



The effects of visual and auditory synchrony on human foraging

Ivan Makarov^{1,2} · Runar Unnthorsson² · Árni Kristjánsson¹ · Ian M. Thornton³

Accepted: 21 December 2023 / Published online: 22 January 2024
© The Psychonomic Society, Inc. 2024

Abstract

Can synchrony in stimulation guide attention and aid perceptual performance? Here, in a series of three experiments, we tested the influence of visual and auditory synchrony on attentional selection during a novel human foraging task. Human foraging tasks are a recent extension of the classic visual search paradigm in which multiple targets must be located on a given trial, making it possible to capture a wide range of performance metrics. Experiment 1 was performed online, where the task was to forage for 10 (out of 20) vertical lines among 60 randomly oriented distractor lines that changed color between yellow and blue at random intervals. The targets either changed colors in visual synchrony or not. In another condition, a non-spatial sound additionally occurred synchronously with the color change of the targets. Experiment 2 was run in the laboratory (within-subjects) with the same design. When the targets changed color in visual synchrony, foraging times were significantly shorter than when they randomly changed colors, but there was no additional benefit for the sound synchrony, in contrast to predictions from the so-called “pip-and-pop” effect (Van der Burg et al., *Journal of Experimental Psychology*, 1053–1065, 2008). In Experiment 3, task difficulty was increased as participants foraged for as many 45° rotated lines as possible among lines of different orientations within 10 s, with the same synchrony conditions as in Experiments 1 and 2. Again, there was a large benefit of visual synchrony but no additional benefit for sound synchronization. Our results provide strong evidence that visual synchronization can guide attention during multiple target foraging. This likely reflects the local grouping of the synchronized targets. Importantly, there was no additional benefit for sound synchrony, even when the foraging task was quite difficult (Experiment 3).

Keywords Foraging · Visual search · Synchrony · Audition · Cross-modal perception · Grouping · Perceptual organization · Pip-and-pop effect

Introduction

The ability to find what we’re looking for when we search a visual scene – whether an experimental display or a real-world environment – is known to be largely determined by attention. Unexpected or novel items may attract attention in a bottom-up fashion, making them easy to detect, so that they “pop-out.” In the absence of such saliency, search is

slower and more effortful, with attention being guided via top-down search templates, scene attributes and/or other factors, such as expected value or prior history (for review, see Kristjánsson & Egeth, 2020; Wolfe & Horowitz, 2017).

The purpose of the current work was to explore whether search for multiple items is altered when targets are linked via temporal synchrony, and if so, what aspects of attention are responsible for such modulation. A second question concerned whether the addition of a nonspatial auditory cue to synchrony would lead to performance benefits, as has been shown in previous single-target visual search studies (Van der Burg et al., 2008).

While classic visual search tasks – involving a single target among a variable set size of distractors – have taught us a great deal about the nature of attentional selection (Treisman & Gelade, 1980; Wolfe & Horowitz, 2017), there has recently been a growing interest in studying more complex scenarios that may better reflect the demands and dynamics

✉ Ivan Makarov
ivm3@hi.is

¹ Faculty of Psychology, School of Health Sciences, University of Iceland, Reykjavik, Iceland

² Faculty of Industrial Engineering, Mechanical Engineering and Computer Science, University of Iceland, Reykjavik, Iceland

³ Department of Cognitive Science Faculty of Media & Knowledge Science, University of Malta, Msida, Malta

of real-life scenes (for review, see Kristjánsson & Dra-schkow, 2021).

Inspired by the large literature on animal foraging (Bond, 1983; Bond & Kamil, 2006; Dawkins, 1983; Dukas & Elner, 1993; Gilliam & Fraser, 1987) our group has been exploring human “foraging”, where people are asked to search for multiple target items, in a variety of two-dimensional 2D (Jóhannesson et al., 2016, 2017; Á. Kristjánsson et al., 2014; Kristjánsson et al., 2018, 2020; Thornton et al., 2019) and 3D (T. Kristjánsson et al., 2022; Prpic et al., 2019) scenarios. Our typical foraging tasks involve arrays of items where multiple targets from more than one category (e.g., red and green circles) are presented among distractor items (e.g., blue and yellow circles) and observers have to select the targets (e.g., with fingers, or with a mouse cursor) and avoid the distractors. Selection difficulty (Á. Kristjánsson et al., 2014) and imposed time constraints (T. Kristjánsson et al., 2018; Thornton et al., 2020) have been shown to greatly affect how participants organize search in these contexts, whereas selection modality (Jóhannesson et al., 2016; Tagu & Kristjánsson, 2022; Thornton et al., 2019) and the need to divide attention (Thornton et al., 2021a) appear to have a modest impact. More generally, foraging tasks are proving to be a useful tool with which to measure a range of cognitive mechanisms in humans (Bella-Fernández et al., 2023; Gil-Gómez De Liaño & Wolfe, 2022; Á. Kristjánsson et al., 2019; Ólafsdóttir et al., 2020, 2021).

Here, in a series of three experiments, we modified our original human foraging task (Á. Kristjánsson et al., 2014), instructing participants to search for multiple, orientation-defined targets while also manipulating whether those targets changed color synchronously or asynchronously. Target items were always embedded in displays containing oriented distractors that changed color asynchronously. These display characteristics – orientation targets with orthogonal color changes – directly mirror those used in previous single-target visual search synchrony studies (Van der Burg et al., 2008).

There is considerable evidence that temporal synchrony – in the stimulus array and/or at the neural level – plays a fundamental role in creating and maintaining stable object representations (see Blake & Lee, 2005, and Treisman, 1999). But what role might such synchrony play in guiding behavior during foraging for multiple targets? Given the single-target nature of traditional visual search studies, direct evidence relevant to this question is limited, but there would seem to be at least two main possibilities.

A first possibility is that the synchrony between target items *globally segregates* them in some way from the distractor items. This could occur, for example, if synchrony automatically attracts attention in a bottom-up fashion, as other unique features, such as color or orientation singletons, are known to do. Evidence from the texture segmentation literature, where rapid detection of temporally defined object boundaries has been shown to occur, suggest this

could happen (Blake & Lee, 2005; Fahle, 1993; Farid, 2002; Guttman et al., 2007; Leonards et al., 1996). For example, Fahle (1993) found that phase differences in the appearance of groups of dots can induce shape perception, and Leonards et al. (1996) reported that temporal and textural cues can cause objects to emerge. Additionally, other, mid-level visual grouping principles could support some form of global segregation. While the classic Gestalt grouping principle of “common fate” was initially defined specifically in terms of motion perception (Wertheimer, 1923), more recent work has suggested that “synchrony” (Alais et al., 1998; Lee & Blake, 1999) may also play a fundamental role in perceptual organization (for review, see Wagemans et al., 2012).

If such global segregation between synchronized targets and asynchronous distractors were to occur in the context of multiple target visual search, we would expect target items to be located very quickly and easily. In terms of search speed, this should lead to extremely *regular* responses to consecutive target items, as is typically found during feature-based foraging (T. Kristjánsson et al., 2020). Essentially, the slope of the function mapping the change in time interval between consecutive targets (inter-target times; ITTs) should be almost flat (< 10 ms/item), mirroring the pattern observed for “pop-out” in classic visual search. In terms of the spatial organization of search, we would expect a very *irregular* pattern, as participants should be able to jump from one target to the next, without the need to use a specific horizontal or vertical scanning path, as tends to happen when displays are serially searched. The level of spatial organization in a foraging display can be indexed using a measure known as best-r (Woods et al., 2013), and global segregation would predict low, and possibly non-significant, correlations between the temporal order of target selection and their position within the search array (see *Methods* for details).

A second alternative is that synchronous targets benefit from some form of *local grouping* rather than global segregation. As a given orientation target is selected, observers may become aware of synchronized items in the local vicinity, guiding their search towards those items. As attention shifts to this new patch of synchrony, additional targets may become available, leading to a connected series of response episodes. The detection of local synchrony itself could rely on either bottom-up or top-down mechanisms.

The important characteristic of this alternative is that the speed of movement from one local patch of synchrony to another would be more sporadic than the scenario where all targets “pop-out” via global segregation. Search speed would still be expected to be faster, relative to the asynchronous, un-grouped condition, but the serial nature of such local grouping would predict that the ITTs would show variation as the trial progresses. Search organization, on the other hand, would be expected to be more regular, as individual targets and/or patches of locally grouped targets would need

to be located via serial search episodes. The slope of the ITT function and the magnitude of the best-r metric should then allow us to distinguish between global segregation and local grouping influences of synchrony on multiple target search.

In real life, humans and animals do not typically rely solely on their visual systems. They can, for example, use audition to understand and represent their perceptual environment. In the laboratory, various demonstrations show how attentional orienting to visual stimuli is strongly affected by auditory information (Driver, 1996; Driver & Spence, 1998; Fujisaki & Nishida, 2007; Ho et al., 2013; Ho & Spence, 2005; T. Kristjánsson et al., 2014). It is known, for example, that sound can guide attention if it shares the same location as a visual stimulus (Driver, 1996; Reisberg, 1978). More relevant to the current work, Van der Burg et al., (2008) suggested that non-spatial sounds occurring during a visual search task can decrease the search times for a single target item changing color in synchrony with a sound, even if this signal provides no information about target orientation, color, or location (the so-called “pip-and-pop” effect).

To explore this effect in the context of multiple target items, in each experiment, we included a condition in which visual synchrony was accompanied by a non-spatial auditory cue, in the form of a series of short “beeps,” timed to coincide with the target color changes. If the influence of such a cue in the current displays is similar to that seen in single-target search – where the “pip” is thought to create “pop-out,” greatly speeding target detection – then we might expect it to help globally segregate the display or at least to greatly speed the local grouping of target items. Either way, the prediction would be for faster and more efficient search in the *vision + sound* synchrony condition than in either the asynchronous or the vision-only synchrony conditions.

The current aims

We tested the influence of visual and auditory synchrony on attentional selection during a multiply target visual foraging task. Experiment 1 was an online experiment where the task was to forage for 10 (out of 20) vertical lines among 60 randomly oriented distractor lines. Items changed color between yellow and blue at random intervals and the presence of target synchrony was a within-subject factor, while the presence of the non-spatial sound was a between-subject factor. Experiment 2 was run in the laboratory with the same design except that all conditions were now run within-subject. In Experiment 3, task difficulty was increased while participants foraged for as many 45° rotated lines as possible among lines of different orientations within 10 s (when the trial ended), with the same synchrony conditions as Experiment 2.

