Modality-specific and multisensory mechanisms of spatial attention and expectation

Authors: Arianna Zuanazzi[1,2], Uta Noppeney[1,3]

[1]Computational Neuroscience and Cognitive Robotics Centre, University of Birmingham, UK; [2]Department of Psychology, New York University, New York, NY, USA; [3]Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands

Corresponding author: Arianna Zuanazzi, Department of Psychology, New York University, New York, NY, USA, email: az1864@nyu.edu

Number of words: 9110

Abstract

In our natural environment, the brain needs to combine signals from multiple sensory modalities into a coherent percept. While spatial attention guides perceptual decisions by prioritizing processing of signals that are task-relevant, spatial expectations encode the probability of signals over space. Previous studies have shown that behavioral effects of spatial attention generalize across sensory modalities. However, because they manipulated spatial attention as signal probability over space, these studies could not dissociate attention and expectation or assess their interaction.

In two experiments, we orthogonally manipulated spatial attention (i.e., task-relevance) and expectation (i.e., signal probability) selectively in one sensory modality (i.e., primary) (experiment 1: audition, experiment 2: vision) and assessed their effects on primary and secondary sensory modalities in which attention and expectation were held constant.

Our results show behavioral effects of spatial attention that are comparable for audition or vision as primary modalities; yet, signal probabilities were learnt more slowly in audition, so that spatial expectations were formed later in audition than vision. Critically, when these differences in learning between audition and vision were accounted for, both spatial attention and expectation affected responses more strongly in the primary modality in which they were manipulated, and generalized to the secondary modality only in an attenuated fashion. Collectively, our results suggest that both spatial attention and expectation rely on modality-specific and multisensory mechanisms.

Keywords

Attention, expectation, multisensory, perceptual inference and learning, space

Introduction

Spatial attention is a top-down mechanism that is critical for the selection of task-relevant information. It facilitates perception (e.g., faster reaction times, greater accuracy) of signals presented at the attended location (Carrasco, 2011; van Ede et al., 2012). By contrast, spatial expectation (signal probability) facilitates perception by encoding the statistical structure of the environment (Summerfield & Egner, 2009; Rohenkohl et al., 2014). In everyday life, spatial attention and expectation are closely intertwined. For instance, observers often allocate attentional resources to locations in space where events are more likely to occur (Summerfield & Egner, 2009; Feldman & Friston, 2010 for further discussion within the predictive coding framework).

Importantly, in our natural multisensory environment our brain is constantly exposed to auditory and visual signals. This raises the critical question of whether allocation of attention and encoding of signal probability are performed in a modality-specific fashion or interactively across sensory modalities. Previous research has suggested that spatial attention relies on cognitive resources that are partially shared across sensory modalities (Eimer & Driver, 2001; Wahn & König 2015, 2017). For instance, Spence & Driver (1996) manipulated spatial attention by presenting signals with a higher probability in the attended relative to unattended hemifield in one modality only (i.e., primary modality). They showed behavioral facilitation for signals presented at the attended location not only for the primary modality (e.g., audition) but also for the secondary modality (e.g., vision) in which spatial attention was not explicitly manipulated. Likewise, neuroimaging studies showed increased activations for signals presented at the attended location not only in the primary modality in which attention was manipulated but also in the secondary modality (Eimer & Schröger, 1998; Eimer, 1999; Macaluso et al., 2002; Santangelo et al., 2009; Zuanazzi & Noppeney,

2019). Crucially, in this past work the attentional effects were greater in the primary than in the secondary modality (Spence & Driver, 1996; Mondor & Amirault, 1998). The attenuated generalization across sensory modalities suggests that attentional resources are not supramodal, but partially shared (Driver & Spence, 1998).

However, this past research conflated attention and expectation by manipulating spatial attention via probabilistic spatial cues or changes in signal probability (Posner, 1980; Spence & Driver, 1996, 1997; Macaluso et al., 2002; Kincade et al., 2005; Bressler et al., 2008; Santangelo et al., 2009). Notably, the Posner cuing paradigm shifts observers' attention via spatial cues that indicate whether a target is for instance more likely to appear the left or right hemifield. As a result, previous research cannot not dissociate the effects of spatial attention and expectation or assess their interaction.

Thus, the first question of this study is whether spatial attention and/or expectation generalize across sensory modalities to a similar extent when they are manipulated independently. A recent neuroimaging study for instance suggested that spatial attention relies mainly on frontoparietal cortices for both primary and secondary modalities, while spatial expectations are formed in sensory systems selectively for the primary modality (Zuanazzi & Noppeney, 2019). As a consequence, we would expect spatial signal probability to be encoded selectively for the primary modality and generalize to a secondary modality only to a limited degree.

A second unresolved question is whether the generalization across sensory modalities depends on whether attention and expectation are manipulated in the auditory or visual modalities as primary manipulation modality – i.e., on the direction of cross-sensory generalization. Previous studies of multisensory attention have indeed shown asymmetric multisensory generalization, depending on which modality was manipulated as primary modality (Ward at al., 2000; Greene et al., 2001; Molholm et al., 2007). Moreover, one may

expect differences in cross-sensory generalization from vision to audition and vice versa because spatial representations and expectations are encoded differently in audition and vision. Visual and auditory systems encode space via different reference frames (i.e., eye-centered vs head-centered) and representational formats. In the visual system spatial location is directly encoded in the sensory epithelium and later in a place code, i.e., via retinotopic organization of primary and higher order visual cortices (e.g., Sereno et al., 1995; Maier & Groh, 2009). In the auditory system, spatial locations are computed from binaural and monoaural cues in the brain stem and are represented in a hemifield code in primary auditory cortices (e.g., Lauter et al., 1985; Maier & Groh, 2009). Further, in everyday life under normal lighting conditions vision usually provides more reliable spatial information than audition and therefore often dominates spatial perception (Spence & Driver, 1997; Aller et al., 2015; Odegaard et al., 2015; Rohe & Noppeney, 2015a, 2015b. 2016, 2018; Aller & Noppeney, 2019; Jones et al., 2019; Meijer et al., 2019). As a result, we would expect the generalization of spatial attention and expectation to depend on whether attention and expectation are manipulated primarily in vision or audition.

Third, in everyday life spatial expectations are formed when observers implicitly learn the statistical structure of their multisensory environment such as the probability of signals occurring at a particular location. Because spatial information is encoded less reliably in audition than vision, this learning may be faster in vision than in audition. Thereby spatial expectations may also be affected by multisensory processes in perceptual learning (e.g., Kim et al., 2008; Batson et al., 2011). For instance, previous studies suggested that perceptual learning in temporal discrimination tasks generalizes across sensory modalities (Warm et al., 1975; Nagarajan et al., 1998; Meegan et al. 2000; Bratzke et al., 2012; Bueti et al., 2012, 2014; but see Lapid et al., 2009). This study will investigate how observers dynamically form spatial expectations by learning signal probability over time (Crist et al., 1997).

To summarize this study was designed to address two questions: First, whether spatial attention and expectations (when manipulated independently) rely on amodal or modality-specific mechanisms. Second, whether the cross-sensory generalization of spatial attention and expectation differ when operating from vision to audition or vice versa, if spatial representations and probability learning of signals over space (i.e., spatial expectations) rely on only partially shared mechanisms. In two experiments (within participants), we orthogonally manipulated spatial attention and expectation selectively in one sensory modality as primary modality (experiment 1: audition, experiment 2: vision). We assessed the effects of spatial attention and expectation in this primary modality and how they generalize crossmodally to signals in a secondary sensory modality, in which spatial attention and expectation were not explicitly manipulated (experiment 1: vision, experiment 2: audition). We expected that both spatial attention and expectations rely on mechanisms that are partially shared across sensory modalities and that the generalization of spatial expectations depends on whether they are manipulated in vision or audition as primary modalities.

Materials and Methods

Participants

Twenty-eight healthy subjects (19 females; mean age = 25.57 years; 24 right-handed) participated in the study (experiment 1 and experiment 2, within participants). A power analysis (G*Power, Faul et al., 2009) showed that at least 14 and 26 subjects were needed to detect effect sizes (i.e., Cohen's d_{av}, calculated as $M_{diff} / ((SD_1 + SD_2) / 2)$) of 0.18 and 0.20 respectively, with a power (1-β) of 0.8 and $\alpha = 0.05$. These effect sizes were estimated in a previous study that we conducted with the same experimental design as experiment 1

(Zuanazzi & Noppeney, 2018) and correspond to the simple main effects resulting from the significant interaction between spatial attention and expectation.

