angle* $\beta = 0.01$. The mean strengths of the central bias for each trial type in the third session for the *colour-shape-spread* group were: *colour-only* $\beta = 0.13$, *shape-only* $\beta = 0.11$, *spread-only* $\beta = 0.1$, *colour-spread* $\beta = 0.05$, *shape-spread* $\beta = 0.05$, and *colour-shape* $\beta = 0.01$.

Open Practices Statement

This experiment was not pre-registered. The raw data files and analysis script are available online at https://osf.io/gj92a/.

Experiment 2: Results

Each row of plots in Figure 4 shows the data that pertains to each possible cue pairing for the *colour-angle-spread* group. In the top row, we plot data from the *colour-only, spread-only,* and *colour-spread* trials. In the second row, we plot data from the *angle-only, spread-only*, and *angle-spread* trials. In the third row, we plot data from the *colour-only, angle-only*, and *colour-angle* trials. The left panel of plots shows variable error using the familiar and novel cues alone across sessions (Figure 4A-C). These plots show that variable error using the familiar spread cue and novel colour cue is stable across sessions but that observers get better using the novel angle cue with increased training and exposure to the task. The right panel of plots shows variable error in each session using the worst of the two single cues (highest variable error), the best of the two single cues (lowest variable error), both cues together, and the optimal variable error using both cues together that would be achieved by taking a reliability-weighted average of estimates from the two single cues (Figure 4D-F). A visual inspection of Figure 4D-F suggests that the median variable error using both cues together may be lower than the best single cue in the third session of the experiment when using the angle and spread cues together but not the other pairs of cues. We also see that the median variable errors using both cues are all higher than the optimal variable error from reliability-weighted averaging, suggesting that even if some pairing of cues resulted in combination, the combination was suboptimal.