Experiment 1

Data for Experiment 1 were collected online during the recent COVID pandemic. Following the methods employed by Thornton et al., (2021a) for web-based experiments, participants located targets by navigating a “sheep” icon using the mouse or trackpad and then selecting them by pressing the spacebar (see Fig. 1). In contrast to our previous foraging studies, however, items did not vanish when selected, but remained visible at all times (in order to keep the synchrony cue constant). An on-screen “score” counter could be used to confirm correct selection, and each target could only be selected once. Having previous targets “remain” on-screen in this way does not substantially affect search performance (Thornton & Horowitz, 2004, 2020a, 2020b) and was important here so that any possible synchrony cue provided by the group of multiple-targets remained constant throughout the trial.

On each trial of Experiment 1, participants were required to forage for 10 out of a possible 20 vertical target lines placed at random among an additional 30 left and 30 right diagonally oriented distractor lines. While Fig. 1 shows a static depiction of a trial, a working, dynamic version can be accessed at <https://maltacogsci.org/synchsearchdemo/>. At the start of each trial, items were colored blue or yellow at random. On *asynchronous* trials (50%), the color of each item switched between blue and yellow with an independently determined frequency (see *Methods* for details).

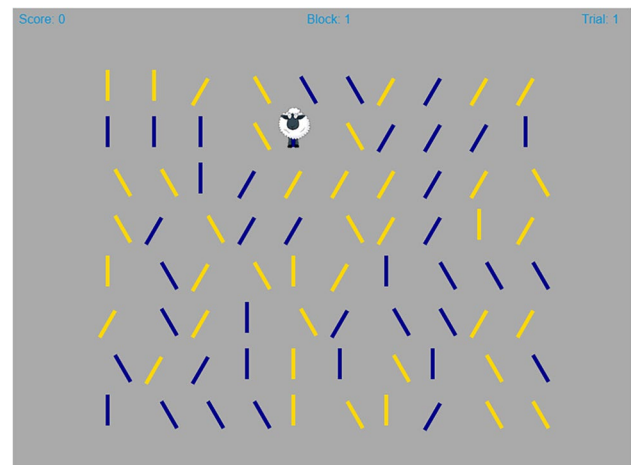


Fig. 1 Example of one trial from Experiments 1 and 2. Eighty stimuli were presented on each trial (20 vertical targets and 60 distractors). Participants’ task was to move the sheep (which served as a cursor) to the vertical lines and then press the spacebar to select targets. The number of targets already selected was shown next to the text “score” on the screen. Block number was denoted in the middle top position of the screen (one block out of four). The number of the current trial was presented in the right top corner. A demo can be accessed at <https://maltacogsci.org/synchsearchdemo/>

On *synchronous* trials, the color of the target items was constrained to start and switch together, while distractors switched color independently. The position of the target items within the display was randomized for each trial.

Our main research question was whether foraging speed would vary as a function of synchrony. We compared reaction time (RT) between *synchronous* and *asynchronous* trials, both in terms of the selection time of the first target (RT1) and the pattern of ITTs observed during the remainder of the trial, the so-called “cruise phase” of foraging (T. Kristjánsson et al., 2020). Within a block of trials, the *synchronous* and *asynchronous* displays were alternated, with the order of the initial trial type randomized for each participant. To reduce the overall duration of the task, given the online context, the presence of sound was a between-subject factor, with half of the participants taking part in the *vision-only* synchrony condition and half in the *vision + sound* synchrony condition.

Methods

Participants

Twenty-four participants took part in Experiment 1, 12 in the *vision only* and 12 in the *vision + sound* condition. There were 11 females (aged between 19 and 29 years, $M = 23.66$ years) and 13 males (aged between 18 and 39 years, $M = 25.21$ years). Participants were recruited online from <https://prolific.co> and the research team had no access to the personal identity of the individuals taking part. All participants reported being native speakers of English and having normal or corrected-to-normal vision. They received approximately £2 for a 15-min session, based on a flat-rate of £8 per hour. Participation was voluntary and there were no potential risks related to taking part in the study.

All participants were asked to explicitly confirm they had read the information and consent form prior to data collection. They were also given the option to download these documents. These procedures conform to the Ethics and Data Protection guidelines of the University of Malta and the University of Iceland for conducting online research.

Sample size and power analysis

Sample size for each condition ($N = 12$) was determined prior to data collection based on a power analysis for detecting group differences in multiple-target search conducted by Thornton et al. (2021b). To verify that this sample size would also allow us to detect differences between *synchronous* and *asynchronous* foraging – the main question of interest here – we used G*Power (Faul et al., 2009) to conduct an *a priori* power analysis based on pilot data collected during development of the current online task. Twelve

participants completed 40 trials of *synchronous* and 40 trials of *asynchronous* foraging, using a blocked design identical to the main study described below (see the Open Science Framework (OSF) page associated with this paper at osf.io/xu5rd/ for raw data and study details). The mean RT difference between completing trials in these two conditions was 457 ms ($SD = 394$), giving rise to an effect size of $d_z = 1.16$. We chose custom settings of assumed alpha for the planned study = 0.05 and a desired power of 0.8. This analysis yielded a minimum sample size of eight participants. We chose to be more conservative and recruited 12 participants – as in our previous work – because the study was run online and because we were unsure of the impact of including a cross-modal cue.

Apparatus

Given the online nature of the study, we were not able to control the precise display or auditory parameters nor the equipment used. Custom code was written in JavaScript so that the task would run via a standard web-browser opened on the participant’s laptop or desktop machine. Previous work from our own group (Thornton et al., 2021a) and others (e.g., Bridges et al., 2020; Chetverikov & Upravitelev, 2016) has demonstrated that the display and RT capabilities of such online experiments can, in principle, provide data comparable with lab-based testing. Here, the code ensured that stimuli were seen in full-screen mode, and that the effective frame rate of the display was approximately 60 Hz.

To minimize possible mouse versus trackpad differences in RTs, participants were allowed to move the cursor with their preferred device but were required to press the spacebar to register a response. We have previously used this same technique to equate response demands across input modalities (Thornton et al., 2019; Thornton et al., 2021a). Please note that the online demo also allows target selection via touch or click events, making it possible for readers to experience the task from a wide variety of devices. These options were not available to experimental participants.

Participants in the *vision + sound* condition were also required to have working speakers/headphones so that they could hear the non-spatial sound during the task. To verify this, a short test of whether the sounds were audible on the participant’s computer was performed prior to the main task (see below). We also note that Experiment 2 (below) was conducted under laboratory conditions to further verify the patterns of results observed online.

Stimuli

Figure 1 shows an example trial display (see <https://maltacog-sci.org/synchsearchdemo/> for an online demo). On each trial, 80 stimuli were presented on a grey background. The size of

the background was 800×600 px (see Experiment 2, *Methods* for example viewing conditions, in terms of degrees of visual angle). Twenty target stimuli (vertically oriented) were randomly distributed among 60 distractors. Half of the distractors were oriented 45° to the left, and half 45° to the right, with a random distribution. The size of each item was 5×40 px and they were placed within a non-visible 10×8 grid. The size of the sheep, which was used as a cursor, was 50×80 px.

At the start of each *asynchronous* trial, half of the 80 items were colored blue and half yellow, at random. Each item was also assigned an independent “time-to-change” value of between 180 ms and 500 ms, randomly determined on each trial to introduce unpredictability about the target change. At the end of this period, the color of the item was reversed (blue-to-yellow or yellow-to-blue), and the time-to-change value was reassigned in the same range. Thus, the color of each item changed randomly and independently.

On *synchronous* trials, the only difference was that all target items shared the same initial color assignment and a single “time-to-change” parameter. This ensured that the group of target items changed from yellow-to-blue and blue-to-yellow together, and with the same temporal frequency. In the *vision + sound* condition, target changes were also accompanied by a short (12-ms) auditory beep (fundamental frequency approximately 1,930 Hz). We note that to reduce the likelihood that playing a sound file had any unwanted impact on the visual timing of the display, on *vision-only* trials, the same code was used, but an empty sound file – with the same duration but with zero fundamental frequency and amplitude – replaced the *vision + sound* beep. The first “beep” appeared within 180–500 ms, and was synchronized with the first target color change. The beep then repeated in synchrony with each of the subsequent target color changes.

Procedure

The experiment was run online via <https://prolific.co>. Participants received instructions digitally through an online information/consent form that described the aim of the research, procedure, their rights, data confidentiality, possible risks, and contact information. After their confirmation, the main part of the experiment began. An introduction/instruction screen repeated the main aspects of the task, and participants were allowed to read this and then click a “Start experiment” button on the screen to begin. There were two training blocks, each containing five trials, and two experimental blocks each containing 20 trials. In the first training block, all trials were *asynchronous* and in the second training block all trials were *synchronous*. However, in the first and second experimental blocks, *asynchronous* and *synchronous* trials were presented in alternation, with the nature of the first trial determined randomly for each participant.

At the end of each correct trial (i.e., following selection of 10 vertical target items), feedback was given on the screen in the form of the total time taken. Targets did not disappear after they were clicked. If participants clicked twice on the same target, nothing happened, the trial continued, and the score did not change. Participants could rest between trials and then used the mouse to click on a “continue” button to start the next trial. If participants accidentally selected a distractor, the trial ended and a text message informed them of their error. Clicking “continue” then repeated a new trial of the same type. The trial counter was not updated until 10 targets were successfully collected, and the trial had been correctly completed. A message screen with a “continue” button was also provided between blocks, to provide information about each phase of the experiment.

For participants in the *vision + sound* condition, an audio pre-test was performed before the start of the main experiment. A series of beeps (between two and five) were played and participants were asked to report whether the number of sounds was odd or even. Correct answers were required on four consecutive trials. This was done to ensure that participants had the necessary equipment to accurately hear the sound stimuli during the experiment. The main experiment proceeded only if the participants responded correctly on all four trials.

Data analysis

We examined four main speed-related dependent variables: the time to find the first target (RT1); the intercept of the ITT function (an estimate of per-item speed during the cruise phase of foraging, computed across the last nine target items); the slope of the ITT function (a measure of foraging efficiency during the cruise phase, computed across the last nine target items); and the total completion time (TCT) for the trial (an estimate of overall speed during the cruise phase of foraging, computed by summing the ITTs for the last nine target items). These dependent variables were only analyzed for trials in which 10 targets were correctly selected. As is typical for multiple target search – particularly with targets defined by a single feature, as here – error rates were extremely low, and initial analysis indicated no significant variation as a function of experimental condition. For the sake of completeness, error data for Experiment 1, and all subsequent experiments, are provided in the respective summary tables, with additional analysis and discussion provided on the OSF page associated with this paper (osf.io/xu5rd/).