All participants had normal or corrected to normal vision and reported normal hearing. All participants provided written informed consent and were naïve to the aim of the study. The study was approved by the local ethics committee of the University of Birmingham (Science, Technology, Mathematics and Engineering (STEM) Ethical Review Committee) and the experiment was conducted in accordance with these guidelines and regulations.

Stimuli and Apparatus

Spatial auditory stimuli of 100 ms duration were created by convolving bursts of white noise (with 5 ms onset and offset ramps) with spatially selective head-related transfer functions (HRTFs) based on the KEMAR dummy head of the MIT Media Lab (http://sound.media.mit.edu/resources/KEMAR.html). Visual stimuli ('flashes') were white discs (radius: 0.88º visual angle, luminance: 196 cd/m2) of 100 ms duration presented on a grey background. Both auditory and visual stimuli were presented at ±10º of horizontal visual angle along the azimuth (0º of vertical visual angle). Throughout the entire experiment, a fixation cross was presented in the center of the screen.

Prior to the beginning of the study, participants were tested for their ability to discriminate left and right auditory stimuli on a brief series of 20 trials. They indicated their spatial discrimination response (i.e., 'left' vs 'right') via a two-choice key press (group mean accuracy was 99% ± 0.4% [across subjects mean ± SEM]).

During the experiment, participants rested their chin on a chinrest with the height held constant across all the participants. Auditory stimuli were presented at approximately 72 dB SPL, via HD 280 PRO headphones (Sennheiser, Germany). Visual stimuli were displayed on a gamma-corrected LCD monitor (2560 x 1600 pixels resolution, 60 Hz refresh rate, 30" Dell

UltraSharp U3014, USA), at a viewing distance of approximately 50 cm from the participant's eyes. Stimuli were presented using Psychtoolbox version 3 (Brainard, 1997; www.psychtoolbox.org), running under Matlab R2014a (Mathworks Inc., Natick, MA, USA) on a Windows machine. Participants' responses were recorded via one key of a small keypad (Targus, USA). Throughout the study, participants' eye-movements and fixation were monitored using Tobii Eyex eyetracking system (Tobii EyeX, Tobii, Sweden, ~60 Hz sampling rate).

Study overview: rationale and analysis strategy

This study included two experiments. Each experiment conforms to a four-factorial design. Because the two experiments were performed within the same participants the study as a whole could also be treated as a five factorial within-subject experiment. However, because (1) experiment 1 and 2 were completed on different days, (2) experiment 1 was a replication of our previous study and (3) the understanding of results of five factorial designs is rather complex, we will initially analyse each of the two experiments separately.

The separate analyses of experiments 1 and 2 allow us to address our first question, i.e., whether spatial attention and expectations rely on modality-specific or at least partially shared mechanisms. While experiment 1 is intended to replicate our findings reported in our previous research, experiment 2 is intended to extend them and demonstrate that this pattern of results does not depend on whether audition or vision is used as a primary manipulation modality. Moreover, showing the same profile across the two experiments also resolves the ambiguity of our previous research, in which a smaller spatial expectation effect for the secondary (i.e. visual) modality in experiment 1 could potentially be explained by differences in sensory modality rather than attenuated cross-sensory generalization.

In a second step we will directly combine data from experiment 1 and 2 to address our second question, i.e. whether the spatial expectation effect generalizes differently from audition to vision than from vision to audition. To address this question, we need to compare the expectation effects for auditory and visual stimuli in the attended hemifield between the two experiments (i.e. this question cannot be addressed by any of the two experiments alone).

Please also note that the Design and Procedure section mostly overlaps with that of our previous paper (Zuanazzi & Noppeney, 2018) to enable the reader to quickly compare our different studies and obtain a convergent picture across all results.

Design and Procedure

Experiment 1 investigated the effect of auditory spatial attention and expectation on detection of auditory (i.e., primary modality) and visual (i.e., secondary modality) targets using a 2 (auditory spatial attention: left vs right hemifield) x 2 (auditory spatial expectation: left vs right hemifield) x 2 (stimulus modality: auditory vs visual) x 2 (stimulus location: left vs right hemifield) factorial design. Hence, experiment 1 manipulated spatial attention and expectation selectively in audition and assessed their direct effects on auditory stimulus processing and indirect generalization to visual stimuli. In experiment 2 primary and secondary modality were reversed (i.e., primary modality: vision; secondary modality: audition); design and procedural details were otherwise comparable to experiment 1. For the data analysis we pooled over stimulus locations (left/right) leading to a 2 (auditory spatial attention: attended vs unattended) x 2 (auditory spatial expectation: expected vs unexpected) x 2 (stimulus modality: auditory vs visual) factorial design.

Spatial attention was manipulated for the primary modality as task-relevance, i.e., the requirement to respond to an auditory (experiment 1) or a visual (experiment 2) target in the

left vs right hemifield. Prior to each run a cue (duration: 2000 ms) informed the observer whether to respond to targets in either their left or right hemifield.

Spatial expectation was manipulated as spatial signal probability for signals in the primary modality across experimental sessions that were performed on different days. Auditory (i.e., primary modality in experiment 1) or visual (i.e., primary modality in experiment 2) signals were presented with a ratio of 2.33/1 (i.e., 70%/30%) in the expected/unexpected hemifield. Observers were not informed about those probabilities but learnt them implicitly. Importantly, spatial attention and expectation were not directly manipulated in the secondary modality, allowing us to assess their cross-sensory generalization. As a result, participants needed to respond to all visual targets that were presented with equal probabilities in their spatial hemifields in experiment 1 (i.e., ratio 1/1 in the expected/unexpected hemifields) (Fig. 1A and 1C). Likewise, they had to respond to all auditory targets that were presented with equal probabilities in experiment 2.

Each experiment included two sessions (i.e., spatial expectation left vs right on different days). Hence, subjects participated in the two experiments on four days: 2 sessions for experiment 1 and 2 sessions for experiment 2 = 4 sessions in total for each participant. Each session included 12 attention runs. Runs were of two types: in run type A (Fig. 1A, 1C and 1D) spatial attention and expectation were congruent (i.e., spatial attention was directed to the hemifield with higher stimulus frequency); in run type B spatial attention and expectation were incongruent (i.e., spatial attention was directed to the hemifield with less frequent stimuli). The overall probability to respond (i.e., response probability) was greater when attention and expectation were congruent and directed to the same hemifield (85%, runs of type A) than when they were directed to different hemifields (65%, runs of type B) (Fig. 1D).

The order of experiments 1 vs 2 and of expectation sessions (i.e., left vs right) was counterbalanced across participants; the order of attention runs (i.e., left vs right) was

241 counterbalanced within and across participants and the order of stimulus locations (i.e., left vs
242 right) and stimulus modalities (sound vs flash) was pseudo-randomized within each
243 participant. Brief breaks were included after every two attention runs to provide feedback to
244 participants about their performance accuracy (averaged across all conditions) in the target
245 detection task and about their eye-movements (i.e., fixation maintenance).
246 Overall, each experiment included 80 trials x 12 attention runs (6 runs of type A and 6 runs of
247 type B, duration: 3 mins/run) x 2 expectation sessions = 1920 trials in total (and 3840 for the
248 whole study). Specifically, each run type included i. 336 targets presented in the expected
249 hemifield (pooled over left and right) and 144 targets in the unexpected hemifield (pooled
250 over left and right) for the primary modality and ii. 240 targets presented in the expected
251 hemifield and 240 targets in the unexpected hemifield (pooled over left and right) for the
252 secondary modality. For further details see Fig. 1C which shows the absolute number of trials
253 for each condition and run type.
254 Each trial (SOA: 2300 ms) included three time windows: i. fixation cross alone (700 ms
255 duration), ii. brief sound or flash (stimulus duration: 100 ms) and iii. fixation cross alone
256 (1500 ms as response interval, see Fig. 1B). Participants responded to the targets in the
257 primary modality presented in the attended hemifield and to all targets in the secondary
258 modality irrespective of hemifield via key press, with their index finger (i.e., the same
259 response for all auditory and visual stimuli) as fast and accurately as possible (Fig. 1B).
260 Prior to each session, participants were familiarized with the stimuli in brief practice runs
261 (with equal spatial signal probability) and trained on target detection performance and
262 fixation (i.e., a warning signal was shown when the disparity between the central fixation
263 cross and the eye-data samples exceeded 2.5 degrees).
264 After the final session (i.e., experiment 1 for 14 participants and experiment 2 for the other 14
265 participants), participants indicated in a questionnaire whether they thought the stimuli in the

primary modality were presented more frequently in one of the two spatial hemifields. Ten out of 14 participants in experiment 1 and 11 out of 14 participants in experiment 2 correctly identified the expectation manipulation in the primary modality. Moreover, 13 out of 14 participants in experiment 1 and 13 out of 14 participants in experiment 2 correctly reported that stimuli in the secondary modality were presented with equal probabilities across the two hemifields. These data suggest that the majority of participants were aware of the manipulation of signal probability at least at the end of the fourth session.

