



Published in final edited form as:

J Neurosci. 2012 March 14; 32(11): 3726–3735. doi:10.1523/JNEUROSCI.4998-11.2012.

Multisensory decision-making in rats and humans

David Raposo^{1,2}, John P. Sheppard^{1,3}, Paul R. Schrater⁴, and Anne K. Churchland¹

¹Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA

²Chamपालimaud Neuroscience Programme, Lisboa, Portugal

³Watson School of Biological Sciences, Cold Spring Harbor, NY 11724, USA

⁴University of Minnesota, Minnesota, MN 55455, USA

Abstract

We report a novel multisensory decision task designed to encourage subjects to combine information across both time and sensory modalities. We presented subjects, humans and rats, with multisensory event streams, consisting of a series of brief auditory and/or visual events. Subjects made judgments about whether the event rate of these streams was high or low. We have three main findings: First, we report that subjects can combine multisensory information over time to improve judgments about whether a fluctuating rate is high or low. Importantly, the improvement we observed was frequently close to, or better than, the statistically optimal prediction. Second, we found that subjects showed a clear multisensory enhancement both when the inputs in each modality were redundant and also when they provided independent evidence about the rate. This latter finding suggests a model where event rates are estimated separately for each modality and fused at a later stage. Finally, because a similar multisensory enhancement was observed in both humans and rats, we conclude that the ability to optimally exploit sequentially presented multisensory information is not restricted to a particular species.

Introduction

A large body of work has shown that animals and humans are able to combine information across time to make decisions in some circumstances (Roitman and Shadlen, 2002; Mazurek et al., 2003; Palmer et al., 2005; Kiani et al., 2008). Specifically, combining information across time can be a good strategy for generating accurate decisions when incoming signals are noisy and unreliable (Link and Heath, 1975; Gold and Shadlen, 2007). For noisy and unreliable signals, combining evidence across sensory modalities might likewise improve decision accuracy, but little is known about whether the framework for understanding combining information over time might extend to combining information across sensory modalities.

The ability of humans to combine multisensory information to improve perceptual judgments for static information has been well-established (for review see Alais et al., 2010). Psychophysical studies have shown that multisensory enhancement requires that information from the two modalities be presented within a temporal “window” (Slutsky and Recanzone, 2001; Miller and D’Esposito, 2005). Physiological studies suggest the same: superior colliculus neurons show enhanced responses for multisensory stimuli only when those stimuli occur close together in time (Meredith et al., 1987). Such a temporally precise mechanism would serve a useful purpose for localizing or detecting objects. There, temporal

synchrony (or near-synchrony) provides an important cue that two sensory signals are related to the same object (Lovelace et al., 2003; Burr et al., 2009).

In other circumstances, multisensory decisions might have more lax requirements for the relative timing of events in each modality. For example, when multiple auditory and visual events arrive in a continuous stream, temporal synchrony of individual events might be difficult to assess. Auditory and visual stimuli drive neurons with different latencies (Pfungst and O'Connor, 1981; Maunsell and Gibson, 1992; Recanzone et al., 2000), making it difficult to determine which events belong together. It is not known whether information in streams of auditory and visual events can be combined to improve perceptual judgments; if they can, this could be evidence for a different mechanism of multisensory integration that has less strict temporal requirements compared to the classic, synchrony-dependent mechanisms.

To invite subjects to combine information across both time and sensory modalities for decisions, we designed an audiovisual rate discrimination decision task. In the task, subjects are presented with a series of auditory and/or visual events and report whether they perceive the event rates to be high or low. Because of differing latencies for the auditory and visual systems, our stimulus would pose a challenge to synchrony-dependent mechanisms of multisensory processing. Nevertheless, we saw a pronounced multisensory enhancement in both humans and rats. Importantly, this enhancement was present whether the events streams were identical and played synchronously or were independently generated, suggesting that the enhancement did not rely on mechanisms that require precise timing. Taken together, these results suggest that some mechanisms of multisensory enhancement might exploit abstract information that is accumulated over the trial duration, and therefore must rely on neural circuitry that does not require precise timing of auditory and visual events.

Materials and Methods

Behavioral task

We presented subjects with a rate discrimination decision task designed to invite subjects to combine information across time and sensory modalities. Each trial consisted of a series of auditory or visual “events” (duration=10 ms long for humans, 15 ms for rats), with background noise between events (Fig. 1a, top). Visual events were flashes of light and auditory events were brief sounds (see below for methodological details particular to each species). The amplitude of the events was adjusted for each subject so that on the single sensory trials performance was approximately 70–80% correct and was matched for audition and vision. We chose these values because previous studies have indicated that multisensory enhancement is the largest when individual stimuli are weak (Stanford et al., 2005). This appears to be particularly true for synchrony-dependent mechanisms of multisensory integration (Meredith et al., 1987).

Trials were generated so that the instantaneous event rate fluctuated over the course of the trial. Each trial was created by sequentially selecting one of two inter-event intervals: either a long duration or a short duration (Fig. 1a, bottom) until the total trial duration exceeded 1000 ms (or occasionally slightly longer/shorter durations for catch trials; see below). Trial difficulty was determined by the proportion of inter-event intervals from each duration. As the proportion of short intervals varies from zero to one, the average rate smoothly changes from clearly low to clearly high. For example, when all of the inter-event intervals were long, the average rate was clearly low (Fig. 1a, left); and similarly when all of the inter-event intervals were taken from the short interval, the average rate was clearly high (Fig. 1a, right). When inter-event intervals were taken more evenly from the two values, the average

of the fluctuating rate was intermediate between the two (Fig. 1a, middle). When the number of long intervals exceeded the number of short intervals, subjects were rewarded for making a “low rate” choice and vice versa. When the number of short and long intervals in a trial were equal, subjects were rewarded randomly. Note that in terms of average rate, this reward scheme places the “low rate”-“high rate” category boundary closer to the lower extreme (all long durations) than to the higher extreme (all short durations) because of the differing duration of the intervals. The strategies of both human and rat subjects reflect this: typically, subjects’ points of subjective equality (PSEs) were closer to the lowest rate and were less than the median of the set of unique stimulus rates presented. Nevertheless, for simplicity, we plot subjects’ choices as a function of stimulus rate. Nearly equivalent results were achieved when we analyzed subjects’ responses as a function of the number of short intervals in a trial rather than stimulus rate.

For single sensory trials, event streams were presented to just the auditory or just the visual system. Visual trials were always 1000 ms long. Auditory trials were usually 1000 ms long but we sometimes included “catch trials” that were 800 or 1200 ms. Catch trials were collected for four human subjects. The purpose of the catch trials was to determine whether subjects’ decisions were based on just event counts or on event counts relative to the total duration over which those counts occurred. Catch trials constituted a total of 2.31% of the total trials. We reasoned that using such a small percentage of the trials for variable durations would make it possible to probe subjects’ strategy without encouraging them to alter it. For these trials, we rewarded subjects based on event counts rather than the rate. This should have increased the likelihood that subjects would have made their decisions based on count, if they were able to detect that stimuli were sometimes of variable duration.

For multisensory trials, both auditory and visual stimuli were present and were played for 1000 ms. To distinguish possible strategies for improvement on multisensory trials, we employed two conditions. First, in the *synchronous condition*, identically timed event streams were presented to the visual and auditory systems simultaneously (Fig. 1b). Second, in the *independent condition*, auditory and visual event streams were on average in support of the same decision (i.e., the proportion of inter-event intervals taken from each duration was the same), but each event stream was generated independently (Fig. 1c). As a result, auditory and visual events frequently did not occur at the same time. Auditory and visual events did not occur simultaneously even for the highest stimulus strengths because we imposed a 30 ms delay between events. Although a 30 ms delay does prevent auditory and visual events from being synchronous at the highest and lowest rates, the delay may be too brief to prevent auditory and visual stimuli from being perceived as synchronous (Fujisaki and Nishida, 2005). To be sure that our conclusions about the independent condition were not affected by these “perceptually synchronous” trials at the highest and lowest rates, we analyzed the independent condition both with and without the easiest trials (see Results). The multisensory effects we observed were very similar regardless of whether or not we included the easy trials.

Because auditory and visual event streams were generated independently, trials for this condition fell into one of three categories:

1. *Matched trials*, where auditory and visual event streams had the same number of events. Example match trials are shown in Figure 1c.
2. *Bonus trials*, where both modalities provided evidence for the same choice, but one modality provided evidence that was 1 events/s stronger than the other (i.e. auditory and visual streams had proportions of short or long durations both above or below 0.5, but were not equal).

3. *Neutral trials*, where only one modality provided evidence for a particular choice while the other modality provided evidence that was so close to the PSE that it didn't support one choice or the other.
4. *Conflict trials*, where each modality provided evidence for a different choice. Conflict trials were rare, were rewarded randomly, and were excluded from further analysis. Because we aimed to match the reliability of the single cue trials, conflict trials were not expected to reveal differential weighting of sensory cues. This is unlike other multisensory paradigms (Fine and Jacobs, 1999; Hillis et al., 2004; Fetsch et al., 2009) in which the reliability of the two single sensory stimuli is explicitly varied.