In our previous foraging studies (e.g., Á. Kristjánsson et al., 2014), we typically included two target categories in each display, making it possible to focus analysis on the pattern of “run” behavior, as participants switched between categories. A run in this context simply refers to the

contiguous selection of items from a single category. Here, as displays only contained a single category, such analysis is not possible. However, in addition to the reaction time analysis already discussed, we did examine one further foraging measure: best-r (e.g., Woods et al., 2013). Best-r is a measure of search organization that correlates the Cartesian coordinates of the targets with the sequence in which they are selected. A high best-r value denotes organised search (irrespective of whether search proceeds in horizontal or vertical sweeps), whereas a low best-r indicates disorganised search (Jóhannesson et al., 2016). In simple cancellation tasks (Mark et al., 2004; Woods et al., 2013), disorganised search can give rise to very low best-r values (< 0.3), whereas fully organised search yields highly consistent measures (> 0.9). In previous studies from our own group, average best-r measures for feature foraging ($M = 0.58$, $SE = 0.02$) have been shown to be consistently higher than for conjunction foraging ($M = 0.44$, $SE = 0.02$), reflecting the difference between regularly scanning through displays, in the former case, and following the distribution of specific target categories, in the latter case (Jóhannesson et al., 2016; Á. Kristjánsson et al., 2020; T. Kristjánsson et al., 2022; Ólafsdóttir et al., 2021). As mentioned in the *Introduction*, this measure may be useful in the current context as a way to distinguish between global segregation and local grouping influences of synchrony. To calculate best-r we first correlated every target position and serial selection position for the x-axis and y-axis separately on each trial. Following this we determined the larger coefficient of the two correlations, irrespective of the sign of the correlation. Finally, these best-r coefficients were averaged across trials for each participant to produce a single measure for each condition.

Returning to our main RT analysis, we treated RT1 and the ITT cruise phase measures separately as previous work suggests that they may tap into different aspects of multiple target search (Horowitz & Thornton, 2008; T. Kristjánsson et al., 2020; Thornton & Horowitz, 2004, 2020a). In particular, RT1 is usually elevated relative to all other responses – incorporating aspects of forward planning and response preparation (Horowitz & Thornton, 2008) – while subsequent responses tend to occur at a more rapid and consistent tempo (T. Kristjánsson et al., 2020).

We speculated that the presence of the non-spatial synchronized sound could speed responses to the first target, or it could slow responses, if participants attempt to coordinate their responses with the on-going sound sequence (Thornton et al., 2019). Similarly, sound could have either facilitative or inhibitory effects during the cruise phase, modulating the speed or efficiency of search, independent of RT1 responses. For all four measures, we calculated median ITTs across the 20 experimental trials in each condition on a participant-by-participant basis. ITT intercepts and slopes were extracted by regressing median ITT by target position on the individual participant

aggregates. The TCT measure was computed by summing the nine ITT values, again using the by-participant aggregates, to provide an overall estimate of cruise-phase speed.

While the intercept of the ITT function and the TCT measure provide a straightforward way to compare response speed across conditions, the slope of the ITT function provides a measure of the efficiency with which successive target items are selected during the cruise phase of a trial. This latter measure was of particular interest here, as it may help to distinguish the nature of any synchrony modulation. In our previous work, cruise phase slopes for feature targets have been shown to be < 10 ms/item (see Tables 3 and 6; T. Kristjánsson et al., 2020). If synchrony, with or without sound, can provide a mechanism to globally segregate target items from distractors in the current displays, we would similarly expect to see very shallow ITT slopes.

We analyzed each of the dependent measures separately using the same 2 (*synchrony presence*: synchronous/asynchronous) $\times 2$ (*synchrony type*: vision only/vision + sound) mixed ANOVA model, where *synchrony presence* was a repeated, within-subject factor, and *synchrony type* was a between-subject factor. In addition to these main null-hypothesis significance tests, we also conducted Bayesian analysis, using the approach outlined by Masson (2011). Specifically, for all tests where the null hypothesis was not rejected, we also estimated the weight of evidence in favor of the null hypothesis, and report both BF_{01} and $p(H_0|D)$ in the text where appropriate. Where only weak evidence was obtained – specifically $BF_{01} < 3.0$ and $p(H_0|D) < 0.75$ (Jeffreys, 1998; Raftery, 1995; Wagenmakers et al., 2011) – this is noted in the text. Full details of all analyses are provided on the OSF page associated with this paper at osf.io/xu5rd/, while in-text reporting is minimized for the sake of brevity.

Data analysis was performed in RStudio, using the *rstatix*, *ez*, and *tidyverse* packages (Kassambara, 2021; Lawrence, 2016; RStudio Team, 2020; Wickham et al., 2019). Analysis scripts and raw data files have also been made available on the OSF page associated with this paper at osf.io/xu5rd/.

Results

Figure 2 summarizes the median time taken to click on and select the first target (RT1) in Experiment 1, averaged across the 20 experimental trials for each condition. Data are presented separately for the *vision-only* and the *vision + sound* conditions. Average RT1 responses remained well above a second in all conditions. There does appear to be a speed advantage for *synchronous* trials, particularly in the *vision + sound* condition (see Table 1).

However, the 2×2 mixed ANOVA revealed only a significant main effect of *synchrony presence*, with responses on *synchronous* trials ($M = 1,105$ ms, $SE = 47$), being

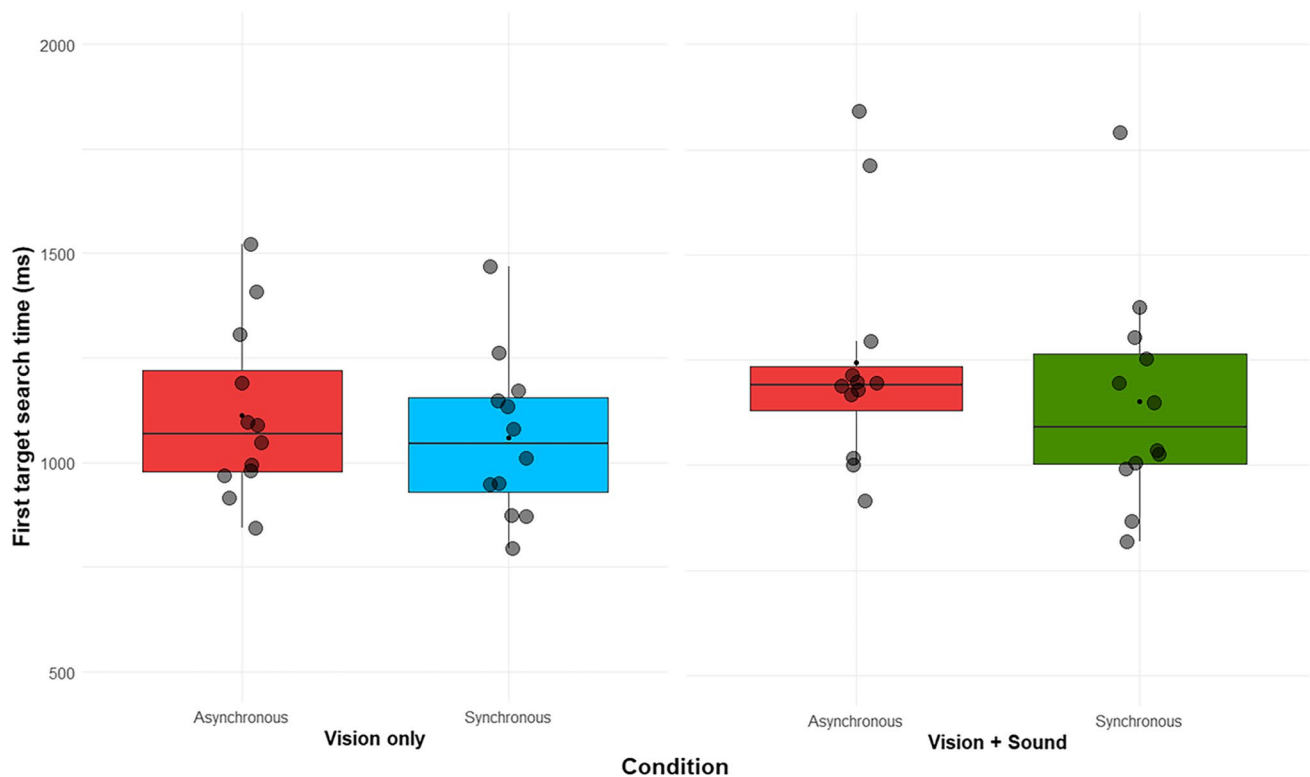


Fig. 2 The search times for the first target (RT1) in Experiment 1. The horizontal lines in the box plots represent medians. The top horizontal line of the boxplots represents the upper quartile (75%) while the bottom horizontal line represents the lower quartile (25%). The

top vertical lines represent the maximum value, while the bottom lines represent minimum values. The dots represent individual participants' median RT1s

Table 1 Descriptive statistics for all dependent measures in Experiment 1. Standard errors are provided in brackets

	Vision only		Vision + Sound	
	Asynchronous	Synchronous	Asynchronous	Synchronous
RT1 (ms)	1,114 (60)	1,060 (55)	1,243 (79)	1,130 (76)
ITT Intercept (ms)	613 (115)	569 (102)	519 (66)	501 (58)
ITT slope (ms)	15 (4)	18 (4)	25 (5)	16 (4)
TCT (ms)	6,207 (883)	5,908 (931)	5,784 (583)	5,222 (560)
Best-r	0.61 (0.05)	0.65 (0.05)	0.54 (0.06)	0.58 (0.06)
Error rate %	2.7 (0.7)	1.9 (0.5)	1.5 (0.4)	1.6 (0.4)

Note. RT1 = response to first target; ITT Intercept = estimate of per-item response time computed as the intercept of the linear regression function mapping median ITT across target position during the cruise phase of a trial (see text for details); ITT Slope = linear regression coefficient, estimating the rate of change in ITT across target position during the cruise phase of the trial; TCT = trial completion time (sum of all cruise phase ITTs); Best-r = a measure of search organisation computed as the average correlation coefficient (r) between the Cartesian coordinates of targets and the sequence in which they are selected; Error rate = % incorrect clicks divided by the sum of incorrect and correct clicks

approximately 70 ms faster than on *asynchronous* trials ($M = 1,179$ ms, $SE = 50$), $F(1, 22) = 5.16$, $MSE = 12448$, $p = 0.033$, $\eta^2 = 0.19$. Neither the main effect of *synchrony type*, $F(1, 22) = 1.48$, $MSE = 98529$, $p = 0.236$, $\eta^2 = 0.06$, $BF_{01} = 2.27$, $p(H_0|D) = 0.69$, nor the *synchrony presence* $\times$ *synchrony type* interaction, $F(1, 22) = 0.34$, $MSE = 12448$, $p = 0.562$, $\eta^2 = 0.02$, $BF_{01} = 4.05$, $p(H_0|D) = 0.8$, were significant. We note that for the main effect of *synchrony type*, evidence in favor of the null hypothesis is relatively weak. However, this potential trend would appear to reflect a general difference in response speed between the two groups of participants, rather than a synchrony effect of interest, as *asynchronous* as well as *synchronous* trials are slower for the *vision + sound* group compared to the *vision only* group (see Table 1).