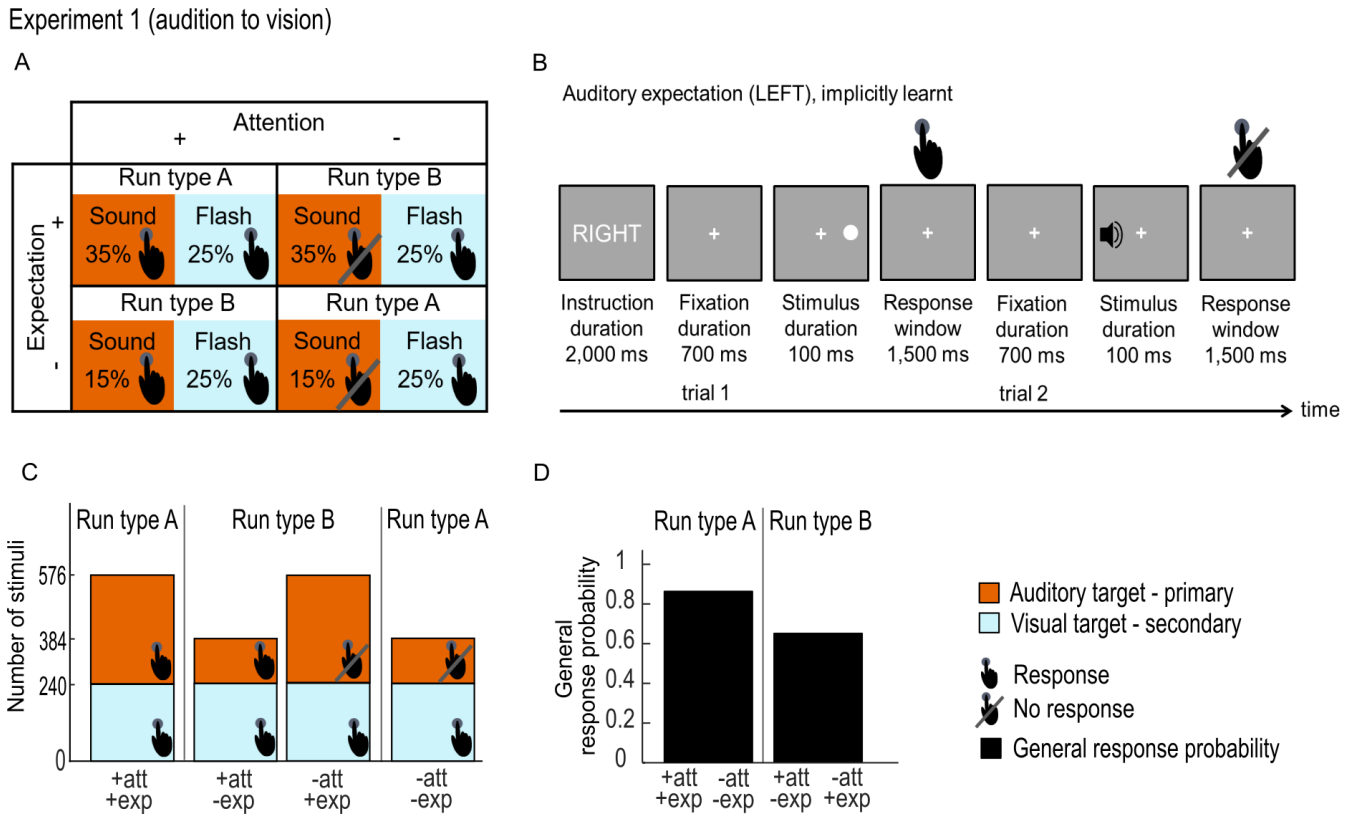

Figure 1: Design and example trials of experiment 1 (audition to vision)

A. Experiment 1: auditory spatial attention and expectation (i.e., signal probability) were manipulated in a 2 (auditory modality – dark orange, vs visual modality – light blue) x 2 (attended hemifield vs unattended hemifield) x 2 (expected hemifield vs unexpected hemifield) factorial design. For illustration purposes, stimulus locations (left/right) were collapsed. Presence vs absence of response requirement is indicated by the hand symbol, spatial signal probability manipulation is indicated by the %. **B.** Experiment 1: example of two trials in a session where auditory stimuli were presented with a probability of 0.7 in the left hemifield and 0.3 in the right hemifield. At the beginning of each run (i.e., 80 trials), a cue informed participants whether to attend and respond to auditory signals selectively in their left or right hemifield throughout the entire run. On each trial participants were presented with an auditory or visual stimulus (100 ms duration) either in their left or right hemifield. They were instructed to respond to auditory stimuli only in the attended hemifield

and to all visual stimuli irrespective of the hemifield as fast and accurately as possible with their index finger. The response window was limited to 1500 ms. Participants were not explicitly informed that auditory signals were more likely to appear in one of the two hemifields. Instead, spatial expectation was implicitly learnt within a session (i.e., day).

C. Experiment 1: number of auditory (dark orange) and visual (light blue) trials in the 2 (attended vs unattended hemifield) x 2 (expected vs unexpected hemifield) design (pooling over left/right stimulus location). Presence vs absence of response requirement is indicated by the hand symbol. The fraction of the area indicated by the 'Response' hand symbol pooled over the two bars of one particular run type (e.g., run type A) represents the response related expectation (i.e., general response probability: the overall probability that a response is required on a particular trial); general response probability is greater for run type A (85%), where attention and expectation are congruent, than for run type B (65%), where attention and expectation are incongruent, as indicated in **D**.

Note. Design and procedure of experiment 2 were comparable to that of experiment 1, with the only difference that vision was the primary modality and audition was the secondary modality. In other words, in experiment 2 attention and expectation were manipulated selectively in vision.

305 **Data analysis**

306 **Eye movement: exclusion criteria**

307 We excluded trials where participants did not successfully fixate the central cross based on a
308 dispersion criterion (i.e., distance of fixation from subject's center of fixation, as defined in
309 calibration trials, > 1.3 degrees for three subsequent samples, Blignaut, 2009). Our
310 eyetracking data confirmed that participants successfully maintained fixation in both
311 experiments with only a small number of trials to be excluded (experiment 1: excluded
312 auditory response trials 1.8% ± 0.5% [across subjects mean ± SEM]; excluded visual
313 response trials 1.7% ± 0.5% [across subjects mean ± SEM]; experiment 2: excluded visual
314 response trials 2.7% ± 1% [across subjects mean ± SEM]; excluded auditory response trials
315 2.7% ± 0.9% [across subjects mean ± SEM]).

316

317 **Response time analysis - separately for experiments 1 and 2**

318 We initially analysed response times separately for primary and secondary modalities and
319 independently for experiments 1 and 2.

320 Response time (RT) analysis was limited to trials with RT within the 1500 ms response
321 window and was performed after pooling over stimulus location (left/right).

322 For the primary modality (i.e., experiment 1 = audition, experiment 2 = vision), subject-
323 specific median RT were entered into a two-sided paired-sample t-test with spatial
324 expectation (expected vs unexpected stimulus) as factor (observers did not respond to targets
325 in the primary modality in the 'unattended' hemifield). For the secondary modality (i.e.,
326 experiment 1 = vision, experiment 2 = audition), subject-specific median response time were
327 entered into a 2 (spatial attention: attended vs unattended stimulus) x 2 (spatial expectation:
328 expected vs unexpected stimulus) repeated measures analysis of variance (ANOVA).

