

Dynamic weighting of multisensory stimuli shapes decision-making in rats and humans

John P. Sheppard

Cold Spring Harbor Laboratory,
Cold Spring Harbor, NY, USA
Watson School of Biological Sciences,
Cold Spring Harbor, NY, USA



David Raposo

Cold Spring Harbor Laboratory,
Cold Spring Harbor, NY, USA
Champalimaud Neuroscience Programme,
Lisbon, Portugal



Anne K. Churchland

Cold Spring Harbor Laboratory,
Cold Spring Harbor, NY, USA



Stimuli that animals encounter in the natural world are frequently time-varying and activate multiple sensory systems together. Such stimuli pose a major challenge for the brain: Successful multisensory integration requires subjects to estimate the reliability of each modality and use these estimates to weight each signal appropriately. Here, we examined whether humans and rats can estimate the reliability of time-varying multisensory stimuli when stimulus reliability changes unpredictably from trial to trial. Using an existing multisensory decision task that features time-varying audiovisual stimuli, we independently manipulated the signal-to-noise ratios of each modality and measured subjects' decisions on single- and multi-sensory trials. We report three main findings: (a) Sensory reliability influences how subjects weight multisensory evidence even for time-varying, stochastic stimuli. (b) The ability to exploit sensory reliability extends beyond human and nonhuman primates: Rodents and humans both weight incoming sensory information in a reliability-dependent manner. (c) Regardless of sensory reliability, most subjects are disinclined to make “snap judgments” and instead base decisions on evidence presented over the majority of the trial duration. Rare departures from this trend highlight the importance of using time-varying stimuli that permit this analysis. Taken together, these results suggest that the brain's ability to use stimulus reliability to guide decision-making likely relies on computations that are conserved across species and operate over a wide range of stimulus conditions.

Introduction

Sensory information bearing on perceptual decisions can vary greatly in its reliability. For example, the same snippet of human speech can provide a reliable auditory signal in quiet conditions but an extremely unreliable signal in a crowded room. This creates a challenge for decision-making: Noise interferes with a subject's perception, reducing decision accuracy. The brain's ability to estimate reliability and use this estimate to guide decisions is not yet well understood.

Multisensory stimuli provide a tractable entry point for understanding how reliability is estimated and used to guide decisions (Shams, 2012). In some circumstances, one modality is naturally more reliable than another and tends to dominate a subject's percept. In spatial localization tasks, vision tends to dominate (Howard & Templeton, 1966), while in timing tasks, auditory and tactile stimuli tend to dominate (Shams, Kamitani, & Shimojo, 2000, 2002; Shams, Ma, & Beierholm, 2005). It would be a mistake to think of sensory perception as being dominated by one modality or the other: Considerable evidence indicates that in most situations, multiple modalities influence a subject's decision. The degree to which a particular modality influences a decision can accurately be predicted by that modality's reliability.

To measure the influence of each of two modalities, subjects are presented with information from two

Citation: Sheppard, J. P., Raposo, D., & Churchland, A. K. (2013). Dynamic weighting of multisensory stimuli shapes decision-making in rats and humans. *Journal of Vision*, 13(6):4, 1–19, <http://www.journalofvision.org/content/13/6/4>, doi:10.1167/13.6.4.

sensory modalities (e.g., Alais & Burr, 2004; Ernst & Banks, 2002). On some trials, the two modalities convey consistent information supporting the same decision outcome, but on “conflict” trials the information in the two modalities differs. By examining subjects’ decisions on conflict trials, researchers infer the perceptual “weights” subjects assign to each modality (Young, Landy, & Maloney, 1993). These weights define the degree to which a particular modality influences a subject’s decisions. Previous studies demonstrate that humans weight incoming stimuli in a way that depends on the reliability of each stimulus. Humans frequently weight available sensory cues in proportion to their reliabilities, a strategy that is statistically optimal because it minimizes uncertainty in the final perceptual estimate (Ernst & Banks, 2002; Hillis, Watt, Landy, & Banks, 2004; Jacobs, 1999; Knill & Saunders, 2003; Young et al., 1993). The ability to weight sensory inputs according to their reliability is referred to here as “dynamic weighting” because subjects must change the weights assigned to each modality on a trial-by-trial basis even when the reliabilities change randomly on successive trials (e.g., Fetsch, Turner, DeAngelis, & Angelaki, 2009; Triesch, Ballard, & Jacobs, 2002).

Existing work on the influence of reliability on multisensory integration leaves several key questions unanswered. First, the vast majority of studies on multisensory integration have been limited to tasks involving static stimuli (Alais & Burr, 2004; Ernst & Banks, 2002). This is a shortcoming because many (if not most) ecologically relevant stimuli in the world vary over rapid timescales ranging from tens or hundreds of milliseconds to a few seconds (Rieke, 1997). For example, moving predators emit a time-varying stream of auditory and visual signals, the dynamics of which are informative about the animal’s trajectory and identity (Maier, Chandrasekaran, & Ghazanfar, 2008; Maier & Ghazanfar, 2007; Thomas & Shiffrar, 2010). The behavioral and neural mechanisms relevant for decisions about time-varying signals within one modality have been extensively studied (Cisek, Puskas, & El-Murr, 2009; Gold & Shadlen, 2007; Huk & Meister, 2012). However, the effect of dynamic stimuli on multisensory integration is only just beginning to be studied. The ecological importance of time-varying multisensory sensory signals for survival and reproduction suggests that animals may be able to rapidly estimate the reliability of such stimuli and appropriately combine the relevant information in a multisensory context.

A second unanswered question is whether only primates are capable of dynamic weighting. Existing studies have focused on human and nonhuman primates (for review, see Alais, Newell, & Mamassian, 2010). Observing dynamic weighting in other species,

such as rodents, would speak to the generality of the computations that underlie this behavior.

A final unanswered question is of critical importance. It concerns whether subjects change the manner in which they accumulate sensory evidence for reliable versus unreliable stimuli. In the visual system, a class of analyses has revealed that subjects often are strategic in how they make use of time (Cisek et al., 2009; Kiani, Hanks, & Shadlen, 2008; Nienborg & Cumming, 2009; Raposo, Sheppard, Schrater, & Churchland, 2012; Sugrue, Corrado, & Newsome, 2004). These analyses suggest that subjects sometimes make fast decisions and ignore large portions of the available evidence. By contrast, the current framework for optimal multisensory integration assumes that subjects make use of all available information. In reality, subjects might make fast decisions when presented with reliable stimuli, causing them to ignore large portions of the available information. This would lead experimenters to underestimate the perceived reliability of such stimuli, leading, in turn, to inaccurate predictions for perceptual weights on multisensory trials. A failure of subjects to meet these assumptions could be interpreted as behavioral deviations from optimal cue weighting, when, in fact, the integration process might be optimal given the subjects’ strategy of evidence accumulation.

The current study addresses these unanswered questions. We extended a task that we have recently developed to study multisensory integration in rodents and humans. Subjects are presented with time-varying auditory and/or visual event streams amidst noise and judge the stimulus event rates. In previous work, we matched the reliability of auditory and visual stimuli and demonstrated that subjects’ accuracy improved on multisensory trials (Raposo et al., 2012). In the present study, we challenged subjects to not only integrate sensory information, but to do so in a way that takes into account the reliability of each stimulus. To achieve this, we systematically varied the sensory reliability of auditory and visual evidence and placed the two modalities in conflict with one another to gain insight into how subjects exploited stimulus reliability as it varied unpredictably from trial to trial.

We report three main conclusions: First, we observed that dynamic weighting extends to the processing of time-varying sensory information. Second, we observed that dynamic weighting is evident in both humans and rodents. Finally, we determined that most subjects are influenced by sensory evidence throughout the trial duration regardless of whether stimuli are reliable or unreliable. Rare exceptions to this trend illustrate the importance of using stimuli that permit this analysis, and may offer insight into previous deviations from optimal cue weighting that are reported in the literature (Fetsch, Pouget, DeAngelis, & Angelaki, 2012; Fetsch et al., 2009; Rosas, Wagemans,

Ernst, & Wichmann, 2005; Rosas, Wichmann, & Wagemans, 2007; Zalevski, Henning, & Hill, 2007). Our findings indicate that dynamic weighting likely relies on neural computations that are shared among many species and are common to the processing of a wide class of stimuli. Overall, this work offers insight into a fundamental feature of the brain: the ability to estimate the reliabilities of multiple incoming streams of sensory evidence and use such estimates to optimize behavior in an uncertain world.

Materials and methods

Overview of rate discrimination task

We examined the choice behavior of rodents and humans on a multisensory rate discrimination task that we have used previously (Raposo et al., 2012 and Figure 1a). Stimuli consisted of 1-s streams of auditory and/or visual events separated by random sequences of short and long interevent intervals of fixed duration (rats: 50, 100 ms; humans: 60, 120 ms). The ratio of short to long intervals over the course of the whole trial determined the average event rate of each trial. Trials consisting solely of long intervals produced the lowest possible event rates; trials consisting of only short intervals produced the highest rates. Event rates ranged from 7–15 events/s in humans (Figure 1b) and 9–16 events/s in rats. Subjects made two-alternative low-/high-rate judgments by comparing the average rate of each trial to an enforced category boundary (humans: 10.5 events/s; rats: 12.5 events/s). Stimuli with more equal proportions of short and long intervals produced intermediate event rates difficult to classify as low or high.

To manipulate sensory reliability, we adjusted the signal-to-noise ratios (SNRs) of the auditory and visual stimuli. The SNR of either modality could be independently manipulated since auditory and visual stimuli were presented amidst background noise (see below; Figure 1b). To gauge how strongly subjects weighted evidence from one sensory modality relative to the other for each pairing of sensory reliabilities, we presented multisensory trials in which the auditory and visual rates conflicted (Figure 1b, left and right columns). To generate a multisensory trial, we first randomly selected a conflict level (for example, 2 events/s). Next, we randomly selected a pair of rates with the specified conflict (for example, 9 and 11 events/s). Between the two event rates chosen, the lower and higher rates were randomly assigned to the visual and auditory stimuli, respectively, leading to “positive” or “negative” cue conflicts in equal proportions. When the randomly selected conflict level was 0, the same rate

was assigned to both modalities. Mixtures of long and short interevent intervals were then sampled randomly to generate independent auditory and visual event streams with the desired trial-averaged event rates.

Note that because event rates fluctuated in time due to the random sequences of short and long intervals in a trial, auditory and visual stimuli arrived at different moments and had different instantaneous rates on multisensory trials even when the trial-averaged event rates were equal (Figure 1b, middle column). At the lowest and highest event rates (when all intervals were either short or long), auditory and visual stimuli still arrived at different times since a brief offset (humans: 20 ms; rats: 0–50 ms randomly selected for each trial) was imposed between the two event streams. Previous work indicates that this configuration leads to multisensory improvements in performance comparable to that observed when auditory and visual events are presented simultaneously (Raposo et al., 2012). This is likely because the window of auditory-visual integration can be very flexible, depending on the task (Powers, Hillock, & Wallace, 2009; Serwe, Kording, & Trommershauser, 2011).