Figure 3 summarizes the cruise phase of foraging by plotting the median ITT as a function of target position and synchrony condition. There are several patterns to note. First, all four functions have a similar positive slope, indicating that target detection times became slower as the trial progressed. Second, the ITT functions for both *synchronous* conditions appear to be shifted down, relative to the *asynchronous* conditions, consistent with the RT advantage already evident in

the RT1 data. Third, participants in the *vision + sound* condition appear to have responded generally more quickly than those in the *vision-only* condition, with the two functions in the lower panel being, at least initially, shifted downwards. This pattern may well reflect the additional time the *sound + vision* group took to make their first response, providing an opportunity for increased forward planning (Horowitz & Thornton, 2008; Thornton & Horowitz, 2004).

The difference between the *synchronous* and *asynchronous* functions in the two panels was best captured by the TCT analysis (see Table 1 and Fig. 4). Here, there was a significant main effect of *synchrony presence*, with foraging on *synchronous* trials ($M = 5,565$ ms, $SE = 536$) proceeding more than 400 ms faster than on *asynchronous* trials ($M = 5,995$ ms, $SE = 519$), $F(1, 22) = 23.98$, $MSE = 92554$, $p < 0.001$, $\eta^2 = 0.52$. Despite this advantage for synchrony, we note that taking approximately 6 s to complete the cruise phase of the trial, with ITTs approaching or exceeding 600 ms/item (see Table 1), indicates that target detection remained quite effortful, given that targets were defined by a unique orientation cue. That is, the addition of synchrony appears to have led to a modest search advantage, rather than making targets immediately available in any sense.

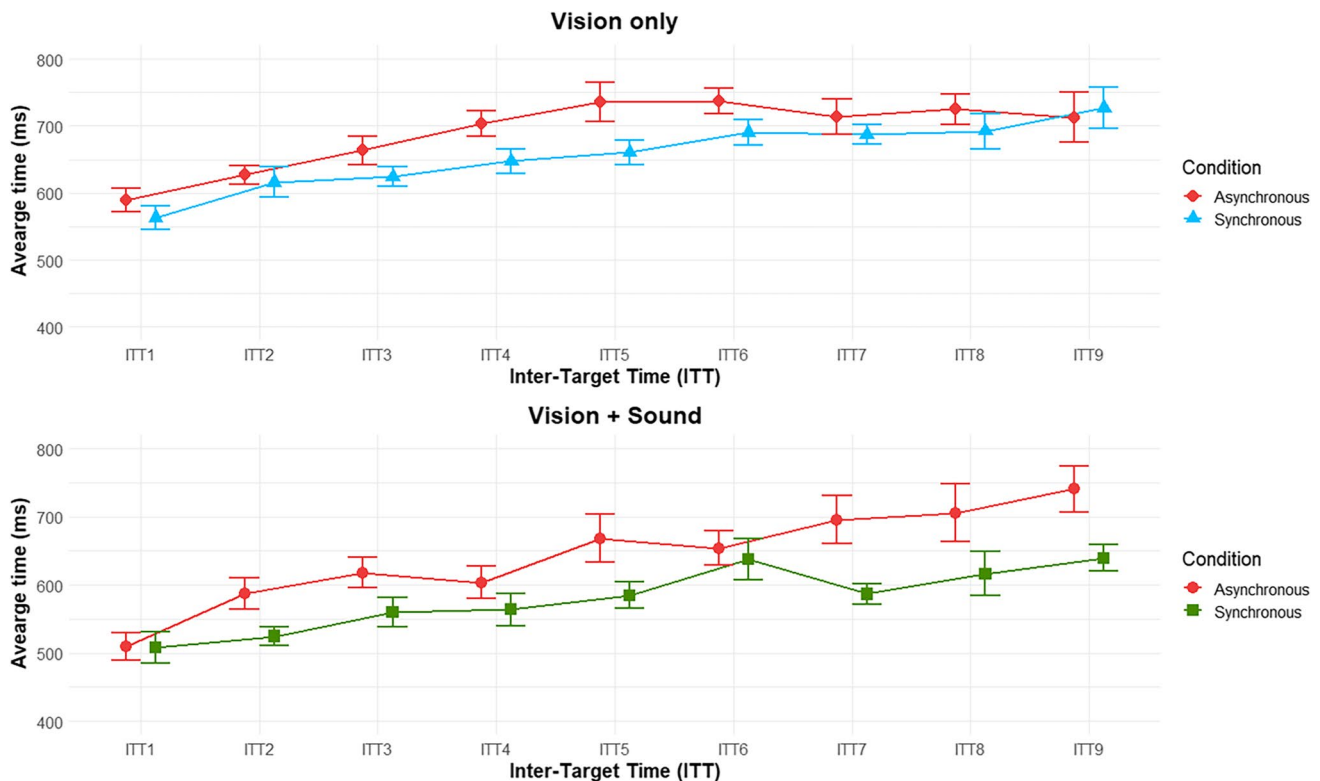


Fig. 3 Inter-target time (ITT) as a function of target selection order. ITTs are calculated as $RT_n - RT_{n-1}$, so there are only nine values, with ITT1 being the difference in time between RT2 and RT1, etc. The upper panel shows results from the *vision-only* condition while the lower panel shows results for the *vision + sound* condition. The

circle dots were used for *asynchronous* trials. The triangle dots were used for *synchronous* trials in the *vision-only* condition. The square dots were used for *asynchronous* trials in the *vision + sound* condition

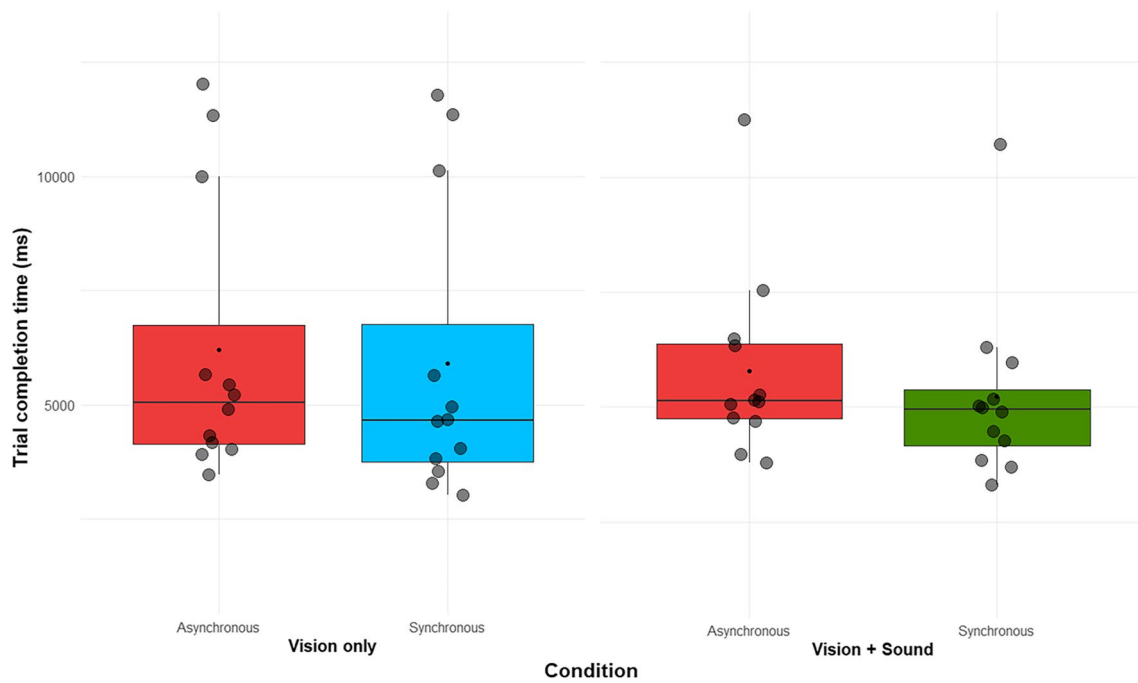


Fig. 4 Total completion time (TCT; sum of ITTs) during the cruise phase of Experiment 1. The horizontal lines in the box plots represent medians. The top horizontal line of the boxplot represents the upper quartile (75%), the bottom horizontal lines represent the lower quar-

tile (25%). The top vertical line represents the maximum value, the bottom line represents the minimum value. The dots represent individual participants' median TCTs

There was no main effect of *synchrony type* for TCT, $F(1, 22) = 0.27$, $MSE = 13710151$, $p = 0.61$, $\eta^2 = 0.01$, $BF_{01} = 4.17$, $p(H_0|D) = 0.81$, and no *synchrony presence* $\times$ *synchrony type* interaction, $F(1, 22) = 2.25$, $MSE = 92554$, $p = 0.15$, $\eta^2 = 0.09$, $BF_{01} = 1.52$, $p(H_0|D) = 0.6$, although we note that evidence for this latter null effect is relatively weak. There does appear to be a trend for the synchrony advantage to be bigger in the *vision + sound* condition, a point we return to in the *Discussion* below.

Analysis of the ITT intercepts (see Table 1) revealed no significant effects (see Online Supplementary Material (OSM) Table 6 at osf.io/xu5rd/), although as with the RT1 analysis, evidence in favor of the null hypothesis for the main effect of *synchrony type* was relatively weak, $F(1, 22) = 1.83$, $MSE = 5832$, $p = 0.19$, $\eta^2 = 0.08$, $BF_{01} = 1.88$, $p(H_0|D) = 0.65$. Again, this potential trend probably reflects general speed differences between the two groups, rather than relating to any synchrony-specific effects of interest.

Returning to the overall ITT patterns in Fig. 3, as already noted above, the slope of all four functions is positive (see Table 1). This indicates that the difficulty in target selection is not constant across the whole trial, but rather increases with target position. Such a pattern is not consistent with the notion of *global segregation* of targets and distractors, as might be expected if synchrony acted as a visual feature (see Tables 3 and 6; T. Kristjánsson et al., 2020), but rather suggests a more serial, effortful search process.