For both experiments 1 and 2, the mean hit rates were very high (> 99% in all conditions, Table 1), indicating that participants accurately performed the detection task. Because of the absence of a substantial number of misses, hit rates were not further analyzed.

Response time analysis - combined for experiments 1 and 2

Next, we investigated whether the effect of spatial expectation (i.e., expected vs unexpected) in the attended hemifield (no response was required for stimuli in the primary modalities presented in the unattended hemifield) depended on i. whether targets were presented in the primary or secondary modalities (i.e., the extent to which spatial expectations generalize across the senses) and ii. the multisensory generalization direction, from audition to vision and from vision to audition (i.e., whether spatial expectations generalize differently depending on whether audition or vision is the primary modality). Hence, we first computed the difference in median RT (ΔRT_{Exp}) between unexpected and expected stimuli presented in the attended hemifield (which corresponds to ΔRT_{Exp} between attended stimuli in run type B and run type A) for targets in each experiment, yielding four conditions: i. auditory targets as primary modality (experiment 1), ii. visual targets as secondary modality (experiment 1), iii. visual targets as primary modality (experiment 2), iv. auditory targets as secondary modality (experiment 2). ΔRT_{Exp} were entered into a 2 (multisensory generalization direction: audition to vision vs vision to audition) x 2 (manipulation: primary/direct vs secondary/indirect modality) repeated measures ANOVA. Please note that the interaction then reflects the difference between targets in the auditory and visual modality.

Time course of response times - combined for experiment 1 and 2

Finally, we assessed how these effects of multisensory generalization direction and direct/indirect manipulation evolved over time. For this, we computed the difference in

354 median RT (ΔRT_{Exp}) between unexpected and expected stimuli, as in the previous analysis,
355 but now separately for the first and second half of the experiment (i.e., one half = 430 trials).
356 These ΔRT_{Exp} (for each half) were entered in a 2 (multisensory generalization direction:
357 audition to vision vs vision to audition) x 2 (manipulation: primary/direct vs
358 secondary/indirect modality) x 2 (time: first vs second half of the experiment) repeated
359 measures ANOVA.
360 Unless otherwise indicated, we only report effects that are significant at $p < 0.05$
361 (Greenhouse-Geisser corrected where appropriate).
362
363
364 **Results**
365
366 **Generalization of attention and expectation effects across modalities – separately for**
367 **experiment 1 and 2**
368 In experiment 1, participants responded to auditory targets presented in their attended
369 hemifield and to all visual targets. In experiment 2, participants responded to visual targets
370 presented in their attended hemifield and to all auditory targets.
371 We observed qualitatively similar effects across experiments 1 and 2. Table 1 shows RT
372 (across participants' mean ± SEM) for targets in the auditory and visual modalities for the
373 two experiments.
374 For the primary modality, the two-sided paired-sample t-tests showed significantly faster RT
375 in the attended hemifield, when this hemifield was expected than unexpected (experiment 1,
376 auditory modality: $t(27) = -2.83$, $p = 0.009$, Cohen's d_{av} [95% CI] = -0.16 [-0.27 , -0.04];
377 experiment 2, visual modality: $t(27) = -9.62$, $p < 0.001$, Cohen's d_{av} [95% CI] = -0.35 [-0.47 ,
378 -0.23]) (Table 1, Fig. 2A and 2B).

379 For the secondary modality, the 2 (attended vs unattended) x 2 (expected vs unexpected)
380 repeated measures ANOVAs revealed a significant main effect of attention (experiment 1,
381 visual modality: $F(1, 27) = 72.08$, $p < 0.001$, η_p^2 [90% CI] = 0.73 [0.55, 0.80]; experiment 2,
382 auditory modality: $F(1, 27) = 36.91$, $p < 0.001$, η_p^2 [90% CI] = 0.58 [0.34, 0.70]). Results
383 showed that participants responded faster to targets presented in their attended than
384 unattended hemifields.
385 Moreover, a significant crossover interaction between attention and expectation was observed
386 (experiment 1, visual modality: $F(1, 27) = 10.09$, $p = 0.004$, η_p^2 [90% CI] = 0.27 [0.06, 0.46];
387 experiment 2, auditory modality: $F(1, 27) = 44.10$, $p < 0.001$, η_p^2 [90% CI] = 0.62 [0.40,
388 0.73]) (Table 1, Fig. 2A and 2B). The simple main effects showed that participants responded
389 significantly faster to targets in their attended hemifield, when this hemifield was expected
390 than unexpected (experiment 1, visual modality: $t(27) = -2.81$, $p = 0.009$, Cohen's d_{av} [95%
391 CI] = -0.08 [-0.15 , -0.02]; experiment 2, auditory modality: $t(27) = -5.96$, $p < 0.001$, Cohen's
392 d_{av} [95% CI] = -0.14 [-0.19 , -0.08]). This significant simple main effect demonstrates that
393 the effects of spatial attention and expectation generalized from primary to secondary
394 modalities, where neither attention not expectation were explicitly manipulated. By contrast,
395 participants responded significantly more slowly to targets in the secondary modality in the
396 unattended hemifield, when this hemifield was expected than unexpected (experiment 1,
397 visual modality: $t(27) = 2.56$, $p = 0.016$, Cohen's d_{av} [95% CI] = 0.09 [0.02, 0.17];
398 experiment 2, auditory modality: $t(27) = 5.53$, $p < 0.001$, Cohen's d_{av} [95% CI] = 0.18 [0.10 ,
399 0.26]) (Table 1, Fig. 2A and 2B). We suggest that simple main effects for expectation show
400 opposite directions because of response inhibition. In the attended hemifield observers need
401 to respond to the stimuli in the primary modality. Hence, if stimuli from the primary modality
402 are frequent (i.e., expected) in the attended hemifield, observers need to respond on a large
403 percentage of trials. By contrast, in the unattended hemifield observers should not respond to

404 the stimuli in the primary modality. Hence, if stimuli in the primary modality are frequent
405 (i.e., expected) in the unattended hemifield, observers need to inhibit their response on a large
406 percentage of trials.
407
408 **Table 1**: Group mean hit rates and reaction times (RT) and for each stimulus modality in
409 each condition that required a response for experiment 1 (primary modality: audition,
410 secondary modality: vision) and experiment 2 (primary modality: vision, secondary modality:
411 audition). In experiment 1, participants responded only to attended auditory targets (and to all
412 visual targets); in experiment 2 participants responded only to attended visual targets (and to
413 all auditory targets). Standard errors (SEM) are given in parentheses.
414

	Auditory modality				**Visual modality**			
Experiment 1	+att +exp	+att -exp	-att +exp	-att -exp	+att +exp	+att -exp	-att +exp	-att -exp
Hit rate (%) (SEM)	99.5 (0.19)	99.5 (0.15)	/	/	99.7 (0.07)	99.7 (0.07)	99.3 (0.16)	99.4 (0.13)
RT (ms) (SEM)	599.7 (20.1)	616.2 (19.2)	/	/	514.4 (16.9)	522.1 (16.9)	583.8 (18.5)	574.4 (18.9)
Experiment 2	+att +exp	+att -exp	-att +exp	-att -exp	+att +exp	+att -exp	-att +exp	-att -exp
Hit rate (%) (SEM)	99.8 (0.08)	99.9 (0.03)	99.2 (0.22)	99.4 (0.18)	99.7 (0.09)	99.7 (0.09)	/	/
RT (ms) (SEM)	527.4 (20.2)	542.1 (20)	605.2 (25.7)	581.3 (24.2)	526.3 (17.9)	561.8 (19.9)	/	/

415
416

The effects of spatial expectation – combined for experiment 1 and 2

In the previous analysis we showed that the effects of expectation on response times generalized from the primary to the secondary modality. Next, we directly compared the effects of spatial expectation across the two experiments, in which either audition was the primary and vision the secondary modality or vice versa. The 2 (multisensory generalization direction: audition to vision vs vision to audition) x 2 (manipulation: primary/direct vs secondary/indirect modality) repeated measures ANOVA revealed a significant main effect of manipulation ($F(1, 27) = 18.03$, $p < .001$, η_p^2 [90% CI] =0.40 [0.16, 0.56]), showing overall greater expectation effects for primary than secondary modalities (dark vs light bars in Fig. 2C). Moreover, we observed a significant main effect of crossmodal generalization direction ($F(1, 27) = 20.71$, $p < .001$, η_p^2 [90% CI] = 0 .43, [0.19, 0.59]), with a greater expectation effect (i.e., greater ΔRT_{Exp}) for vision to audition than for audition to vision (experiment 2 vs experiment 1 in Fig. 2C). Critically, this generalization effect may be greater from vision to audition than vice versa, because observers learn signal probabilities and hence form spatial expectations faster when vision is the primary modality (dark blue bar in Fig. 2C). Alternatively, the expectation effect generalizes more effectively from vision to audition than vice versa (light orange bar in Fig. 2C).