Psychophysical methods

Subject performance was characterized using psychometric functions. For each type of visual, auditory, or multisensory stimulus, we calculated the subject’s proportion of high-rate decisions as a function of the trial-averaged event rate (see Figures 2, 3). Standard errors were computed for choice proportions using the binomial distribution. We fit psychometric functions using the `psignifit` version 3 toolbox (<http://psignifit.sourceforge.net/>) for Matlab® (MathWorks, Natick, MA). Four-parameter psychometric functions f were fit to the choice data via maximum likelihood estimation (Wichmann & Hill, 2001):

$$f(r, \mu, \sigma, \gamma, \lambda) = \gamma + (1 - \gamma - \lambda) \left(1 + \operatorname{erf} \left(\frac{r - \mu}{\sigma\sqrt{2}} \right) \right), \quad (1)$$

where r is the trial-averaged event rate (averaged between visual and auditory rates for multisensory trials), μ and σ are the first and second moments of a cumulative Gaussian function, γ and λ are guessing and lapse rates (i.e., lower and upper asymptotes of the psychometric function, constrained such that $0 \leq \gamma, \lambda \leq 0.1$), and erf is the error function. σ is referred to as the psychophysical threshold and provides a metric of subjects’ performance; smaller σ indicates a steeper psychometric function and hence improved discrimination performance (Ernst & Banks, 2002; Hillis, Ernst, Banks, & Landy, 2002; Hillis et al., 2004; Jacobs, 1999;

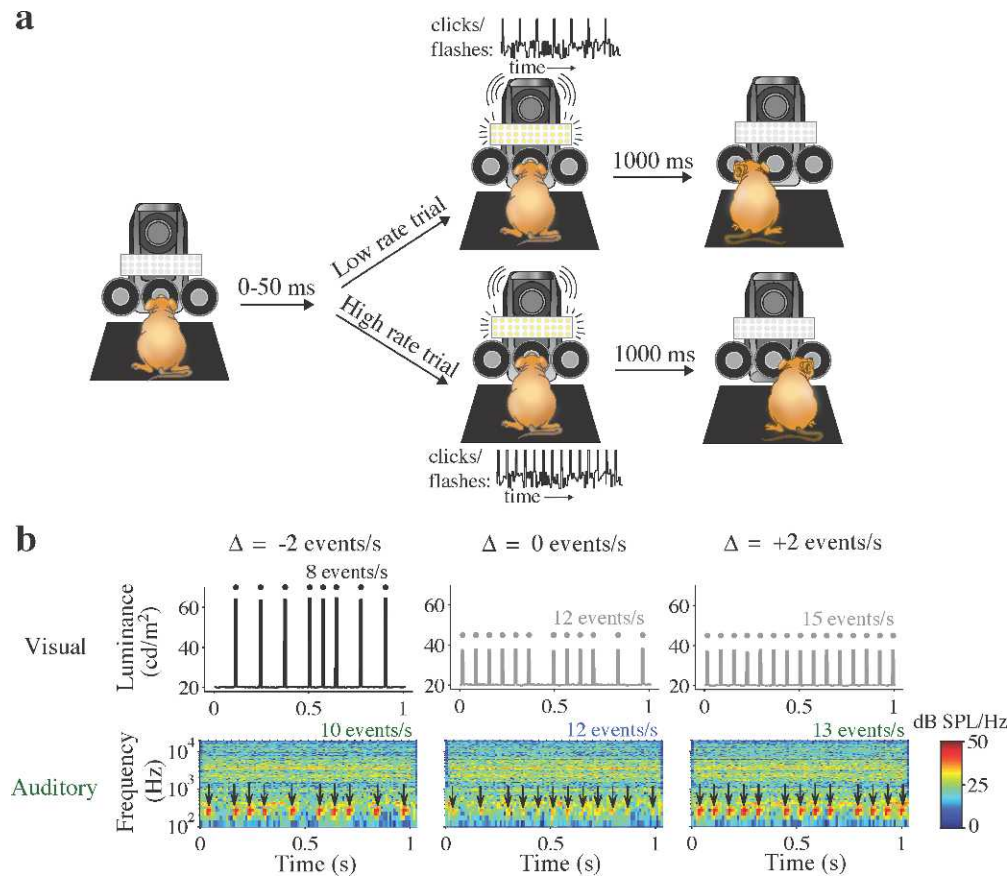


Figure 1. Rate discrimination decision task. (a) Schematic of rat behavioral setup. Rats were trained to perform a multisensory rate discrimination task via operant conditioning. Rats initiated trials by inserting their snouts into a central port (left), triggering an infrared sensor. After a randomized delay period, 1-s stimuli consisting of auditory and/or visual event streams were delivered through a speaker and LED panel (middle). Rats were required to remain in the central port until the end of the stimulus, and were provided with 20 μ L of water reward when they selected the correct choice port (low-rate trials: left port; high-rate trials: right port). Trials in which the rat did not remain in the central port for the full stimulus were punished with a 4-s timeout period before allowing initiation of a new trial. (b) Example stimuli presented in human version of task. Time courses indicate arrival of events over the course of the 1-second stimuli. 10 ms events were separated by either short (60 ms) or long (120 ms) intervals. Top: Values on the ordinate indicate average luminance of events and background noise presented during high-reliability (black) and low-reliability (gray) visual trials. Bottom: Spectrograms indicate spectral power of sound pressure fluctuations during auditory stimulus presentation (color bar indicates signal power across frequency bands in units of dB SPL/Hz). Arrows indicate event (220 Hz tone) arrival times for low- (middle) and high- (left, right) reliability auditory stimuli. Auditory events in rat experiments consisted of white noise bursts (not shown). Note logarithmic scaling of frequency range. Multisensory trials consisted of visual (top) and auditory (bottom) stimuli presented together, and included different pairings of auditory and visual reliabilities. Multisensory trials included cue conflict levels ranging from -2 (left) to $+2$ (right) events/second. Auditory and visual event streams were generated independently on all multisensory trials.

Young et al., 1993). Standard errors on model parameters were obtained via bootstrapping with 2000 resamples of the choice data.

For a multisensory trial with auditory and visual event rate estimates \hat{r}_A and \hat{r}_V , we model the subject's final rate estimate \hat{r} on a multisensory trial as

$$\hat{r} = \hat{r}_A w_A + \hat{r}_V w_V, \quad (2)$$

where w_A and w_V are linear weights such that $w_A + w_V = 1$ (Ernst & Banks, 2002; Young et al., 1993). Assuming

\hat{r}_A and \hat{r}_V are unbiased, minimum variance in \hat{r} is achieved by assigning sensory weights proportional to the relative reliabilities (i.e., reciprocal variances) of the estimates obtained from the two modalities (Landy & Kojima, 2001; Young et al., 1993). As in previous work, we estimated the single-sensory reliabilities (R) as the squared reciprocal of subjects' psychophysical thresholds obtained from their single sensory psychometric data: $R = 1/\sigma^2$ (Ernst & Banks, 2002; Young et al., 1993). Given only the single sensory thresholds σ_A

and σ_V , an optimal maximum likelihood estimator assigns sensory weights

$$w_A = \frac{\sigma_V^2}{\sigma_A^2 + \sigma_V^2}$$

and

$$w_V = \frac{\sigma_A^2}{\sigma_A^2 + \sigma_V^2}. \quad (3)$$

Predicted weights can be compared to estimates of the subjects' actual sensory weights obtained from multisensory stimuli in which the conflict between auditory and visual event rates is systematically varied. In this conflict analysis, we assess each subject's point of subjective equality (PSE; the average event rate for which the subject is equally likely to make low- and high-rate decisions, estimated from the psychometric function) as a function of the sensory conflict Δ :

$$\Delta = r_V - r_A, \quad (4)$$

where r_V and r_A are the presented visual and auditory rates. Rearranging Equation 2, we obtain

$$w_V = \frac{\hat{r} - r_A}{\Delta}. \quad (5)$$

Neglecting choice biases, \hat{r} will equal the category boundary rate (r_{CB}) when the average event rate, r_{Mean} , is equal to PSE_{AV} (where $r_{Mean} = r_A + \Delta/2$ and PSE_{AV} is the PSE for the given multisensory trial type; see Fetsch et al., 2012; Young et al., 1993). Substituting these terms into Equation 5 and solving for PSE_{AV} yields

$$PSE_{AV} = \Delta \left(\frac{1}{2} - w_V \right) + r_{CB}. \quad (6)$$

One can thus estimate the perceptual weights by measuring subjects' PSEs for multisensory trials presented across a range of conflict levels (Young et al., 1993), since the slope obtained from simple linear regression of PSE_{AV} against Δ provides an empirical estimate of w_V :

$$\hat{w}_V = 1/2 - slope. \quad (7)$$

We used this approach to compare the empirically estimated weights to the weights predicted from subjects' single sensory thresholds (henceforth, observed and predicted weights; Figure 3e and f, Figure 4). Standard errors for the predicted and observed weights were estimated by propagating the uncertainty associated with σ_A , σ_V , and PSE_{AV} . Statistical comparisons of observed and predicted weights in individual subjects were performed using Z-tests.

Estimating the timecourses of sensory evidence accumulation

We performed an additional analysis to assess the influence on subjects' decisions of transient event rates occurring at specific times throughout each trial. The quantity we computed is closely related to the choice-triggered average (Kiani et al., 2008; Nienborg & Cumming, 2009; Sugrue et al., 2004), but is adapted for our rate stimuli. The details of the analysis are described elsewhere (Raposo et al., 2012). Briefly, we first select trials that have neutral event rates (i.e., equal to the category boundary rate) on average for all time points falling outside a sliding temporal window of interest. Next, we define the "excess rate" for each temporal window as the difference in the mean window-averaged event rates between the selected trials that preceded high- versus low-rate decisions; repeating this analysis for local windows centered at each time point produces timecourses that indicate how the excess rate fluctuates across the trial duration. Here, we used this analysis to compare excess rate timecourses for reliable versus unreliable stimuli. Window sizes were 333 ms (humans) or 280 ms (rats), which were specified so that trials could be selected with neutral average rates outside the windows for the presented event streams. The difference in window size for humans versus rats arises because slightly higher event rates were used for rats. The shorter window size in rats allowed us to compute a slightly longer excess rate timecourse for rats as compared to the humans.

Because our goal was to assess the influence of dynamic sensory evidence on subjects' decisions under different levels of sensory reliability, we limited our analysis of excess rates to unisensory trials. Approximately 40% and 50% of single sensory trials possessed the required neutral rates outside the sliding windows for at least one point within the trial for humans and rats, respectively. For each trial type, this resulted in 102 ± 10 and 845 ± 82 (mean \pm SE) trials per window on average for the example human and rat (Figure 5c, d), accounting for ~ 980 of 2,200 and $\sim 7,500$ of 14,500 total single sensory trials. We also computed excess rate curves for which data were pooled across all human and rat subjects (Figure 6); on average, the composite human and rat datasets included 888 ± 77 and $3,486 \pm 345$ trials per window for each trial type.