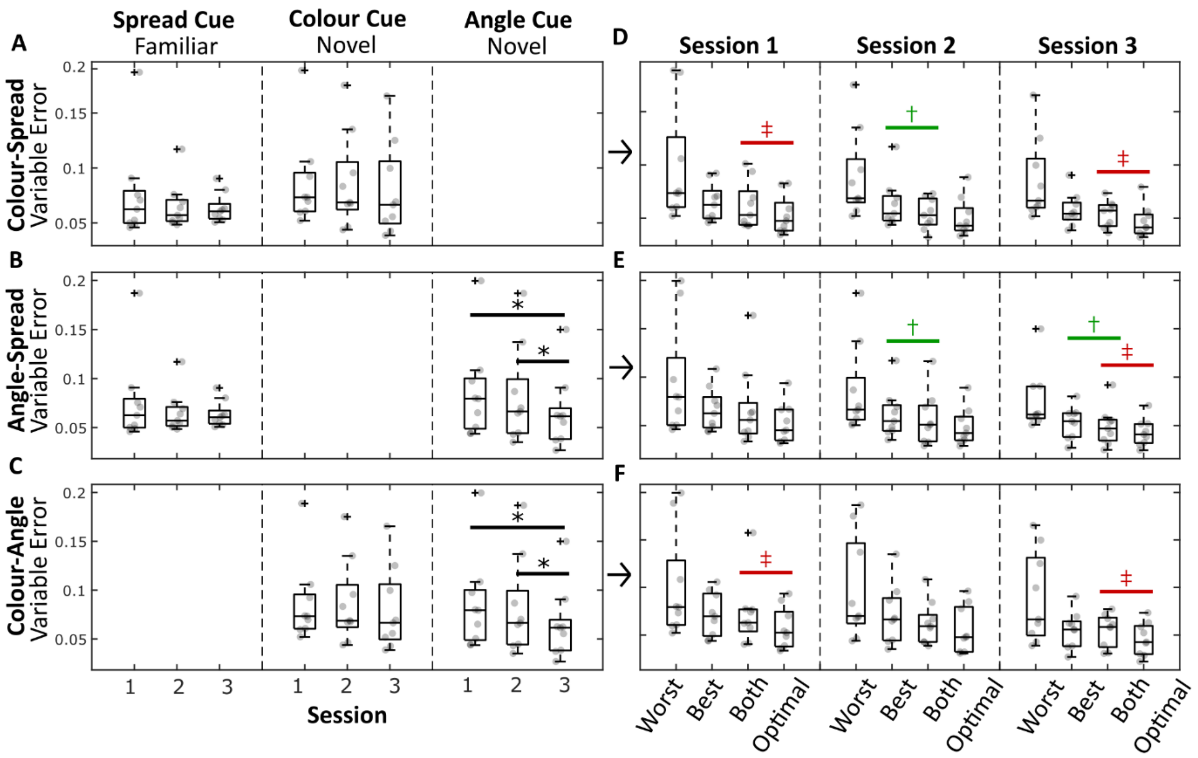

*Figure 4: Results of the colour-angle-spread group in Experiment 2. (A-C) Variable errors using the familiar and novel cues alone for each group of observers across sessions. (D-F) Variable errors for each group of observers in each session using the worst single cue, the best single cue, both cues together, and the optimal variable error that could be achieved using both together by taking a reliability-weighted average of estimates from each single cue. The whiskers of the boxplots extend to adjacent values (the most extreme data points that are not more than 1.5 times the interquartile range above or below the upper and lower quartiles or that are not outliers). Outliers are indicated by black crosses and the black line across the box is the median value. Grey circles show individual variable errors for each observer. * indicates significant difference at the 5% significance level when testing for a difference in variable error across sessions. † indicates significant difference at the 5% significance level when testing for evidence of combination (best > both). ‡ indicates significant difference at the 5% significance level when testing for a difference from optimal (both ≠ optimal).*

Figure 5 shows the data in the same way for the *colour-shape-spread* group. These plots show that variable error using all cues was stable across sessions for this group of observers (Figure 5A-C). A visual inspection of Figure 5D-F suggests that the median variable error using both cues together may be lower than the best single cue in the second and third session for all cue pairs and that median variable errors using both cues seem to approach the optimal variable error from reliability-weighted averaging, suggesting combination may be optimal for this group of observers.

