Journal of Mathematical Psychology 99 (2020) 102449

Contents lists available at ScienceDirect

Journal of Mathematical Psychology

journal homepage: www.elsevier.com/locate/jmp

Revealing multisensory benefit with diffusion modeling

Carolyn A. Murray [a,*], E. Sebastian Lelo de Larrea-Mancera [b], Arit Glicksohn [a], Ladan Shams [a,1], Aaron R. Seitz [b,1]

[a] Department of Psychology, University of California, Los Angeles, 1285 Franz Hall, Los Angeles, CA, 90095, United States of America
[b] Department of Psychology, University of California, Riverside, 900 University Avenue, Riverside, CA, 92521, United States of America

ARTICLE INFO

Article history:
Received 9 December 2019
Received in revised form 18 August 2020
Accepted 24 August 2020
Available online xxxx

Keywords:
Multisensory processing
Drift diffusion model
Perceptual processing
Multisensory benefits

ABSTRACT

Multisensory information can benefit perceptual, memory, and decision-making processes. These benefits commonly manifest in superior detection and discrimination of multisensory stimuli, as well as improved perception and subsequent memory of unisensory representation of an object previously encoded in a multisensory context. However, the vast majority of studies to date analyze accuracy, sensitivity and/or reaction time data independently to compare multisensory and unisensory conditions. Considering the well-established speed-accuracy trade-off, we asked whether some multisensory benefits go unnoticed when measured using traditional methods that do not take both reaction time and accuracy into account simultaneously, and whether an approach combining them can more reliably characterize and quantify the broad extent of multisensory interactions across perception and cognition. While drift diffusion models have been previously shown to be effective in addressing the speed-accuracy trade-off and providing a reliable and accurate measure of multisensory benefits, one impediment of this approach is the requirement of a large number of trials to estimate model parameters and to characterize effects. This may be prohibitive in many experimental paradigms. Several model variants attempt to reduce the required number of trials, either by averaging across participants or limiting the search space for the parameters. Here, we employed a hierarchical drift diffusion model, that utilizes Bayesian priors, allowing parameter estimation with smaller sample sizes while still making subject-specific parameter estimates. We analyzed data in perceptual detection and discrimination tasks across multiple sensory combinations, to investigate if the diffusion model would provide a sensitive and reliable measure of multisensory benefits. Results indicate that across visual, auditory and tactile modality combinations, the diffusion model was either as or more sensitive than traditional accuracy, sensitivity, or reaction time measures, and was the only measure that consistently detected multisensory benefits in a statistically significant fashion. We recommend the use of diffusion modeling approaches when assessing the outcomes of multisensory experiments, especially as they become more computationally efficient.

1. Introduction

The extent to which we are able to integrate information across the senses to reach a behavioral decision and the nature of this integration has been the focus of many studies across the sensory modalities (for some reviews see Driver & Spence, 2000; Rosenblum, Dias, & Dorsi, 2017; Shams & Seitz, 2008). More often than not, the world around us concentrates contingent information about objects that can be picked up by our distinct senses (Shinn-Cunningham, 2008). The multisensory integration that can be achieved by our neural system in this object-oriented context ultimately serves to provide a more accurate picture of what is it that we are experiencing, especially in the face of uncertainty regarding the origins and reliability of information (Gold & Shadlen, 2007; Raposo, Sheppard, Schrater, & Churchland, 2012; Stein, Stanford, Wallace, & Jiang, 2004). Multisensory integration has been defined differently by different authors, without a clear consensus about the behavioral signature of multisensory integration. Regardless of the criterion used for multisensory integration, it is clear that even when senses do not fully integrate (for example, they do not meet the criterion of optimal integration in a given dimension such as accuracy), there is often evidence for a crossmodal interaction, whereby information from multiple modalities benefits performance compared to the unisensory conditions. Multisensory benefits can also take the form of enhanced attention to both senses, enhanced processing of a dimension in one modality due to a congruent dimension in a second modality, and more.

* Corresponding author.
E-mail address: camurray@ucla.edu (C.A. Murray).

1 Equal contribution.

https://doi.org/10.1016/j.jmp.2020.102449
0022-2496/

Typically, inferences regarding the presence or absence of a multisensory benefit are guided by behavioral data based on accuracies and/or reaction times. However, these measures are known to interact. The speed-accuracy tradeoff is one such example, where pressure to complete a task in a short time frame causes participants to respond less accurately (Fitts, 1966; Wickelgren, 1977). Of perhaps greater interest to the area of multisensory research, is the different behavior participants display in different sensory conditions. For example, responses to auditory stimuli are typically faster than those to visual information (Brebner & Welford, 1980; Shelton & Kumar, 2010), and many spatial tasks, such as localization tasks, show higher response accuracy in unisensory visual condition than unisensory auditory condition (Bushara et al., 1999). This makes a comparison between multisensory conditions and the unisensory conditions more difficult to quantify without a single, combined measure that takes these factors into account. Many studies lack such a combined measure, leaving a reliance on separate measures of response time and accuracy, potentially obscuring the interpretation of the results. To help solve this problem, we propose a methodological approach to combine these measures, allowing for an improved ability to uncover these interactions even when traditional approaches fail to do so.

One such model that has made its way into the psychological literature is the *diffusion decision model* (Ratcliff, 1978), sometimes also called the *drift diffusion model*. The model utilizes trial-by-trial accuracy and reaction time data to estimate parameters that capture a dynamic decision-making process, wherein decision-making is a process of evidence accumulation over time. Under this model, the decision-making process is a dynamic and noisy process of evidence accumulation from some starting point towards one of two decision boundaries. When the accumulated evidence reaches a boundary threshold, sufficient evidence has been gathered to make a decision. The parameters of the model capture key aspects of this process, including the distance between the decision thresholds, called the boundary separation (a), the starting point for evidence accumulation (z), and the non-decision time in this process (t; see Fig. 1 for a graphical representation). Of greatest interest for the current study, however, is the *drift rate (v)*, which can be thought of as the average rate of evidence accumulation (Ratcliff & McKoon, 2008). Higher drift rates typically lead to reaching a decision boundary more rapidly, and are considered an indication that participants are better at extracting information from the evidence than if the drift rate was lower. As the effects of multisensory stimulus presentation are often modeled in terms of reducing variance in sensory representations (Gingras, Rowland, & Stein, 2009), which in turn alters the signal-to-noise ratio, this could be a key parameter to uncover multisensory benefits in this model.

Previous studies have shown that parameters of the drift diffusion model measure can track multisensory benefits (Diederich, 2008; Drugowitsch, DeAngelis, Angelaki, & Pouget, 2015; Drugowitsch, DeAngelis, Klier, Angelaki, & Pouget, 2014), but such modeling approaches have not yet become well-employed. We suspect this is, at least in part, due to the large number of trials that standard implementations of these models need to converge on a solution. Standard drift diffusion models commonly need several hundred trials in each experimental condition to fit model parameters reliably (see Drugowitsch et al., 2014; Gomez, Ratcliff, & Perea, 2007; Leite & Ratcliff, 2010; Van Zandt, Colonius, & Proctor, 2000 for examples), which can be prohibitive for their use. Methods do exist that either combine trials across participants to obtain a large enough number of trials (Mahani, Bausenhart, Ahmadabadi, & Ulrich, 2019), or constrain the search space for parameters to allow estimation to converge more quickly (Nidiffer, Diederich, Ramachandran, & Wallace, 2018), however these are not widely employed.