To make statistical comparisons of the magnitude of excess rate timecourses between low- and high-reliability stimuli on unisensory auditory and visual trials (see Figures 5, 6), we computed the mean excess rates by averaging all window-averaged event rates included across all sliding windows within the trial duration. Standard errors for the mean excess rates averaged across the trial duration were computed by bootstrapping with 10,000 resamples of the single-sensory

choice data. Statistical comparisons of average excess rates between low- and high-reliability trials were performed using two-tailed *Z*-tests.

Because visual and auditory events arrived at discrete moments within each trial, the raw excess rate curves contained peaks and troughs related to the arrival of events. Since these peaks and troughs are artifacts related to the event and interevent interval lengths and do not reflect meaningful signatures of the subjects' decisions, we removed these high-frequency artifacts from the plotted curves in Figures 5, 6 by applying a 130-ms moving-average filter. This filter width was chosen because it was small relative to the 1-s trial duration, corresponding to the time-scale of individual event arrival times. Smoothing was performed for clarity of display only; all reported statistics were performed on the raw, unfiltered excess rate data.

Human experiments

We collected data from seven adult human participants (two male) with normal hearing and normal or corrected-to-normal vision. One subject was an author; the remaining six were naïve to the purposes of the experiment. We recruited subjects through fliers posted at Cold Spring Harbor Laboratory and paid them for their participation. All research performed on human subjects was conducted in compliance with the Declaration of Helsinki.

Experiments were conducted in a sound-isolating booth (Industrial Acoustics, Bronx, NY). Subjects sat in front of a computer monitor and speaker, and registered their decisions with a keyboard. Prior to each trial, subjects were cued to fixate a central black target offset from the visual stimulus. Stimulus presentation began 500 ms following the onset of the fixation target. Stimulus presentation and data acquisition were performed in Matlab using the Psychophysics Toolbox package (Pelli, 1997).

Auditory events consisted of 220 Hz pure tones played from a single speaker attached to the left side of the monitor, directly adjacent to the visual stimulus. Auditory events ranged in intensity from 35 to 57 dB SPL, calibrated using a Brüel & Kjær pressure-field microphone (Nærum, Denmark) placed ~ 45 cm from the speaker. Average auditory event intensity across trials was 41.1 ± 3.8 and 50.1 ± 5.9 dB SPL for low- and high-reliability auditory stimuli, respectively. Corrupting white noise was sampled from a Gamma distribution (shape, scale parameters = 5, 2) and overlaid atop the auditory signal; noise levels averaged 54.9 ± 0.1 dB SPL across trials, leading to unreliable events that were 13.9 ± 3.8 dB below background noise levels and reliable events that were 4.7 ± 6.1 dB below background noise levels. Stimulus events were never-

theless detectable in noise because their power was concentrated at a single frequency. Waveforms were sampled at 44.1 kHz.

Visual stimuli were presented on a Dell M991 CRT monitor (100 Hz refresh rate) (Dell, Round Rock, TX), and consisted of a flashing square occupying $10^\circ \times 10^\circ$ of visual angle at 17.2° azimuth. The visual stimulus was positioned at the leftmost edge of the display to minimize its distance from the speaker (3.5 cm). Visual stimuli were presented atop Gaussian white noise sampled independently for every pixel within the stimulus. To estimate the SNR of visual stimuli, we measured luminance directly at the monitor surface using an Extech HD400 light meter (Nashua, NH). Background visual noise was sampled from the same distribution across visual trials and averaged 20.2 ± 0.2 candela/m² in luminance. Visual event luminance ranged from 14.5 to 66.5 cd/m²; average visual signal-to-noise ratios were 1.0 ± 0.4 and 3.6 ± 1.5 dB for low- and high-reliability visual trials, respectively.

Human behavioral sessions lasted approximately 50 minutes and typically included 6 blocks of 140 trials each. Initially, subjects were trained for 4–6 sessions to familiarize themselves with the task. Initial training consisted exclusively of single sensory trials (auditory or visual in equal proportions), during which auditory feedback was provided after each response: a high tone (6400 Hz) indicated a correct choice while a low tone (200 Hz) indicated an incorrect choice.

During initial training, we first familiarized subjects with the single sensory version of the task and then implemented “tracking” blocks featuring an adaptive staircase procedure (Kesten, 1958) in which auditory and visual SNRs were adjusted to achieve target performance levels for the easiest trials (7, 8, 14, 15 Hz). For both sensory modalities, target accuracy levels were approximately 70% and 90% for low- and high-reliability trials, respectively. After initial training, we introduced testing blocks into the behavioral sessions, during which both single- and multi-sensory trials were randomly interleaved (trial composition: 20% auditory, 20% visual, 60% multisensory). Once training was complete, behavioral sessions began with a warm-up block followed by one tracking block (all single-sensory trials) and subsequently four testing blocks. Auditory and visual SNRs were only changed during tracking blocks, and were held fixed across all testing blocks in a behavioral session. Multisensory cue conflicts of 0, ± 1 and ± 2 events/s were presented. No response feedback was provided during testing blocks for five of the seven human subjects. For the last two participants in our study, we added response feedback during testing blocks in order to minimize subjects' uncertainty about the category boundary separating low- and high-rate trials. In doing this, we intended to minimize biases observed with these particular subjects prior to

incorporating response feedback, which were initially greater in magnitude than in our previous subjects; we therefore omitted the initial testing sessions conducted without response feedback in these two subjects from our analysis. Random feedback was provided on multisensory conflict trials in which the auditory and visual rates fell on opposite sides of the category boundary. We observed no systematic differences between the subjects who received response feedback during testing sessions as compared to the other humans. Only data collected during testing blocks were included in our analyses.

Rat experiments

Data were collected from five adult male Long Evans rats (240–310 g, Taconic Farms, Hudson, NY) trained to do a similar version of the task. Animals initiated trials and registered decisions by inserting their snouts into choice ports outfitted with infrared sensors (Island Motion, Tappan, NY). These methods are similar to those we have reported previously (see Figure 1d in Raposo et al., 2012). All animal experiments were approved by the Cold Spring Harbor Animal Care and Use Committee, and complied with the NIH Guide for the Care and Use of Laboratory Animals as well as the ARVO Statement for the Use of Animals in Ophthalmic and Vision Research.

Auditory events consisted of white noise bursts produced by a single speaker (Harmon Kardon, Stamford, CT) positioned behind the central port; the speaker was calibrated with a pressure-field microphone (Brüel & Kjær, Naerum, Denmark). Auditory stimuli in these experiments were played amidst white noise (sampled from a uniform distribution with an average intensity of 16.4 dB SPL); however, this level was very low relative to the auditory event intensity. Low- and high-reliability auditory stimuli were produced by varying the loudness of auditory events in the range of 74–88 dB. Across rats, event loudness was 76.0 ± 1.6 dB (mean \pm SD) and 86.9 ± 0.8 dB SPL for the two reliability levels, never exceeding 88.4 dB SPL.

Visual stimuli were produced via a centrally positioned panel of 96 LEDs (6 cm high \times 17 cm wide). The bottom of the LED panel was \sim 4 cm above the rats' eyes. Since visual stimuli were presented in a dark sound-isolating booth, rats did not need to fixate on the LED panel to perceive the flashing visual stimuli. Furthermore, the stationary, full-field nature of the LED flashes made it unlikely that small changes in eye or head position would substantially alter perception of the visual stimuli. Unlike the human experiments, visual stimuli included only a single reliability level and were not corrupted with added visual noise. Visual stimuli were delivered to the LED panel via the same

sound card used to deliver auditory stimuli, but using the other channel. At the approximate level of the rats' eyes, visual stimulus luminance in the behavioral rigs ranged from 112 to 144 cd/m^2 (rig 1: 112 cd/m^2 ; rig 2: 144 cd/m^2 ; rig 3: 116 cd/m^2). Across all trials, average luminance was 124.5 ± 14.6 cd/m^2 (mean \pm SD).

Naïve animals typically required 1–2 months of behavioral training before reaching stable performance levels on all trial types. Training began with short, reliable single sensory trials in one modality only. After good performance was achieved, trial durations were gradually extended to 1000 ms. Next, multisensory trials were added; once stable multisensory performance was achieved, the remaining single sensory modality was introduced until rats stably performed the task for all trial types. For two of the five rats, training began instead with multisensory stimuli in which the auditory and visual event streams were synchronized (see Raposo et al., 2012); for these rats, single-sensory auditory and visual stimuli were introduced subsequently, after which the synchronized multisensory stimuli were replaced with multisensory stimuli featuring independently generated auditory and visual event streams. For all five rats, all auditory stimuli were limited to the high-reliability condition during initial training; low-reliability auditory stimuli were subsequently introduced on both single- and multi-sensory trials once stable performance was attained for high-reliability stimuli. The long training times were required to minimize biases and achieve the desired performance levels for all trial types simultaneously. The ability to learn the category boundary for a unisensory condition came easily to most animals. Further, the transition between unisensory and multisensory stimuli was typically seamless. This suggests that although the task is not ecological in nature, it makes use of existing circuitry for combining auditory and visual signals.

We began collecting testing data for the sensory weighting experiments only after stable performance was reached on all single- (low-reliability auditory, high-reliability auditory, visual) and multi-sensory (visual/low-reliability auditory, visual/high-reliability auditory) trial types. Testing sessions included interleaved trials of all types, including multisensory trials with conflicting auditory and visual rates. Only completed trials in which rats waited the required 1000 ms in the center port and registered a decision (i.e., entered a reward port) within 2 s following stimulus offset were included in the analysis. Trials where the rats withdrew early from the center port or waited too long to make a choice ($>$ 2 s after stimulus offset) were not rewarded. Rats received feedback on all trials since water rewards motivated them to perform the task. Multisensory conflict trials in which auditory and visual event rates fell on opposite sides of the category

boundary were rewarded randomly. Multisensory cue conflicts of 0 and ± 2 events/s were presented.

Results

We examined the decisions of rat and human subjects on a rate discrimination decision task in which we systematically varied both the stimulus strength (i.e., the trial-averaged event rate) and the reliability (i.e., SNR) of auditory and visual stimuli. We first describe subjects' performance on single sensory trials. Next, we describe results obtained on multisensory trials. When the reliability of auditory and visual stimuli differed, rats and humans assigned more weight to the reliable modality. Finally, we consider whether stimulus reliability affected the moments within the trial that influenced subjects' decisions. For most subjects, stimuli throughout the trial influenced the decision, an effect consistently observed for both low- and high-reliability stimuli.

Perceptual weights change with stimulus reliability

We first quantified subjects' performance by computing their probabilities of high-rate decisions across the range of trial event rates and fitting psychometric functions to the choice data using standard psychophysical techniques (Materials and methods). On single sensory trials, estimated psychophysical thresholds were comparable across modalities for matched reliability trials but significantly smaller for high- relative to low-reliability trials of either modality, as highlighted in a representative human subject (Figure 2a, green and black lines steeper than blue and gray lines; $\hat{\sigma} \pm SE$: high-reliability auditory: $1.01 \pm 0.12 < \text{low-reliability auditory: } 3.20 \pm 0.21, p < 10^{-5}$; high-reliability visual: $1.24 \pm 0.11 < \text{low-reliability visual: } 3.40 \pm 0.30, p < 10^{-5}$). Rats were similarly presented low-reliability and high-reliability auditory stimuli, but only a single reliability level was used for the visual stimuli. Rats' thresholds also differed significantly between the two auditory reliability levels, as demonstrated in an example rat (Figure 2b, green line steeper than blue line; high-reliability auditory: $1.93 \pm 0.35 < \text{low-reliability auditory: } 5.32 \pm 0.95, p = 0.0004$), with an intermediate threshold for visual trials (black line; visual: 2.72 ± 0.31). In both species, we attempted to minimize bias; however, achieving zero bias for all three (rat) or four (human) single sensory trial types proved challenging. Analyses that could in principle be affected by subject bias were always repeated in subsampled data where biases were minimal (below).