Analysis of the slope data further confirms this impression (see Table 1), as there were no significant differences between conditions, as a function of neither *synchrony presence*, $F(1, 22) = 0.73$, $MSE = 211$, $p = 0.4$, $\eta^2 = 0.03$, $BF_{01} = 3.31$, $p(H_0|D) = 0.77$, nor *synchrony type*, $F(1, 22) = 0.69$, $MSE = 231$, $p = 0.41$, $\eta^2 = 0.01$, $BF_{01} = 3.36$, $p(H_0|D) = 0.77$. The *synchrony presence* $\times$ *synchrony type* interaction was also not significant, $F(1, 22) = 1.88$, $MSE = 211$, $p = 0.18$, $\eta^2 = 0.08$, $BF_{01} = 1.83$, $p(H_0|D) = 0.65$, although evidence for this null effect was relatively weak. We note, however, that any apparent trend here appears to be driven by particularly steep slopes in the *asynchronous* condition during *vision + sound* trials, rather than any potential synchrony advantage (see Table 1).

In terms of search organisation, the average best-r coefficient was relatively high ($M = 0.60$, $SE = 0.2$), consistent with levels we have seen previously for feature foraging. There was a slight modulation of foraging organisation by *synchrony presence*, with best-r being higher on *synchronous* trials ($M = 0.62$, $SE = 0.04$), compared to *asynchronous* trials ($M = 0.58$, $SE = 0.04$), $F(1, 22) = 5.27$, $MSE = 0.003$, $p = 0.031$, $\eta^2 = 0.19$. The overall level of search organization – together with the slight synchrony advantage – suggests that participants were scanning through displays quite regularly, a pattern more consistent with a *local grouping* account of target detection, rather than what would be expected from *global segregation*. No

other effects were significant (see OSM for details at osf.io/xu5rd/).

As expected, error rates were extremely low (see Table 1) and were not modulated by any of the experimental manipulations (see OSM for details at osf.io/xu5rd/).

Discussion

Our main research question in Experiment 1 was whether foraging behavior would vary as a function of visual and auditory synchrony. We found clear evidence that the *presence* of synchrony affects foraging. Specifically, the time to find the first target (RT1) and the time to complete the cruise phase of the trial (TCT) were reduced, and the level of search organisation increased (best-r), when target color changed synchronously, compared to when changes were asynchronous. Examination of the pattern of ITTs during the cruise phase of the trial suggests that synchrony provides relatively modest benefits via *local grouping*, rather than changing the nature of target detection via *global segregation*. That is, search for synchronized targets remained relatively inefficient, with cruise phase slopes exceeding 10 ms/item and not varying across conditions (T. Kristjánsson et al., 2020). The level of search organization was also relatively high, again favoring a *local grouping* explanation.

Somewhat surprisingly – given evidence from the “pip-and-pop” effect in single target search – the addition of a cross-modal auditory cue to target synchrony did not appear to enhance performance in any way in the current experiment. However, our use of a between-subject design – as well as the variability in auditory environments inherent in running the study online – may have reduced the ability to detect subtle cross-modal advantages, given the presence of a robust visual synchrony cue. As noted above, several null effects involving the factor *synchrony type* were associated with only weak supporting evidence in our Bayesian analyses. There were also several trends to suggest overall speed differences between the two groups of participants. To address these concerns, we decided to replicate Experiment 1 under more controlled, laboratory conditions using a within-subject design, before drawing further conclusions about the influence of cross-modal synchrony.

Experiment 2

Experiment 2 was designed as a replication of Experiment 1, with the addition of several methodological changes aimed at increasing the likelihood of detecting a cross-modal synchrony effect, should one exist. First, rather than running the study online, we brought participants into the lab so as to create a fixed visual/auditory display environment. Second, we used a within-subject, blocked design. Each participant

completed blocks of *asynchronous, vision-only* (synchronous) and *vision + sound* (synchronous) trials, with block order randomized individually. Third, we greatly increased the repetitions for each trial type (from 20 to 72) by dividing the experiment into two sessions spread across consecutive days.

Methods

Participants

There were 12 participants in Experiment 2, eight females (aged between 22 and 33 years, $M = 24.5$ years) and four males (aged between 18 and 23 years, $M = 21.25$ years). All participants reported being fluent speakers of English and having normal or corrected-to-normal vision. In contrast to Experiment 1, recruitment was by convenience sampling from the wider University of Iceland community. Participation was completely voluntary, with no payment provided. All participants provided written, informed consent and were aware that they were free to end participation at any time. The experiment was run in accordance with the requirements of the local ethics committee.

Apparatus

The stimuli were presented on an ASUS Nvidia G-Sync monitor (model PG279Q) with screen dimensions of 60 × 34 cm and an effective resolution of 1,920 × 1,080 pixels. The monitor was connected to a Windows 10 computer running Google Chrome (Version 108.0.5359.124 (Official Build) (64 bit)), and the monitor refresh rate was fixed at 60 Hz. Participants were seated approximately 60 cm from the screen and had an MI-2500X optical mouse from Trust and a K120 logitech keyboard in front of them. As in Experiment 1, the mouse was used to move a sheep icon on the screen, and participants could then use the spacebar to collect the targets. The sound in the synchronized-sound condition was presented through the computer speakers. The experiment was written in JavaScript and the experimental code involved only minor changes from Experiment 1 to accommodate a within-subject, blocked design with increased repetitions.

Stimuli

The stimuli were identical to those used in Experiment 1 (see Experiment 1, *Methods* and Fig. 1 for details). However, as Experiment 2 was run under laboratory conditions with a fixed viewing distance (60 cm) on a standard monitor (1,920 × 1,080), here we also report visual angle. On each trial, 80 stimuli were presented on a grey background. The size of the background was 800 × 600 px, which subtended approximately 24 × 18° visual angle. Twenty target stimuli

(vertically oriented) were randomly distributed among 60 distractors. Half of the distractors were oriented 45° to the left, and half 45° to the right, randomly distributed on the screen. The size of each item was 5×40 px, subtending approximately $0.2 \times 1.2^\circ$ visual angle, and items were placed within a non-visible 10×8 grid. The size of the sheep, which was used as a cursor, was 50×60 px, subtending approximately $1.4 \times 1.6^\circ$ visual angle.

Procedure

Participants received instructions through a physical information/consent form that described the aim of the research, procedure, their rights, data confidentiality, possible risks, and researcher contact information. Moreover, participants were invited to address any questions to the experimenter. After providing written informed consent, the main part of the experiment started. Data collection in Experiment 2 was split into six sessions, typically spread over two consecutive days, with three sessions per day. Each session included three blocks of trials, one for each experimental condition: *asynchronous*, *vision only (synchronous)*, and *vision + sound (synchronous)*. The blocks were presented in a random order, but within a block trial-type was constant. There were three practice trials and 12 experimental trials within each block, taking approximately 10 min per block to complete. In total

there were 270 trials ($6 \text{ sessions} \times 3 \text{ blocks} \times 15 \text{ trials}$) with 72 trials of each experimental condition available for data analysis (compared to 20 in Experiment 1). All other aspects of task and design were identical to Experiment 1.

Data analysis

The dependent variables in Experiment 2 were identical to those in Experiment 1. For the speed-related measures, we calculated median RTs separately for each participant across the 72 experimental trials in each condition. All dependent measures were analyzed using a separate one-way repeated-measures ANOVA with the factor *condition*, having three levels: *asynchronous*; *vision only (synchronous)*; and *vision + sound (synchronous)*. Pairwise t-tests with Bonferroni corrections were used to analyze the difference between these condition levels.

Data analysis was performed using the same statistical packages as in Experiment 1, and all analysis scripts and raw data files have been made available on the OSF page associated with this paper at osf.io/xu5rd/.

Results

Figure 5 summarizes the median time taken to locate the first target (RT1) in Experiment 2, averaged across the 72

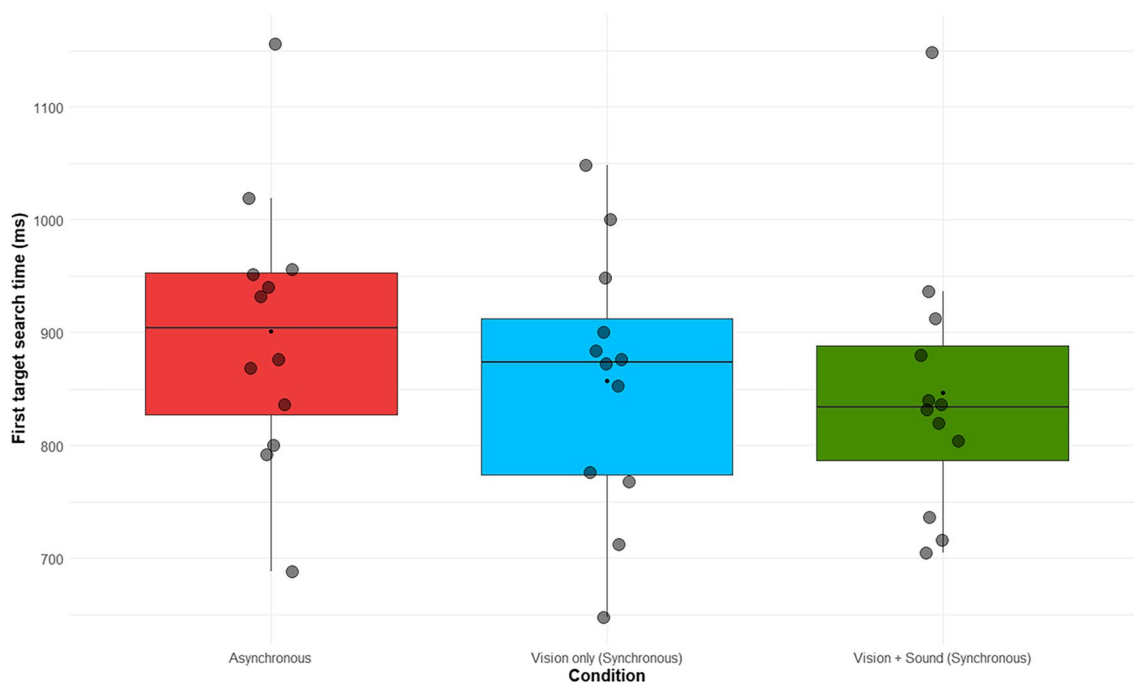


Fig. 5 The selection times for the first target (RT1) in Experiment 2. The horizontal lines in the box plots represent medians. The top horizontal line of the boxplot represents the upper quartile (75%), the bottom horizontal line represents the lower quartile (25%). The top verti-

cal line represents the maximum value, the bottom line represents the minimum value. The dots represent individual participants' median RT1s

Table 2 Descriptive statistics for all dependent measures in Experiment 2. Standard errors are provided in brackets

	Asynchronous	Vision only (Synchronous)	Vision + Sound (Synchronous)
RT1 (ms)	901 (35)	857 (33)	847 (34)
ITT Intercept (ms)	385 (16)	356 (20)	344 (14)
ITT slope (ms)	20 (3)	16 (3)	18 (2)
TCT (ms)	4357 (196)	3932 (166)	3897 (132)
Best-r	0.51 (0.06)	0.49 (0.05)	0.51 (0.05)
Error rate %	2.9 (0.5)	2.6 (0.5)	3.2 (0.6)

Note. RT1 = response to first target; ITT Intercept = estimate of per-item response time computed as the intercept of the linear regression function mapping median ITT across target position during the cruise phase of a trial (see text for details); ITT Slope = linear regression coefficient, estimating the rate of change in ITT across target position during the cruise phase of the trial; TCT = trial completion time (sum of all cruise phase ITTs); Best-r = a measure of search organisation computed as the average correlation coefficient (r) between the Cartesian coordinates of targets and the sequence in which they are selected; Error rate = % incorrect clicks divided by the sum of incorrect and correct clicks

experimental trials in each condition. Compared to Experiment 1 (Fig. 2), foraging times in the laboratory were considerably faster, with average RT1 responses remaining well below 1,000 ms in all conditions (see Table 2). The overall data spread was also reduced. More importantly, the same tendency for faster *synchronous* responses was replicated.