As a method to use fewer trials per condition while still managing to fit individual participant parameters, we utilized the Hierarchical Drift Diffusion Model (HDDM; Wiecki, Sofer, & Frank, 2013). This variant of the drift diffusion model uses Bayesian priors to begin a Markov Chain Monte Carlo (MCMC) process to converge on model estimates in fewer trials. The model also fits parameters per subject, and then creates an estimate of the population distribution from the participant parameters, allowing for estimates at both individual and population levels. This particular variant of the model has previously been used to show that drift rate reflects the strength of audiovisual integration in a detection task (Regenbogen, Johansson, Andersson, Olsson, & Lundström, 2016). The current study aims to expand this investigation to a systematic investigation of multiple perceptual tasks and sensory combinations, and to examine whether the HDDM can reliably and accurately detect multisensory interactions with moderate sample sizes at least as well as traditional accuracy, reaction time, and sensitivity measures. This would support the general use of such a modeling approach to more fully characterize the results of multisensory experiments.

2. General methods

2.1. Participants

A total of 67 participants were recruited across 4 separate studies. All participants were undergraduate students at either the University of California, Los Angeles (UCLA), or at the University of California, Riverside (UCR). All reported having normal or corrected to normal vision and hearing, and no history of neurological issues that would reduce tactile sensitivity. Written informed consent was collected from each participant and experimental procedures were reviewed and approved by the UCLA and UCR Institutional Review Boards. Full details about the participants in each experiment have been included in the relevant experiments.

2.2. Task overview

In each experiment, participants were presented with a combination of pseudo-randomly interleaved unisensory and bisensory stimuli. Across three experiments, audiovisual, visuotactile, and audiotactile bisensory combinations were used. Visual stimuli were squares of dynamic salt-and-pepper noise, and auditory and tactile stimuli were Gaussian white noise, delivered over headphones or a vibro-tactile stimulator, respectively. In each experiment, participants were asked to complete two separate 2AFC tasks (Fig. 2). In what we will call the *detection task*, participants were provided with a stimulus and were asked to determine if the stimulus "pulsed". Pulses were rhythmic changes in contrast between the light and dark pixels in the stimuli in the visual condition and rhythmic changes in amplitude in the auditory and tactile conditions. In what we will call the *discrimination task*, the stimulus in a given trial would always pulse, and participants were asked to determine if the pulse had been a "slow" or "fast" pulse. All trials lasted a maximum of 3000 ms from stimulus onset.

Data was analyzed for each participant in terms of average accuracy and reaction time in each modality. Signal detection theory sensitivity (d'; Macmillan & Creelman, 2005) was also calculated for each participant, to provide a measure of performance that would not be affected by response biases, although, like accuracy measure, would not consider response times simultaneously. Drift rates were also estimated per-participant for each sensory condition. In each of these measures, bisensory performance was compared to the best unisensory performance to investigate if a multisensory benefit could be detected in each case.

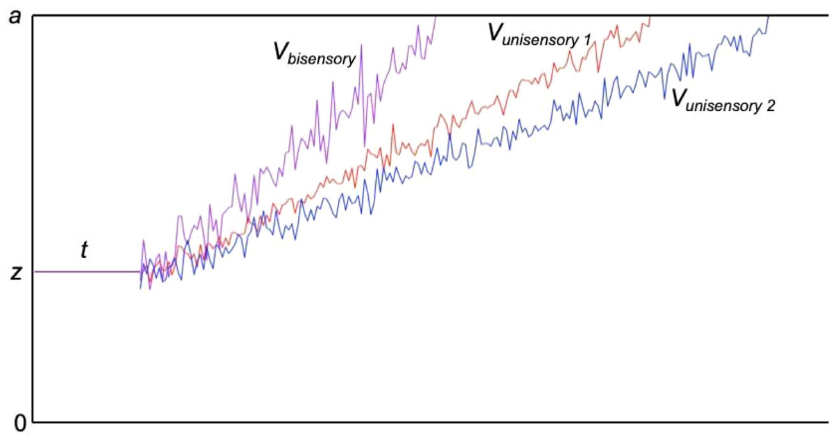

Fig. 1. A schematic of the basic components of a drift diffusion model where the hypothesis that the drift rate in the bisensory condition is faster than those of unisensory conditions is displayed. The four major components included in models used to fit the current experimental data are shown. Drift rate (v) is the average slope of the noisy evidence schematically portrayed here.

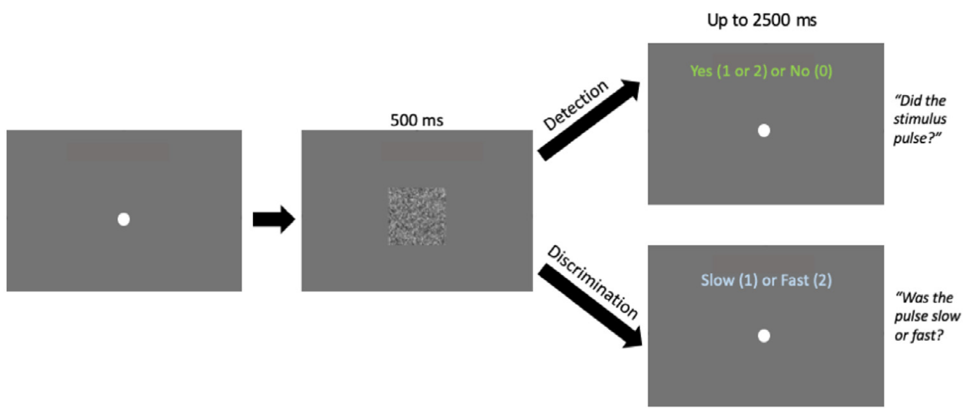

Fig. 2. Experimental procedure for general tasks. The stimulus shown here is an example of one of our visual stimuli. Participants were made aware whether the block was a detection or discrimination block and were then presented with a unisensory or bisensory stimulus for 500 ms. For up to 2500 ms after the stimulus disappeared from the screen, the participant was able to respond to the stimulus for that trial.

2.3. Materials

In the audiovisual task, adapted from Raposo et al. (2012), participants were presented with visual, auditory, or audiovisual stimuli. Visual stimuli were squares of dynamic salt-and-pepper noise, half the width and height of the 18″ CRT monitors used, lasting 500 ms. The arrangement of light and dark pixels in the stimulus changed at a 100 Hz frequency. Participants were asked to detect or discriminate *pulses*, which were changes in contrast polarity of each pixel, relative to the gray background, at either 8 or 12 Hz (which were "slow" and "fast" pulses, respectively). Overall contrast between the visual stimulus and the background was adjusted following each detection or discrimination mini-block for each participant, to keep participant accuracy on detection and discrimination between 60 and 80%. In experiment 1a, if mini-block performance was above or equal to 80% stimulus intensity was decreased 10%, else if mini-block performance was less than or equal to 60% stimulus intensity was increased by 10%. Auditory stimuli were Gaussian white noise lasting 500 ms. Pulses in this modality were fluctuations in sound amplitude occurring at 8 and 12 Hz (which were, again, considered slow and fast pulses). Overall difference in sound amplitude was adjusted for each participant in a similar manner than the visual stimuli in experiment 1a to keep accuracy between 60 and 80%, with the additional rule of a 20% decrease in stimulus intensity when mini-block performance was 100%. In experiment 1b, the accuracy cutoffs were the same, but the change in stimulus intensity was altered to 1.2% in order to more precisely match task difficulty across different sensory modalities. Audiovisual stimuli were constructed by combining auditory and visual stimuli on screen, in synchrony, with a visual contrast and auditory amplitude modulation that matched unisensory stimuli in intensity. All stimuli were created and presented using MATLAB (Mathworks Inc., Natick MA), with the use of Psychophysics Toolbox (Brainard, 1997).