Next, we examined decisions on multisensory trials. As in previous experiments (Raposo et al., 2012), both rats' and humans' performances improved on multisensory trials, and the performances were frequently close to the optimal prediction. The magnitude of the multisensory improvement ($\sigma_{\text{predicted}}/\sigma_{\text{observed}}$; Materials and methods) was unrelated to the magnitude of the cue conflict (mean correlations averaged across trial types, 95% CIs; humans: $r = 0.07 [-0.13, 0.26]$, rats: $r = 0.03 [-0.25, 0.31]$). We took advantage of cue conflict trials and asked whether subjects' multisensory decisions reflected the relative reliabilities of the auditory and visual stimuli as estimated from subjects' single sensory psychophysical thresholds. On conflict trials, the trial-averaged event rates for auditory and visual stimuli differed (Materials and methods, Figure 1b). To assess the relative weights subjects assigned to the auditory and visual stimuli, we compared subjects' decisions on multisensory trials across a range of conflict levels for each of the possible reliability pairings (Materials and methods).

Both humans' and rats' decisions on multisensory trials were influenced by the relative reliabilities of the auditory and visual stimuli. The effects of stimulus reliability on subjects' decisions can be visualized by comparing subjects' choice data on trials with different levels of conflict between the auditory and visual event rates. When auditory and visual reliabilities are matched, subjects should weight both modalities equally. Indeed, on matched reliability trials, conflict in the event rates did not systematically bias subjects' decisions towards either cue. When sensory reliabilities were unequal, however, subjects preferentially weighted the more reliable modality, and their PSEs were systematically shifted towards this cue on conflict trials (Figure 3a, b; red, blue curves). These results are in agreement with previous observations from experiments using static stimuli (e.g., Ernst & Banks, 2002; Jacobs, 1999). The shifts in the psychometric functions for the example rat subject were smaller than in the human (Figure 3b). The smaller magnitude of the shift in the rat relative to the human reflects the fact that the single sensory thresholds (and thus the sensory reliabilities) were more disparate between the two modalities in the human than in the rat (i.e., compare human and rat psychometric curves in Figure 2a, c).

The magnitude and direction of the shift in PSE depended on the magnitude and direction of the stimulus conflict as well as the relative reliabilities of the two modalities. Figure 3c and d displays the example subjects' estimated PSEs as a function of conflict level for two multisensory trial types. For multisensory trials featuring low-reliability visual and high-reliability auditory stimuli in the example human, linear regression of PSE against conflict level (Δ) produced slopes significantly greater than zero (Figure

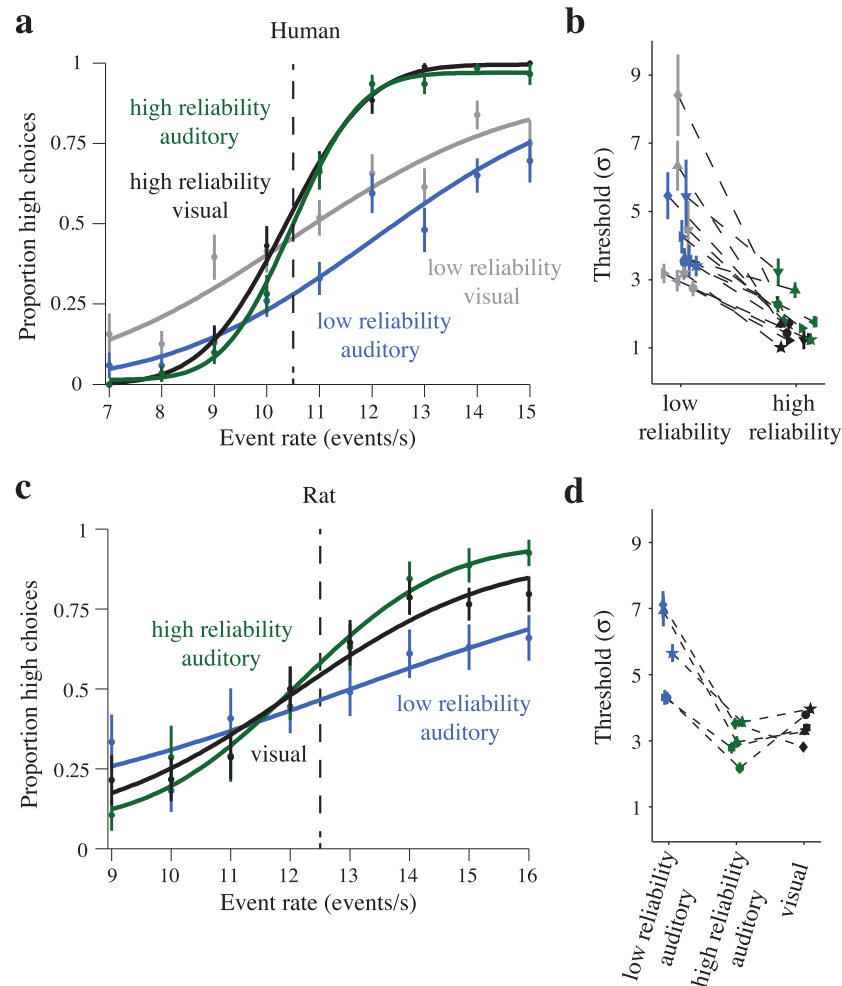


Figure 2. Single sensory performance on rate discrimination task depends on sensory reliability. (a) Performance of an individual human subject, displayed as the proportion of high-rate decisions plotted against the trial-averaged event rate. Data are presented separately for each single sensory trial type. Lines indicate psychometric functions fit via maximum likelihood estimation. Data were combined across multiple behavioral sessions (2,161 trials). (b) Psychophysical thresholds obtained from seven human subjects for each single sensory trial type (low/high reliability auditory: blue/green; low/high reliability visual: gray/black). Symbols depict individual subjects. (c) Single sensory performance in an individual rat, pooled from two consecutive sessions (975 trials). (d) Single sensory thresholds obtained across cohort of 5 rat subjects (symbols). Thresholds in (b) and (d) were estimated from data combined across multiple behavioral sessions (humans/rats: 19,143/62,363 total single sensory trials). Star symbols indicate the example human and rat subjects used in (a) and (c). Error bars indicate standard errors in all panels.

3c, left; slopes, 95% CIs: 0.36 [0.26, 0.46]). On the other hand, the slope of this regression was significantly less than zero for multisensory trials featuring high-reliability visual and low-reliability auditory stimuli (Figure 3c, right; slopes, 95% CIs: -0.46 [-0.59 , -0.33]). The positive and negative slopes of the regression lines indicate stronger and weaker weighting of the auditory stimulus (respectively) relative to the visual stimulus; thus, this subject weighted the high-reliability modality more strongly than the low-reliability modality in either case. Similarly, slopes of the PSE versus Δ regression lines differed significantly between the two multisensory trial types in the example

rat, reflecting the relative reliabilities of auditory and visual stimuli (Figure 3d: visual/high-reliability auditory: 0.09 [0.005, 0.18]; visual/low-reliability auditory: -0.22 [-0.32 , -0.12], $p < 10^{-5}$).

The changes in subjects' PSEs across the range of cue conflicts agreed well with predictions based upon the sensory reliabilities we inferred from subjects' performance on single sensory trials. To test whether subjects' cue weighting approximated statistically optimal behavior, we compared the observed sensory weights estimated from the slopes of the regression lines with the theoretical weights predicted by subjects' thresholds on the corresponding unisensory auditory and visual

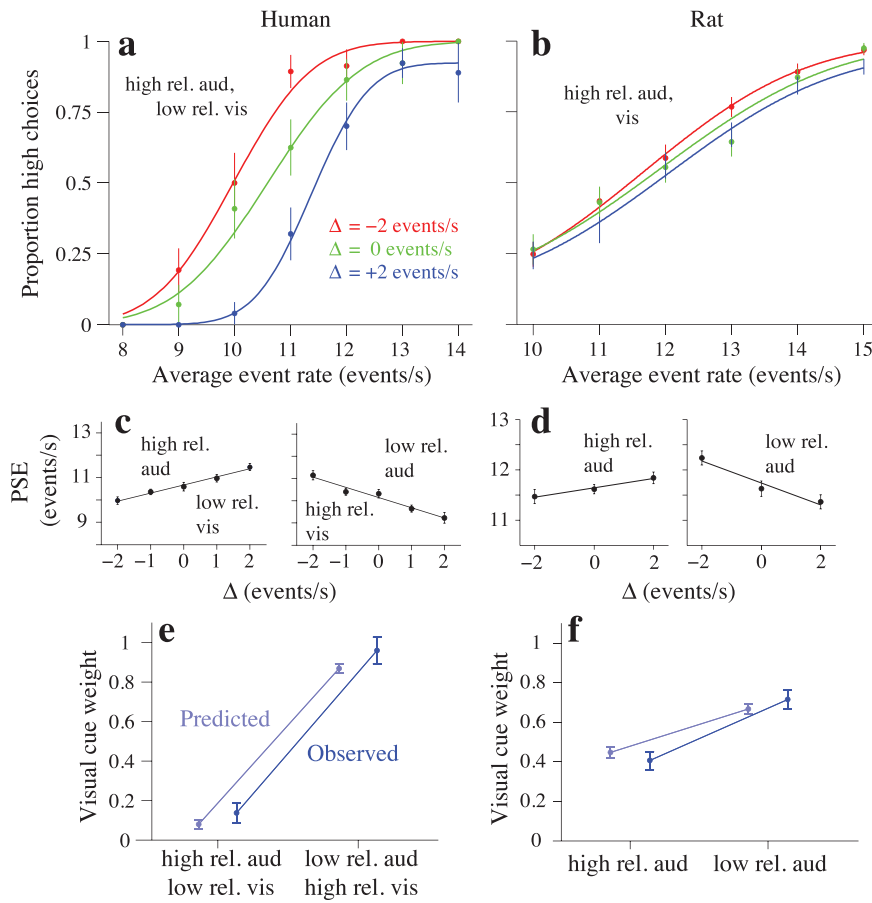


Figure 3. Subjects weight auditory and visual evidence in proportion to sensory reliability. (a) Performance on multisensory trials in an individual human pooled over multiple sessions (values on abscissae indicate mean trial event rates averaged between auditory and visual stimuli). Colors indicate level of conflict between modalities ($\Delta = \text{visual rate} - \text{auditory rate}$). Presented human data were obtained from the low-reliability visual/high-reliability auditory condition. (b) Same as (a) but for one rat. Data were obtained from the visual/high-reliability auditory condition. (c) Points of subjective equality (PSEs) from multisensory trials plotted as a function of conflict level for different pairings of auditory and visual stimulus reliabilities, shown for the same subject as in (a). Fitted lines were obtained via linear regression. Plotted data correspond to trials consisting of low- and high-reliability auditory stimuli paired with high- and low-reliability visual stimuli, respectively. Analogous fits were obtained for the other pairings of auditory and visual reliabilities presented to human subjects (see Figure 4b). (d) Same as (c) but for the single rat subject in (b). (e) Comparisons of the observed visual weights to the values predicted from the example human's single sensory thresholds. Data pertain to the same two multisensory trial types reported in (c). $N = 3,861$ trials. (f) same as (e) but for the rat in (b) and (d). $N = 4,018$ trials. Error bars indicate standard errors in all panels.

trials (Materials and methods; Young et al., 1993). The observed and predicted weights were in close agreement for all multisensory trial types in both example subjects (Figure 3e, f).