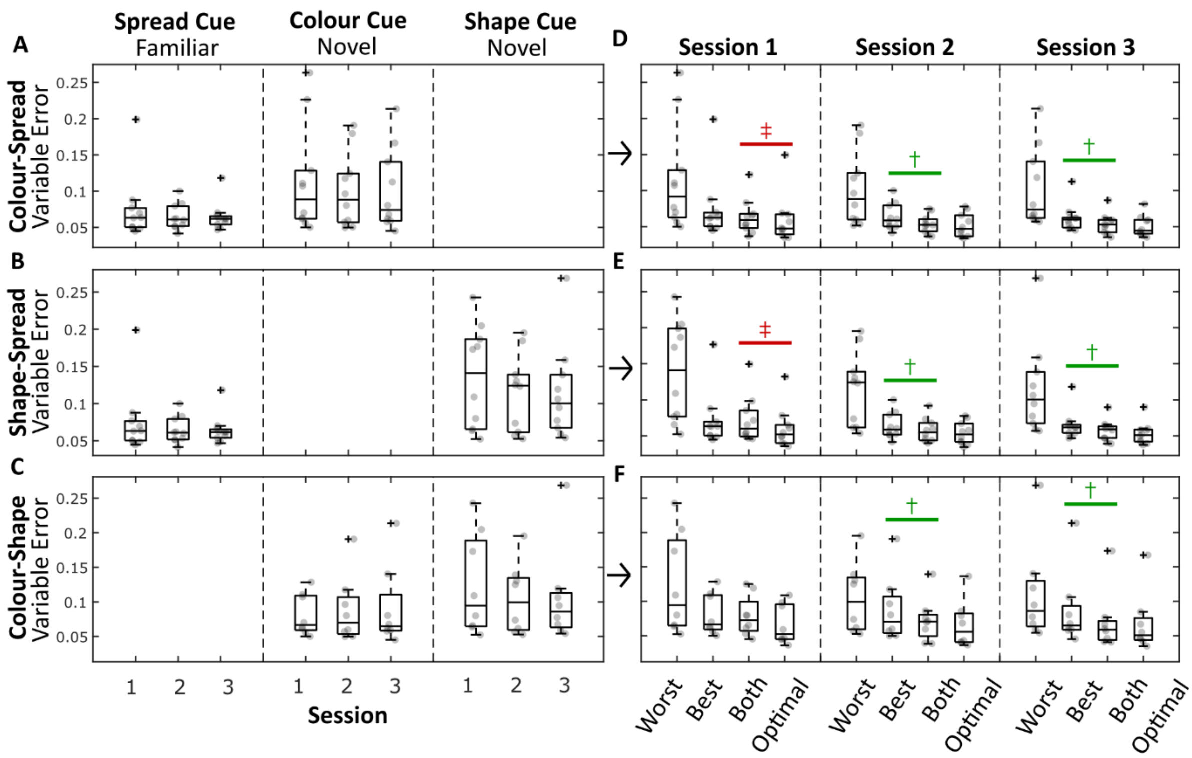

*Figure 5: Results of the colour-shape-spread group in Experiment 2. (A-C) Variable errors using the familiar and novel cues alone for each group of observers across sessions. (D-F) Variable errors for each group of observers in each session using the worst single cue, the best single cue, both cues together, and the optimal variable error that could be achieved using both together by taking a reliability-weighted average of estimates from each single cue. The whiskers of the boxplots extend to adjacent values (the most extreme data points that are not more than 1.5 times the interquartile range above or below the upper and lower quartiles or that are not outliers). Outliers are indicated by black crosses and the black line across the box is the median value. Grey circles show individual variable errors for each observer. * indicates significant difference at the 5% significance level when testing for a difference in variable error across sessions. † indicates significant difference at the 5% significance level when testing for evidence of combination (best > both). ‡ indicates significant difference at the 5% significance level when testing for a difference from optimal (both ≠ optimal).*

Observers quickly learned to use the novel cues and variability using some of the cues decreased with repeated training and exposure to the task in the colour-angle-spread group

Nine of the ten *colour-angle-spread* observers passed the learning criterion in all three sessions of the experiment. The remaining observer passed the learning criterion in the second and third sessions. Six of the ten *colour-shape-spread* observers passed the learning criterion in all three sessions. Of the remaining four, three of them passed the criterion in the second and third sessions, but one only passed the learning criterion in the second but not third session. Thus, overall, observers quickly learned the mappings between the novel cues and location and could use the novel cues to complete the task.

The *colour-angle-spread* observers reduced their variable error over time using the colour cue ($\chi^2(2) = 6.89, p = .032$, Figure 4A) and angle cue ($\chi^2(2) = 14.6, p = .001$, Figure 4B), but not the spread cue ($\chi^2(2) = 2.89, p = .236$, Figure 4A). Using the angle cue, variable errors reduced significantly from session one to three ($W = 55, p = .002$) and two to three ($W = 54, p = .004$). None of the pairwise comparisons were significant for the colour cue, but the median variable error showed the same trend of reducing across sessions.

The *colour-shape-spread* observers did not reduce variable error over time for any of the cues (spread cue: $\chi^2(2) = 1.8, p = .407$, Figure 5A; colour cue: $\chi^2(2) = 0.25, p = .882$, Figure 5B; shape cue: $\chi^2(2) = 1, p = .607$, Figure 5C)

Novel and familiar cues were consistently combined in the colour-shape-spread group but not the colour-angle-spread group, and novel colour and shape cues were combined while novel colour and angle cues were not

Table 2 summarises the results for the *colour-angle-spread* group. In the first session, this group did not show evidence of combination for any cue pairing but were only suboptimal in *colour-spread* and *colour-angle* trials (rows 1-3 in Table 2; Figure 5). In the second session, they showed evidence of combination in *colour-spread* and *angle-spread* trials but not *colour-angle* and did not differ from optimal for any trial type (rows 4-6 in Table 2; Figure 5). In the third session, the *colour-angle-spread* group only showed evidence of combination in *angle-spread* trials and were suboptimal in all trial types (rows 7-9 in Table 2; Figure 5).