Single sensory and multisensory trials were interleaved. For all three rats and for six human subjects, we collected data for subjects from the independent and synchronous conditions on separate days (Figs 2,3,4 and 6). However, we included one control condition where 4 different human subjects were presented with interleaved blocks of the synchronous and independent conditions (Fig 5).

Human subjects

We report data from ten human volunteers (five male, five female, age 22–60) with normal or corrected-to-normal vision and normal hearing. Two subjects were authors; the remaining eight subjects were naïve about the experiment. Volunteers were recruited through fliers posted at Cold Spring Harbor Laboratory. Experiments were conducted in a large sound-isolating booth (Industrial Acoustics, Bronx NY).

Visual stimuli were displayed on a cathode ray tube monitor (Dell M991) using a refresh rate of 100Hz. Subjects were seated comfortably in front of the monitor; the distance from their eyes to the screen was approximately 51 cm. Stimulus presentation and data acquisition were controlled by the Psychophysics Toolbox (Pelli) running on Matlab (Mathworks, Natick, MA) on a Quad Core Macintosh computer. Subjects were instructed to fixate a central black spot ($0.22 \times 0.24^\circ$ of visual angle); eye position was not monitored since small deviations in eye position are unlikely to impact rate judgments about a large flashing stimulus. After a delay (500 ms), the stimulus presentation began.

Auditory events were pure tones (220 Hz) that were played from a single speakers attached to the left side of the monitor. Speakers were generic mini USB speakers (4.8×7.2 cm; Hewlett Packard, azimuth = 22.6°) that produced 78 dB-SPL in the range of 200–240 Hz at the position of the subject (tested using a pressure-field microphone, Brüel & Kjær). Waveforms were created in software at a sampling rate of 44 KHz and delivered to speakers through a digital (TOSlink optical) audio output using the PsychPortAudio function in the Psychophysics Toolbox. The visual stimulus, a flashing square that subtended $10^\circ \times 10^\circ$ of visual angle (azimuth = 17.16°), was positioned eccentrically so that its leftmost edge touched the left side of the screen. This configuration meant that auditory and visual stimuli were separated by only 3.5 cm (the width of the plastic frame of the CRT). The top of the speaker was collinear with the top of the flashing visual square. We positioned the stimuli close together because spatial proximity has been previously shown to encourage multisensory binding (Slutsky and Recanzone, 2001; Kording et al., 2007). The timing of auditory and visual events was checked regularly using a photodiode and a microphone connected to an oscilloscope. Both auditory and visual events were played amid background white noise. For the visual stimulus, the white noise was restricted to the region of the screen where the stimulus was displayed.

Subjects reported their decisions by pressing one of two keys on a standard keyboard. They received auditory feedback about their choices: correct choices resulted in a high tone (6400

Hz) and incorrect choices resulted in a low tone (200 Hz). We provided feedback to the human subjects so that their experience with the task would be as similar as possible to the rats. Feedback for the rats was essential since the liquid rewards the rats receive motivate them to do the task (see below). The two intervals used to generate the stimuli were 60 ms and 120 ms. Individual events were 10 ms. The resulting trials had fluctuating rates whose averages ranged from 7 Hz to 15 Hz.

We trained subjects for 4–6 days so that they could learn the association between stimulus rate and the correct choice. We began by presenting very salient unimodal stimuli, both visual and auditory; initially, the rates used were uniformly low or uniformly high. Over the course of several of days, we used a staircase procedure to gradually make the task more difficult by lowering the amplitude (contrast or volume) of the stimuli. We also began to include trials with mixtures of long and short inter-event intervals so that the overall rate was harder to judge, as well as multisensory trials. Subjects typically achieved a high performance level on both unisensory and multisensory trials within 3 days. We typically used an additional 2–3 days for training during which we made slight adjustments to the auditory and visual background noise so that the two modalities were equally reliable (as measured by the subjects' thresholds on single modality trials). After this additional training, subjects' performance was typically well-matched for the two modalities and also very stable from session to session. Once trained, subjects began each additional day with a 100 trial warm-up session. They then completed approximately 5 blocks of 160 trials on a given day.

Animal subjects

We collected data from three adult male Long Evans rats (250–350g, Taconic Farms) that were trained to do freely-moving two-alternative forced-choice behavior in a sound isolating booth (Industrial Acoustics, Bronx, NY). Trials were initiated when the rats poked their snouts into a centrally positioned port (Fig. 1d). Placing their snouts in this port broke an infrared beam; this event triggered onset of the visual or auditory stimulus.

Auditory events were played from a single, centrally positioned speaker and consisted of pure tones (15 KHz, rat 1 and rat 2) or bursts of white noise (rat 3), with sinusoidal amplitude modulation. Speakers were generic electromagnetic dynamic speakers (Harman Kardon) calibrated using a pressure-field microphone (Brüel & Kjær) to produce 75 dB-SPL in the range of 5–40 kHz at the position of the subject. Waveforms were created in software at a sampling rate of 200 KHz and delivered to speakers through a *Lynx L22* sound card (Lynx Studio Technology).

Visual stimuli were presented on a centrally-positioned panel of 96 LEDs that spanned 6 cm high by 17 cm wide (118° horizontal angle × 67° vertical angle). The bottom edge of the panel was approximately 4 cm above the animals' eyes. Note that because the interior of the acoustic box is dark, it was not necessary for the rats to look directly at the LED panel in order to see the stimulus events. The use of a large, stationary visual stimulus ensured that small head or eye movements during the stimulus presentation period did not grossly distort the incoming visual information. The LEDs were driven by output from the same sound card that we used for auditory stimuli; the sound card has two channels (typically used for a left and right speaker), which can be controlled independently. Auditory and visual stimuli were sent to different channels on the sound card so that we could use different auditory/visual stimuli for the independent condition. The timing of auditory and visual events was checked regularly using a photodiode and a microphone connected to an oscilloscope.

Animals were required to stay in the port for 1000 ms. Mean wait times were very close to 1000 ms for all 3 animals (Means and standard errors for all animals: rat 1: 953.8 ± 1.4 ms,

rat 2: 984.0 ± 2.4 ms rat 3: 960.1 ± 2.3 ms). Withdrawal from the center port before the end of the stimulus presentation resulted in a 2000–3000 ms “time out” during which a new trial could not be initiated. These trials were excluded from further analysis. Animals typically displayed equal numbers of early withdrawal trials for auditory, visual and multisensory stimuli. When animals successfully waited for the entire duration, they then had up to 2000 ms to report their decision by going to one of two eccentrically positioned reward ports, each of which was arbitrarily associated with either a high rate (right side) or a low rate (left side). The time delay between leaving the center port and arriving at the reward port varied across animals (Means and standard errors for all animals: rat 1: 774.3 ± 1.6 ms, rat 2: 391.5 ± 0.8 ms rat 3: 479.7 ± 0.9 ms). When the rats correctly went to the port corresponding to the presented rate, they received a drop of water (15–20 μ L) delivered directly into the port through tubing connected to a solenoid. Open times of the solenoid were regularly calibrated to ensure that equal amounts of water were delivered to each side.

The two intervals used to generate the stimuli were slightly different for the three rats. For rat 1, intervals had durations of 30 ms and 70 ms. The resulting trials had fluctuating rates whose averages ranged from 12 Hz to 22 Hz. For rat 2, intervals had durations of 17 and 67 ms. The resulting trials had fluctuating rates whose averages ranged from 12 Hz to 32 Hz. Shorter intervals (and therefore higher rates) were used with the first two animals because we had originally feared that the lower rates used on the human task would be difficult for the rats to learn (lower rates require longer integration times because the information arrives more slowly). However, this proved incorrect: For rat 3, the intervals were very similar to those used in humans: intervals were 50 ms and 100 ms; the resulting trials had fluctuating rates whose averages ranged from 9 Hz to 15 Hz. No major differences in multisensory enhancement were observed in the three animals. The main consequence of using different rates for each animal is that it prevented us from pooling data across animals for certain analyses (Fig. 4d and Fig. 6b). For these analyses, we report results from each animal individually.

Animals were first trained using auditory stimuli alone. Once they had achieved proficiency with the task, we introduced a small proportion of multisensory trials where the auditory and visual events were played synchronously. Shortly thereafter, we introduced some trials that contained the visual stimulus only. Performance on these trials was typically near chance for the first few days and improved rapidly thereafter. Once proficiency was achieved on the visual task, animals were presented with equal numbers of auditory, visual, and multisensory trials interleaved in a single block.

All experimental procedures were in accordance with the NIH Guide for the Care and Use of Laboratory Animals and were approved by the Cold Spring Harbor Animal Care and Use Committee.