The weighting of multisensory stimuli seen in the example human and rat was typical: Nearly all humans and rats weighted sensory information in a manner that reflected the relative reliabilities of auditory and visual stimuli (Figure 4). For each subject, we computed the difference in w_V between multisensory trials consisting of high-reliability auditory/low-reliability visual versus low-reliability auditory/high-reliability visual stimulus pairings. This change was significantly greater than

zero for six of seven individual humans (Figure 4a, $p < 0.007$, one-tailed Z -tests). This indicates that nearly all human subjects increasingly relied on the visual or auditory evidence when its reliability was increased relative to the other modality. The increase in w_V between multisensory trials containing low- versus high-reliability auditory stimuli was likewise significant in four out of five rats (Figure 4d; $p < 0.032$, Z -tests). The remaining human and rat also showed changes in w_V in the expected direction, but the changes did not reach significance ($p > 0.19$).

Having established that both humans and rats dynamically changed their perceptual weights on

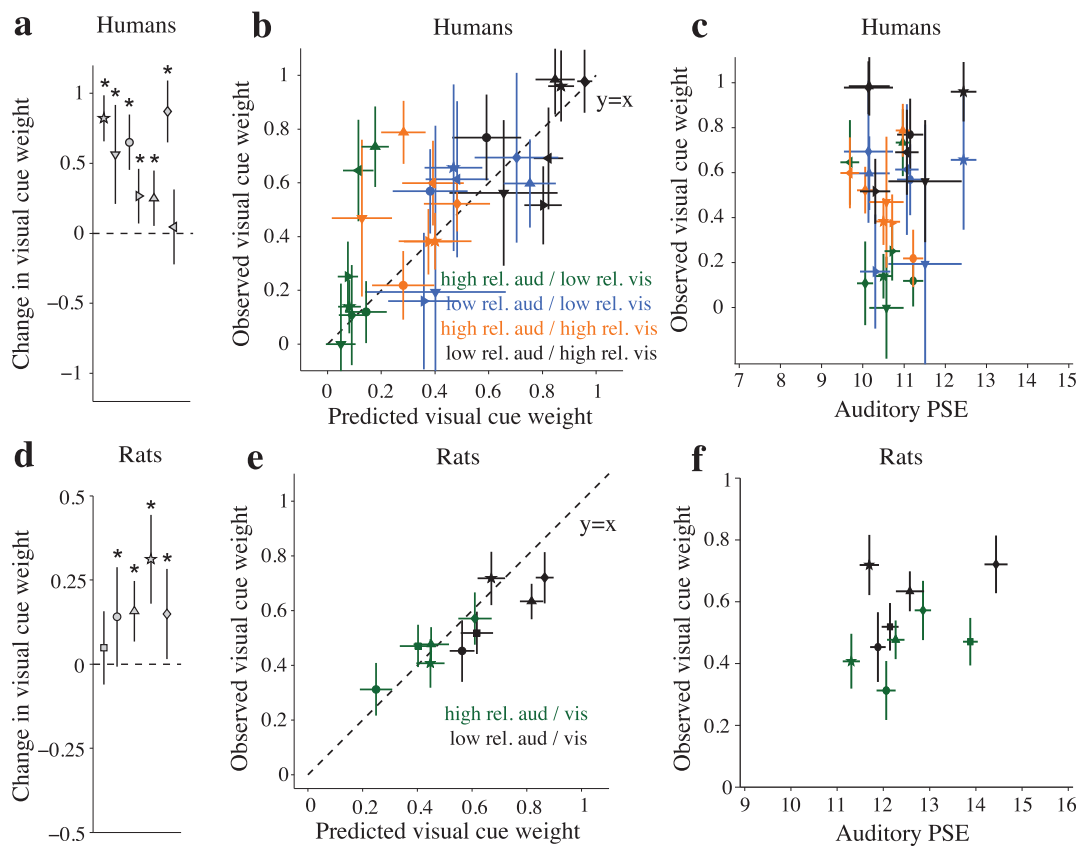


Figure 4. Reliability-based sensory weighting is observed consistently across subjects. Cue weights were estimated from data pooled over multiple behavioral sessions (humans/rats = 23,873/17,984 total multisensory trials). (a) Data points indicate the change in observed cue weights observed in seven individual human subjects, computed as the differences in subjects' visual cue weights between high-reliability visual/low-reliability auditory and low-reliability visual/high-reliability auditory trials. * indicates significant change in visual cue weights ($p < 0.05$, within-subjects one-tailed Z-tests). (b) Scatterplot compares observed visual cue weights (ordinate) to predicted values (abscissa) for all multisensory trial types in the individual human subjects. Legend indicates colors corresponding to each multisensory trial type. (c) Comparison of the observed visual cue weights (ordinate) to the PSE for unisensory auditory trials (abscissa). Color conventions are the same as in (b). (d) Same as (a) but for five individual rats. (e) Same as (b) but showing data for five rats. (f) Same as (c) but for five rats. Error bars indicate 95% confidence intervals in all panels.

multisensory trials in a manner that reflected the relative reliabilities of the auditory and visual evidence, we examined the degree to which these changes matched the statistically optimal predictions. These predictions are based on the sensory reliabilities inferred by subjects' performance on single sensory trials (Material and methods). In humans, observed visual weights were generally closely matched to predictions within the individual subjects (Figure 4b); seven of 28 comparisons (7 subjects \times 4 multisensory trial types) exhibited significant deviations between predicted and observed weights ($p < 0.05$, Z-tests). The observed deviations were distributed across four subjects. Interestingly, six of the seven deviations involved overweighting of visual evidence relative to predictions when the auditory reliability was high (Figure 4b, green and orange symbols). The remaining 21 comparisons for the other human subjects revealed no significant differences between observed and pre-

dicted weights. A limitation of our analysis is that we cannot rule out the possibility that some of the apparent deviations from optimality were, in fact, false positives arising from the large number of comparisons; however, all seven deviations remained robust to multiple comparisons correction after allowing for a false discovery rate of 20% (Benjamini & Hochberg, 1995).

In rats, as in humans, the perceptual weights for many individual subjects were close to the optimal predictions (Figure 4e). In general, rats came closest to the optimal prediction on high-reliability auditory trials (observed visual weights did not differ significantly from predictions in any of the rats; $p > 0.18$, two-tailed Z-tests). Deviations from the optimal prediction were observed more frequently on trials where the auditory stimulus reliability was low. On such trials, the perceptual weights for three of five rats differed significantly from optimality ($p < 0.05$, Z-tests). One

other rat's perceptual weights also differed from the optimal prediction on such trials, though the effect was only marginally significant ($p = 0.07$). In all four of these cases, observed visual weights were lower than predicted, suggesting that rats may systematically under-weight visual evidence relative to the optimal prediction when auditory reliability is low (Figure 4e: black square, triangle, diamond, and circle). This contrasted with the deviations from optimality observed in humans, in which subjects occasionally over-weighted visual evidence when auditory reliability was high (above). Note that these observations do not mean that the rats ignored the visual stimulus; when auditory reliability was low, rats generally relied more heavily on the visual stimulus than on the auditory stimulus (i.e., $w_V > 0.5$). The deviations from optimality here imply that the rats would have made better use of the available information had they relied even more heavily on the visual stimulus than observed.

It is unlikely that our changing cue weights were driven by unisensory biases. First, unisensory biases have previously been shown to have very little effect on weights measured during multisensory trials (Fetsch et al., 2012). This is because although non-negligible single sensory biases are assumed to systematically shift the PSE on multisensory trials in proportion to the relative reliabilities of either cue (e.g., Fetsch et al., 2012), such biases should shift the PSE in an identical manner on conflict trials and nonconflict trials. Therefore, our estimates of w_V and w_A , which are generated by taking the slope of the line relating PSE and cue conflict (Figure 3c), should not be affected by unisensory bias under the classic cue integration framework (Young et al., 1993). Nevertheless, we took two additional steps to guard against the possibility that our results were confounded by single sensory bias. The first step was to examine whether w_V (and by extension w_A) was related to unisensory bias. We found that for every multisensory trial type considered individually in both rats and humans, w_V was unrelated to the PSE measured from corresponding unisensory auditory trials (all p -values > 0.05 , across-subject Pearson's correlations, Figure 4c, f). In other words, subjects who had a slight bias on auditory trials were just as likely as any other subject to demonstrate a particular cue weight. This was also true for the relationship between w_V and visual PSE (all p -values > 0.05). These observations are consistent with theoretical predictions and provide reassurance that discrepant biases on the two unisensory conditions had no systematic effects on cue weights. Our second step to guard against artifacts from unisensory bias was to recompute w_V for humans and rats after restricting the included data to behavioral sessions for which subjects' single-sensory PSEs were all equal within a tolerance of ± 2 events/

second (i.e., less than the range of cue conflicts presented), and obtained nearly equivalent results in both species.

We performed one final analysis on subject bias to provide insight into its source. Specifically, we tested whether subject bias could be interpreted as a (misguided) prior expectation about the stimulus rate. We compared the measured bias between easy trials (i.e., trials with average rates > 2 events/second from category boundary) and difficult trials (i.e., < 2 events/second from category boundary) across subjects for every unisensory and multisensory trial type. We found no significant differences for any of the eight human trial types ($p > 0.05$, t -tests), and only one significant difference among the five rat trial types (difficult trial bias $>$ easy trial bias, uncorrected $p = 0.03$). This single significant observation was of little practical consequence, corresponding roughly to a 2% difference in the rats' probability of making a particular choice. Because a prior expectation on the stimulus would tend to influence the decision more when the sensory evidence is weak (Battaglia et al., 2010; Weiss, Simoncelli, & Adelson, 2002), our observations of similar biases on trials with strong and weak evidence argues against the possibility that biases were driven by a non-uniform prior expectation on stimulus rate. An alternative explanation for the stimulus biases is that they were driven by uncertainty about which event rate corresponded to the category boundary (Vicente, Mendonca, Pouget, & Mainen, 2012).