Time course of the effects of spatial expectation - combined for experiment 1 and 2

To disentangle between these two possibilities, we investigated how signal probability is learnt over time when audition (experiment 1) or vision (experiment 2) are the primary modality by repeating the previous analysis with the additional factor of time (i.e., first vs second half of experiment). The 2 (multisensory generalization direction: audition to vision vs vision to audition) x 2 (manipulation: primary/direct vs secondary/indirect modality) x 2 (time: first vs second half of the experiment) repeated measures ANOVA performed on

ΔRT_{Exp} revealed significant main effects of multisensory generalization direction ($F(1, 27) = 22.70$, $p < 0.001$, η_p^2 [90% CI] = 0.46 [0.21, 0.61]), manipulation ($F(1, 27) = 13.77$, $p < 0.001$, η_p^2 [90% CI] = 0.34 [0.10, 0.51]) as well as a significant interaction between multisensory generalization direction x manipulation x time ($F(1, 27) = 11.38$, $p = 0.002$, η_p^2 [90% CI] = 0.29 [0.07, 0.48]).

We unpacked the 3-way ANOVA into two 2-ways ANOVAs for further analysis, one for each half of the experiment, with factors multisensory generalization direction (audition to vision vs vision to audition) and manipulation (primary/direct vs secondary/indirect modality).

For the first half of the experiment, our results revealed a significant main effect of multisensory generalization direction ($F(1, 27) = 8.34$, $p = 0.008$, η_p^2 [90% CI] = 0.24 [0.04, 0.42]), manipulation ($F(1, 27) = 28.84$, $p < 0.001$, η_p^2 [90% CI] = 0.52 [0.27, 0.65]) and a significant interaction between multisensory generalization direction and manipulation ($F(1, 27) = 12.83$, $p = 0.001$, η_p^2 [90% CI] = 0.32 [0.09, 0.50]) (left bar plot in Fig. 2D). Post-hoc comparisons indicated that ΔRT_{Exp} in the primary modality of experiment 2 (i.e., vision) were significantly greater than ΔRT_{Exp} in the primary modality of experiment 1 (i.e., audition) ($t(27) = 5.72$, $p < 0.001$, Cohen's d_{av} [95% CI] = 1.05, [0.59, 1.50], dark blue vs dark orange bars in the left bar plot of Fig. 2D), and greater than the effects of expectation in the secondary modality of experiment 2 (i.e., audition) ($t(27) = 4.97$, $p < 0.001$, Cohen's d_{av} [95% CI] = 1.03, [0.53, 1.51], dark blue vs light orange bars in the left bar plot of Fig. 2D). Moreover, the effects of expectation in the secondary modality of experiment 2 (i.e., audition) were significantly greater than those in the secondary modality of experiment 1 (i.e., vision) ($t(27) = 2.14$, $p = 0.042$, Cohen's d_{av} [95% CI] = 0.48, [0.02, 0.94], light orange vs light blue bars in the left bar plot of Fig. 2D).

467 For the second half of the experiment, our results only revealed a significant main effect of
468 manipulation ($F(1, 27) = 11.41$, $p = 0.002$, η_p^2 [90% CI] = 0.30 [0.07, 0.48]) (right bar plot in
469 Fig. 2D) but no significant main effect of multisensory generalization direction ($p = 0.20$) or
470 significant interaction between multisensory generalization direction and manipulation ($p =$
471 0.65) was found.
472
473 To summarize, these results show that: (1) an effect of generalization direction (i.e., audition
474 to vision vs vision to audition) was found only for the first half of the experiment. Here, the
475 effects of expectation generalized crossmodally in an attenuated fashion only from vision to
476 audition in experiment 2 (i.e., primary modality: vision) but no difference between audition
477 and vision was found in experiment 1 (i.e., primary modality: audition). By contrast, in the
478 second half of the experiment we did not observe an effect of multisensory generalization
479 direction. Instead, the effects of expectation generalized from the primary to the secondary
480 modality in an attenuated fashion similarly when vision or audition were the primary
481 modality. As shown in figure 2D, this difference between first and second halves can be
482 explained by the fact that observers form spatial expectations (i.e., learn signal probability
483 over space) more slowly in audition than vision. Yet, once expectations are learnt in audition,
484 the crossmodal generalization is comparable for audition and vision. In other words, the
485 effect of generalization direction that we observed in our analysis that did not yet account for
486 learning effects (i.e., our second analysis) can be explained away by the speed with which
487 observers learn signal probabilities and form spatial expectations in their primary modality.
488 Signal probabilities are learnt faster in vision than audition, but once spatial expectations are
489 formed, they generalize similarly from vision to audition and vice versa.
490

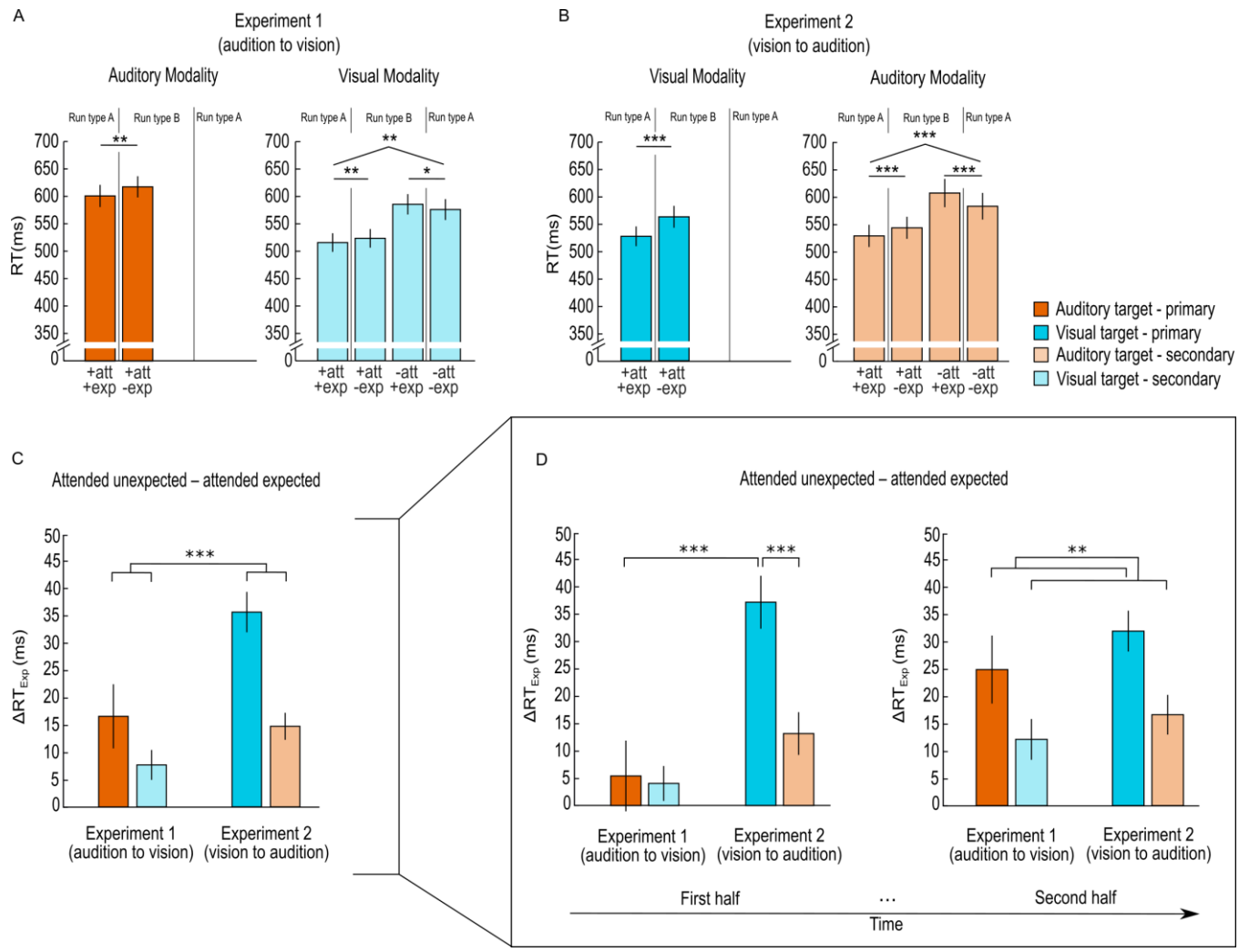

Figure 2: Behavioural results of experiment 1 and 2.