Visuotactile used identical stimuli to the audiovisual experiment, but instead of playing through audio-headphones the Gaussian white noise and its associated amplitude fluctuations, the stimuli were presented to participants through vibro-tactile electromagnetic solenoid-type stimulators powered by a vibro-tactile amplifier tactamp 4.2 from Dancer Design. In this case, participants were presented with visual, tactile, or visuotactile stimuli, and asked to perform the detection and discrimination tasks identical to those in the audiovisual experiments. Audiotactile stimuli were identical to the one generated for audiovisual and visuotactile experiments, but no visual stimuli were presented on screen. Both headphones and vibro-tactile stimulators were used to deliver the stimuli. This time, participants were presented with audio, tactile, or audiotactile stimuli during detection and discrimination tasks. During both visuotactile and audiotactile tasks, the sound from the vibro-tactile stimulators was masked by an external speaker playing the white noise at an individual comfort level.

2.4. Procedure

The task was organized such that blocks of detection and discrimination trials were alternated, and participants completed two blocks of each task per session. In each block, participants experienced 40 mini-blocks of 5 trials each for a total of 200 trials per block and 400 trials per task. In the first audiovisual experiment, 25% of the trials were visual only, 25% were auditory only, and 50% were audio-visual, all interleaved pseudorandomly. In all other experiments, unisensory trials made up 40% of the total trials, and the remaining 60% of trials were bisensory, all interwoven pseudorandomly. Before each block, instructions were presented to let participants know if they were supposed to respond for the detection or discrimination task. Task difficulty was adjusted by increasing or decreasing the contrast of a pulsation based on the criteria reported above on participant performance.

In each discrimination block, participants were asked to judge if the stimulus presented to them was pulsing at a relatively fast or slow rate. Fast pulsations occurred at 12 Hz, and slow pulsations occurred at 8 Hz. Participants were instructed to press a button "1" for slow or "2" for fast pulsations. There was a 50% chance that a stimulus presented to them was pulsing at either rate and occurred equally in each sensory modality. In the case of detection blocks, participants were asked to judge if the stimulus presented to them was pulsing or not. Participants were instructed to press either button "1" or "2" if the stimulus was pulsing or a "0" if there was no pulsation. We used different keys for each type of response to avoid interference of inter-block stimulus–response mappings. There was a 50% chance that a stimulus presented to them was pulsing. Prior to completing the main task, participants completed a practice run on each task for at least 10 trials in each unisensory modality. In some cases, additional verbal instruction and a second practice block was delivered to ensure the tasks were adequately understood.

2.5. Hierarchical drift diffusion model fitting

Drift diffusion modeling utilized the Hierarchical Drift Diffusion Model (HDDM) toolbox for Python (Wiecki et al., 2013). Data was split by task (detection vs. discrimination) for modeling, and two different models were compared for each task. Reaction times (RT) were trimmed such that any RT that was 20 ms or fewer were removed before analysis, accounting for fewer than 0.5% of trials in any experiment. Additionally, only trials from the second half of each session were used, to ensure participant thresholds and task difficulty were largely stable throughout the data. Behavioral and modeling data both used this subset of trials.

Accuracy-coded drift diffusion models were fit to the data, such that the upper boundary of the model represented correct responses and the lower boundary was incorrect responses, and included terms boundary separation (a), non-decision time (t), drift rate (v), and an outlier term for any extreme reaction times in the data. Drift rate was allowed to vary with sensory condition, as we predicted any difference in performance should emerge as a result of a change in evidence accumulation. We did not predict significant differences in boundary separation as we did not predict these would vary with sensory condition, given participants had the same speed and accuracy instructions for all conditions. We also did not expect non-response time to vary with condition, as the motor responses to hit buttons were not linked to sensory condition but rather the "yes/no" or "slow/fast" distinction. The modeling process started with priors on all parameters as set by the toolbox (Wiecki et al., 2013), which, for our free parameters, were as follows:

$$a \sim G\,(1.5, 0.75)$$

$$t \sim G(0.4, 0.2)$$

$$v \sim N(2, 3)$$

where G represents a gamma distribution and N represents a normal distribution. Additionally, the starting point for the drift process (z), was not allowed to vary freely, and instead used the prior values from the model, such that z $\sim$ N(0.5, 0.5), placing z halfway between 0 and a.

The model conducted a 6000-sample Markov Chain Monte Carlo (MCMC) simulation by running 8000 samples with a burn-in of 2000 samples. Comparison of these models was conducted using the deviance information criterion (DIC; Spiegelhalter, Best, Carlin, & Van Der Linde, 2002), which functions similarly to other information criteria but is specialized for hierarchical models. DIC results for all models were compared to single-drift versions of the same model, to ensure the additional complexity of separate drift rates provided better fit to the data, and, in all of the models assessed, this was the case. In addition, we performed a posterior predictive check, which assessed the ability of the model to recapture the behavioral data using its own parameters across the 10th through 90th quantiles of the data, which allowed us to check the fit of the model to the data across the entire distribution of reaction times. This posterior predictive check indicated that 95% confidence intervals around the model's estimates did recapture the observed reaction times in the data. Parameter recovery on the model indicated that the parameters output by the model could be reliably recovered using synthesized data created using the parameters from the model. These tests indicated a good fit of the model to our data, with less than 9% deviation from the input on all model parameters. Model convergence was assessed with a Gelman–Rubin statistic ($<$1.02; Gelman & Rubin, 1992) calculated across 5 models. This statistic compares within- and between-model variability in the estimates, and provided evidence for convergence of the models on stable solutions.

2.6. Data analysis

Multisensory benefit ("MSB") in accuracy, signal detection theory sensitivity (d′), and drift rate was calculated for each participant as the proportion change in performance in the bisensory condition above that of the best unisensory condition (Rach, Diederich, & Colonius, 2011). As such, the best unisensory performance was subtracted from the multisensory performance in the same experiment, and this difference was divided by the best unisensory performance (Eq. (1)).

$$MSB = \frac{(bisensory) - max(unisensory)}{max(unisensory)} \tag{1}$$

For reaction time, the equation was changed somewhat to reflect fast reaction times as superior performance, such that the average bisensory reaction time was subtracted from the faster of the average unisensory reaction times, and divided by the faster of the average unisensory reaction times (Eq. (2)).

$$MSB = \frac{min\,(unisensory) - bisensory}{min(unisensory)} \tag{2}$$

In addition to analyzing average reaction time for each participant per condition, we also assessed *inverse efficiency scores* (Rach et al., 2011), where the average reaction time (RT) is adjusted by the average detection rate for a stimulus. This adjusted RT measure, hereafter shortened as RT*, helps to separate improved performance due to speed accuracy tradeoffs from accuracy changes that reflect improved performance (Rach et al., 2011). Benefit to the adjusted RT was also assessed for multisensory benefit, using the formula in Eq. (2).