Subjects accumulate sensory evidence similarly for low- and high-reliability stimuli

We previously demonstrated that our stimuli invite animals to accumulate sensory evidence over the majority of the trial (Raposo et al., 2012). We demonstrated this by computing a quantity, excess rate, which reveals the degree to which a particular moment in time influences the subject's eventual decision. Our quantity, excess rate, is similar to the quantity computed in a choice-triggered average (Kiani et al., 2008; Nienborg & Cumming, 2009; Sugrue et al., 2004) in that it considers the influence on subjects' decisions of events occurring within local temporal windows (Materials and methods). When excess rate > 0 at a particular time point, we conclude that stimuli at that time influence the decision. By comparing the time-course of the excess rate curves between different trial types, we gain insight into the animals' strategies. Here, we evaluate whether animals use time differently for high- versus low-reliability stimuli. If reliable and unreliable stimuli lead to similar strategies for evidence accumulation, the excess rate curves should be elevated above zero for the majority of the trial duration

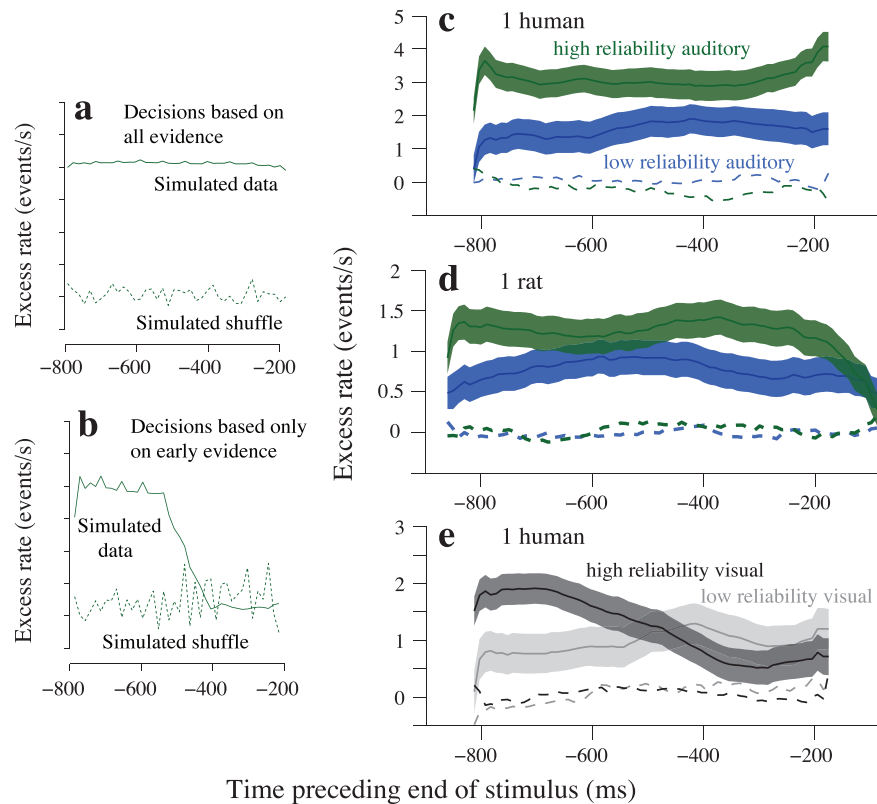


Figure 5. Excess rates: Predictions and individual subject examples. (a) Simulated excess rate curve for a decision process that is based on all evidence presented over a 1000-ms trial duration. Solid curve reflects stimuli from 2000 simulated trials assigned to a “high rate decision” or “low rate decision” pool based on the value of a decision variable at the end of the trial. Dashed curve reflects a shuffle control (Materials and methods). (b) Same as (a) but for a decision process that only considers evidence arriving in the first 500 ms of the trial. (c) Excess rate results for a single human subject on auditory trials. Abscissae indicate centers of sliding windows (milliseconds preceding end of stimulus). Shaded regions provide confidence bounds (mean \pm SE) on excess rate curves at each time point. Colors indicate reliability. Dashed lines: shuffle controls. (d) Results for a single rat on auditory trials. Conventions are the same as in (b). (e) Visual performance in a second human subject demonstrating an atypical evidence accumulation strategy. Note that the high-reliability visual trace is only elevated early in the trial for this subject.

regardless of stimulus reliability (Figure 5a). On the other hand, if reliable stimuli lead subjects to make faster decisions compared to less reliable stimuli, the excess rate curve for reliable stimuli will approach zero late in the trial (Figure 5b).

Because individual subjects may employ different strategies, we computed the excess rate for each subject separately. Examples are shown in Figure 5c through e. The excess rate curves for auditory trials in an example human (same subject used in Figures 2, 3) were elevated above 0 for nearly the entire trial duration (Figure 5c), indicating that this subject based decisions on evidence presented at all times. This confirms our previous observations in human subjects (Raposo et al., 2012). Further, the timecourses of the curves were similar for low- and high-reliability stimuli. This observation is novel: It refutes the possibility that the subject made faster decisions on high-reliability trials. If the subject had done so, the excess rate curve should have approached zero by midtrial for the high-

reliability stimuli, as shown in the simulated example (Figure 5b). An individual rat subject showed a similar profile of evidence accumulation on auditory trials (Figure 5d). Excess rates were elevated above zero at all timepoints, and showed a similar timecourse for reliable and unreliable stimuli. However, this rat differed from the human example in one respect: The excess rate curves decreased during the last 200 ms of the trial for high-reliability stimuli. We and others have previously observed a decline in excess rate during this period (Nienborg & Cumming, 2009; Raposo et al., 2012); we speculated that sensory stimuli influence the decision less during this time because the animals devote some cognitive resources to planning the full body movements required to communicate their decisions. Here, we note that the declines in excess rate differentially impacted the excess rate curves in the rats depending on sensory reliability: high-reliability trials suffered the greatest declines in excess rate relative to their levels earlier in the trial (Figure 5d). However,

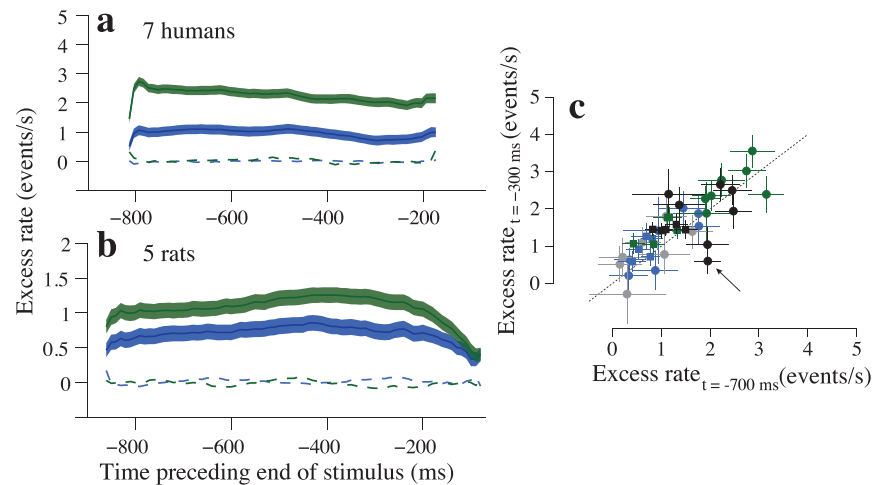


Figure 6. Pooled data indicate that most subjects integrate sensory evidence over the entire course of the trial. (a) Pooled auditory data from all seven humans. Line/color conventions are the same as in Figure 5. (b) Pooled auditory data for the five rats. (c) Scatter plot with data for all subjects comparing the value of excess rates early in the trial (700 ms before stimulus offset) and late in the trial (300 ms before stimulus offset). Color indicates unisensory trial type. Blue: low-reliability auditory; green: high-reliability auditory; gray: low-reliability visual; black: high-reliability visual. Shape indicates species. Circles: humans. Squares: rats. Arrow highlights a single human subject with unusual behavior for high-reliability visual trials. This is the same subject represented in Figure 5e.

because the observed declines in excess rate begin at roughly the same moments in time for high- and low-reliability trials (see Figure 6, below), the steeper declines for high-reliability stimuli likely reflect the fact that excess rates were higher for those stimuli throughout earlier parts of the trial (differences in excess rate magnitude are discussed in more detail below).

A final example illustrates a second human subject with a noticeably different strategy (Figure 5e). For this subject, the excess rate timecourses for visual stimuli differed strikingly depending on whether sensory reliability was high or low. For high-reliability (but not low-reliability) visual trials, the subject's excess rate curve decreased dramatically over the course of the trial (Figure 5e, dark gray trace). This temporal profile resembles our simulated example of decisions based only on evidence presented early in the trial (Figure 5b). Interestingly, this individual also overweighted visual evidence on multisensory trials (Figure 4b, triangles). Inability to accumulate evidence stably throughout the trial could thus potentially be a factor associated with nonoptimal cue combination on our task, though larger cohorts would be needed to draw conclusions regarding these uncommon departures from optimality (see Discussion).

Auditory data pooled across all human and rat subjects are shown in Figure 6. For both the composite human (Figure 6a) and rat (Figure 6b) datasets, excess rate curves were elevated over the entire course of the trial, and showed similar timecourses for reliable and unreliable stimuli. A similar timecourse was evident for visual stimuli (data not shown), except that the outlying

human subject depicted in Figure 5f drove a decrease late in the trial for high-reliability stimuli. A scatter plot including all subjects and single sensory trial types highlights that deviations from this tendency were rare (Figure 6c). Points near the $x = y$ line indicate subjects whose excess rate values were similar early and late in the trial. Most subjects' results fell close to this line. Points below the $x = y$ line indicate subjects who had lower excess rate curves late in the trial compared to early in the trial. The human subject with the unusual excess rate timecourses (Figure 5e) is indicated by an arrow.

In addition to our observations regarding the timecourses of excess rates for reliable and unreliable stimuli, we also evaluated the magnitude of excess rates. We found that excess rates were greater in magnitude for reliable compared to unreliable stimuli in both the humans and rats (Figure 6a, b: compare green and blue curves). This effect was significant for the data pooled across all seven humans (Figure 6a; low-reliability auditory: 0.98 [0.78, 1.18], high-reliability auditory: 2.33 [2.13, 2.53], 95% CIs; $p < 10^{-5}$) and also in the data pooled across the five rats (Figure 6b; low-reliability auditory: 0.74 [0.62, 0.86]; high-reliability auditory: 1.11 [0.99, 1.23]; $p < 10^{-5}$). These differences in excess rate magnitude were expected; they indicate that fluctuations in event rate occurring throughout the trial more strongly discriminated subjects' decisions when sensory reliability was high, whereas subjects' abilities to discriminate higher and lower event rates suffered on low-reliability trials across the entire stimulus duration, likely due to increased levels of sensory noise.

Discussion

In this study, we measured how humans and rats weighted time-varying auditory and visual stimuli on a rate estimation task. This was accomplished using established methods for estimating subjects' perceptual weights and applying the methods to a novel paradigm we have developed to study decision-making in humans and rodents. We report three main findings: (a) Dynamic weighting of sensory inputs extends to time-varying stimuli. (b) Dynamic weighting of sensory stimuli is not restricted to primates. (c) Regardless of reliability, most subjects based decisions on sensory evidence presented throughout the trial duration. Subjects who accumulated sensory evidence in a nonstandard way were rare, but underscore the importance of using stimuli that allow one to investigate the temporal dynamics of multisensory processing. The importance of each of these conclusions is described below.