Data analysis

We computed psychometric functions using completed trials (where the animal waited the required 1000 ms). For each actual rate, we computed the proportion of trials where subjects reported that the stimulus rate was high (Fig. 2a,b, Fig. 4a–d, Fig. 5b). Standard errors for these proportions were computed using the binomial distribution. We then fit the choice data with a cumulative Gaussian function using *psignifit* version 3 (<http://psignifit.sourceforge.net/>), a Matlab software package designed to implement the maximum-likelihood method of Wichmann and Hill (2001a). The psychophysical threshold and point of subjective equality (PSE) (also known as the bias) were taken as the standard deviation (σ) and mean (μ), respectively, of the best-fitting cumulative Gaussian function (Fetsch et al., 2009). Errors on these parameters were generated via bootstrapping based on 1000 simulations. We express parameters from the fits as the estimate and 95% confidence

interval. Vertical error bars in Figs. 2c, 4e and 5a reflect these confidence intervals. To make comparisons between thresholds or PSEs for different psychometric functions (single sensory versus multisensory, for example), we generated t-statistics from the estimated parameters and their standard errors and computed p-values with the Students' t cumulative distribution function on matlab (tcdf.m).

As in previous studies (Young et al., 1993; Jacobs, 1999; Ernst and Banks, 2002; Alais and Burr, 2004; Hillis et al., 2004), we took the thresholds computed from the single sensory trials as an estimate of each single cue's reliability. Single sensory thresholds were typically quite similar, but there was some individual variation. Next, we used these reliabilities to predict the "optimal" threshold that one would observe if the multisensory estimate were more precise because it had been generated by combining each single sensory estimate weighted in proportion to its reliability:

$$\frac{1}{\sigma_{Multisensory}^2} = \frac{1}{\sigma_{Auditory}^2} + \frac{1}{\sigma_{Visual}^2} \quad (1)$$

Combining single sensory estimates using this rule will generate a multisensory estimate with the lowest possible variance (Jacobs, 1999; Landy and Kojima, 2001). Therefore this prediction provides a benchmark, and makes it possible to connect this approach to previous multisensory (and within-sensory cue combination) behavior. Previous studies have reported that subjects' performance is close to the optimal prediction (Ernst and Banks, 2002; Alais and Burr, 2004; Fetsch et al., 2009). To estimate errors on the predicted thresholds, we propagated the errors associated with the single cue thresholds. Horizontal error bars in Figs. 2c, 4e and 5a reflect 95% confidence intervals computed from these standard errors. We also used the standard error when computing a statistic for the Student's t-test to make comparisons between the predicted and measured multisensory thresholds.

For most analyses, psychometric functions were computed separately for each subject. For the analysis shown in Fig. 4b, 5b, however, we pooled data across human subjects. This was because some bonus trials and neutral trials were sufficiently rare that psychometric functions computed for individual subjects were very noisy. The rat data for this analysis (Fig. 4d) could not be pooled across animals, however, because different rates were used for the two animals. We therefore show data for a single rat and report the values for the other animals in the text. Note that the analysis of "neutral trials" in Fig. 4b (orange trace) included those trials where the evidence from one modality was close to the experimenter-defined category boundary between high and low rates (10.5 Hz). The PSEs for the subjects in those figures were very close to this value. However, for the group of subjects in Fig. 5b, PSEs were slightly lower (9.5). For this reason, we defined "neutral trials" as those that were close to the subjects' PSE rather than the true category boundary. Otherwise, we might have erroneously characterized an (for example) 11 Hz stimulus as "neutral" when, for this group of subjects, an 11 Hz trial was taken as evidence for a high-rate choice. Similar effects on thresholds were obtained whether we used 9.5 or 10.5 Hz as the dividing line, but using the latter value resulted in an apparent high frequency bias on neutral trials relative to matched trials that likely does not reflect a true change in subjects' PSE for those trials.

To identify over what time period the subjects accumulated evidence, we performed an analysis similar to the choice-triggered average (Sugrue et al., 2004; Kiani et al., 2008). We performed this analysis using 300 ms time windows but saw similar results with larger and smaller windows. We selected trials that had an average event rate of 10 Hz (7 events in the 700 ms outside the 300 ms window of interest, Fig. 6a). We then computed the average rate of trials within the window of interest for trials that ended in high versus low choices.

Because the stimulus rates outside the window of interest were identical, this analysis allowed us to ask whether the stimuli inside the window of interest influenced the outcome of the trial. Because this analysis selects trials based on their average rate during a specified time period, it necessarily eliminates a large fraction of the data. An average event rate of 10 Hz was observed on approximately 1800 out of 7600 trials, for example. However, this analysis did not depend on a particular event rate: similar results were obtained when we used higher or lower rates. The values we chose for the final analysis were selected because they maximized the total number of included trials. To estimate the significance of this result, we performed a bootstrapping analysis: at each time point, we computed the difference in rate between trials randomly assigned to the “high” or “low” group. This procedure was repeated 1000 times. We then computed the standard deviation of the resulting distribution and took that value as the standard error on the difference between frequencies in the “high” and “low” groups. Further, to provide a comparison for the magnitude of the difference between frequencies in the “high” and “low” groups, we assigned trials randomly to these groups rather than by what the subject actually chose. We then computed the excess rate and its standard error in the same manner as described above (Fig. 6c, dashed lines). This quantity serves as a baseline to which the choice-triggered average can be compared. Because stimulus events occurred only at specified times during the trial, responses over many trials tend to have obvious fluctuations (Fig. 6b and c). These are more evident for lower rate stimuli than for higher rate stimuli, which is why they are more apparent in Fig. 6b compared to Fig. 6c.

Results

We examined whether subjects could combine information about the event rate of a stimulus when the information was presented in two modalities. Combining evidence should produce lower multisensory thresholds relative to single sensory thresholds. We first describe results from the synchronous condition where the same stream of events was presented to the auditory and visual systems, and the events occurred simultaneously (Fig. 1b). Figure 2a shows results for the synchronous condition for a representative human subject: the subject's threshold was lower for multisensory trials compared to single sensory trials (Fig. 2a, blue line is steeper than green, black lines), demonstrating that subjects made more correct choices on multisensory trials. The difference between the single sensory and multisensory thresholds was highly significant (auditory: $p=0.0001$, visual: $p<0.0003$); the change in threshold was not significantly different from the optimal prediction (measured: 1.75, CI: [1.562–1.885]; predicted: 1.69, CI: [1.54–1.84], $p=0.38$). In contrast, the PSE for multisensory trials was similar to those seen on single sensory trials (auditory: $p=0.06$, visual: $p=0.52$).

The example was typical: all subjects we tested showed an improvement on the multisensory condition, and this improvement was frequently close to optimal (Fig. 2c, circles are close to the $x=y$ line indicating optimal performance). On average, this improvement was not accompanied by a change in PSE (mean PSE difference between visual and multisensory: 0.54, CI: [0.48–1.56], $p=0.23$, mean PSE difference between auditory and multisensory: 0.05, CI: [-0.66–0.75], $p=0.87$).

Significant multisensory enhancement was also observed in all three rats. Fig. 2b shows results for a single rat. The difference in thresholds between the single sensory and the multisensory trials was highly significant (auditory: $p<10^{-5}$, visual: $p<10^{-5}$). The improvement exceeded that predicted by optimal cue combination (measured: 1.97, CI: [1.72–2.21]; predicted: 2.70, CI: [2.57–2.83], $p<10^{-5}$). This improvement was also seen in the remainder of our rat cohort (Fig. 2c, squares); the improvement was significantly greater

than the optimal prediction in one of the two additional animals (*rat 1*, $p < 10^{-5}$); the improvement nearly reached significance in a third (*rat 2*, $p = 0.06$).

Multisensory enhancement on our task could be driven by decisions based upon estimates of event rate of the stimulus, or by estimates of the number of events. Human subjects have previously been shown to be adept at estimating counts of sequentially presented events, even when they occur rapidly (Cordes et al., 2007). In principle, either strategy could give rise to uncertain estimates in the single sensory modalities that could be improved with multisensory stimuli. To distinguish these possibilities, we included a small number of random “catch” trials that were either longer or shorter than the standard trial duration. Consider an example trial that has 11 events (Fig. 3a, arrow). If subjects use a counting strategy, they would make the same proportion of “high” choices whether the 11 events are played over 800 ms, 1000 ms or 1200 ms. We did not observe these results in our data. Rather, subjects made many fewer “high” choices when the same number of events were played over a longer (blue trace) compared to a shorter (red trace) duration.

These findings argue that subjects normalize the total number of events by the duration of the trial. In fact, the example subject in Fig. 3a–b normalized quite accurately: he made the same proportion of “high” choices for a given rate (Fig 3c, arrow) whether that rate consisted of a small number of events played over a short duration; (red trace), or a larger number of events played over a longer durations (green, blue trace). The tendency to make choices based on stimulus rate rather than stimulus count was evident when examined in a single subject (top) and in the population of four subjects (bottom). Note that our subjects exhibited such a rate strategy despite the fact that we rewarded them based on the absolute event count, not the event rate (see Methods).