Bar plots represent across subjects' mean (±SEM) RT for each of the six conditions with response requirements for experiment 1 (primary modality: audition; secondary modality: vision, **A**) and 2 (primary modality: vision; secondary modality: audition, **B**), pooling over left/right stimulus location. Overall slower RT are observed for runs type B than runs type A, which reflects differences in general response probability (see Fig. 1D) **C.** Bar plots represent across subjects' mean (±SEM) ΔRT for effects of spatial expectation (attended unexpected – attended expected hemifield) in the primary (dark bars) and secondary modalities (light bars) for experiment 1 and 2. **D.** Effects of response probability (attended unexpected – attended

expected hemifield) over time (i.e., first and second half) for audition and vision as primary (dark bars) or secondary (light bars) modality. Brackets and stars indicate significance of main effects and interactions. *$p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. Audition: orange; vision: blue.

507 **Discussion**

508 The current study investigated how observers allocate attention and form expectations (by
509 learning signal probabilities) over space across audition and vision. We orthogonally
510 manipulated spatial attention as response requirement and expectation as stimulus probability
511 over space selectively in the primary modality and assessed their effects on behavioral
512 responses to targets presented in the primary and secondary modalities. Across two
513 experiments we alternated the assignment of vision and audition to primary or secondary
514 modality. This allowed us to compare behavioral effects of spatial attention and expectation
515 in audition and vision and their crossmodal generalization.

516 Regardless of sensory modality we observed a significant main effect of spatial attention for
517 targets in the secondary modality, in which attention was not directly manipulated. Auditory
518 spatial attention partially generalized to the visual modality and vice versa. These findings
519 converge with a large body of behavioral and neuroimaging work suggesting that attentional
520 resources are allocated interactively across the senses (Spence & Driver, 1996; Eimer &
521 Schröger, 1998; Eimer, 1999; Macaluso et al., 2002; Santangelo et al., 2009; Zuanazzi &
522 Noppeney, 2019).

523 Likewise, in the attended hemifield observers were faster at target detection in the primary
524 and secondary modalities when the hemifield was expected than unexpected (i.e., high > low
525 signal probability). Again, this response facilitation for expected (relative to unexpected)
526 spatial locations were observed irrespective of whether vision or audition served as primary
527 modality. Yet, in the unattended hemifield, in which we could assess effects of expectation
528 only for the secondary modality, we observed the opposite pattern, i.e., observers were slower
529 at target detection for expected than unexpected hemifields. Combining these two results, we
530 observed a significant interaction between spatial attention and expectation, for both vision
531 and audition as secondary modalities. In a previous study (Zuanazzi & Noppeney, 2018) we

532 argued that this interaction profile between spatial attention and expectation is explained by
533 attention and expectation jointly co-determining general response probability (i.e., the
534 probability to respond regardless of the hemifield in which the signal is presented). More
535 specifically, in runs in which attention and expectation are directed to the same hemifield
536 (runs of type A, Fig. 1A, 1C and 1D), participants have to respond to 85% of trials of the
537 entire run, but only to 65% trials in runs of type B in which attention and expectation are
538 directed to different hemifields (i.e., general response probability, Fig. 1A, 1C and 1D).
539 Hence, faster response times may result from an increase in alertness, arousal or motor
540 preparation that is needed to respond on a large proportion of trials (i.e., attended/expected
541 and unattended/unexpected conditions, run type A, Fig. 2A and 2B) (Mars et al., 2007;
542 Bestmann et al., 2008). By contrast, in runs in which attention and expectation are directed to
543 different hemifields (runs of type B, Fig. 1A, 1C and 1D), observers need to inhibit their
544 response to the frequent stimuli of the primary modality in the expected hemifield and hence
545 respond more slowly to targets in the secondary modality.
546 Critically, response probability does not depend on whether the signal is auditory or visual,
547 but it is calculated as the probability that any signal is responded to. If the expectation effects
548 result purely from amodal mechanisms (e.g., general alertness, arousal, motor preparation
549 etc.) associated with changes in response probability, the expectation effects in the attended
550 hemifield should be equal for primary and secondary modalities. By contrast, if expectations
551 (i.e., auditory or visual signal probability) are formed at least partially in a modality-specific
552 fashion, we should observe expectation effects (i.e., ΔRT_{Exp}) that are greater for the primary
553 modality (where expectation was explicitly manipulated) than for the secondary modality
554 (i.e., an attenuated crossmodal generalization).
555 To arbitrate between these two hypotheses, we analyzed ΔRT_{Exp} in a 2 (multisensory
556 generalization direction) x 2 (manipulation) repeated measures ANOVA. This analysis

557 revealed a significant main effect of 'manipulation', i.e., whether the stimulus was presented
558 in the primary modality (in which expectation was explicitly manipulated) or in the
559 secondary modality. Consistent with additional modality-specific mechanisms of expectation,
560 observers showed a greater expectation effect for targets in the primary than the secondary
561 modality. These results strongly suggest that implicitly learned spatial expectations modulate
562 perceptual decision making via both modality-specific and amodal response mechanisms.
563 This duality of modality-specific and amodal mechanisms converge with recent
564 neuroimaging findings which showed effects of expectation selective for auditory stimuli as
565 primary modality in auditory cortices and higher-order frontoparietal systems (Zuanazzi &
566 Noppeney, 2019). Potentially, the activation increases in auditory cortices may reflect
567 prediction error signals based on modality-specific expectations (Friston, 2005), while higher
568 frontoparietal systems may be associated with additional response-related processes.

569 Surprisingly, this repeated measures ANOVA also revealed a main effect of
570 multisensory generalization direction with greater effects in experiment 2 (i.e., vision to
571 audition) than experiment 1 (i.e., audition to vision). Our time course analysis showed that
572 this difference between experiments does not reflect genuine differences in the effectiveness
573 with which spatial expectations are implicitly learnt and generalize from audition to vision
574 and vice versa, but reflect differences in the speed with which spatial expectations are learnt
575 in audition and vision. In the second half of the experiment in which the expectation effect in
576 the primary modality is comparable between auditory (experiment 1) and visual (experiment
577 2) targets, we no longer observe a significant effect of multisensory generalization direction.
578 In other words, observers are slower at learning spatial expectations (i.e., signal probability)
579 in audition than in vision. But, once they have formed spatial expectations of comparable
580 precision for audition and vision, they also generalize similarly from audition to vision and
581 vice versa.