There was a significant main effect of *condition*, $F(2, 22) = 8.3$, $MSE = 1,203$, $p = 0.002$, $\eta^2 = 0.43$. Pairwise comparisons showed significant differences between the *asynchronous* condition and both the *vision-only*, $t = 3.27$, $p = 0.02$, and *vision + sound*, $t = 4.35$, $p < 0.0001$, *synchronous* conditions. There was no difference between the two *synchronous* conditions ($t = 0.62$, $p > 0.05$, $BF_{01} = 2.96$, $p(H_0|D) = 0.75$). Again, this indicates there were no benefits from the synchronous sound, just as in Experiment 1.

Figure 6 summarizes the cruise phase of foraging in Experiment 2. As with the RT1 responses, participants in Experiment 2 were considerably faster during the cruise phase of the trial, compared to either group in Experiment 1 (see Fig. 3). There was again a positive slope to all three functions, indicating that target detection times increase as the trial progresses. The clearest pattern, however, is the separation of the two synchronous conditions from the asynchronous condition, replicating the general advantage seen in Experiment 1. Here, it is even clearer that RTs in the *vision only* and *vision + sound* conditions largely overlap.

In contrast to Experiment 1, the TCT analysis and ITT intercept analysis yielded identical results. This probably

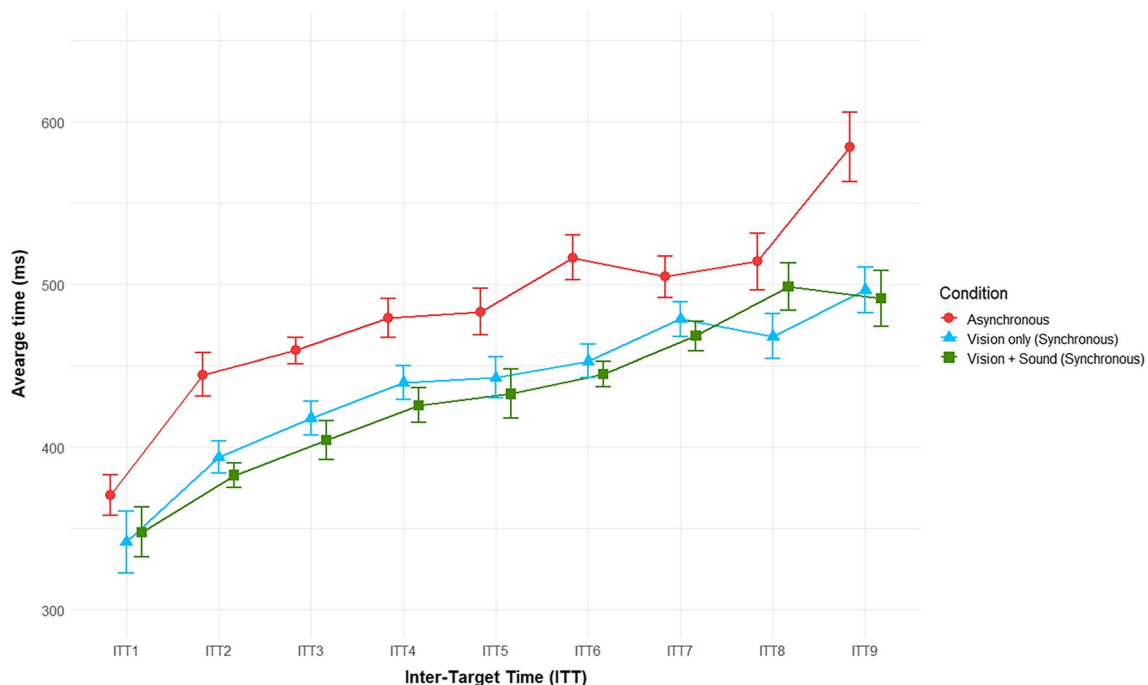


Fig. 6 Inter-target time (ITT) as a function of condition and target position in Experiment 2. ITTs are calculated as $RT_n - RT_{n-1}$, so there are only nine values, with ITT1 being the difference in time between RT2 and RT1, etc.

reflects the increased repetitions and more experienced observers in Experiment 2. For the sake of brevity – and consistency with Experiment 1 – we report the TCT in the text, referring the reader to the OSM for the details of the ITT intercept analysis (OSM can be found at osf.io/xu5rd/).

Figure 7 shows the median time taken to collect the subsequent nine targets (TCT) in Experiment 2. Overall foraging times were more than a second faster than in Experiment 1, while the difference between *asynchronous* trials and the two *synchronous* conditions is also more pronounced (compare Tables 1 and 2). There appears to be very little difference in TCTs between the two *synchronous* conditions.

Analyses of the TCT measure showed a significant main effect of *condition*, $F(2, 22) = 18.5$, $MSE = 42514$, $p < 0.0001$, $\eta^2 = 0.63$. Pairwise t-tests indicated significant differences between the *asynchronous* condition and both the *vision-only*, $t = 5.12$, $p < .001$ and *vision + sound*, $t = 4.72$, $p < .001$, *synchronous* conditions. As in Experiment 1, there was no difference between the two *synchronous* conditions ($t = 0.51$, $p = 0.62$, $BF_{01} = 3.1$, $p(H_0|D) = 0.76$).

As can be seen in Fig. 6, there was again no difference between the slopes of the ITT functions for any of the conditions, $F(2, 22) = 2.6$, $MSE = 14.84$, $p = 0.1$, $\eta^2 = 0.19$, $BF_{01} = 1.8$, $p(H_0|D) = 0.64$, although we should note that evidence in favor of this null effect is relatively weak. Nevertheless, the slopes of these functions are all strongly

positive, indicating relatively effortful search, and if there is any advantage for the synchrony conditions, this is more likely to be for the *vision-only* condition, rather than the *vision + sound* condition (see Table 2).

Best- r values were slightly lower than seen in Experiment 1 (compare Tables 1 and 2), but with averages of approximately 0.5, still higher than expected if participants were selecting targets at random via global segregation (see Experiment 1 for further details on best- r). However, the effect of *synchrony* seen in Experiment 1 was not replicated, $F(2, 22) = 0.61$, $MSE = 0.091$, $p = 0.52$, $\eta^2 = 0.53$, $BF_{01} = 12.7$, $p(H_0|D) = 0.93$. We can only speculate, but it is possible that the more experienced observers used in Experiment 2 and/or the large increase in trial repetitions led to more idiosyncratic and less organised search strategies. Error rates were again extremely low, and were not modulated by synchrony (see Table 2 and supplementary analysis in the OSM at osf.io/xu5rd/).

Discussion

The goal of Experiment 2 was to replicate and extend Experiment 1 by introducing methodological changes that might increase the likelihood of detecting a cross-modal synchrony advantage, if one existed. Changing to a within-subject design, run under laboratory conditions with increased trial

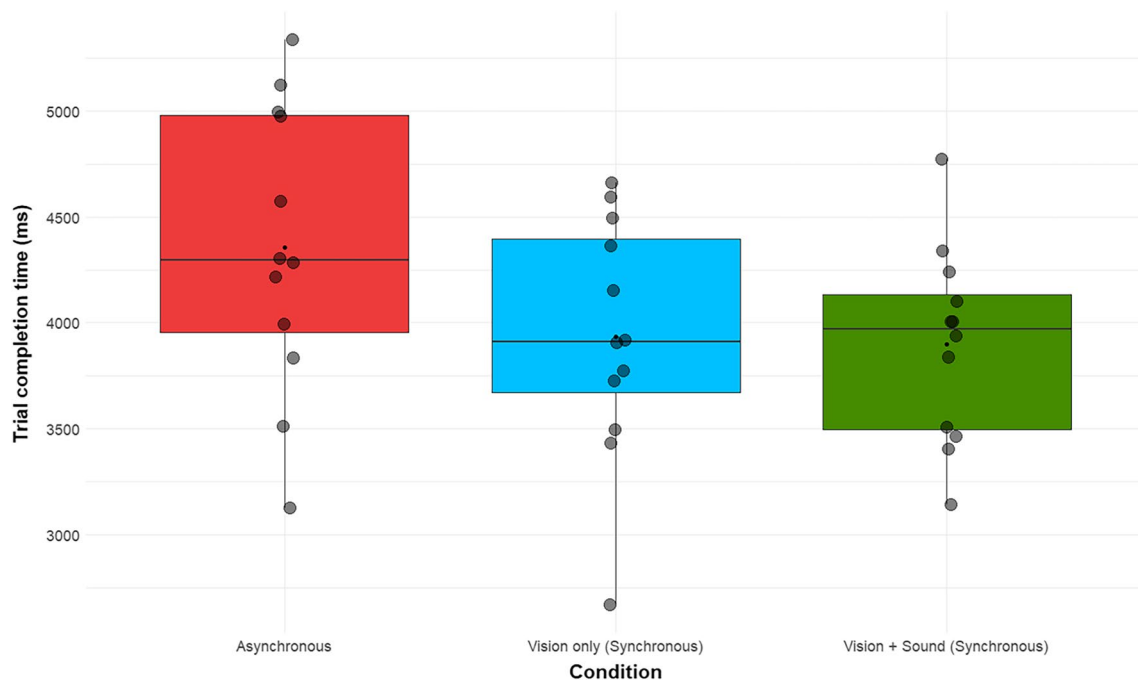


Fig. 7 Total completion time (TCT; sum of ITTs) during the cruise phase of Experiment 2. The horizontal lines inside the box plots represent medians. The top horizontal lines of the boxplots represent the upper quartiles (75%), the bottom horizontal line represent the lower

quartiles (25%). The top vertical lines represent the maximum values, the bottom lines represent the minimum values. The dots represent individual participants' median completion times

repetitions, had a clear effect on overall performance levels. Participants responded more quickly and with less variability than in Experiment 1 for all dependent variables except the cruise-phase slopes (compare Tables 1 and 2). We also very clearly replicated the general synchrony advantage seen in Experiment 1, with performance on *synchronous* trials being better than on *asynchronous* trials. The presence of strongly positive cruise-phase slopes and spatially organized search patterns adds further support to the idea that synchrony in this task improves *local grouping* of targets, rather than causing them to be segregated as a whole from the distractors, where they would essentially “pop out.”