The multisensory enhancement observed for our task (Fig. 2) might simply have resulted because the auditory and visual stimuli were presented amidst background noise, and therefore were difficult to detect. Thus, in the synchronous condition, multisensory information may have enhanced subjects’ performance by increasing the effective signal-to-noise ratio for each event by providing both auditory and visual events at the same time. To evaluate this possibility, we tested subjects on a condition designed to prevent multisensory information from facilitating event detection: we achieved this by generating auditory and visual event streams independently. We term this the “independent condition”. On each trial, we used the same ratio of short to long events for the auditory and visual stimuli. First, we restrict our analysis to the case where the resulting rates were identical, or nearly identical (Methods). Importantly, the auditory and visual events did not occur at the same time, and could have different sequences of long and short intervals (Fig. 1c). Because the events frequently did not occur simultaneously, subjects had little opportunity to use the multisensory stimulus to help them detect individual events. In spite of this, subjects’ performance still improved on the multisensory condition compared to the single sensory condition. This is evident in the performance of a single human and rat subject (Fig 4a,c). For both the human and the rat, thresholds were significantly lower on multisensory trials compared to visual or auditory trials (Human: auditory: $p = 0.0002$, visual: $p < 10^{-5}$; Rat: auditory: $p < 10^{-5}$, visual: $p < 10^{-5}$). The change in threshold was close to the optimal prediction for the human (measured threshold: 2.08, CI:[1.70–2.46], optimal: 2.39, CI:[2.60–2.17], $p = 0.09$) and was lower than the optimal prediction for the rat (measured threshold: 2.64, CI:[2.47–2.81], optimal: 3.26, CI:[3.10–3.38], $p < 10^{-5}$). This example was typical: all subjects we tested showed an improvement on the multisensory condition, and multisensory thresholds were frequently slightly lower than the optimal prediction (Fig. 4e, many circles are below the $x = y$ line); the improvement was significantly greater than the optimal prediction for one additional rat (*rat 1*, $p < 10^{-5}$). On average, this improvement was not accompanied by a change in PSE, for neither the humans (Mean PSE difference between

visual and multisensory condition: 0.45, CI:[-1.93-1.04], $p=0.47$, Mean PSE difference between auditory and multisensory conditions: 0.62, CI:[-0.4-1.65], $p=0.18$) nor for the rats (Mean PSE difference between visual and multisensory condition: 0.48, CI:[-0.33-1.29], $p=0.13$, Mean PSE difference between auditory and multisensory conditions: 0.55, CI:[-2.04-3.14], $p=0.46$). To ensure that subjects' improvement on the independent condition was not driven by changes at the highest and lowest rates (where stimuli might be perceived as synchronous), we repeated this analysis excluding trials at those rates. The multisensory improvement was still evident for the example human and was again very close to the optimal prediction (measured threshold: 2.45, CI:[1.85-3.05], optimal: 2.79, CI:[2.40-3.18], $p=0.82$). The multisensory improvement for the example rat was also present, and was still better than the optimal prediction (measured threshold: 2.75, CI:[2.41-3.09], optimal: 3.53, CI:[3.19-3.88], $p=0.0008$). For the collection of human subjects, we found that thresholds were lower for multisensory trials compared to visual, $p=0.039$, or auditory trials, $p=0.003$). For the rats, thresholds were lower for multisensory trials compared to visual, $p=0.02$, or auditory trials, $p=0.04$. Excluding trials with the highest/lowest rates did not cause significant changes in the average ratio of multisensory to single sensory thresholds for either modality or either species ($p<0.05$).

Subjects might have shown a multisensory improvement on the independent condition for two trivial reasons: first, the presence of two modalities together might have been more engaging and therefore recruited additional attentional resources compared to single sensory trials. Second, events in the independent condition might sometimes occur close enough in time to aid in event detection. We performed two additional analyses that rule out both of these possibilities.

First, we examined subsets of trials from the independent condition where the rates were different for the auditory and visual trials (bonus trials, see Methods). For these trials, evidence from one modality (say, vision) provided stronger evidence about the correct decisions than the other modality (say, audition). For example, the auditory stimulus might be 10 Hz, a rate that is quite close to threshold, but still in favor of a low rate choice, while the visual stimulus is 9 Hz (Fig. 4c, arrow, cyan line). We compared such trials possessing different auditory and visual rates with "matched-evidence" trials where both stimuli had the same rate (Fig. 4c, arrow, blue line). If subjects exhibit improved event detection on the multisensory condition because of near-simultaneous events, they should perform worse on the bonus-evidence trials (at least for low rates), because the likelihood of events occurring at the same time is lower when there are fewer events (10 and 10 events for matched trials, 10 and 9 events for bonus trials). To the contrary, we found that performance improved on bonus evidence trials: in the example, the subject made fewer high choices on the bonus trials at low rates (Fig. 4b, arrow, cyan trace below blue trace), demonstrating improved performance. Accuracy was improved at the higher rates as well, leading to a significantly lower threshold for bonus trials (matched: threshold = 2.3, CI:[2.00-2.57]; bonus: threshold = 1.3, CI:[1.12-1.38], $p < 10^{-5}$). The enhanced performance seen in this subject was typical: 5 of 6 subjects showed lower thresholds for bonus evidence trials, and this reached statistical significance ($p<0.05$) in three individual subjects. Data from an example rat subject supported the same conclusion (Fig. 4d): performance on bonus trials was better than performance on matched trials (matched: threshold = 2.6, CI:[2.42-2.78], bonus: threshold = 1.4, CI:[0.71-2.09], $p = 0.02$). Bonus trials were collected from one of the remaining two rats; for this rat also, performance on bonus trials was better than performance on matched trials (rat 2; matched: threshold = 4.4 ± 0.22 ; bonus: threshold = 2.6 ± 0.70 , $p=0.008$).

Next, we examined subsets of neutral trials for which the rate of one modality (say, vision) was so close to the PSE that it didn't provide compelling evidence for one choice or the other. If multisensory trials are simply more engaging and help subjects pay attention,

performance should be the same on matched trials and neutral trials. To the contrary, we found that performance was worse for neutral trials compared to matched trials: the example subject made many more errors on neutral trials and had elevated thresholds (matched: threshold = 2.3, CI:[2.00–2.57]; neutral trials: threshold = 4.2, CI:[3.56–4.57], $p < 10^{-5}$). The decreased performance seen on neutral trials was typical: all subjects showed higher thresholds for neutral trials, and this was statistically significant ($p < 0.05$) in 5 subjects. Data from an example rat subject support the same conclusion (Fig. 4d): performance on the neutral trials was worse than performance on matched trials (threshold = 2.6, CI:[2.42–2.78]; neutral: threshold = 5.0, CI:[3.80–6.20], $p < 10^{-5}$). Neutral trials were collected from one of the remaining two rats; for this rat also, performance on neutral trials was worse than performance on matched trials 4.4, CI:[4.03–4.76]; neutral: threshold = 6.6, CI:[5.09–8.11], $p = 0.01$).

Note that our dataset included a small proportion of trials where the auditory and visual cues were in support of different decisions (conflict trials). Because we made every effort to match the reliability of the two single sensory cues, we did not expect conflict trials to be systematically biased towards the rate of one modality over the other. Indeed, subjects' PSEs were very similar for a variety of cue conflict levels and did not vary systematically with the level of conflict (weighted linear regression, slope = 0.06, CI:[–0.07–0.19]).

Because we typically collected data from only the independent condition or only the synchronous condition on a given day, we could not rule out the possibility that subjects developed different strategies for the two conditions. If true, then their performance should decline when the trials from the two conditions were mixed within a session. To test this, we collected data from 4 additional subjects on a version of the task where the synchronous and independent conditions were presented in alternating blocks of 160 trials over the course of each day. Because the condition switched so frequently within a given experimental session, subjects would have a difficult time adjusting their strategy. Therefore, a comparable improvement on the two tasks can be taken as evidence that subjects can use a similar strategy for both conditions. Indeed, we found that subjects showed a clear multisensory enhancement on both conditions, even when they were presented in alternating blocks (Fig. 5a, both triangles and circles close to the $x=y$ line). Further, this group of subjects showed the same enhancement on bonus trials as did subjects who were tested in the more traditional configuration (Fig. 5b; matched: threshold = 3.7, CI:[2.93–4.45]; bonus: threshold = 1.9, CI:[1.67–2.04], $p < 10^{-5}$). Individual subjects all showed reduced thresholds for the bonus condition; this difference reached significance for one individual subject. This group of subjects showed increased thresholds for the neutral trials, but this difference did not reach statistical significance (Fig. 5b; matched-evidence: threshold = 3.7, CI:[2.93–4.45]; neutral: threshold = 6.4, CI:[4.91–7.89], $p = 0.0008$). Three out of four individual subjects showed increased thresholds for neutral trials relative to matched trials.