Our conclusions about time-varying stimuli demonstrate that dynamic weighting extends to a much broader class of stimuli than have been studied previously. Time-varying stimuli represent a particularly important subset of multisensory stimuli because they are encountered in many, if not most, ecological situations, including predator movement and conspecific vocalizations (Ghazanfar et al., 2007; Maier et al., 2008; Maier & Ghazanfar, 2007; Sugihara, Diltz, Averbeck, & Romanski, 2006; Thomas & Shiffrar, 2010). Observing dynamic weighting on a decision task featuring stochastic, time-varying stimuli suggests that these behaviors are supported by highly flexible neural computations. Given that dynamic weighting in humans has also been reported for many other within- and cross-modal cue integration tasks (albeit typically involving static stimuli; see Trommershauser, Kording, & Landy, 2011 for review), the underlying computations may be a widespread feature of neural circuits across many brain areas involved in processing different types of sensory stimuli.

Extending dynamic weighting to rodents indicates that the ability to estimate stimulus reliability for dynamic stimuli is conserved across diverse species in the mammalian lineage. Although previous behavioral studies on multisensory integration have been conducted in rats (Hirokawa, Bosch, Sakata, Sakurai, & Yamamori, 2008; Hirokawa et al., 2011; Sakata, Yamamori, & Sakurai, 2004), they have not systematically varied stimulus reliability in a way that made it possible to estimate perceptual weights. The dynamic weighting we observed in rats suggests that the ability to estimate reliability and use such estimates to guide decisions likely relies on neural mechanisms common across many species. Further, by establishing dynamic weighting for rodents, we open the possibility of using

this species to examine the underlying neural circuits that drive this behavior.

Comparing the timecourses of decision-making for reliable versus unreliable stimuli is critical because it allows us to identify differences in the strategies subjects use to accumulate evidence for their decisions. Previous studies of visual decision-making using dynamic random-dot stimuli revealed that decisions are sometimes influenced most by information presented at the beginning of the trial (Kiani et al., 2008). Those results argue that, for some stimuli, subjects might accumulate evidence up to a threshold level, even in tasks such as ours where experimenters impose a fixed trial duration. This could have important consequences for comparisons of responses to reliable versus less reliable stimuli, which are standard in multisensory paradigms. Specifically, it raises the possibility that evidence accumulation might not only stop part way through the trial, but might stop at different times depending on stimulus reliability. This could have profound implications for the framework of optimal integration, which does not consider the possibility that subjects use time in a reliability-dependent manner. Our analyses revealed that most subjects were influenced by sensory evidence occurring at similar times within the trial for both low- and high-reliability stimuli. This provides reassurance that, at least for fixed-duration stimuli that invite evidence accumulation over time, subjects generally harvest information in a similar manner regardless of sensory reliability.

The tendency to integrate incoming sensory stimuli over an entire trial likely depends strongly on the type of stimulus used. Decisions that are made about suprathreshold stimuli are likely made very quickly, because integrating over longer time periods would provide only negligible improvements in accuracy. In future studies, we plan to explicitly vary the degree to which different times within the trial are informative for making the correct decision. Such manipulations might lead subjects to have shorter, more concentrated integration times. However, our current stimulus was designed to encourage subjects to integrate over long time periods and our analysis indicates that subjects did indeed integrate over the entire trial. Stimuli like ours that are dynamic and time-varying will tend to invite long integration times, and can therefore offer insight into cognitive processes that operate on more flexible timescales than sensorimotor reflexes.

Although we found that both humans and rats accumulated evidence over most of the trial, we did observe some deviations from this trend. The largest deviation was evident in a single human subject: for high-reliability visual trials, this subject tended to ignore evidence late in the trial (Figure 5e, dark gray trace). Interestingly, this subject also showed pronounced deviations from optimality on multisensory

trials featuring high-reliability visual stimuli (Figure 4b, triangles). Indeed, this subject had the largest deviations from the optimal prediction of any subject tested, rat or human. The current experimental design does not provide direct evidence as to whether deviations from optimal cue integration arose from unequal utilization of evidence across the stimulus duration. Nevertheless, the coincidence is intriguing. One possibility is that this subject's tendency to make “snap” decisions on high-reliability visual trials led us to under-estimate his true reliability for those stimuli. When visual stimuli were presented in a multisensory context, they may then have influenced the decision more than we predicted. By our analysis, this appeared as a deviation from optimality, but it could alternatively have reflected our inability to characterize sensory reliability for subjects with unusual evidence accumulation strategies. In any case, these observations serve as a cautionary tale: Although subjects may generally accumulate information stably over time on perceptual tasks, this is by no means guaranteed and may lead to marked departures from optimality in the context of cue integration. Indeed, it is possible that these effects may explain deviations from optimality that have been reported in other human papers as well (Rosas et al., 2005; Rosas et al., 2007; Zalevski et al., 2007). However, the stimuli used in past human studies did not permit the opportunity to test, as we have here, how subjects accumulated information over time during the trial. This underscores the importance of utilizing time-varying stimuli in studies of multisensory decision-making.

Like most human subjects, rats' decisions were also influenced by evidence presented throughout the entire trial duration. There was a tendency for rats to down-weight sensory evidence arriving near the end of the trial, but this was evident for both reliable and unreliable stimuli of either modality (Figure 6b), and so is unlikely to have produced any systematic effects on subjects' perceptual weights. A more systematic deviation in the rats was evident in the magnitude of their perceptual weights: Rats tended to slightly underweight reliable visual stimuli when they were presented alongside less reliable auditory stimuli (Figure 4d). Note that rats did not ignore the visual stimulus; indeed, it influenced their decisions more strongly than the auditory stimulus on these trials. However, the predicted optimal solution would have been to weight the visual stimulus even more than observed in the rats' decisions. Although deviations from optimality in our human cohort were somewhat more idiosyncratic, humans tended to show the opposite deviations: They occasionally overweighted visual evidence when auditory reliability was high. These contrasting deviations from optimality between rats and humans are intriguing given the differences in visual acuity between rodents and humans (Busse et al., 2011; Chalupa &

Williams, 2008; Prusky, West, & Douglas, 2000). One possibility is that humans and rodents have natural tendencies to over- or under-weight visual inputs (respectively) because of lifetimes of experience with either high- or low-acuity visual systems. However, the abilities of both rats and humans to use time-varying auditory and visual stimuli and weight them according to sensory reliability is clear from this study (Figure 4a, c). The ease with which rats dynamically reweighted inputs, even when reliability levels changed unpredictably from trial to trial, suggests that rodents, like primates, possess flexible neural circuits that are designed to exploit all incoming sensory information regardless of its modality.

What neural mechanisms might underlie this ability to flexibly adjust perceptual weights? Although a wealth of multisensory experiments have been carried out in anesthetized animals (Jiang, Wallace, Jiang, Vaughan, & Stein, 2001; Meredith, Nemitz, & Stein, 1987; Stanford, Quessy, & Stein, 2005), many fewer have been carried out in behaving animals; as a result, much about the underlying neural mechanisms for optimal integration remain unknown. Here, our subjects reweighted sensory inputs even when the relative reliabilities varied from trial to trial, suggesting that the dynamic weighting could not have resulted from long-term changes in synaptic strengths (for instance, between primary sensory areas and downstream targets). The required timescales of such mechanisms are far too long to explain dynamic weighting. One possibility is that populations of cortical neurons automatically encode stimulus reliability due to the firing rate statistics of cortical neurons. Assuming Poisson-like firing statistics, neural populations naturally reflect probability distributions (Salinas & Abbott, 1994; Sanger, 1996). Unreliable stimuli may generate population responses with reduced gain and increased variability at the population level (Beck et al., 2008; Deneve, Latham, & Pouget, 2001; Ma, Beck, Latham, & Pouget, 2006). Such models of probabilistic population coding offer an explanation for how dynamic cue weighting might be automatically implemented as a circuit mechanism without changes in synaptic strengths. A plausible circuit implementation of such a coding scheme has been recently described in the context of multisensory integration (Ohshiro, Angelaki, & DeAngelis, 2011); this model allows for random connectivity among populations of sensory neurons and achieves sensitivity to stimulus reliability using well-established mechanisms of divisive normalization (Carandini, Heeger, & Movshon, 1997; Heeger, 1993; Sclar & Freeman, 1982).

A competing explanation for multisensory enhancement is that it arises from synchronous activity between areas responsive to each individual sensory modality (for review, see Senkowski, Schneider, Foxe, & Engel,

2008). Indeed, classic work in the superior colliculus suggests that precise timing of sensory inputs is crucial for multisensory enhancement of neural responses (Meredith et al., 1987), and psychophysical effects can likewise require precise timing of the relevant inputs (Lovelace, Stein, & Wallace, 2003; Shams et al., 2002). By contrast, multisensory improvements on our task do not require synchronous auditory and visual stimuli (Raposo et al., 2012 and current study). Our subjects' ability to combine independent streams of stochastic auditory and visual information bearing on a single perceptual judgment is testament to the flexibility of multisensory machinery in the mammalian brain.

Conclusions

Multisensory research has the potential to provide insight into a very general problem in neuroscience: How can the brain make the best decisions possible in light of the inevitable uncertainty inherent to sensory signals from our environments? In natural behaviors, subjects must often make inferences based on noisy stimuli that change rapidly over time (Rieke, 1997). Our findings make clear that diverse mammalian species possess the circuitry needed to use reliability to guide decisions involving such time-varying stimuli. Further, by demonstrating this ability in both humans and rats, our experiments suggest that dynamic weighting may rely on neural computations that are conserved across these species.

Keywords: decision-making, multisensory integration, rodent, sensory reliability, cue weighting, psychophysics

Acknowledgments

We thank Stephen Shea, Greg DeAngelis, Anthony Zador, Alexandre Pouget and Joshua Dubnau for comments and useful discussions, and Petr Znamenskiy, Santiago Jaramillo, Haley Zamer, Amanda Brown and Barry Burbach for technical assistance. We also thank Matteo Carandini, Chris Burgess, Matthew Kaufman, and Amanda Brown for providing feedback on an earlier version of this manuscript. Funding for this work was provided to AKC through NIH EY019072, EY022979, the John Merck Fund, the McKnight Foundation, and the Marie Robertson Memorial Fund of Cold Spring Harbor Laboratory. JPS was supported by a National Defense Science and Engineering Graduate Fellowship from the U.S. Department of Defense, and by the Watson School of Biological Sciences through NIH training grant 5T32GM065094. DR was supported by doctoral

fellowship SFRH/BD/51267/2010 provided through FCT Portugal.

Commercial relationships: none.

Corresponding author: John P. Sheppard.

Email: sheppard@cshl.edu.

Address: Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA.