Table 2: Statistical tests for evidence of combination and a difference from optimal for the colour-angle-spread group in Experiment 2. A one-tailed Wilcoxon Signed-Rank test was used to test for evidence of combination and a two-tailed test was used to test for a difference from optimal. The columns "Best > Both" and "Both > Optimal" show the number of participants whose individual data satisfy the inequality out of the total number of participants included in the analysis of that session.

Row No.	**Cue Pairing**	**Session**	**Best > Both**	W	p	**Combine?**	**Both > Optimal**	W	p	**Suboptimal?**
1	*Colour-spread (N-F)*	1	7/9	34	.102	**No**	9/9	45	.004	**Yes**
2	*Angle-spread (N-F)*	1	7/9	30	.213	**No**	8/9	38	.074	**No**
3	*Colour-angle (N-N)*	1	6/9	27	.326	**No**	8/9	40	.039	**Yes**
4	*Colour-spread (N-F)*	2	7/10	48	.019	**Yes**	8/10	43	.131	**No**
5	*Angle-spread (N-F)*	2	9/10	45	.042	**Yes**	7/10	43	.131	**No**
6	*Colour-angle (N-N)*	2	7/10	36	.216	**No**	8/10	41	.193	**No**
7	*Colour-spread (N-F)*	3	6/10	42	.08	**No**	8/10	47	.049	**Yes**

8	*Angle-spread (N-F)*	3	9/10	45	.042	Yes	9/10	53	.006	Yes
9	*Colour-angle (N-N)*	3	3/10	22	.722	**No**	9/10	54	.004	Yes

Table 3 summarises the results for the *colour-shape-spread* group. In the first session, this group also did not show evidence of combination for any cue pairing but were only suboptimal in *colour-spread* and *shape-spread* trials (rows 1-3 in Table 3; Figure 6). In the second session, they showed evidence of combination and did not differ from optimal for any trial type (rows 4-6 in Table 3; Figure 6). This was also true in the third session (rows 7-9 in Table 3; Figure 6).

Table 3: Statistical tests for evidence of combination and a difference from optimal for the colour-shape-spread group in Experiment 2. A one-tailed Wilcoxon Signed-Rank test was used to test for evidence of combination and a two-tailed test was used to test for a difference from optimal. The columns "Best > Both" and "Both > Optimal" show the number of participants whose individual data satisfy the inequality out of the total number of participants included in the analysis of that session.

Row No.	**Cue Pairing**	**Session**	**Best > Both**	W	p	**Combine?**	**Both > Optimal**	W	p	**Suboptimal?**
1	*Colour-spread (N-F)*	1	5/8	28	.098	**No**	8/8	36	.008	Yes
2	*Shape-spread (N-F)*	1	5/8	23	.273	**No**	8/8	36	.008	Yes
3	*Colour-shape (N-N)*	1	4/6	13	.344	**No**	5/6	18	.156	**No**
4	*Colour-spread (N-F)*	2	8/10	51	.007	Yes	5/10	32	.695	**No**
5	*Shape-spread (N-F)*	2	9/10	53	.003	Yes	9/10	46	.064	**No**
6	*Colour-shape (N-N)*	2	8/10	51	.007	Yes	8/10	46	.064	**No**

7	*Colour-spread (N-F)*	3	9/10	51	.007	**Yes**	6/10	42	.16	**No**
8	*Shape-spread (N-F)*	3	8/9	37	.049	**Yes**	6/9	39	.055	**No**
9	*Colour-shape (N-N)*	3	9/9	45	.002	**Yes**	6/9	25	.82	**No**

Experiment 2: Summary

We found that observers quickly learned to use the novel cues to location. Although use of some novel cues improved over time (location estimate variability reduced), observers were able to use the cues in the first session of the experiment, implying that they had leaned the association after only a small number of training trials. Observers were able to combine the newly learned novel cues with a familiar cue to improve precision (reduce variability) regardless of the pair of cues that they learned, but combination of novel and familiar cues was inconsistent for the *colour-angle-spread* group and often suboptimal. While the colour-shape group combined the two novel cues with each other to improve precision, the *colour-angle-spread* group did not.