Our stimulus was deliberately constructed so that the stimulus rate fluctuated over the course of the trial. We exploited these moment-to-moment fluctuations in rate in order to determine which time points in each trial influenced the subjects' final decisions. To explain this analysis, consider a group of trials that are selected because they have the same average rate during the first 800 ms of the trial (Fig. 6a, *top*). By examining the average stimulus rate in the last 200 ms of the trial, and comparing it for trials that end in "high" versus "low" choices, we can determine whether rates during that late interval influenced the subjects' final choice. In the schematic example, trials preceding "low" choices (Fig. 6a, red trace) have a lower rate during the final 300 ms compared to trials preceding "high" choices (Fig. 6a, blue trace). We denote differences in rate within such windows as the "*excess rate*" supporting one choice versus the other. The same process can be repeated for other windows within the trial (e.g., Fig. 6a, *bottom*). Systematically varying the temporal window makes it

possible to generate a time-varying weighting function, termed the *choice-triggered average*, that describes the degree to which each moment in the trial influenced the final outcome of the decision and can be informative about the animal's underlying strategy. For example, if the choice-triggered average is only elevated very late in the trial, this suggests that the subjects either do not pay attention early in the trial, or do not retain the information (a leaky integrator) and simply base their decision on what happened at the end of the trial. By contrast, if the choice-triggered average is elevated for the entire duration of the trial, this suggests that subjects exploit information presented at every moment. The choice-triggered average that we computed was elevated over the course of the entire trial for human subjects (Fig. 6b), and over about 600 ms for rodents (Fig. 6c). This suggests that subjects integrate evidence over time for the majority of the trial. The integration time appears to be longer for humans compared to rats: the choice-triggered average for rats was not elevated during the first 200 or last 100 ms of the trials. On average, humans' choice-triggered averages were significantly larger for multisensory trials compared than auditory-only or visual only trials (Multisensory: 1.63 ± 0.05 , Visual: 1.16 ± 0.04 , Auditory: 1.19 ± 0.03 , Multisensory > Auditory, $p < 10^{-5}$; Multisensory > Visual, $p < 10^{-5}$). The multisensory choice-triggered average was also elevated in the example rat (Multisensory: 1.99 ± 0.08 , Visual: 1.05 ± 0.04 , Auditory: 1.29 ± 0.04 , Multisensory > Auditory, $p < 10^{-5}$; Multisensory > Visual, $p < 10^{-5}$). Similar results were observed in the other two rats: the choice-triggered averages exhibited similar shapes and the magnitude of the multisensory trials was larger than that of the single sensory trials (*rat 2*: Multisensory > Auditory, $p < 10^{-5}$; Multisensory > Visual, $p < 10^{-5}$; *rat 3*: Multisensory > Auditory, $p < 10^{-5}$; Multisensory > Visual, $p < 10^{-5}$).

Discussion

We report three main findings. First, we found that subjects can combine multisensory information for decisions about an abstract quantity, event rate, which arrives sequentially over time. Second, we found that subjects showed multisensory enhancement both when the sensory inputs presented to each modality were redundant and also when they were generated independently. This latter finding suggests a model where event rates are estimated independently for each modality and then are fused into a single estimate at a later stage. Finally, because a similar, near-optimal multisensory enhancement was observed in humans and rats, we conclude that the neural mechanisms underlying this multisensory enhancement are very general and are not restricted to a particular species.

Our results differ from previous observations about multisensory integration in a number of ways. First, our stimulus is unique in that the relevant information, event rate, is not available all at once but must be accumulated over time. Most prior studies of multisensory integration have not explicitly varied the incoming evidence with respect to time (Ernst et al., 2000; Ernst and Banks, 2002; Alais and Burr, 2004). Some previous studies have presented time varying rates in a multisensory context (Recanzone, 2003), but have not, to our knowledge, asked whether subjects can exploit the multisensory information to improve performance. Second, we have shown multisensory enhancement in both humans and in rodents. Most of the multisensory integration studies in animals have been carried out using non-human primates (Avillac et al., 2007; Gu et al., 2008). There have been some studies of multisensory integration in rodents (Sakata et al., 2004; Hirokawa et al., 2008), but our study goes further than these in several ways: First, previous studies did not establish animals' thresholds or PSEs. Our approach allowed us to determine the animals' psychophysical thresholds and PSEs and therefore compare changes on multisensory trials to a maximum likelihood prediction. Further, previous studies in rodents also did not compare human data alongside the animals, making it difficult to know whether the two species use similar strategies when combining multisensory information.

Our results are perhaps consistent with a different mechanism for multisensory integration than has been thus far observed physiologically. Recordings from the superior colliculus in anaesthetized animals have made it clear that temporal synchrony (or near-synchrony) of individual stimulus events is a requirement for multisensory integration (Meredith et al., 1987). Because we observed multisensory integration in the absence of temporal synchrony, the circuitry in the superior colliculus probably does not underlie the improvement we observed. Instead, our observations point to mechanisms that estimate more abstract quantities, such as the average rate over a long time interval.

Critical features of the task we used may have invited modality-independent combination of evidence. First of all, auditory and visual events probably arrived in the brain with different latencies even on the synchronous condition (Pfungst and O'Connor, 1981; Maunsell and Gibson, 1992; Recanzone et al., 2000). More importantly, the stimulus rates were sufficiently high that connecting individual auditory and visual events was likely not feasible for most subjects. Previous research has shown that when auditory and visual events are embedded in periodic pulse trains, the detection of temporal synchrony falls to chance levels at only 4 Hz (Fujisaki and Nishida, 2005). A separate study found that discrimination thresholds for time interval judgments are about 5 times longer for multisensory stimuli than for auditory or visual stimuli alone (Burr et al., 2009). Taken together, these studies and ours suggest that the brain faces a major challenge when trying to associate specific auditory or visual events that are arriving quickly. A reasonable solution to this problem is to generate separate estimates for each modality and then combine them at a later stage, perhaps in an area outside the superior colliculus that receives both auditory and visual inputs, such as parietal cortex (Reep et al., 1994; Mazzoni et al., 1996). This type of strategy may not be necessary for stimuli that lack discrete events. It remains to be seen whether other continuously varying stimuli would likewise be integrated over time at an early stage and then combined later on (for example Fetsch et al., 2009).

Several of our conclusions warrant caution. First, we have argued that our subjects use data presented over the entire duration of the trial. Because we presented stimuli for a fixed duration (1000 ms), however, we make this conclusion with caution. The choice-triggered average we report suggests that subjects use information over long periods of time in the trial, but a reaction time paradigm is necessary to make this conclusion with complete confidence.

Secondly, although we conclude that rats and humans are both quite capable of multisensory integration, there are small differences in the behavior of the two species. Specifically, humans' decisions were influenced by stimuli at all times during the trial (Fig. 6b), whereas rats' decisions were influenced mainly by the middle 650 ms (Fig. 6c). The weak influence of stimuli at the very beginning of the trial is consistent with a "leaky integrator" that accumulates evidence but leaks it away according to a time constant that is shorter than the trial. An alternative explanation is that the rats had little time to prepare for the onset of the stimulus. The stimulus began as soon as the rats initiated a nose poke into the center port. The human subjects, by contrast, began each trial with a brief fixation period before the stimulus began, providing them with some preparation time. Accordingly, humans' decisions were clearly influenced by stimuli very early on in the trial. Several explanations are consistent with the weak influence of stimuli near the end of the trial. One possibility is that the rats accumulated evidence up to a threshold level or bound that was frequently reached before the end of the trial. If this were the case, stimuli arriving after the bound was reached would not influence the animals' decisions, leading to the pattern of results that we observed. A signature of this has been previously reported in monkeys (Kiani et al., 2008). A second possibility is that stimuli late in the trial don't influence the rats' choices because the rats use the last 200 ms of the trial to prepare the full-body movement that is required to

report their decisions. Human observers, by contrast, made much smaller movements to report their decisions, so might not have needed the additional movement preparation time. Further, human subjects almost never responded before the stimulus was over. Rodents' more frequent early responses are consistent with the possibility that they used part of the stimulus presentation time to plan a movement. Note that other aspects of the choice triggered average were quite similar for the two species. For example, we observed a reliable difference in the magnitude of the choice-triggered average for single sensory and multisensory trials. This suggests that multisensory stimuli exert more influence over the choice compared to single sensory stimuli, a conclusion that is in agreement with the overall improvements we observed on multisensory trials (e.g., Figs. 2, 4 and 5).

A final caveat is the observation that our subjects, particularly the rats, frequently showed multisensory enhancements that were larger than one would expect based on maximum likelihood combination. Individual sessions with super-optimal enhancements have been observed previously in animals (Fetsch et al., 2009; Fetsch et al., 2011), so our observations are not without precedent. Nevertheless, the tendency toward supra-optimality is more prevalent and consistent in our dataset than in previous ones. The most likely explanation is that performance on single-sensory trials provides an imperfect estimate of a modality's reliability. The apparent supraoptimal cue combination likely indicates that, at least for a few subjects, we underestimated the reliability of the single-sensory stimulus. A possible explanation is that the animals had different levels of motivation on single sensory versus multisensory trials. One reason for this might be as follows: the rats are, in general, very sensitive to overall reward rate. For example, we have observed improved overall performance on a given modality when we decrease the proportion of easy trials for that modality. This suggests that the animals may strive for a particular reward rate, and adjust their motivation levels when they exceed or fall short of that rate. Because single-sensory trials yield lower average reward rates compared to multisensory trials, animals might have decreased motivation on those trials, particularly when they are interleaved with higher reward-rate multisensory trials. Although this explanation is a speculative one, we favor it over other possibilities, such as the possibility that there is an additive noise source: high-level decision noise, for example. Additive noise would indeed cause underestimates of the subjects' reliability on the single sensory condition; however, it has been previously demonstrated to have only a very small effect on the relationship between the measured and predicted behavior on the multisensory condition (Knill and Saunders, 2003; Hillis et al., 2004). Indeed, if high-level decision noise were present and constant across single sensory and multisensory trials (Hillis et al. 2004), it would result in an over-estimate of multisensory improvement, whereas we observed multisensory improvements greater in magnitude than predicted by the maximum-likelihood model.