This lack of clear *global* segregation was equally true in the *sound-only* and *sound + vision* conditions of Experiment 2. In terms of cross-modal synchrony, then, there was still no evidence of any additional advantage from adding sound. This replication of the patterns seen in Experiment 1 would seem to rule out our methodological concerns over the online, between-subject design of that experiment. In Experiment 3, we explored whether increasing the overall difficulty of the foraging task might further encourage participants to make use of the cross-modal synchrony cues.

Experiment 3

The purpose of Experiment 3 was to increase the overall difficulty of the multiple target search. Our hypothesis was that this might encourage the use of all possible synchrony cues for target detection, in particular the additional auditory cue. We increased difficulty in three ways. First, we changed the spatial orientation of the targets. In Experiments 1 and 2, targets were canonically oriented, vertical lines. Such 0° targets should be very easy to locate, relative to the diagonal distractors (Kong et al., 2017; Wolfe et al., 1982). Here, we also chose a canonical orientation for the targets, but one that has been shown to be much less effective in guiding search, specifically 45° diagonally oriented rectangles (Kong et al., 2017). Second, we introduced a heterogeneous field of distractors. In Experiments 1 and 2, the use of fixed 45° distractors might have facilitated distractor grouping, making target detection easier (Chetverikov et al., 2016; Duncan & Humphreys, 1989). In Experiment 3, the orientation of each distractor was therefore independently determined (see *Methods* for details).

Finally, we introduced a strict time limit on each trial. In Experiments 1 and 2, trials continued until 10 targets were collected. In Experiment 3, participants had to collect as many targets as possible during a fixed 10-s period. Previous work from our group (T. Kristjánsson et al., 2018; Thornton et al., 2020) has shown that imposing time limits can modulate the pace and nature of foraging behavior, in

particular, encouraging foraging patterns consistent with increased cognitive/attentional load (Thornton et al., 2020).

Experiment 3 was run online using a within-subject design. During experimental blocks, each participant saw trials from all three conditions (*asynchronous*, *vision only*, *vision + sound*) in a random order. The main dependent variable was now the average number of targets collected per trial. However, we also report reaction time analysis for comparison with Experiments 1 and 2.

Methods

Participants

Twelve participants took part in Experiment 3, seven females (aged between 19 and 39 years, $M = 28.2$ years) and five males (aged between 22 and 37 years, $M = 24.86$ years). Participants were recruited online from <https://prolific.co>, and the research team therefore had no access to the personal identity of the individuals taking part. All participants reported being native speakers of English and having normal or corrected-to-normal vision. They received approximately £2 for a 15-min session, based on a flat-rate of £8 per hour. Participation was voluntary and there was no potential risk to taking part in the study. All participants were asked to explicitly confirm that they had read the information and consent form prior to data collection. They were also given the option to download these documents. These procedures conform to the Ethics and Data Protection guidelines of the University of Malta and the University of Iceland for conducting online research.

Apparatus

Experiment 3 was run online using the same methods as detailed in Experiment 1 (including the sound test).

Stimuli

Figure 8 shows an example trial display. A dynamic version of the task can also be viewed at <http://maltacogsci.org/synchsearchdemo/Exp3Demo.html>. The basic layout and physical characteristics of the display (i.e., number of stimuli, stimuli-size, background, sheep-cursor, color categories, color-change tempo) and the non-spatial sound parameters were the same as in Experiments 1 and 2. In Experiment 3, however, the target items were 20 diagonally oriented rectangles (oriented 45° clockwise or counter-clockwise from vertical). The orientation of each of the 60 distractor rectangles was randomly and independently assigned on each trial, with the following two constraints. First, no item could be closer than 30° to vertical. Second, no item could be within 30° of the target orientation. For eight participants, targets were

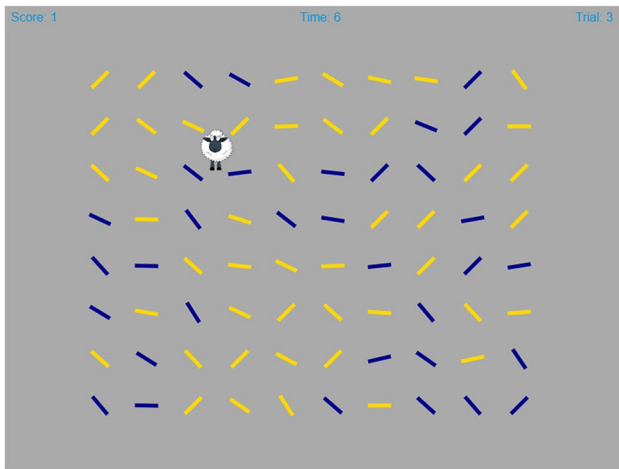


Fig. 8 Example of one trial from Experiment 3. Eighty stimuli were presented on each trial (20 targets and 60 distractors). The participants' task was to move the sheep (which served as a cursor) to the targets (the 45° lines) and press the spacebar to select targets. The number of targets that were already selected were reported next to the score. The time left before the trial ended was reported in the middle top position of the screen (here 6 s left out of 10). The number of completed trials was presented in the right top corner

oriented 45° clockwise while for the other eight participants targets were oriented 45° counter-clockwise.

Procedure

The procedure in Experiment 3 was the same as for Experiment 1, except for the following modifications. First, the main task of the experiment was now to collect as many targets as possible within a fixed 10-s period. Second, Experiment 3 involved a fully within-subject design. There were three practice blocks, where participants were introduced to each type of trial, in consecutive blocks, using a fixed order (*asynchronous*, *vision only*, *vision + sound*). This was followed by three experimental blocks in which the three trial types appeared six times each in a random order, for a total of 18 trials per block. Each participant thus completed a total of 54 experimental trials. The same auditory test as used in Experiment 1 was now performed by all participants before the main study began (see *Methods* for Experiment 1).

Data analysis

The main dependent variable in Experiment 3 was the average number of targets found during the fixed 10-s trial period. However, we also report the time to find the first target (RT1), ITT slopes and intercepts, as well as best-r analysis to allow comparison with Experiments 1 and 2.

For RT1, as before, we calculated median RT separately for each participant across the experimental trials in each

condition. The average number of targets found was also calculated separately for each participant across the 18 experimental trials in each condition. Given the design of Experiment 3, with the number of selected targets expected to vary considerably from trial to trial, ITT slopes and intercepts were calculated on a trial-by-trial basis and then averaged within participant and condition. Best-r values were calculated as in Experiments 1 and 2. For all dependent measures, data were analyzed using one-way repeated-measures ANOVAs with the factor *condition*, which included three levels (*asynchronous*, *vision only (synchronous)* and *vision + sound (synchronous)*). Pairwise t-tests with Bonferroni corrections were used to analyze the differences between conditions. Data analyses were performed with the same program and packages as in previous experiments. Analysis scripts and raw data files have been made available on the OSF page associated with this paper at osf.io/xu5rd/.

Results

Figure 9 summarizes the median time taken to locate the first target (RT1) in Experiment 3, averaged across the 18 experimental trials for the three conditions. Compared to Experiment 1, selection times in Experiment 3 were slower, with RT1 responses remaining above 1,250 ms in all conditions (compare Tables 1 and 3). This shows that our manipulation for making the foraging harder was effective. The main effect of condition was not significant, $F(2, 22) = 2.9$, $MSE = 37269$, $p = 0.08$, $\eta^2 = 0.21$, $BF_{01} = 1.35$, $p(H_0|D) = 0.57$, although evidence for the null effect was quite weak. Interestingly, the trend in the data is for responses on *vision-only* trials to be faster than *vision + sound* trials – arguing against a cross-modal advantage, a pattern quite different to the within-subject results of Experiment 2 (Fig. 5). We return to this point in the *Discussion* below.

Table 3 provides details on the ITT slope and intercept measures. As in Experiments 1 and 2, slope values were all positive and there were no significant differences between conditions, $F(2, 22) = 1.20$, $MSE = 39144$, $p = 0.35$, $\eta^2 = 0.09$, $BF_{01} = 7.67$, $p(H_0|D) = 0.88$. Consistent with the RT1 analysis, ITT intercepts were all considerably slower than in Experiments 1 and 2, reflecting the difficulty of the search task. The pattern across conditions was the same as in the previous experiments. There was a significant main effect of condition, $F(2, 22) = 6.98$, $MSE = 340808$, $p = 0.004$, $\eta^2 = 0.39$. Pairwise comparisons showed significant differences between the *asynchronous* condition and both the *vision-only*, $t = 3.47$, $p = 0.016$, and *vision + sound*, $t = 3.03$, $p = 0.03$, synchronous conditions. There was no difference between the *vision-only* and *vision + sound* conditions ($t = 0.38$, $p = 0.99$, $BF_{01} = 3.27$, $p(H_0|D) = 0.77$).