The MSB for accuracy, d′, reaction time, RT*, and drift rate was calculated per participant and averaged across individuals for analysis. Average MSB for accuracy, d′, and reaction time were analyzed with t-tests with p-values adjusted for multiple comparisons using the Holm procedure (Holm, 1979). Because the HDDM procedure violated the independence of observations assumption necessary for a t-test, we instead used Bayesian hypothesis testing (conducted using the BEST package in R; Kruschke, 2013) to assess via 10 000-sample MCMC simulation if the MSB for drift rate was significant.

In addition to the diffusion model, the data was investigated using a race model (Miller, 1982). This model compares observed bimodal reaction times to those that would be predicted by optimally combining the unisensory reaction times. Violations of this model such that reaction times for bimodal conditions are significantly faster than those predicted by the probability sum of the unimodal components. Such violations would indicate RT facilitation above that expected with fully independent unisensory inputs, and would indicate crosstalk between these senses (see Colonius & Diederich, 2017 for an overview in multisensory contexts). In a similar fashion to the drift diffusion model, this model examines deviations across the entire distribution of reaction times, though it only utilizes reaction time information. Comparison to the race model helps establish if simultaneous use of reaction time and accuracy data are important for establishing multisensory benefit. We have specifically chosen to use an extension of the race model which uses a permutation test of the model to control for Type I error rate (Gondan, 2010; Gondan & Minakata, 2016). For each experiment, the race model was assessed at every 5th quantile in the data, with 10,001 permutations computed to create an estimate of the distribution of the probability sum per participant. The results of this test are normed such that negative t_{max} values indicate multisensory performance regularly below the probability sum of the unisensory components, and sufficiently large positive values indicate a violation of the race model inequality.

3. Experiment 1, audiovisual integration

3.1. Participants

Experiment 1 was split into two halves based on different staircasing procedures for thresholding used for the participants and the proportions of unisensory and multisensory trials. In experiment 1a, participants were 15 undergraduate students (11 females) from the University of California, Los Angeles, with an average age of 20.20 years (SD = 1.15 years). In experiment 1b, participants were 18 undergraduate students (13 females) from the University of California, Los Angeles, with an average age of 19.56 years (SD = 1.04 years). Participants in both experiments had normal or corrected-to-normal vision and reported no hearing issues. Participants in both experiments were compensated with course credit for their participation. Prior to the start of the experiment, participants signed an informed consent and were presented with written and verbal instructions of both tasks.

3.2. Results

In the detection task portion of experiment 1a, participants showed no significant MSB in mean accuracy (M = 0.011, SD = 0.067, $t(14) = 0.640$, $p = 0.99$), reaction time (M = 0.007, SD = 0.076, $t(14) = 0.347$, $p = 0.99$), or signal detection sensitivity ($M = -0.019$, $SD = 0.212$, $t(14) = -0.339$, $p = 0.99$). RT*, where RT was adjusted by detection rate to assess intersensory facilitation, likewise showed a non-significant benefit (M = 0.045, SD = 0.094, $t(14) = 1.850$, p = .342). However, there was a significant MSB for audiovisual drift rates over the unisensory drift rates (M = 0.200, SD = 0.078, 95% credible interval = [0.044, 0.355]), indicating a multisensory advantage was present in the data (Fig. 3a). The discrimination task in experiment 1a showed a similar pattern of results (Fig. 3b), such that there was no significant MSB in accuracy ($M = 0.004$, $SD = 0.060$, $t(14) =$ 0.241, p = 0.99), sensitivity ($M = -0.002$, $SD = 0.336$, $t(14) =$ -0.023, $p = 0.99$), reaction time ($M = -0.008$, $SD = 0.095$, $t(14)$ $= -0.331$, $p = 0.99$), or RT* ($M = 0.063$, $SD = 0.093$, $t(14) =$ 2.648, p = .076). However, drift rate did show a significant MSB in the performance ($M = 0.186$, $SD = 0.072$, 95% credible interval = [0.042, 0.327]).

Experiment 1b (Fig. 3c and d) showed a somewhat different pattern of results. In the detection task, a significant multisensory benefit was observed in response accuracy (M = 0.043, SD = 0.049, $t(17) = 3.771$, p = .006) and RT* ($M = 0.040$, $SD = 0.045$, $t(17) = 3.761$, p = .006). Unadjusted reaction time (M = −0.006, SD = 0.069, t(17) = −0.383, p = 0.707) and signal detection sensitivity (M = 0.111, SD = 0.206, t(17) = 2.295, p = 0.070) did not show a multisensory benefit. A benefit of multisensory presentation of stimuli on drift rate was also apparent in the detection task (M = 0.135, SD = 0.024, 95% credible interval = [0.087, 0.181]). The discrimination portion of this experiment showed no significant advantages in response accuracy (M = 0.046, SD = 0.079, $t(17) = 2.451$, p = 0.051) or sensitivity (M = 0.158, SD = 0.363, $t(17) = 1.845$, p = 0.083), but there was a significant benefit in reaction time (M = 0.047, SD = 0.058, $t(17) = 3.423$, p = .010) and in RT* (M = 0.104, SD = 0.088, $t(17) = 5.037$, $p < .001$). Further, drift rate, still showed a significant benefit for multisensory stimuli in the discrimination task (M = 0.403, SD = 0.092, 95% credible interval = [0.247, 0.611]) tasks.

Tests of the race model inequality for both portions of experiment 1a indicate that there was no significant deviance from predicted sensory facilitation for either detection ($t_{max} = -2.898$, p > 0.99) or discrimination tasks ($t_{max} = -1.493$, p > .99). The same was found for detection ($t_{max} = -3.522$, p > .99) and discrimination tasks ($t_{max} = 0.319$, p = .716) in experiment 1b. As such, we do not observe multisensory benefit in reaction time that exceeds the expectation of statistical facilitation between multiple sensory signals.

4. Experiment 2, visuotactile integration

4.1. Participants

Participants in experiment 2 were 17 undergraduate students (7 females) from the University of California, Riverside, with an average age of years 22.75 years (SD = 5.43 years). All had normal or corrected-to-normal vision and reported no tactile issues. Participants were compensated with 10 dollars per hour of their participation. Prior to the start of the experiment, participants signed an informed consent and were presented with written and verbal instructions of both tasks.