To conclude, our behavioral evidence argues for the existence of neural circuits that make it possible to flexibly fuse information across time and sensory modalities. Our observations suggest that this kind of multisensory integration may use different circuitry compared to the synchrony-dependent mechanisms that have been previously reported. The existence of many mechanisms for multisensory integration likely reflects the fact that multisensory stimuli in the world probably activate neural circuits on a variety of timescales. As a result, many different mechanisms, each of them suited to the particular constraints of a class of stimuli, may operate in parallel in the brain.

Acknowledgments

We thank Haley Zamer, Barry Burbach, Santiago Jaramillo, Petr Znamenskiy and Amanda Brown for technical assistance. We would also like to thank Chris Fetsch for giving feedback on an early version of the manuscript. This work was supported by NIH EY019072 and a grant from the John Merck Fund.

References

- Alais D, Burr D. The ventriloquist effect results from near-optimal bimodal integration. *Curr Biol*. 2004; 14:257–262. [PubMed: 14761661]
- Alais D, Newell FN, Mamassian P. Multisensory processing in review: from physiology to behaviour. *Seeing Perceiving*. 2010; 23:3–38. [PubMed: 20507725]
- Avillac M, Ben Hamed S, Duhamel JR. Multisensory integration in the ventral intraparietal area of the macaque monkey. *J Neurosci*. 2007; 27:1922–1932. [PubMed: 17314288]
- Burr D, Silva O, Cicchini GM, Banks MS, Morrone MC. Temporal mechanisms of multimodal binding. *Proceedings*. 2009; 276:1761–1769.
- Cordes S, Gallistel CR, Gelma R, Latham PE. Nonverbal arithmetic in humans: Light from noise. *Perception and psychophysics*. 2007; 69:1185–1203. [PubMed: 18038956]
- Ernst MO, Banks MS. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*. 2002; 415:429–433. [PubMed: 11807554]
- Ernst MO, Banks MS, Bulthoff HH. Touch can change visual slant perception. *Nature neuroscience*. 2000; 3:69–73.
- Fetsch CR, Turner AH, DeAngelis GC, Angelaki DE. Dynamic reweighting of visual and vestibular cues during self-motion perception. *J Neurosci*. 2009; 29:15601–15612. [PubMed: 20007484]
- Fetsch CR, Pouget A, Deangelis GC, Angelaki DE. Neural correlates of reliability-based cue weighting during multisensory integration. *Nature neuroscience*. 2011
- Fine I, Jacobs RA. Modeling the combination of motion, stereo, and vergence angle cues to visual depth. *Neural computation*. 1999; 11:1297–1330. [PubMed: 10423497]
- Fujisaki W, Nishida S. Temporal frequency characteristics of synchrony-asynchrony discrimination of audio-visual signals. *Experimental brain research. Experimentelle Hirnforschung*. 2005; 166:455–464. [PubMed: 16032402]
- Gold JI, Shadlen MN. The neural basis of decision making. *Annual review of neuroscience*. 2007; 30:535–574.
- Gu Y, Angelaki DE, Deangelis GC. Neural correlates of multisensory cue integration in macaque MSTd. *Nature neuroscience*. 2008; 11:1201–1210.
- Hillis JM, Watt SJ, Landy MS, Banks MS. Slant from texture and disparity cues: optimal cue combination. *Journal of vision*. 2004; 4:967–992. [PubMed: 15669906]
- Hirokawa J, Bosch M, Sakata S, Sakurai Y, Yamamori T. Functional role of the secondary visual cortex in multisensory facilitation in rats. *Neuroscience*. 2008; 153:1402–1417. [PubMed: 18440715]
- Jacobs RA. Optimal integration of texture and motion cues to depth. *Vision research*. 1999; 39:3621–3629. [PubMed: 10746132]
- Kiani R, Hanks TD, Shadlen MN. Bounded integration in parietal cortex underlies decisions even when viewing duration is dictated by the environment. *J Neurosci*. 2008; 28:3017–3029. [PubMed: 18354005]
- Knill DC, Saunders JA. Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vision research*. 2003; 43:2539–2558. [PubMed: 13129541]
- Kording KP, Beierholm U, Ma WJ, Quartz S, Tenenbaum JB, Shams L. Causal inference in multisensory perception. *PLoS ONE*. 2007; 2:e943. [PubMed: 17895984]
- Landy MS, Kojima H. Ideal cue combination for localizing texture-defined edges. *J Opt Soc Am A Opt Image Sci Vis*. 2001; 18:2307–2320. [PubMed: 11551065]
- Link SW, Heath RA. A sequential theory of psychological discrimination. *Psychometrika*. 1975; 40:77.
- Lovelace CT, Stein BE, Wallace MT. An irrelevant light enhances auditory detection in humans: a psychophysical analysis of multisensory integration in stimulus detection. *Brain Res Cogn Brain Res*. 2003; 17:447–453. [PubMed: 12880914]
- Maunsell JH, Gibson JR. Visual response latencies in striate cortex of the macaque monkey. *Journal of neurophysiology*. 1992; 68:1332–1344. [PubMed: 1432087]

- Mazurek ME, Roitman JD, Ditterich J, Shadlen MN. A role for neural integrators in perceptual decision making. *Cereb Cortex*. 2003; 13:1257–1269. [PubMed: 14576217]
- Mazzoni P, Bracewell RM, Barash S, Andersen RA. Spatially tuned auditory responses in area LIP of macaques performing delayed memory saccades to acoustic targets. *Journal of neurophysiology*. 1996; 75:1233–1241. [PubMed: 8867131]
- Meredith MA, Nemitz JW, Stein BE. Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. *J Neurosci*. 1987; 7:3215–3229. [PubMed: 3668625]
- Miller LM, D’Esposito M. Perceptual fusion and stimulus coincidence in the cross-modal integration of speech. *J Neurosci*. 2005; 25:5884–5893. [PubMed: 15976077]
- Palmer J, Huk AC, Shadlen MN. The effect of stimulus strength on the speed and accuracy of a perceptual decision. *Journal of vision*. 2005; 5:376–404. [PubMed: 16097871]
- Pelli D. The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision*. 1997; 10:437–442. [PubMed: 9176953]
- Pfingst BE, O’Connor TA. Characteristics of neurons in auditory cortex of monkeys performing a simple auditory task. *Journal of neurophysiology*. 1981; 45:16–34. [PubMed: 7205342]
- Recanzone GH. Auditory influences on visual temporal rate perception. *Journal of neurophysiology*. 2003; 89:1078–1093. [PubMed: 12574482]
- Recanzone GH, Guard DC, Phan ML. Frequency and intensity response properties of single neurons in the auditory cortex of the behaving macaque monkey. *Journal of neurophysiology*. 2000; 83:2315–2331. [PubMed: 10758136]
- Reep RL, Chandler HC, King V, Corwin JV. Rat posterior parietal cortex: topography of corticocortical and thalamic connections. *Experimental brain research. Experimentelle Hirnforschung*. 1994; 100:67–84. [PubMed: 7813654]
- Roitman JD, Shadlen MN. Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *J Neurosci*. 2002; 22:9475–9489. [PubMed: 12417672]
- Sakata S, Yamamori T, Sakurai Y. Behavioral studies of auditory-visual spatial recognition and integration in rats. *Experimental brain research. Experimentelle Hirnforschung*. 2004; 159:409–417. [PubMed: 15249987]
- Slutsky DA, Recanzone GH. Temporal and spatial dependency of the ventriloquism effect. *Neuroreport*. 2001; 12:7–10. [PubMed: 11201094]
- Stanford TR, Quessy S, Stein BE. Evaluating the operations underlying multisensory integration in the cat superior colliculus. *J Neurosci*. 2005; 25:6499–6508. [PubMed: 16014711]
- Sugrue LP, Corrado GS, Newsome WT. Matching behavior and the representation of value in the parietal cortex. *Science (New York, NY)*. 2004; 304:1782–1787.
- Young MJ, Landy MS, Maloney LT. A perturbation analysis of depth perception from combinations of texture and motion cues. *Vision research*. 1993; 33:2685–2696. [PubMed: 8296465]