General Discussion

It is clear that a mature perceptual system can learn new mappings between novel cues and properties of the environment (Di Luca et al., 2010; Ernst, 2007; Haijiang et al., 2006; Harrison & Backus, 2012; Michel & Jacobs, 2008; Negen et al., 2018), with a limited number of studies suggesting that novel cues can be integrated into the normal perceptual experience by combining them with familiar cues in a “Bayes-like” way to increase perceptual precision (Ernst, 2007; Gibo et al., 2017; Michel & Jacobs, 2008; Negen et al., 2018). Here, we trained observers to use abstract novel cues to estimate the horizontal location of hidden objects on a computer screen. In Experiment 1, observers benefitted from a suboptimal but significant gain in precision using novel and familiar cues together, extending previous reports of novel-familiar cue combination. We found evidence of a reduction in variable error from combining novel and familiar cues in the third session of the experiment for all four of the abstract novel cues we tested. In Experiment 2, we tested for the first time whether two novel cues may also be combined with each other. We found that one pair of novel cues could be combined to improve precision but the other could not, even after three sessions of repeated training. Taken together, our results add to the current literature on the integration of novel cues into the normal perceptual experience by showing that abstract novel cues to location are quickly learned and combined with familiar cues to increase perceptual precision, but that whether two novel cues to location are combined may depend on the choice of cues.

Why might some pairs of novel cues be easier to combine than others?

Whether or not two cues are combined can depend on the strength of the belief that the two cues are coupled (Ernst, 2006) or that they come from the same source (Körding et al., 2007). It is possible that, in Experiment 2, the colour-shape group were able to combine the two novel cues, but the colour-angle group were not because our observers were more likely to expect a coupling or correspondence between colour and shape than they were between colour and angle size. There are many natural associations between different shapes and colours, but it is harder to think of similar associations between different angle sizes and colours. Indeed, in the colour perception literature there several reports of object shape modulating colour perception, such as when a grey banana appears slightly yellow (Hansen et al., 2006; Olkkonen et al., 2008; Witzel et al., 2011; Witzel & Hansen, 2015), an effect that can also be conceptualised within a reliability-weighted averaging framework where shape is an extra cue to colour (Witzel et al., 2018). This could explain why observers combined colour and shape cues but not colour and angle size cues in Experiment 2.

Why is combination of novel and familiar cues often suboptimal?

To take a reliability-weighted average of novel and familiar cues, observers must learn the novel cue's reliability. Obtaining an accurate estimate of the novel cue's reliability may require more time (feedback) than is offered in our experiments. In contrast, this is not an issue in experiments where an observer is presented with two familiar cues, where we can expect that, through a lifetime of repeated exposure, they have good internal estimates of the cue reliabilities. Such an explanation is in line with the inability of children to combine cues before the age of 10 (Gori et al., 2008; Nardini et al., 2010) unless they receive explicit training (Negen et al., 2019). In our task, variable error using some of the novel cues decreases over time, so not only might repeated exposure be needed to develop good internal estimates of the cue reliabilities, but the learning the correct reliabilities is made harder by the fact that they are still to stabilise.

Another possibility is that optimal combination in not possible for the type of information provided to observers in our task. In classic cue combination experiments, low-level sensory cues are combined to increase perceptual precision and enhance discrimination (Alais & Burr, 2004; Ernst & Banks, 2002; Knill & Saunders, 2003). In other words, observers are able to account for low-level sensory noise when combining cues. However, there is evidence to suggest that the brain may not be able to perform the same calculation across more complex, higher-level information (Jarvstad et al., 2014; Summerfield & Tsetsos, 2012; Wu et al., 2009). Indeed, the results of a recent study suggest that as we displayed the novel cues in our experiments in a way that required "cognitive integration" of the eight novel stimulus values, this could cause the suboptimalities we see in our data (Castañón et al., 2019), see also Dakin et al. (2005). However, we must also note that even low-level sensory cue combination is not always optimal (Rahnev & Denison, 2018).