4.2. Results

Results for the visuotactile detection task (Fig. 4a) showed a significant multisensory benefit for response accuracy (M = 0.097, SD = 0.072, $t(16) = 5.578$, $p < .001$) and sensitivity (M = 0.274, SD = 0.269, $t(16) = 4.190$, p = 0.001). There was no significant advantage in reaction time (M = 0.021, SD = 0.055, $t(16) = 1.557$, p = 0.139), however, RT* did show a significant multisensory benefit ($M = 0.085$, $SD = 0.063$, $t(16) =$ 5.520, $p < .001$). Drift rate also revealed a multisensory performance benefit ($M = 0.510$, $SD = 0.058$, 95% credible interval = [0.395, 0.623]). The discrimination task showed a similar pattern of results (Fig. 4b), where significant multisensory benefit was

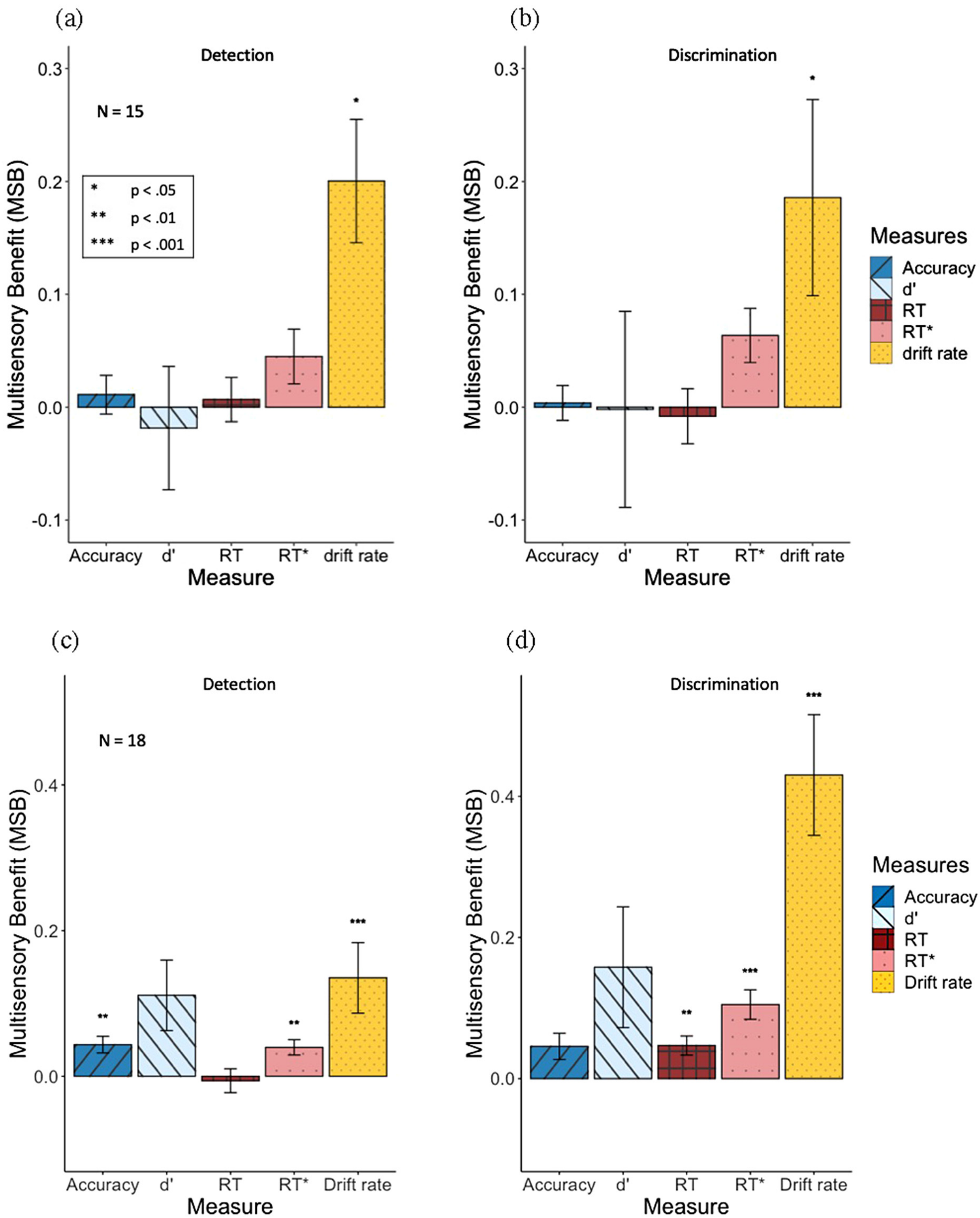

Fig. 3. Advantage in accuracy, d′, RT, and RT* for (a) the detection and (b) discrimination portions of experiment 1a, as well as the (c) detection and (d) discrimination tasks in experiment 1b.

observed in sensitivity ($M = 0.134$, $SD = 0.190$, $t(16) = 2.912$, $p = .041$), but not in reaction time data (M = 0.020, SD = 0.056, t(16) = 1.442, p = .168). Accuracy was marginally significant ($M = 0.033$, $SD = 0.051$, $t(16) = 2.668$, $p = .051$), as was RT* ($M = 0.030$, $SD = 0.047$, $t(16) = 2.594$, $p = .051$). Drift rates, again, indicated there was a significant benefit obtained from multisensory stimulus presentation ($M = 0.279$, $SD = 0.048$, 95% credible interval = [0.183, 0.376]).

Race model inequality tests indicated no significant difference between observed and predicted multisensory reaction time in either detection ($t_{max} = -2.564$, p > .99) or discrimination ($t_{max} = -2.495$, p > .99).

5. Experiment 3, audiotactile integration

5.1. Participants

Participants in experiment 3 were 17 undergraduate students (11 females) from the University of California, Riverside, with an average age of 21.18 years (SD = 2.34 years). All reported

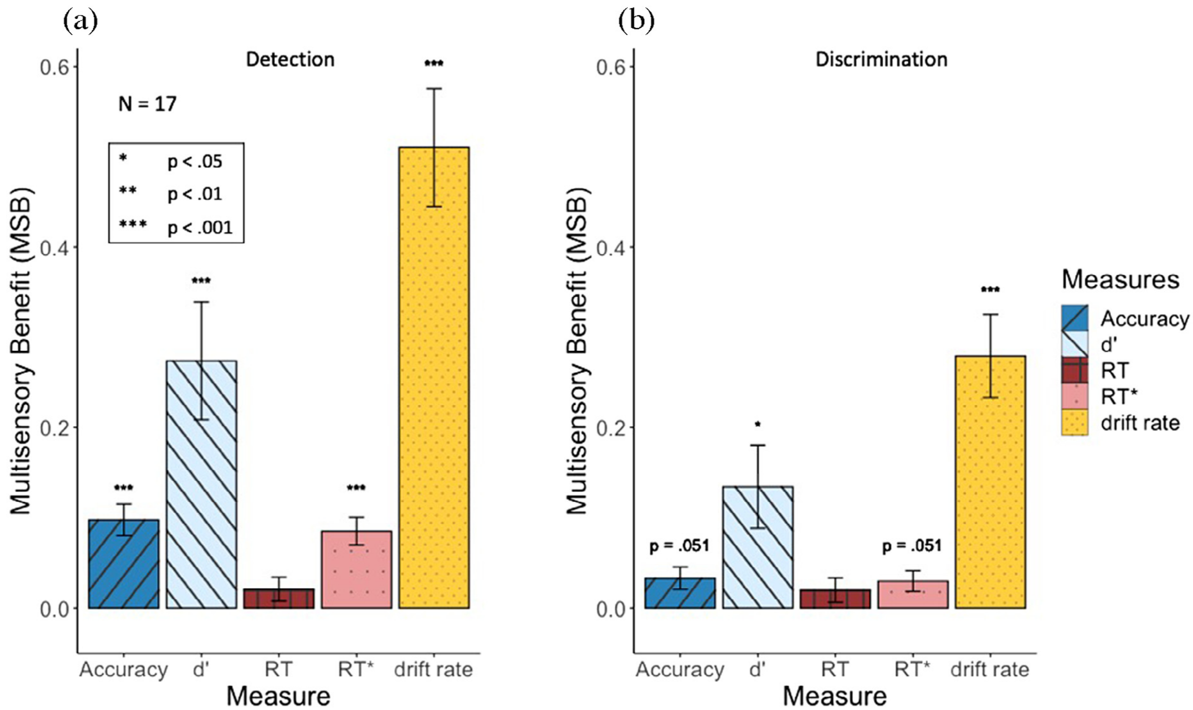

Fig. 4. For the visuotactile experiment, the average advantage in accuracy, d′, RT, RT*, and drift rate for (a) detection and (b) discrimination tasks. Visuotactile data showed advantages in accuracy, sensitivity, inverse efficiency score, and drift rate, while also showing no advantage for participant RT.