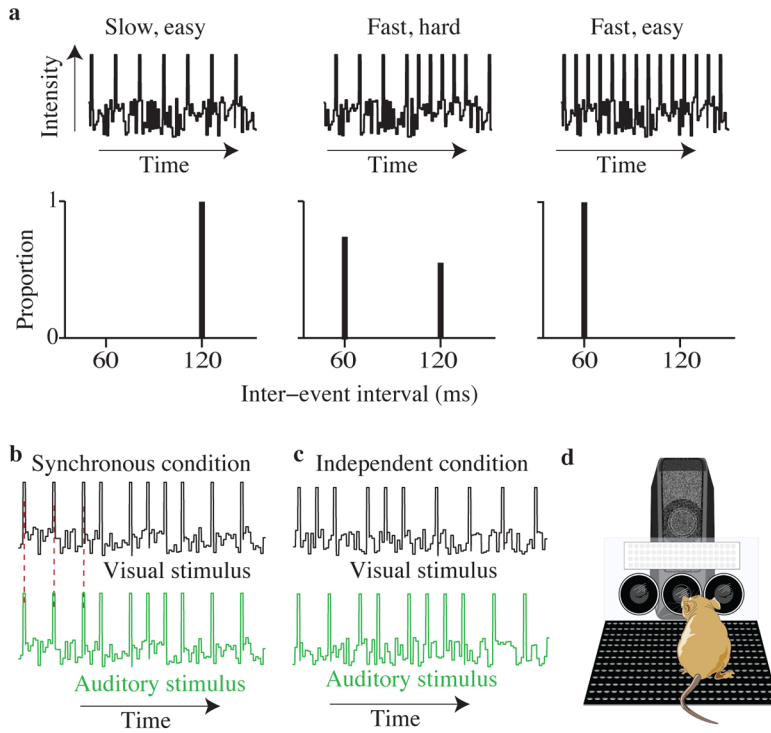


Figure 1. Schematic of rate discrimination decision task and experimental set up
a. Each trial consists of a stream of events (auditory or visual) separated by long or short intervals (top row). Events are presented in the presence of ongoing white noise. For easy trials, all inter-event intervals are either long (left column) or short (right column). More difficult trials are generated by selecting inter-event intervals of both values (middle column). Values of inter-event intervals reflect those used for all human subjects. **b.** Example auditory and visual event streams for the synchronous condition. Peaks represent auditory or visual events. Red dashed lines indicate that auditory and visual events are simultaneous. **c.** Example auditory and visual event streams for the independent condition. **d.** Schematic drawing of rodent in operant conditioning apparatus. Circles are the “ports” where the animal pokes his nose to initiate stimuli or receive liquid rewards. The white rectangle is the panel of LEDs. The speaker is positioned behind the LEDs.

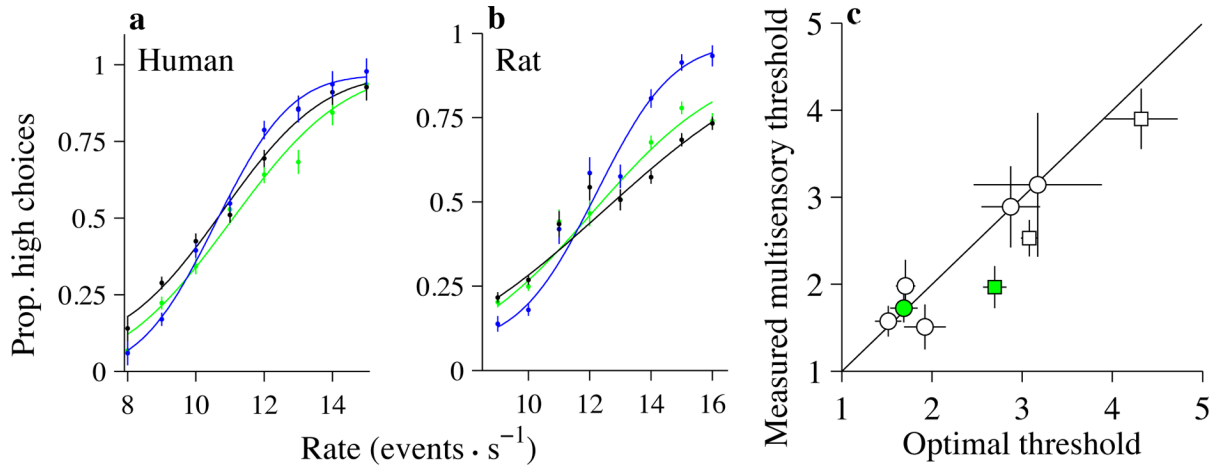


Figure 2. Subjects' performance is better on multisensory trials

a. Performance of a single subject (*a*) plotted as the fraction of responses judged as “high” against the event rate. Green trace: auditory only; black trace: visual only; blue trace: multisensory. Error bars indicate standard error of the mean (binomial distribution). Smooth lines are cumulative Gaussian functions fit via maximum likelihood estimation. N=7680 trials. **b.** Same as a but for one rat (*rat 3*), N=12,459 trials. **c.** Scatter plot comparing the observed thresholds on the multisensory condition compared to those predicted by the single cue conditions for all subjects. Circles: human subjects. Squares: rats. Green symbols are for the example rat and human shown in a,b. Black solid line shows $x=y$; points above the line show a suboptimal improvement. Error bars indicate 95% confidence intervals.

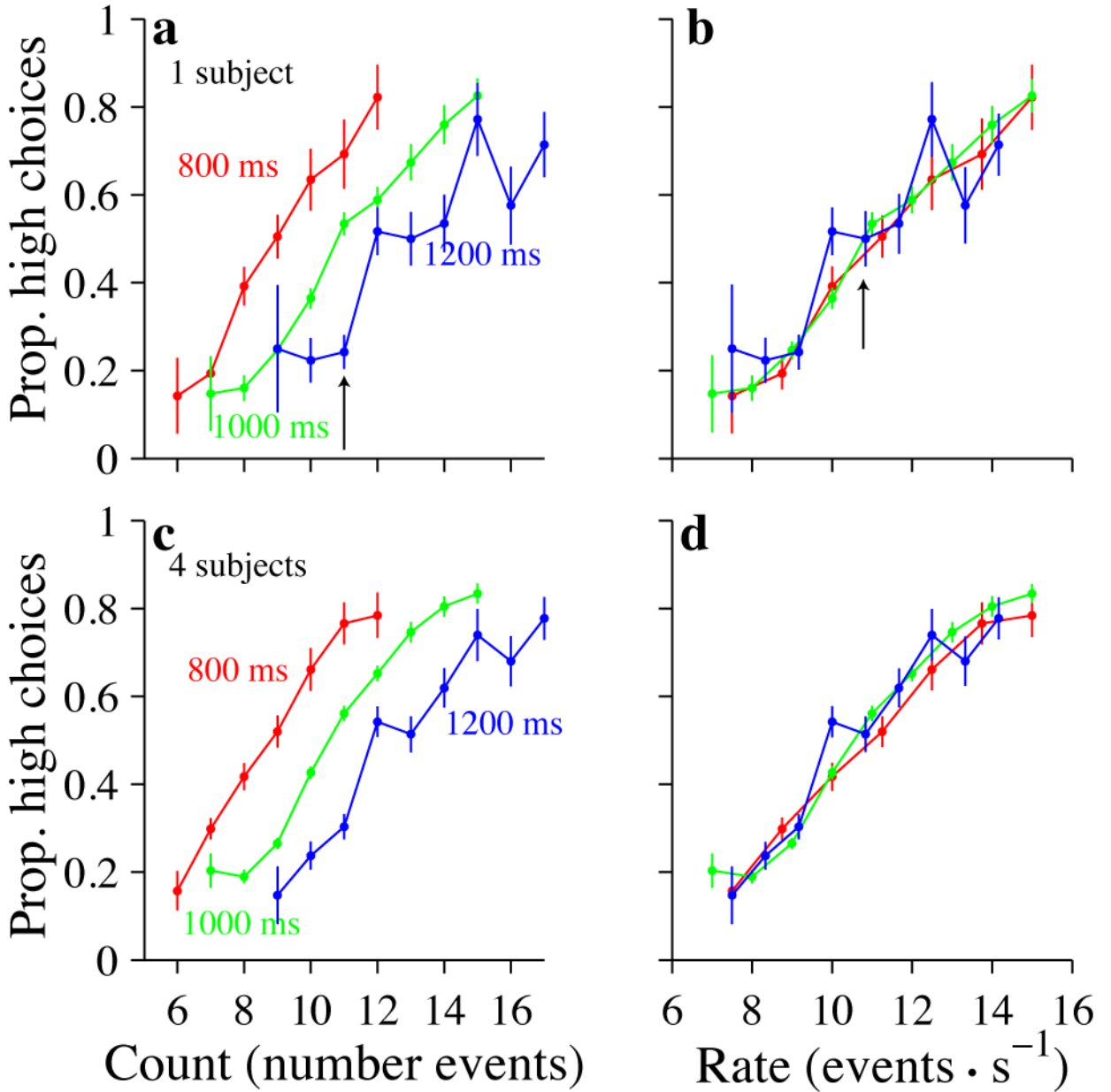


Figure 3. Subjects make decisions according to event rates, not event counts

a. Example subject (*mk*). Abscissa indicates the number of event counts. Each colored line shows the subject's performance for trials where the event count on the abscissa was presented over the duration specified by the labels. For a given event count (eg, 11 events, black arrow) the subject's choices differed depending on trial duration. N=3656 trials **b.** Same data and color conventions as in a except that the abscissa indicates event rate instead of count. For a given event rate (eg, 11 events/second, black arrow), the subject's choices were very similar for all trial durations. **c,d.** Data for 4 subjects; conventions are the same as in a,b. N=9727 trials.