Limitations

As explained in the methods section, the standard deviation of the Gaussian distribution from which the eight stimulus values were drawn varied for each novel cue to account for the fact that the ability of observers to average the eight stimulus values varied with novel cue type. We determined the different values that we used in pilot testing such that, on average across pilot observers, variability using each novel cue and the familiar cue alone was roughly matched. As can be seen in Figure 2, the values that we used did not transfer across observer groups. The values that worked in piloting to match cue variabilities did not extend to the main experiments, where observers were generally worse with the novel cues compared to the familiar cue. Future experiments could attempt to match the cue variabilities better by scaling the cues individually for each observer based on some pre-testing.

In a previous paper, we discussed the issues surrounding the use of continuous responses to test for combination of multiple cues using measures of variability (Aston et al., 2021). That paper focused on the need to account for central biases in continuous responses and how that could be done, introducing a method we adopted in the analyses of the data presented here. In that paper, we also discussed the effects of additional response noise (e.g., motor noise). We showed that if the additional noise is equivalent across all trial types (single and combined cue trials), then it does not disrupt a researcher's ability to detect a reduction in variability using both cues compared to the best single – what we termed the "combination effect" (see equation 3 in Aston et al., 2021). However, the equivalence between the optimal prediction and measured variability using both cues (where the optimal prediction is calculated from the measured single cue variabilities) is not preserved. Specifically, the calculated optimal prediction will suggest that variability could be lower than is possible (see footnote 3 in Aston et al., 2021). Here, this means that while we can be confident in our ability to detect a reduction in variability using both cues compared to the best single cue, we cannot be confident in our ability to test for optimal combination (or deviance from it), as our optimal predictions may be lower than can be achieved by our observers. Future experiments could seek to separate out measures of variability in continuous response data into the parts due to sensory error and additional sources of noise. For more discussion of early vs late or motor noise during cue combination, see Hillis et al. (2004) and Knill and Saunders (2003).

Conclusion

Overall, our results provide extensive evidence that novel cues can be learned and combined with familiar cues to enhance perception, but mixed evidence for whether perceptual and decision-making systems can extend this ability to the combination of multiple novel cues with only short-term training. Whether the ability can be extended to the case of two novel cues may depend on the choice of cues.

Acknowledgements

We would like to thank Sophie Barnes, Abbey Fletcher, and Josefin Rosman for their help with data collection. We would also like to thank Anya Hurlbert for use of the Konica Minolta CS2000 and James Negen for many useful conversations. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 820185) and a Leverhulme Trust Research Project Grant (RPG-2017-097).

References

Abboud, S., Hanassy, S., Levy-tzedek, S., Maidenbaum, S., & Amedi, A. (2014). EyeMusic : Introducing a " visual " colorful experience for the blind using auditory sensory substitution. *Restorative Neurology and Neuroscience*, *32*, 247–257. https://doi.org/10.3233/RNN-130338

Alais, D., & Burr, D. (2004). Ventriloquist Effect Results from Near-Optimal Bimodal Integration. *Current Biology*, *14*(3), 257–262. https://doi.org/10.1016/S0960-9822(04)00043-0

Aston, S., Negen, J., Nardini, M., & Beierholm, U. (2021). Central tendency biases must be accounted for to consistently capture Bayesian cue combination in continuous response data. *Behavior Research Methods*. https://doi.org/10.3758/s13428-021-01633-2

Auvray, M., Hanneton, S., & O'Regan, J. K. (2007). Learning to perceive with a visuo-auditory substitution system: Localisation and object recognition with "The vOICe." *Perception*, *36*(3), 416–430. https://doi.org/10.1068/p5631

Bach-y-Rita, P., Collins, C. C., Saunders, F. A., White, B., & Scadden, L. (1969). Vision Substitution by Tactile Image Projection. *Nature*, *221*, 963. https://doi.org/10.1038/221963a0

Bejjanki, V. R., Knill, D. C., & Aslin, R. N. (2016). Learning and inference using complex generative models in a spatial localization task. *Journal of Vision*, *16*(2016), 1–13. https://doi.org/10.1167/16.5.9.doi