normal hearing and reported no tactile issues. Participants were compensated with 10 dollars per hour of their participation. Prior to the start of the experiment, participants signed an informed consent and were presented with written and verbal instructions of both tasks.

5.2. Results

In the audiotactile detection task (Fig. 5a), the data revealed no significant benefit for multisensory stimuli compared to unisensory stimuli in accuracy ($M = -0.052$, $SD = 0.135$, $t(16) = -1.600$, $p = 0.388$), sensitivity ($M = -.102$, $SD = 0.415$, $t(16) = -1.013$, $p = 0.652$), reaction time ($M = -0.010$, $SD = 0.043$, $t(16) = -0.987$, $p = .652$), or RT* ($M = -0.010$, $SD = 0.043$, $t(16) = -1.897$, $p = .304$). Analysis of model drift rates also showed no significant MSB ($M = -0.015$, SD $= 0.092$, 95% credible interval $= [-0.120, 0.165]$).

Audiotactile discrimination results (Fig. 5b) showed a significant MSB for multisensory presentation of stimuli in accuracy ($M = 0.148$, $SD = 0.112$, $t(16) = 5.419$, $p < .001$), sensitivity ($M = 0.601$, $SD = 0.543$, $t(16) = 4.449$, $p < .001$), reaction time ($M = 0.069$, $SD = 0.051$, $t(16) = 5.580$, $p < .001$), and RT* ($M = -0.077$, $SD = 0.168$, $t(16) = 5.828$, $p < .001$). The model drift rates also indicated a significant MSB ($M = 0.983$, $SD = 0.123$, 95% credible interval $= [0.741, 1.229]$).

Race model results for experiment 3 indicated no significant deviation from probability sum predictions in either detection ($t_{\max} = -6.771$, p $> .99$) or discrimination tasks ($t_{\max} = 1.066$, p $= .462$).

6. Discussion

The results presented here suggest that diffusion models are sensitive to multisensory benefits, across both detection and discrimination tasks, and across different multiple sensory combinations. In each of the presented experiments, drift rate was found to be at least as sensitive as measures of accuracy, sensitivity, reaction time, inverse efficiency scores, and the race model. Across all of these experiments, drift rate was consistently the most reliable measure of multisensory benefit. The next most consistent measure was RT*, the only other measure that combined RT and accuracy, highlighting the need for a combined measure. However, drift rate in experiment 1a did pick up a benefit even when RT* did not, indicating a benefit of the diffusion model. While multisensory benefit may manifest in a variety of different measures of performance, here, we show that in the majority of cases, the benefit is manifested in the change in drift rate, and less consistently in other measures commonly used in the literature such as accuracy, reaction time, or sensitivity.

Is it possible that drift rate shows a benefit when it should not (i.e., when in reality there is not a true multisensory benefit to processing)? In principle, this possibility cannot be ruled out, the same way that any other measure (such as accuracy and reaction time) can by random chance exhibit a benefit. In the absence of a ground truth about the presence of enhanced neural processing, one cannot rule out the possibility of false positives in any measure. Strictly speaking, ground truth about multisensory benefit is never available, even if one could track the activity of all neurons in the brain in different conditions. However, the literature on multisensory processing in the last two decades has established that multisensory redundant signals generally result in enhanced processing compared to unisensory conditions, as shown in a variety of perceptual tasks, settings, and sensory combinations (for examples of reviews and overviews, see Calvert, Spence, & Stein, 2004; Ernst & Bülthoff, 2004; Groh, 2014; Murray & Wallace, 2011; Shams & Kim, 2010; Trommershauser, Kording, & Landy, 2011). Further, the results shown here were replicated across multiple experiments and sensory combinations. Therefore, it is reasonable to expect that the redundant multisensory signals in the current experiments also result in superior processing over those of unisensory conditions. Therefore, we consider it reasonable to conclude that the drift rate is reflective of a true benefit in processing compared to unisensory conditions. Still,

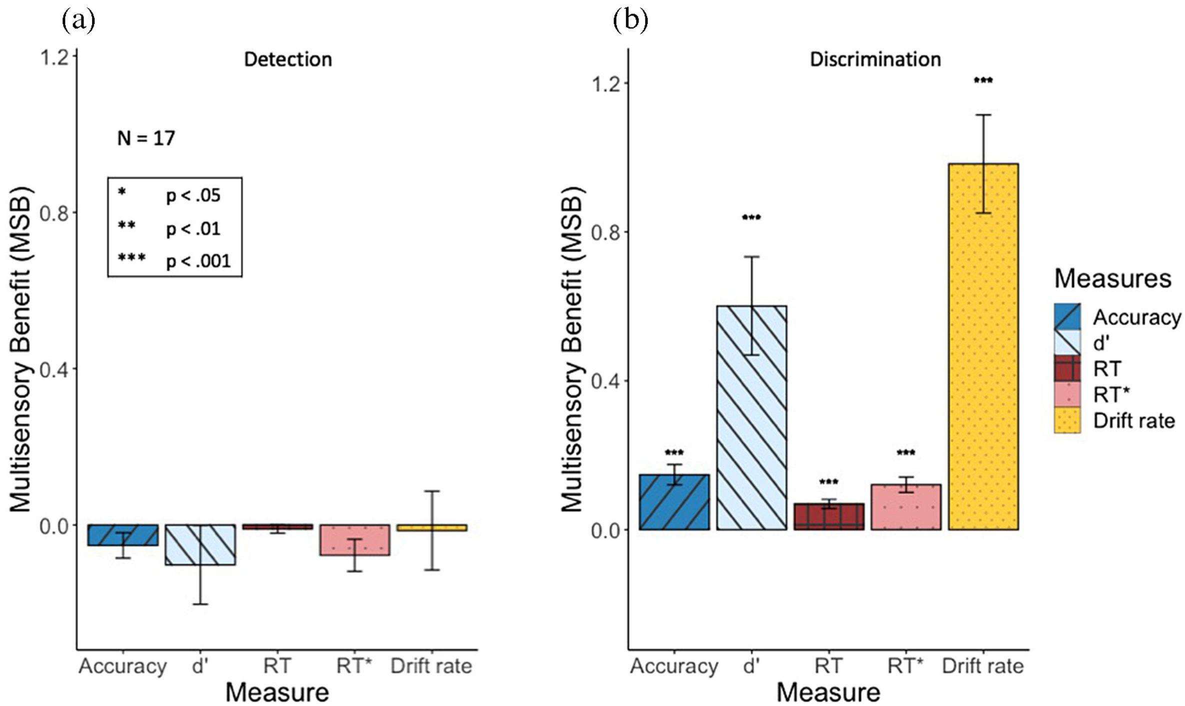

Fig. 5. For the audiotactile experiment, the average advantage in accuracy, d′, RT, RT*, and drift rate for (a) detection and (b) discrimination tasks. Audiotactile detection showed no multisensory advantage in any of the measures used, and audiotactile discrimination showed an advantage in all of the measures used.

future work should examine how the relative false positive rates of diffusion models compare to alternative methods.