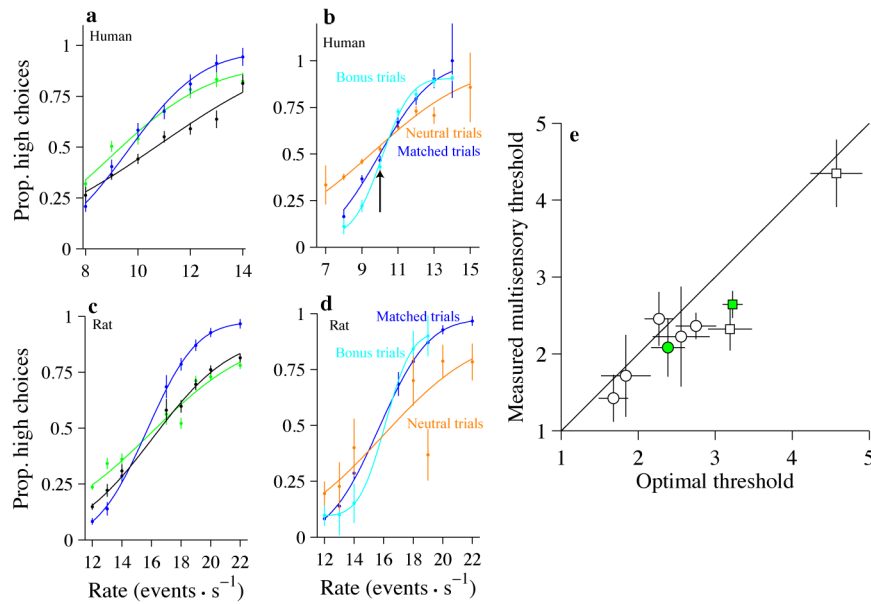


Figure 4. The multisensory enhancement is still present for the independent condition

a. Performance of a single subject (*aj*). Conventions are the same as in Fig. 2a. $N=4255$ trials. **b.** A comparison of accuracy for matched trials (blue trace), bonus trials (cyan trace) and neutral trials (orange trace). Abscissa plots the rate of the auditory stimulus. Data are pooled across 6 humans. $N=1957$ (Matched condition), 2933 (Bonus condition), 3825 (Neutral condition). **c.** Same as a, but for a single rat (*rat 1*). $N=13,116$ trials. **d.** Same as b, but for a single rat. $N=3725$ (Matched condition), 244 (Bonus condition), 247 (Neutral condition) **e.** Scatter plot for all subjects comparing the observed thresholds on the multisensory condition compared to the predicted thresholds. Conventions are the same as in Fig. 2c. Error bars indicate 95% confidence intervals.

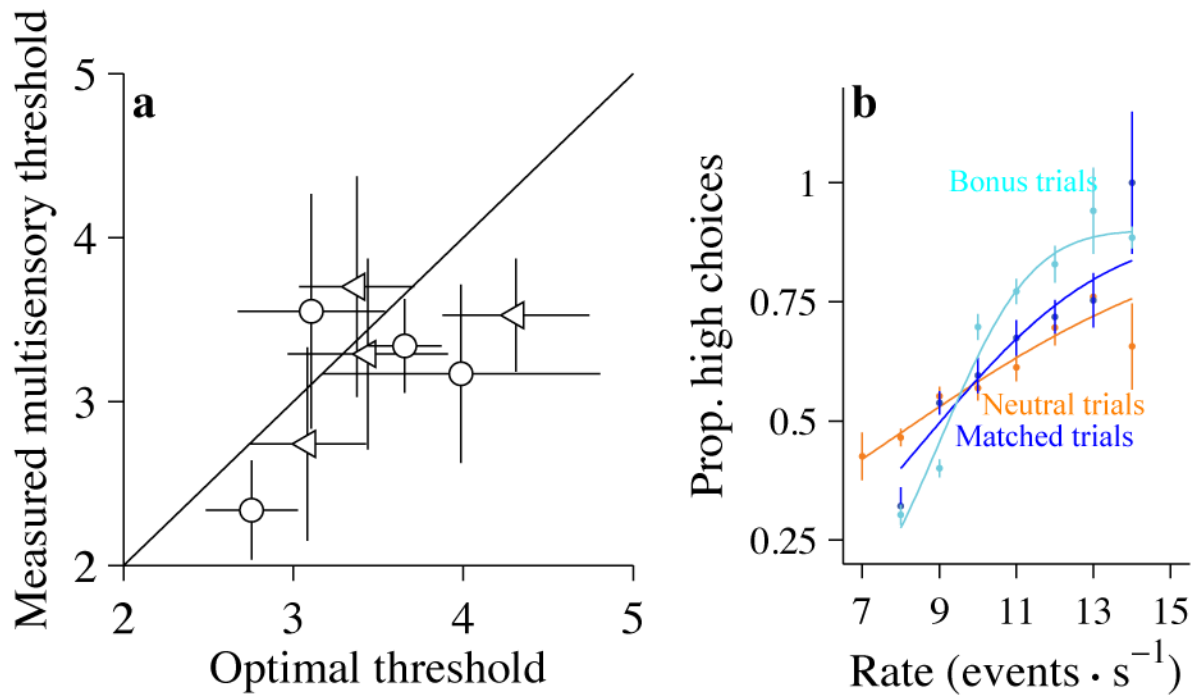


Figure 5. Human subjects' performance is better on multisensory trials even when the synchronous and independent conditions are presented in alternating blocks
a. Scatter plot for all subjects comparing the observed thresholds on the multisensory condition compared to the predicted thresholds. Data from the synchronous (circles) and independent (triangles) conditions are shown together. Error bars indicate 95% confidence intervals. **b.** Subjects perform better on bonus trials compared to the matched trials, and slightly worse on neutral trials. Conventions are the same as Fig. 4c. N=1390 (Matched condition), 2591 (Bonus condition), 2464 (Neutral condition).

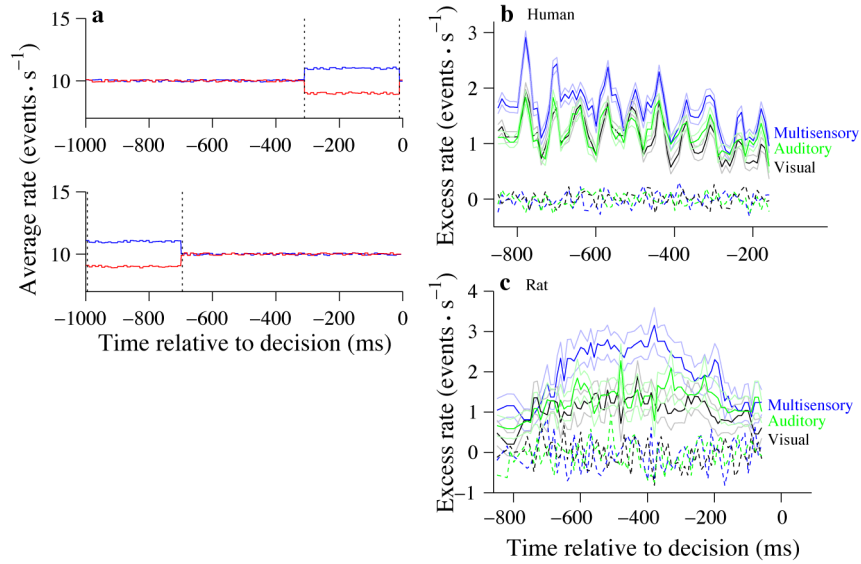


Figure 6. Subjects' decisions reflect evidence accumulated over the course of the trial
a. Schematic of average stimulus frequencies for trials supporting opposing decisions. *Top:* Trials were selected if their average stimulus rate from 0 to 700 ms was 10 Hz (7 events over 700 ms). Trials were then grouped according to whether the subject chose low (red) or high (blue) on each trial. Average stimulus rate within the bin of interest (700 – 1000 ms, dashed lines) was then compared for stimuli preceding left and right choices. *Bottom:* Same as in top panel except that the window of interest occurred early in the trial (0 – 300 ms). **b.** Solid traces: Difference in average event rate for trials that preceded left versus right choices for all time points in a trial. Color conventions are the same as in Fig. 2a,b. Thin lines: Standard error of the mean computed via bootstrapping. Dashed traces: Difference in average rate for trials assigned randomly to two groups. Data were pooled from 6 human subjects. Trial numbers differed slightly for each time point; approximately 1800 trials were included at each point. **c.** Same as b but for an individual rat. Trial numbers differed slightly for each time point; approximately 3200 trials were included at each point.