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 10:433-436. http://www.ncbi.nlm.nih.gov/pubmed/9176952

Castañón, S., Moran, R., Ding, J., Egner, T., Bang, D., & Summerfield, C. (2019). Human noise blindness drives suboptimal cognitive inference. *Nature Communications*, *10*(1), 1719. https://doi.org/10.1038/s41467-019-09330-7

Chambers, C., Sokhey, T., Gaebler-Spira, D., & Kording, K. P. (2018). The development of Bayesian integration in sensorimotor estimation. *Journal of Vision*, *18*(12), 8. https://doi.org/10.1167/18.12.8

Dakin, S. C., Mareschal, I., & Bex, P. J. (2005). Local and global limitations on direction integration assessed using equivalent noise analysis. *Vision Research*, *45*(24), 3027–3049. https://doi.org/https://doi.org/10.1016/j.visres.2005.07.037

Di Luca, M., Ernst, M. O., & Backus, B. T. (2010). Learning to Use an Invisible Visual Signal for Perception. *Current Biology*, *20*(20), 1860–1863. https://doi.org/10.1016/j.cub.2010.09.047

Ernst, M. O. (2006). A Bayesian view on multimodal cue integration. *Human Body Perception from the inside Out*, *131*, 105–131. https://doi.org/10.1016/j.cub.2011.11.039

Ernst, M. O. (2007). Learning to integrate arbitrary signals from vision and touch. *Journal of Vision*, *7*(5), 7. https://doi.org/10.1167/7.5.7

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429–433. https://doi.org/10.1038/415429a

Fahle, M., & Poggio, T. (Eds). (2002). *Perceptual Learning* (M. Fahle & T. Poggio (eds.)). MIT Press, Cambridge, MA.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Gibo, T. L., Mugge, W., & Abbink, D. A. (2017). Trust in haptic assistance: weighting visual and haptic cues based on error history. *Experimental Brain Research*, *235*(8), 2533–2546.

https://doi.org/10.1007/s00221-017-4986-4

Gori, M., Del Viva, M., Sandini, G., & Burr, D. C. (2008). Young Children Do Not Integrate Visual and Haptic Form Information. *Current Biology*, *18*(9), 694–698. https://doi.org/10.1016/j.cub.2008.04.036

Haijiang, Q., Saunders, J. A., Stone, R. W., & Backus, B. T. (2006). Demonstration of cue recruitment: Change in visual appearance by means of Pavlovian conditioning. *Proceedings of the National Academy of Sciences*, *103*(2), 483–488. https://doi.org/10.1073/pnas.0506728103

Hansen, T., Olkkonen, M., Walter, S., & Gegenfurtner, K. R. (2006). Memory modulates color appearance. *Nature Neuroscience*, *9*(11), 1367–1368. https://doi.org/10.1038/nn1794

Harrison, S. J., & Backus, B. T. (2012). Associative learning of shape as a cue to appearance : A new demonstration of cue recruitment. *Journal of Vision*, *12*(2012), 1–17. https://doi.org/10.1167/12.3.15.Introduction

Hillis, J. M., Watt, S. J., Landy, M. S., & Banks, M. S. (2004). Slant from texture and disparity cues: Optimal cue combination. *Journal of Vision*, *4*(12), 1. https://doi.org/10.1167/4.12.1

Jarvstad, A., Hahn, U., Warren, P. A., & Rushton, S. K. (2014). Are perceptuo-motor decisions really more optimal than cognitive decisions? *Cognition*, *130*(3), 397–416. https://doi.org/10.1016/j.cognition.2013.09.009

Kiryakova, R. K., Aston, S., Beierholm, U. R., & Nardini, M. (2020). Bayesian transfer in a complex spatial localization task. *Journal of Vision*, *20*(6), 17. https://doi.org/10.1167/jov.20.6.17

Kleiner, M., Brainard, D. H., & Pelli, D. (2007). *What's new in Psychtoolbox-3?* Perception 36 ECVP Abstract Supplement.