582 The difference in perceptual learning rates between vision and audition may result
583 from how the brain forms spatial representations in vision and audition (Neumann et al.,
584 1986). While visual cortices are retinotopically organized and hence directly represent spatial
585 location in a place code (i.e., based on space, Sereno et al., 1995; Maier & Groh, 2009),
586 primary auditory cortices are tonotopically organized (i.e., based on frequency, Lauter et al.,
587 1985, Middlebrooks &, 1991; Maier & Groh, 2009). In audition, spatial locations are
588 computed only indirectly from binaural amplitude and latency differences and from
589 monoaural filtering cues. Moreover, visual objects tend to be more permanent across time,
590 whereas source sounds are often transient and dynamic (Neisser, 1976; Neumann et al.,
591 1986). Most importantly, in everyday life vision provides typically (i.e., under optimal
592 lighting conditions) more reliable spatial information than audition (Dacey et al., 1992;
593 Knudsen & Brainard, 1995; Stephen et al., 2002; Talsma et al., 2008; Mengotti et al., 2018;
594 see also Molholm et al., 2007). In the current study, the high spatial reliability of the visual
595 stimulus (i.e., a white disc) may also have contributed to the shorter time for participants to
596 learn spatial signal probabilities and thus become aware of their manipulation in vision.
597 Critically, participants' awareness of such manipulation is evidenced by our questionnaires'
598 results. Alternatively, even in absence of explicit awareness, the distribution of events or
599 targets across space could have been learnt faster for more reliable visual than auditory
600 signals (Miller & Pachella, 1973; Jabar & Anderson, 2015). To better understand the fine-
601 grained temporal aspects of spatial expectation or signal probability learning across sensory
602 modalities, future neuroimaging studies will need to characterize the time course of spatial
603 learning across sensory systems.

604 In summary, our results suggest that the brain allocates spatial attention and forms
605 spatial expectation to some extent interactively across audition and vision (Eimer &
606 Schröger, 1998; Eimer, 1999; Macaluso, 2010; Zuanazzi & Noppeney, 2019). With respect to

spatial attention, our results corroborate previous research. With respect to spatial expectation, we show that they rely on modality-specific and amodal mechanisms. In support of modality-specific mechanisms we demonstrate that spatial expectations in the attended hemifield generalize from the primary to the secondary modality only in an attenuated fashion. In support of amodal response-related mechanisms, we demonstrate that, for both primary and secondary modalities, response times are closely related to the general response probability and associated processes of arousal, alertness and motor preparation. Critically, our learning analysis suggests that observers learn spatial probabilities more slowly in audition than vision, which may be related to their different spatial reliabilities. Once observers have formed comparable spatial expectations in audition, these generalize equally effectively from audition to vision as from vision to audition. In other words, our results demonstrate crossmodal interactions of perceptual learning (i.e., expectations building) in spatial perception but also show differences between sensory modalities in terms of the speed with which signal probabilities over space are learnt.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Funding

This research was funded by ERC-2012-StG_20111109 multsens

References

Aller, M., Giani, A., Conrad, V., Watanabe, M., & Noppeney, U. (2015). A spatially collocated sound thrusts a flash into awareness. *Frontiers in Integrative Neuroscience*, 9(February), 1–8. https://doi.org/10.3389/fnint.2015.00016

Aller, M., & Noppeney, U. (2019). To integrate or not to integrate: Temporal dynamics of hierarchical Bayesian causal inference. *PLoS biology* (Vol. 17). https://doi.org/10.1371/journal.pbio.3000210

Batson, M. A., Watanabe, T., Seitz, A. R., & Beer, A. L. (2011). Spatial shifts of audio-visual interactions by perceptual learning are specific to the trained orientation and eye. *Seeing and Perceiving*, 24(6), 579–594. https://doi.org/10.1163/187847611X603738

Bestmann, S., Harrison, L. M., Blankenburg, F., Mars, R. B., Haggard, P., Friston, K. J., & Rothwell, J. C. (2008). Influence of uncertainty and surprise on human corticospinal excitability during preparation for action. *Current Biology*, 18(10), 775–780. http://doi.org/10.1016/j.cub.2008.04.051

Blignaut, P. (2009). Fixation identification: the optimum threshold for a dispersion algorithm. Attention, *Perception & Psychophysics*, 71(4), 881–895.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10(4), 433–436. http://doi.org/10.1163/156856897X00357

Bratzke, D., Seifried, T., & Ulrich, R. (2012). Perceptual learning in temporal discrimination: Asymmetric cross-modal transfer from audition to vision. *Experimental Brain Research*, 221(2), 205–210. https://doi.org/10.1007/s00221-012-3162-0

Bressler, S. L., Tang, W., Sylvester, C. M., Shulman, G. L., & Corbetta, M. (2008). Top-down control of human visual cortex by frontal and parietal cortex in anticipatory visual spatial attention. *The Journal of Neuroscience*, 28, 10056–10061. https://doi.org/10.1523/JNEUROSCI.1776-08.2008

Bueti, D., & Buonomano, D. V. (2014). Temporal perceptual learning. *Timing and Time Perception*, 2(3), 261–289. https://doi.org/10.1163/22134468-00002023

Bueti, D., Lasaponara, S., Cercignani, M., & Macaluso, E. (2012). Learning about time: Plastic changes and interindividual brain differences. *Neuron*, 75(4), 725–737. https://doi.org/10.1016/j.neuron.2012.07.019

Carrasco, M. (2011). Visual attention: The past 25 years. *Vision Research*, 51(13), 1484–1525. http://doi.org/10.1016/j.visres.2011.04.012

Crist, R. E., Kapadia, M. K., Westheimer, G., & Gilbert, C. D. (1997). Perceptual learning of spatial localization: Specificity for orientation, position, and context. *Journal of Neurophysiology*, 78(6), 2889–2894. https://doi.org/10.1152/jn.1997.78.6.2889

Dacey, D. M., & Petersen, M. R. (1992). Dendritic field size and morphology of midget and parasol ganglion cells of the human retina. *Proceedings of the National Academy of Sciences of the United States of America*, 89(20), 9666–9670. http://doi.org/10.1073/pnas.89.20.9666

Driver, J., & Spence, C. (1998). Attention and the crossmodal construction of space. *Trends in Cognitive Sciences*, 2(7), 254–262. http://doi.org/10.1016/S1364-6613(98)01188-7

Eimer, M. (1999). Can attention be directed to opposite locations in different modalities? An ERP study. *Clinical Neurophysiology*, 110(7), 1252–1259. https://doi.org/10.1016/S1388-2457(99)00052-8

Eimer, M., & Driver, J. (2001). Crossmodal links in endogenous and exogenous spatial attention: evidence from event-related brain potential studies. *Neuroscience and Biobehavioral Reviews*, 25(6), 497–511. http://doi.org/S0149-7634(01)00029-X [pii]

Eimer, M., & Schroger, E. (1998). ERP effects of intermodal attention and cross-modal links in spatial attention. *Psychophysiology*, 35(3), 313–27. https://doi.org/10.1017/S004857729897086X

Feldman, H., & Friston, K. J. (2010). Attention, uncertainty, and free-energy. *Frontiers in Human Neuroscience*, 4(December), 1–23. http://doi.org/10.3389/fnhum.2010.00215

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. *Behavior Research Methods*, 41(4), 1149–60. http://doi.org/10.3758/BRM.41.4.1149

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 360(1456), 815–836. https://doi.org/10.1098/rstb.2005.1622

Greene, A. J., Easton, R. D., & LaShell, L. S. R. (2001). Visual-auditory events: Cross-modal perceptual priming and recognition memory. *Consciousness and Cognition*, 10(3), 425–435. https://doi.org/10.1006/ccog.2001.0502

Jabar, S. B., & Anderson, B. (2015). Probability shapes perceptual precision: A study in orientation estimation. *Journal of Experimental Psychology: Human Perception and Performance*, 41(6), 1666–1679. http://doi.org/10.1037/xhp0000121

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. https://doi.org/10.1016/j.neurobiolaging.2019.08.017

Kim, R. S., Seitz, A. R., & Shams, L. (2008). Benefits of stimulus congruency for multisensory facilitation of visual learning. *PLoS ONE*, 3(1). https://doi.org/10.1371/journal.pone.0001532

Kincade, J. M., Abrams, R. A., Astafiev, S. V., Shulman, G. L., & Corbetta, M. (2005). An event-related functional magnetic resonance imaging study of voluntary and stimulus-driven orienting of attention. *The Journal of Neuroscience*, 25(18), 4593–4604. https://doi.org/10.1523/JNEUROSCI.0236-05.2005

Knudsen, E., & Brainard, M. S. (1995). Creating a unified representation of visual and auditory space in the brain. *Annual Review of Neuroscience*, 18(1), 19–43. http://doi.org/10.1146/annurev.neuro.18.1.19

Lapid, E., Ulrich, R., & Rammsayer, T. (2009). Perceptual learning in auditory temporal discrimination: No evidence for a cross-modal transfer to the visual modality. *Psychonomic Bulletin and Review*, 16(2), 382–389. https://doi.org/10.3758/PBR.16.2.382

Lauter, J. L., Herscovitch, P., Formby, C., & Raichle, M. E. (1985). Tonotopic organization in human auditory cortex revealed by positron emission tomography. *Hearing Res*, 20, 199–205.