Interestingly, in the experiment 1a discrimination task, no benefit was seen in either task for any measure but the drift rate. The pattern of these results appears to reveal a difference in the more accurate and the faster unisensory conditions, and a multisensory condition that captures the best of these separate components. Of the 15 participants in the discrimination task, 14 of those were, on average, more accurate in the auditory condition. Of those 14 participants, 11 responded, on average, more quickly in the visual condition. This means that the majority of our participants showed higher accuracy in their slower modality. As drift rate combines these into a single measure, we believe that this reflects that the multisensory condition resembles the speed of the faster unisensory modality and the accuracy of the more accurate unisensory modality, and only then in combination does this benefit emerge.

The significant benefit to RT and RT* without a parallel race model inequality violation, observed across multiple experiments, may also initially appear somewhat contradictory. However, it should be noted that these measures are quantifying benefit against different baselines. RT and RT* are comparing multisensory performance to a single unisensory performance, whereas the race model inequality compares multisensory performance to an additive combination of the unisensory components. Thus, such results can occur when performance in the multisensory condition exceeds those of unisensory conditions but not the probability sum (indicating no coactivation between these senses).

That the model can robustly detect multisensory benefits, across a number of sensory combinations or patterns of benefits, makes drift diffusion models a potentially consistent tool for analyzing the results of multisensory studies. This is especially true as such models become more efficient, requiring fewer trials to fit model parameters. The current experiments had between 100 and 200 trials per condition, though the model has been used with fewer (as in Regenbogen et al., 2016). While still an impediment to some experimental designs, this does reduce how prohibitive the number of trials required for this type of modeling can be, allowing for greater use of this technique. Drift diffusion models have also, traditionally, imposed design restrictions on experiments, as they are typically only used to analyze tasks with binary decisions and are limited to one-step decisions (Ratcliff & McKoon, 2008; Voss, Nagler, & Lerche, 2013). Extensions of the model that address these issues do exist that allow for more alternatives in the decision, up to and including continuous response scales (Krajbich & Rangel, 2011; Ratcliff, 2018; Smith, 2016), allow for multiple steps in a decision-making process (Pleskac & Busemeyer, 2010; Resulaj, Kiani, Wolpert, & Shadlen, 2009), and allow for participant bias to change through sequential trials (Nguyen, Josić, & Kilpatrick, 2019), though such variants are less commonly used.

This is not to say, however, that this particular model is the only option for use in analysis of multisensory stimuli. While the HDDM uses Bayesian priors on its parameters to allow the model to converge more rapidly on parameter values, it is not the only Bayesian drift diffusion model available and drift diffusion modeling is not the only technique available to simultaneously investigate reaction time and accuracy. Previous studies on this topic have, in fact, shown that this particular Bayesian hierarchical method does not necessarily outperform alternative models in terms of efficiency (Lerche, Voss, & Nagler, 2017). Other groups have created similar hierarchical models (Vandekerckhove, Tuerlinckx, & Lee, 2011) that can be implemented using similar Bayesian optimization techniques via Gibbs sampling (Wabersich & Vandekerckhove, 2014). Other non-hierarchical models have also attempted to improve efficiency by limiting the range in which model parameters can fall (Diederich & Busemeyer, 2003; Nidiffer et al., 2018). Indeed, the hierarchical solution provided here is only one of the possible solutions to making this model more computationally efficient, and such methods are being worked into traditional DDMs as well as hierarchical variants. More traditional variants of sequential sampling models are also beginning to become more efficient in their processing, including the *Ornstein–Uhlenbeck model*, a variant of a sequential sampling model that combines evidence accumulation with a decay term that acts to bring the evidence accumulation back towards the starting point (Diederich, 1995; Ratcliff & Smith, 2004), or the *compatibility bias model* (Yu, Dayan, & Cohen, 2009), that investigates decision-making processes through the framework of Bayesian causal inference modeling.

A significant benefit that can be reaped from any of these models, though, beyond just the ability to investigate a speed-accuracy tradeoff in multisensory experiments, is the ability to break down the decision-making process to investigate what portion of the decision-making process is influenced by the presentation of multisensory information. The current experiment focused on drift rate as a parameter of interest because we hypothesized that the evidence accumulation process would be most affected by the inclusion of a second sensory modality. However, the parameters available through this type of modeling allow a large number of features of the decision-making process to be investigated. The compatibility bias model has been used to investigate how multisensory information may be differentially used by older and younger adults, and found that reduced reaction times in older adults were caused in part by more conservative decision making, in terms of larger boundary separation, and slower non-decision response time (Jones, Beierholm, Meijer, & Noppeney, 2019). This allows this type of decision-making model to additionally provide more insight into where in the decision-making process multisensory stimuli may exert influence, allowing us to better categorize where the benefits of multisensory stimuli may arise from.

Given the potential benefits of this type of modeling, then, we would generally advocate for greater use of such sequential sampling models, especially those that allow for a smaller number of trials to be used effectively. This could provide greater insight into both when and how multisensory stimuli benefit performance on a large number of tasks, with a reliability that has the ability to surpass current methods that independently investigate reaction speed and response accuracy.

Acknowledgments

We would like to thank Michael J. Frank, Peggy Seriès, Frank Karvelis, and Jozsef Fiser for their feedback on modeling procedure and advice on fitting and analyzing the drift diffusion model. We would also like to thank Ashkan Arjomandi for help with proofreading this manuscript.

Funding

This project was funded by the National Science Foundation, USA grant BCS-1057625.