References

- Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology*, *14*(3), 257–262.
- Alais, D., Newell, F. N., & Mamassian, P. (2010). Multisensory processing in review: From physiology to behaviour. *Seeing and Perceiving*, *23*(1), 3–38.
- Battaglia, P. W., Di Luca, M., Ernst, M. O., Schrater, P. R., Machulla, T., & Kersten, D. (2010). Within- and cross-modal distance information disambiguate visual size-change perception. *PLoS Computational Biology*, *6*(3), e1000697.
- Beck, J. M., Ma, W. J., Kiani, R., Hanks, T., Churchland, A. K., Roitman, J., et al. (2008). Probabilistic population codes for Bayesian decision making. *Neuron*, *60*(6), 1142–1152.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B (Methodological)*, *57*(1), 289–300.
- Busse, L., Ayaz, A., Dhruv, N. T., Katzner, S., Saleem, A. B., Scholvinck, M. L., et al. (2011). The detection of visual contrast in the behaving mouse. *The Journal of Neuroscience*, *31*(31), 11351–11361.
- Carandini, M., Heeger, D. J., & Movshon, J. A. (1997). Linearity and normalization in simple cells of the macaque primary visual cortex. *The Journal of Neuroscience*, *17*(21), 8621–8644.
- Chalupa, L. M., & Williams, R. W. (2008). *Eye, retina, and visual system of the mouse*. Cambridge, MA: MIT Press.
- Cisek, P., Puskas, G. A., & El-Murr, S. (2009). Decisions in changing conditions: The urgency-gating model. *The Journal of Neuroscience*, *29*(37), 11560–11571.
- Deneve, S., Latham, P. E., & Pouget, A. (2001). Efficient computation and cue integration with noisy population codes. *Nature Neuroscience*, *4*(8), 826–831.
- Ernst, M. O., & Banks, M. S. (2002). Humans integrate

- visual and haptic information in a statistically optimal fashion. *Nature*, 415(6870), 429–433.
- Fetsch, C. R., Pouget, A., DeAngelis, G. C., & Angelaki, D. E. (2012). Neural correlates of reliability-based cue weighting during multisensory integration. *Nature Neuroscience*, 15(1), 146–154.
- Fetsch, C. R., Turner, A. H., DeAngelis, G. C., & Angelaki, D. E. (2009). Dynamic reweighting of visual and vestibular cues during self-motion perception. *The Journal of Neuroscience*, 29(49), 15601–15612.
- Ghazanfar, A. A., Turesson, H. K., Maier, J. X., van Dinther, R., Patterson, R. D., & Logothetis, N. K. (2007). Vocal-tract resonances as indexical cues in rhesus monkeys. *Current Biology*, 17(5), 425–430.
- Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual Review of Neuroscience*, 30, 535–574.
- Heeger, D. J. (1993). Modeling simple-cell direction selectivity with normalized, half-squared, linear operators. *Journal of Neurophysiology*, 70(5), 1885–1898.
- Hillis, J. M., Ernst, M. O., Banks, M. S., & Landy, M. S. (2002). Combining sensory information: Mandatory fusion within, but not between, senses. *Science*, 298(5598), 1627–1630.
- 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, 967–992, <http://www.journalofvision.org/content/4/12/1>, doi:10.1167/4.12.1. [PubMed] [Article]
- Hirokawa, J., Bosch, M., Sakata, S., Sakurai, Y., & Yamamori, T. (2008). Functional role of the secondary visual cortex in multisensory facilitation in rats. *Neuroscience*, 153(4), 1402–1417.
- Hirokawa, J., Sadakane, O., Sakata, S., Bosch, M., Sakurai, Y., & Yamamori, T. (2011). Multisensory information facilitates reaction speed by enlarging activity difference between superior colliculus hemispheres in rats. *PLoS One*, 6(9), e25283.
- Howard, I. P., & Templeton, W. B. (1966). *Human spatial orientation*. London, UK: Wiley.
- Huk, A. C., & Meister, M. L. (2012). Neural correlates and neural computations in posterior parietal cortex during perceptual decision-making. *Frontiers in Integrative Neuroscience*, 6, 86.
- Jacobs, R. A. (1999). Optimal integration of texture and motion cues to depth. *Vision Research*, 39(21), 3621–3629.
- Jiang, W., Wallace, M. T., Jiang, H., Vaughan, J. W., & Stein, B. E. (2001). Two cortical areas mediate multisensory integration in superior colliculus neurons. *Journal of Neurophysiology*, 85(2), 506–522.
- Kesten, H. (1958). Accelerated stochastic-approximation. *Annals of Mathematical Statistics*, 29(1), 41–59.
- Kiani, R., Hanks, T. D., & Shadlen, M. N. (2008). Bounded integration in parietal cortex underlies decisions even when viewing duration is dictated by the environment. *The Journal of Neuroscience*, 28(12), 3017–3029.
- 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.
- Landy, M. S., & Kojima, H. (2001). Ideal cue combination for localizing texture-defined edges. *Journal of The Optical Society of America A: Optics, Image, Science, and Vision*, 18(9), 2307–2320.
- Lovelace, C. T., Stein, B. E., & Wallace, M. T. (2003). An irrelevant light enhances auditory detection in humans: A psychophysical analysis of multisensory integration in stimulus detection. *Brain Research. Cognitive Brain Research*, 17(2), 447–453.
- Ma, W. J., Beck, J. M., Latham, P. E., & Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature Neuroscience*, 9(11), 1432–1438.
- Maier, J. X., Chandrasekaran, C., & Ghazanfar, A. A. (2008). Integration of bimodal looming signals through neuronal coherence in the temporal lobe. *Current Biology*, 18(13), 963–968.
- Maier, J. X., & Ghazanfar, A. A. (2007). Looming biases in monkey auditory cortex. *The Journal of Neuroscience*, 27(15), 4093–4100.
- Meredith, M. A., Nemitz, J. W., & Stein, B. E. (1987). Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. *The Journal of Neuroscience*, 7(10), 3215–3229.
- Nienborg, H., & Cumming, B. G. (2009). Decision-related activity in sensory neurons reflects more than a neuron's causal effect. *Nature*, 459(7243), 89–92.
- Ohshiro, T., Angelaki, D. E., & DeAngelis, G. C. (2011). A normalization model of multisensory integration. *Nature Neuroscience*, 14(6), 775–782.
- Pelli, D. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10(4), 437–442.
- Powers, A. R., 3rd, Hillock, A. R., & Wallace, M. T. (2009). Perceptual training narrows the temporal window of multisensory binding. *The Journal of Neuroscience*, 29(39), 12265–12274.

- Prusky, G. T., West, P. W., & Douglas, R. M. (2000). Behavioral assessment of visual acuity in mice and rats. *Vision Research*, *40*(16), 2201–2209.
- Raposo, D., Sheppard, J. P., Schrater, P. R., & Churchland, A. K. (2012). Multisensory decision-making in rats and humans. *The Journal of Neuroscience*, *32*(11), 3726–3735.
- Rieke, F. (1997). *Spikes: Exploring the neural code*. Cambridge, MA: MIT Press.
- Rosas, P., Wagemans, J., Ernst, M. O., & Wichmann, F. A. (2005). Texture and haptic cues in slant discrimination: Reliability-based cue weighting without statistically optimal cue combination. *Journal of the Optical Society of America A, Optics, Image, Science and Vision*, *22*(5), 801–809.
- Rosas, P., Wichmann, F. A., & Wagemans, J. (2007). Texture and object motion in slant discrimination: Failure of reliability-based weighting of cues may be evidence for strong fusion. *Journal of Vision*, *7*(6):3, 1–21, <http://www.journalofvision.org/content/7/6/3>, doi:10.1167/7.6.3. [PubMed] [Article]
- Sakata, S., Yamamori, T., & Sakurai, Y. (2004). Behavioral studies of auditory-visual spatial recognition and integration in rats. *Experimental Brain Research*, *159*(4), 409–417.
- Salinas, E., & Abbott, L. F. (1994). Vector reconstruction from firing rates. *Journal of Computational Neuroscience*, *1*(1-2), 89–107.
- Sanger, T. D. (1996). Probability density estimation for the interpretation of neural population codes. *Journal of Neurophysiology*, *76*(4), 2790–2793.
- Sclar, G., & Freeman, R. D. (1982). Orientation selectivity in the cat's striate cortex is invariant with stimulus contrast. *Experimental Brain Research*, *46*(3), 457–461.
- Senkowski, D., Schneider, T. R., Foxe, J. J., & Engel, A. K. (2008). Crossmodal binding through neural coherence: Implications for multisensory processing. *Trends in Neurosciences*, *31*(8), 401–409.
- Serwe, S., Kording, K. P., & Trommershauser, J. (2011). Visual-haptic cue integration with spatial and temporal disparity during pointing movements. *Experimental Brain Research*, *210*(1), 67–80.
- Shams, L. (2012). Early integration and Bayesian causal inference in multisensory perception. In M. M. Murray & M. T. Wallace (Eds.), *The neural bases of multisensory processes*. Boca Raton, FL: CRC Press.
- Shams, L., Kamitani, Y., & Shimojo, S. (2000). Illusions. What you see is what you hear. *Nature*, *408*(6814), 788.
- Shams, L., Kamitani, Y., & Shimojo, S. (2002). Visual illusion induced by sound. *Brain Research. Cognitive Brain Research*, *14*(1), 147–152.
- Shams, L., Ma, W. J., & Beierholm, U. (2005). Sound-induced flash illusion as an optimal percept. *NeuroReport*, *16*(17), 1923–1927.
- Stanford, T. R., Quessy, S., & Stein, B. E. (2005). Evaluating the operations underlying multisensory integration in the cat superior colliculus. *The Journal of Neuroscience*, *25*(28), 6499–6508.
- Sugihara, T., Diltz, M. D., Averbeck, B. B., & Romanski, L. M. (2006). Integration of auditory and visual communication information in the primate ventrolateral prefrontal cortex. *The Journal of Neuroscience*, *26*(43), 11138–11147.
- Sugrue, L. P., Corrado, G. S., & Newsome, W. T. (2004). Matching behavior and the representation of value in the parietal cortex. *Science*, *304*(5678), 1782–1787.
- Thomas, J. P., & Shiffrar, M. (2010). I can see you better if I can hear you coming: Action-consistent sounds facilitate the visual detection of human gait. *Journal of Vision*, *10*(12):14, 1–11, <http://www.journalofvision.org/content/10/12/14>, doi:10.1167/10.12.14. [PubMed] [Article]
- Triesch, J., Ballard, D. H., & Jacobs, R. A. (2002). Fast temporal dynamics of visual cue integration. *Perception*, *31*(4), 421–434.
- Trommershauser, J., Kording, K., & Landy, M. S. (2011). *Sensory cue integration*. New York: Oxford University Press.
- Vicente, M., Mendonca, A., Pouget, A., & Mainen, Z. (2012). *Odor detection vs. mixture categorization: Crucial differences in behavioral and learning strategies*. Paper presented at The 9th Annual Computational and Systems Neuroscience (Co-syne) Meeting, Salt Lake City, UT, USA.
- Weiss, Y., Simoncelli, E. P., & Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nature Neuroscience*, *5*(6), 598–604.
- Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics*, *63*(8), 1293–1313.
- Young, M. J., Landy, M. S., & Maloney, L. T. (1993). A perturbation analysis of depth perception from combinations of texture and motion cues. *Vision Research*, *33*(18), 2685–2696.
- Zalevski, A. M., Henning, G. B., & Hill, N. J. (2007). Cue combination and the effect of horizontal disparity and perspective on stereoacuity. *Spatial Vision*, *20*(1-2), 107–138.