Macaluso, E., Frith, C. D., & Driver, J. (2002). Supramodal effects of covert spatial orienting triggered by visual or tactile events. *Journal of Cognitive Neuroscience*, 14(3), 389–401. https://doi.org/10.1162/089892902317361912

Macaluso, E. (2010). Orienting of spatial attention and the interplay between the senses. *Cortex*, 46(3), 282–297. http://doi.org/10.1016/j.cortex.2009.05.010

Meegan, D. V, & Aslin, R. N. (2000). Motor timing learned. *Nature Neuroscience*, 3(9).

Maier, J. X., & Groh, J. M. (2009). Multisensory guidance of orienting behavior. *Hearing Research*, 258(1–2), 106–112. https://doi.org/10.1016/j.heares.2009.05.008

Mars, R. B., Bestmann, S., Rothwell, J. C., & Haggard, P. (2007). Effects of motor preparation and spatial attention on corticospinal excitability in a delayed-response paradigm. *Experimental Brain Research*, 182(1), 125–129. http://doi.org/10.1007/s00221-007-1055-4

Meijer, D., Veselič, S., Calafiore, C., & Noppeney, U. (2019). Integration of audiovisual spatial signals is not consistent with maximum likelihood estimation. *Cortex*. https://doi.org/10.1016/j.cortex.2019.03.026

Middlebrooks, J. C., & Green, D. M. (1991). Sound localization by human listeners. *Annu. Rev. Psychol.*, (42), 135–159.

Mengotti, P., Boers, F., Dombert, P. L., Fink, G. R., & Vossel, S. (2018). Integrating modality-specific expectancies for the deployment of spatial attention. *Scientific Reports*, 8(1), 1210. https://doi.org/10.1038/s41598-018-19593-7

Miller, J. O., & Pachella, R. G. (1973). Locus of the stimulus probability effect. *Journal of Experimental Psychology*, 101(2), 227–231. http://doi.org/10.1037/h0035214

Molholm, S., Martinez, A., Shpaner, M., & Foxe, J. J. (2007). Object-based attention is multisensory: Co-activation of an object's representations in ignored sensory modalities. *European Journal of Neuroscience*, 26(2), 499–509. https://doi.org/10.1111/j.1460-9568.2007.05668.x

Mondor, T. A., & Amirault, K. J. (1998). Effect of same- and different-modality spatial cues on auditory and visual target identification. *Journal of Experimental Psychology*, 24(3), 745–755.

Nagarajan, S. S., Blake, D. T., Wright, B. A., Byl, N., & Merzenich, M. M. (1998). Practice-related improvements in somatosensory interval discrimination are temporally specific but generalize across skin location, hemisphere, and modality. *Journal of Neuroscience*, 18(4), 1559–1570. https://doi.org/10.1523/jneurosci.18-04-01559.1998

Neisser, U. (1977). Cognition and reality. *The American Journal of Psychology*, 90(3), 541–543.

Neumann, O., Van der Heijden, A. H., & Allport, D. A. (1986). Visual selective attention: Introductory remarks. *Psychological Research*, 48(4), 185–188. http://doi.org/10.1007/bf00309082

Odegaard, B., Wozny, D. R., & Shams, L. (2015). Biases in visual, auditory, and audiovisual perception of space. *PLoS Computational Biology*, 11(12), 1–23. https://doi.org/10.1371/journal.pcbi.1004649

Posner, M. I. (1980). Orienting of attention. *Q J Exp Psychol*, 32(1), 3–25. http://doi.org/10.1080/00335558008248231

Rohe, T., & Noppeney, U. (2015a). Cortical hierarchies perform Bayesian Causal Inference in multisensory perception. *PLOS Biology*, 13(2), e1002073. https://doi.org/10.1371/journal.pbio.1002073

Rohe, T., & Noppeney, U. (2015b). Sensory reliability shapes perceptual inference via two mechanisms. *Journal of Vision*, 15, 1–38. https://doi.org/10.1167/15.5.22.doi

Rohe, T., & Noppeney, U. (2016). Distinct computational principles govern multisensory integration in primary sensory and association cortices. *Current Biology*, 26(4), 509–514. https://doi.org/10.1016/j.cub.2015.12.056

Rohe, T., & Noppeney, U. (2018). Reliability-weighted integration of audiovisual signals can be modulated by top-down control. *Eneuro*, 5, 1–20. http://doi.org/10.1523/ENEURO.0315-17.2018

Rohenkohl, G., Gould, I. C., Pessoa, J., & Nobre, A. C. (2014). Combining spatial and temporal expectations to improve visual perception. *Journal of Vision*, 14(4), 1–13. http://doi.org/10.1167/14.4.8

Santangelo, V., Olivetti Belardinelli, M., Spence, C., & Macaluso, E. (2009). Interaction between voluntary and stimulus-driven spatial attention mechanism across sensory modalities. *Journal of Cognitive Neuroscience*, 21(12), 2384–2397. https://doi.org/10.1162/jocn.2008.21178

Sereno, M. I., Dale, a M., Reppas, J. B., Kwong, K. K., Belliveau, J. W., Brady, T. J., … Tootell, R. B. H. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science*, 268(May), 889–893. http://doi.org/10.1126/science.7754376

Spence, C., & Driver, J. (1996). Audiovisual links in endogenous covert spatial attention. *Journal of Experimental Psychology: Human Perception and Performance*, 22(4), 1005–1030. http://doi.org/10.1037/0096-1523.22.4.1005

Spence, C., & Driver, J. (1997). Audiovisual links in exogenous covert spatial orienting. *Perception & Psychophysics*, 59(1), 1–22. http://doi.org/10.3758/BF03206843

Stephen, J. M., Aine, C. J., Christner, R. F., Ranken, D., Huang, M., & Best, E. (2002). Central versus peripheral visual field stimulation results in timing differences in dorsal stream sources as measured with MEG. *Vision Research*, 42(28), 3059–3074. http://doi.org/10.1016/S0042-6989(02)00415-7

Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences*, 13(9), 403–409. http://doi.org/10.1016/j.tics.2009.06.003

Talsma, D., Kok, A., Slagter, H. A., & Cipriani, G. (2008). Attentional orienting across the sensory modalities. *Brain and Cognition*, 66(1), 1–10. https://doi.org/10.1016/j.bandc.2007.04.005

van Ede, F., de Lange, F. P., & Maris, E. (2012). Attentional cues affect accuracy and reaction time via different cognitive and neural processes. *The Journal of Neuroscience*, 32(30), 10408–10412. https://doi.org/10.1523/JNEUROSCI.1337-12.2012

Wahn, B., & König, P. (2015). Audition and vision share spatial attentional resources, yet attentional load does not disrupt audiovisual integration. *Frontiers in Psychology*, 6(July), 1–12. http://doi.org/10.3389/fpsyg.2015.01084

Wahn, B., & König, P. (2017). Is attentional resource allocation across sensory modalities task-dependent? *Advances in Cognitive Psychology*, 13(1), 83–96.

Ward, L. M., McDonald, J. J., & Lin, D. (2000). On asymmetries in cross-modal spatial attention orienting. *Perception & Psychophysics*, 62(6), 1258–1264. http://doi.org/10.3758/BF03212127

Warm, J. S., Stutz, R. M., & Vassolo, P. A. (1975). Intermodal transfer in temporal discrimination. *Perception & Psychophysics*, 18(4), 281–286.

Zuanazzi, A., & Noppeney, U. (2018). Additive and interactive effects of spatial attention and expectation on perceptual decisions. *Scientific Reports*, 1–12. http://doi.org/10.1038/s41598-018-24703-6

Zuanazzi, A., & Noppeney, U. (2019). Distinct neural mechanisms of spatial attention and expectation guide perceptual inference in a multisensory world. *The Journal of Neuroscience*, 39(12), 2873–18. https://doi.org/10.1523/jneurosci.2873-18.2019.