References

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436.
Brebner, J. T., & Welford, A. T. (1980). Introduction: an historical background sketch. In A. T. Welford (Ed.), *Reaction Times* (pp. 1–23). New York: Academic Press.
Bushara, K. O., Weeks, R. A., Ishii, K., Catalan, M. J., Tian, B., Rauschecker, J. P., et al. (1999). Modality-specific frontal and parietal areas for auditory and visual spatial localization in humans. *Nature Neuroscience*, *2*(8), 759–766.
Calvert, G., Spence, C., & Stein, B. E. (Eds.), (2004). *The handbook of multisensory processes*. Cambridge, MA: MIT press.
Colonius, H., & Diederich, A. (2017). Measuring multisensory integration: from reaction times to spike counts. *Scientific Reports*, *7*, 3023.
Diederich, A. (1995). Intersensory facilitation of reaction time: Evaluation of counter and diffusion coactivation models. *Journal of Mathematical Psychology*, *39*(2), 197–215.
Diederich, A. (2008). A further test of sequential-sampling models that account for payoff effects on response bias in perceptual decision tasks. *Perception & Psychophysics*, *70*(2), 229–256.
Diederich, A., & Busemeyer, J. R. (2003). Simple matrix methods for analyzing diffusion models of choice probability, choice response time, and simple response time. *Journal of Mathematical Psychology*, *47*(3), 304–322.
Driver, J., & Spence, C. (2000). Multisensory perception: Beyond modularity and convergence. *Current Biology*, *10*(20), 10–12.
Drugowitsch, J., DeAngelis, G. C., Angelaki, D. E., & Pouget, A. (2015). Tuning the speed-accuracy trade-off to maximize reward rate in multisensory decision-making. *Elife*, *4*, Article e06678.
Drugowitsch, J., DeAngelis, G. C., Klier, E. M., Angelaki, D. E., & Pouget, A. (2014). Optimal multisensory decision-making in a reaction-time task. *Elife*, *3*, e03005.
Ernst, M. O., & Bülthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences*, *8*(4), 162–169.
Fitts, P. M. (1966). Cognitive aspects of information processing: III. Set for speed versus accuracy. *Journal of Experimental Psychology*, *71*(6), 849–857.
Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*(4), 457–472.
Gingras, G., Rowland, B. A., & Stein, B. E. (2009). The differing impact of multisensory and unisensory integration on behavior. *Journal of Neuroscience*, *29*(15), 4897–4902.
Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual Review of Neuroscience*, *30*, 535–574.
Gomez, P., Ratcliff, R., & Perea, M. (2007). A model of the go/no-go task. *Journal of Experimental Psychology: General*, *136*(3), 389–413.
Gondan, M. (2010). A permutation test for the race model inequality. *Behavior Research Methods*, *42*(1), 23–28.
Gondan, M., & Minakata, K. (2016). A tutorial on testing the race model inequality. *Attention, Perception, & Psychophysics*, *78*(3), 723–735.
Groh, J. M. (2014). *Making space: How the brain knows where things are*. Harvard University Press.
Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*(2), 65–70.
Jones, S. A., Beierholm, U., Meijer, D., & Noppeney, U. (2019). Older adults sacrifice response speed to preserve multisensory integration performance. *Neurobiology of Aging*, *84*, 148–157.
Krajbich, I., & Rangel, A. (2011). Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. *Proceedings of the National Academy of Sciences*, *108*(33), 13852–13857.
Kruschke, J. K. (2013). Bayesian estimation supersedes the t test. *Journal of Experimental Psychology: General*, *142*(2), 573–603.
Leite, F. P., & Ratcliff, R. (2010). Modeling reaction time and accuracy of multiple-alternative decisions. *Attention, Perception, & Psychophysics*, *72*(1), 246–273.
Lerche, V., Voss, A., & Nagler, M. (2017). How many trials are required for parameter estimation in diffusion modeling? A comparison of different optimization criteria. *Behavior Research Methods*, *49*(2), 513–537.
Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). New Jersey: Lawrence Erlbaum Associates.
Mahani, M. A. N., Bausenhart, K. M., Ahmadabadi, M. N., & Ulrich, R. (2019). Multimodal Simon effect: A multimodal extension of the diffusion model for conflict tasks. *Frontiers in Human Neuroscience*, *12*, 507.
Miller, J. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive Psychology*, *14*(2), 247–279.
Murray, M. M., & Wallace, M. T. (Eds.), (2011). *The neural bases of multisensory processes*. CRC Press.
Nguyen, K. P., Josić, K., & Kilpatrick, Z. P. (2019). Optimizing sequential decisions in the drift–diffusion model. *Journal of Mathematical Psychology*, *88*, 32–47.
Nidiffer, A. R., Diederich, A., Ramachandran, R., & Wallace, M. T. (2018). Multisensory perception reflects individual differences in processing temporal correlations. *Scientific Reports*, *8*, 14483.
Pleskac, T. J., & Busemeyer, J. R. (2010). Two-stage dynamic signal detection: A theory of choice, decision time, and confidence. *Psychological Review*, *117*(3), 864–901.
Rach, S., Diederich, A., & Colonius, H. (2011). On quantifying multisensory interaction effects in reaction time and detection rate. *Psychological Research*, *75*(2), 77–94.
Raposo, D., Sheppard, J. P., Schrater, J. P., & Churchland, A. K. (2012). Multisensory decision-making in rats and humans. *Journal of Neuroscience*, *32*(11), 3726–3735.
Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59–108.
Ratcliff, R. (2018). Decision making on spatially continuous scales. *Psychological Review*, *125*(6), 888–935.
Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation*, *20*, 873–922.

Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review*, *111*(2), 333–367.

Regenbogen, C., Johansson, E., Andersson, P., Olsson, M. J., & Lundström, J. N. (2016). Bayesian-based integration of multisensory naturalistic perithreshold stimuli. *Neuropsychologia*, *88*, 123–130.

Resulaj, A., Kiani, R., Wolpert, D. M., & Shadlen, M. N. (2009). Changes of mind in decision-making. *Nature*, *461*, 263–266.

Rosenblum, L. D., Dias, J. W., & Dorsi, J. (2017). The supramodal brain: Implications for auditory perception. *Journal of Cognitive Psychology*, *29*(1), 65–87.

Shams, L., & Kim, R. (2010). Crossmodal influences on visual perception. *Physics of Life Reviews*, *7*(3), 269–284.

Shams, L., & Seitz, A. (2008). Benefits of multisensory learning. *Trends in Cognitive Sciences*, *12*(11), 411–417.

Shelton, J., & Kumar, G. P. (2010). Comparison between auditory and visual simple reaction times. *Neuroscience & Medicine*, *1*(1), 30–32.

Shinn-Cunningham, B. G. (2008). Object-based auditory and visual attention. *Trends in Cognitive Sciences*, *12*(5), 182–186.

Smith, P. L. (2016). Diffusion theory of decision making in continuous report. *Psychological Review*, *123*(4), 425–451.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, *64*(4), 583–639.

Stein, B. E., Stanford, T. R., Wallace, J. W. V., & Jiang, W. (2004). Crossmodal spatial interactions in subcortical and cortical circuits. In C. Spence, & J. Driver (Eds.), *Crossmodal Space and Crossmodal Attention* (pp. 25–50). Oxford: Oxford University Press.

Trommershauser, J., Kording, K., & Landy, M. S. (Eds.), (2011). *Sensory Cue Integration*. Oxford University Press.

Van Zandt, T., Colonius, H., & Proctor, R. W. (2000). A comparison of two response time models applied to perceptual matching. *Psychonomic Bulletin & Review*, *7*(2), 208–256.

Vandekerckhove, J., Tuerlinckx, F., & Lee, M. D. (2011). Hierarchical diffusion models for two-choice response times. *Psychological Methods*, *16*(1), 44–62.

Voss, A., Nagler, M., & Lerche, V. (2013). Diffusion models in experimental psychology. *Experimental Psychology*, *60*(6), 385–402.

Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica*, *41*(1), 67–85.

Wiecki, T. V., Sofer, I., & Frank, M. J. (2013). HDDM: Hierarchical Bayesian estimation of the drift-diffusion model in Python. *Frontiers in Neuroinformatics*, *7*, 14.

Yu, A. J., Dayan, P., & Cohen, J. D. (2009). Dynamics of attentional selection under conflict: Toward a rational Bayesian account. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(3), 700–717.