Knill, D. C., & Saunders, J. A. (2003). Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vision Research*, *43*(24), 2539–2558. https://doi.org/10.1016/S0042-6989(03)00458-9

Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., & Shams, L. (2007). Causal Inference in Multisensory Perception. *PLOS ONE*, *2*(9), e943. https://doi.org/10.1371/journal.pone.0000943

Körding, K. P., & Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature*, *427*(6971), 244–247. https://doi.org/10.1038/nature02169

Maidenbaum, S., Hanassy, S., Abboud, S., Buchs, G., Chebat, D. R., Levy-Tzedek, S., & Amedi, A. (2014). The “EyeCane”, a new electronic travel aid for the blind: Technology, behavior & swift learning. *Restorative Neurology and Neuroscience*, *32*(6), 813–824. https://doi.org/10.3233/RNN-130351

Michel, M. M., & Jacobs, R. A. (2008). Learning optimal integration of arbitrary features in a perceptual discrimination task. *Journal of Vision*, *8*, 1–16. https://doi.org/10.1167/8.2.3.Introduction

Nardini, M., Bedford, R., & Mareschal, D. (2010). Fusion of visual cues is not mandatory in children. *Proceedings of the National Academy of Sciences*, *107*(39), 17041–17046. https://doi.org/10.1073/pnas.1001699107

Negen, J., Chere, B., Bird, L.-A., Taylor, E., Roome, H. E., Keenaghan, S., Thaler, L., & Nardini, M. (2019). Sensory cue combination in children under 10 years of age. *Cognition*, *193*, 104014. https://doi.org/https://doi.org/10.1016/j.cognition.2019.104014

Negen, J., Wen, L., Thaler, L., & Nardini, M. (2018). Bayes-Like Integration of a New Sensory Skill with Vision. *Scientific Reports*, *8*(1), 16880. https://doi.org/10.1038/s41598-018-35046-7

Olkkonen, M., Hansen, T., & Gegenfurtner, K. R. (2008). Color appearance of familiar objects: Effects of object shape, texture, and illumination changes. *Journal of Vision*, *8*(5), 13. https://doi.org/10.1167/8.5.13

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 10:437-442.

Rahnev, D., & Denison, R. N. (2018). Suboptimality in perceptual decision making. *Behavioral and Brain Sciences*, *41*, e223. https://doi.org/DOI: 10.1017/S0140525X18000936

Summerfield, C., & Tsetsos, K. (2012). Building bridges between perceptual and economic decision-making: Neural and computational mechanisms. *Frontiers in Neuroscience*, *6*(MAY), 1–20. https://doi.org/10.3389/fnins.2012.00070

Tassinari, H., Hudson, T. E., & Landy, M. S. (2006). Combining Priors and Noisy Visual Cues in a Rapid Pointing Task. *Journal of Neuroscience*, *26*(40), 10154–10163. https://doi.org/10.1523/JNEUROSCI.2779-06.2006

Vilares, I., Howard, J. D., Fernandes, H. L., Gottfried, J. A., & Kording, K. P. (2012). Differential representations of prior and likelihood uncertainty in the human brain. *Current Biology*, *22*(18), 1641–1648. https://doi.org/10.1016/j.cub.2012.07.010

Witzel, C., & Hansen, T. (2015). Memory effects on color perception. In A. J. Elliot, A. Franklin, & M. D. Fairchild (Eds.), *Handbook of Color Psychology* (pp. 641–659). Cambridge University Press. https://doi.org/DOI: 10.1017/CBO9781107337930.032

Witzel, C., Olkkonen, M., & Gegenfurtner, K. R. (2018). A Bayesian Model of the Memory Colour Effect. *I-Perception*, *9*(3). https://doi.org/10.1177/2041669518771715

Witzel, C., Valkova, H., Hansen, T., & Gegenfurtner, K. R. (2011). Object knowledge modulates colour appearance. *I-Perception*, *2*, 13–49.

Wu, S. W., Delgado, M. R., & Maloney, L. T. (2009). Economic decision-making compared with an equivalent motor task. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(15), 6088–6093. https://doi.org/10.1073/pnas.0900102106