Figure 10 summarizes the number of targets found in Experiment 3, averaged across the 18 experimental trials

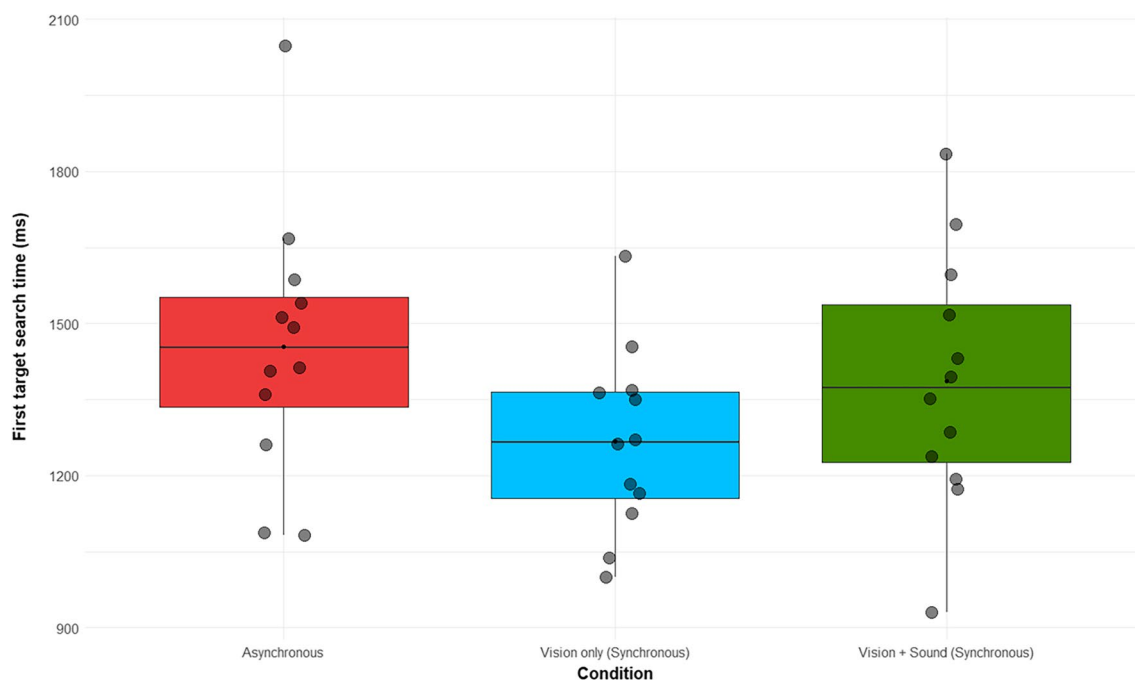


Fig. 9 The selection times for the first target (RT1) in Experiment 3. The horizontal lines in the box plots represent medians. The top horizontal line of the boxplots represent the upper quartile (75%), the bottom horizontal lines represent the lower quartile (25%). The top verti-

cal lines represent the maximum value, the bottom line represents the minimum value. The dots represent individual participants' median selection times

Table 3 Descriptive statistics for all dependent measures in Experiment 3. Standard errors are provided in brackets

	Asynchronous	Vision only (Synchronous)	Vision + Sound (Synchronous)
RT1 (ms)	1455 (75)	1268 (52)	1387 (72)
numTargets	9.84 (0.5)	12.27 (0.61)	11.91 (0.61)
ITT Intercept (ms)	822 (59)	639 (57)	669 (55)
ITT Slope (ms)	32 (16)	7 (15)	14 (10)
Best-r	0.69 (0.03)	0.64 (0.04)	0.66 (0.04)
Error rate %	3.05 (0.72)	2.34 (0.54)	1.8 (0.43)

Note. RT1 = response to first target; numTargets = number of targets found; ITT Intercept = estimate of per-item response time computed as the intercept of the linear regression function mapping median ITT across target position during the cruise phase of a trial (see text for details); ITT Slope = linear regression coefficient, estimating the rate of change in ITT across target position during the cruise phase of the trial; Best-r = a measure of search organisation computed as the average correlation coefficient (r) between the Cartesian coordinates of targets and the sequence in which they are selected; Error rate = % incorrect clicks divided by the sum of incorrect and correct clicks

in the three conditions. On average, participants were able to find a little over 10 items in 10 s, a much slower foraging rate than in Experiments 1 and 2, where items were collected every 600 ms and 500 ms, respectively. Again, this suggests that our target-difficulty manipulation was successful. More importantly, the pattern of results here mirrors that seen in Experiments 1 and 2, but now with a different dependent measure and a harder task. That is, we see better foraging performance in the *synchronous* compared to the *asynchronous* conditions, but notably no additional advantage to having a cross-modal, auditory cue to synchrony.

There was a significant main effect of condition, $F(2, 22) = 34$, $MSE = 0.6$, $p = 0.001$, $\eta^2 = 0.76$. Pairwise comparisons showed significant differences between the *asynchronous* condition and both the *vision-only*, $t = 9.11$, $p < 0.001$, and *vision + sound*, $t = 5.54$, $p < 0.001$, synchronous conditions. There was no difference between the *vision-only* and *vision + sound* conditions ($t = 1.17$, $p = 0.801$, $BF_{01} = 1.97$, $p(H_0|D) = 0.66$). While evidence for this latter null effect is relatively weak, we note that the trend in the data is for slightly better performance in the *vision-only* condition (see Fig. 10).

Best-r values were higher than seen in Experiments 1 and 2 (compare Tables 1, 2, and 3). This most likely reflects that the task in Experiment 3 was hard to complete, leading to more organized foraging (Jóhannesson et al., 2016). However, the effect of *synchrony* in Experiment 3 was in

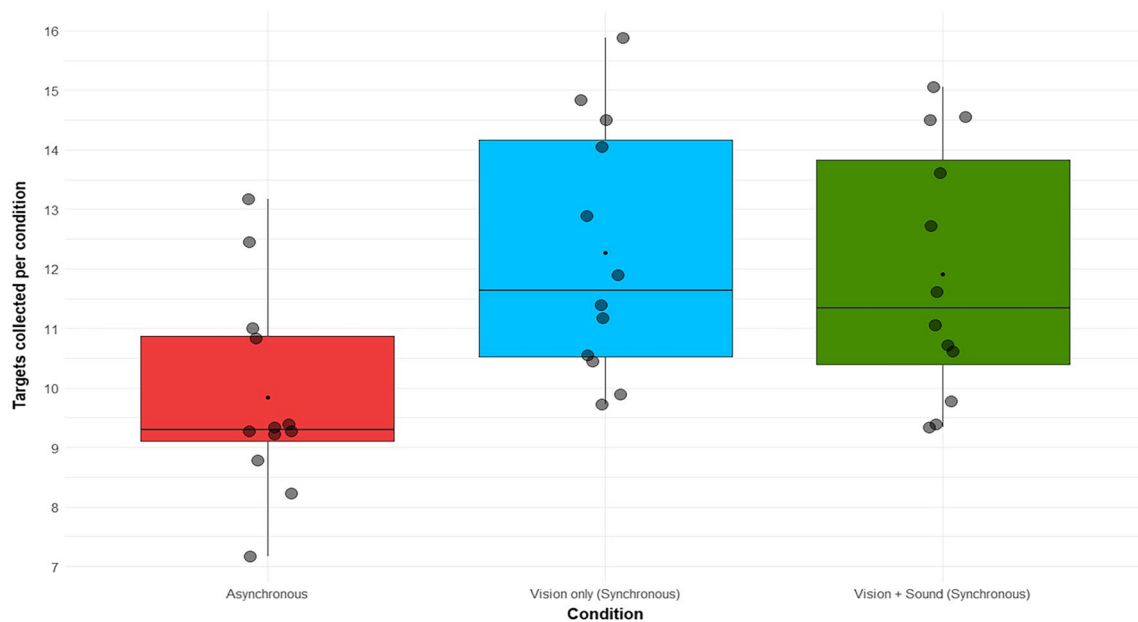


Fig. 10 Average number of targets collected per condition in Experiment 3. The horizontal lines in the box plots represent medians. The top horizontal lines of the boxplots represent the upper quartiles (75%), the bottom horizontal lines represent the lower quartiles

(25%). The top vertical lines represent the maximum value, the bottom lines represent the minimum value. The dots represent individual participants' median results

line with the results of Experiment 2, $F(2, 22) = 1.2$, $MSE = 0.042$, $p = 0.52$, $\eta^2 = 0.02$, $BF_{01} = 6.85$, $p(H_0|D) = 0.087$, with no significant difference between conditions. Error rates were slightly elevated compared to the previous study – to be expected given the increased difficulty of target detection – but were still overall relatively low (between 2% and 3%), and did not differ significantly as a function of condition (see Table 3 and supplementary analysis in OSM at osf.io/xu5rd/).

Discussion

The main question we asked in Experiment 3 was whether making foraging harder compared to previous experiments would facilitate the use of the available cross-modal auditory synchronization cue. We increased difficulty by using 45° lines – rather than 0° lines – as targets embedded among randomly oriented distractors, while imposing a 10-s time limit on each trial. These difficulty manipulations appear to have been successful, as both detection of the initial target item (RT1) and the subsequent foraging rates suggest that participants foraged much more slowly than in our other online experiment (Experiment 1).

However, in terms of our main dependent measure – the number of targets found within 10 s – we still found no evidence that adding an additional auditory cue to synchrony had any performance benefit. Rather, we replicated the previous two experiments in showing a general visual

synchrony advantage, with performance in the *vision-only* and the *vision + sound* conditions being essentially identical.

The pattern of RT1 results in Experiment 3 was quite different from the previous experiments. It seems likely that the lack of significant differences between conditions in Experiment 3 is related to the overall slowing of foraging, compared to Experiment 1, which was also run online. In addition to the increased difficulty of target detection, such slowing could also have been caused by the interleaved trial-types in Experiment 3. Participants may have needed additional time to establish the nature of each trial, prior to initiating the foraging. With slowed foraging times, differences between conditions could have become obscured.

Another explanation for the slower RT1 responses in Experiment 3 could be that participants used a different strategy than in Experiments 1 and 2, given the task change. Previous work from our group has shown that the first selection during foraging is often much slower when participants need to plan a sequence of movements (Thornton & Horowitz, 2004, 2020a). Here, participants may have tried to locate more than one target before commencing a sequence of more rapid responses, delaying the time of initial selection. Such a tendency may even have been more pronounced on *vision + sound* trials if participants tried to attend to the sound sequence. Of course, we can only speculate about the pattern of RT1 responses here, and as noted, we only reported them to allow overall speed comparisons with Experiments 1 and 2.

General discussion

In the current paper, we asked three main questions. First, can visual synchrony improve performance during multiple target foraging? Second, and relatedly, we asked *how* such potential improvement might come about. Would synchrony act as a unique feature, making it possible to *globally segregate* target and distractors, creating a form of pop-out? Or instead, would synchrony enhance *local grouping* of the targets, helping observers guide their next selection as they searched the display? Third, and finally, we investigated whether an additional *cross-modal* cue, in the form of a non-spatial sound occurring in synchrony with the linked color changes of targets, would provide further performance benefits in the context of multiple target foraging.

Regarding the first question, the answer seems very clear. In all three experiments we observed improved foraging performance on *synchronous* compared to *asynchronous* trials. In Experiments 1 and 2, these advantages emerged in how quickly an initial target was found, as well as overall completion time. In Experiment 3, participants were able to locate more targets during a fixed period with *synchronous* compared to *asynchronous* displays. Clearly, visually synchronized targets seem to improve multiple target foraging.

To answer our second question, it is useful to consider more closely the nature of the observed performance advantages. We had wondered whether synchrony between target items would globally segregate them from distractors, so that they would essentially “pop out.” This would predict a qualitative difference between foraging performance in the *synchronous* and *asynchronous* trials. For example, in Experiments 1 and 2, we would have expected extremely rapid RT1 responses, flatter ITT slopes, and relatively disorganized search in the *synchronous* conditions. In Experiment 3, we would have expected large differences in the number of collected items. But interestingly, this was not what we found. In Experiments 1 and 2, the RT1 response advantages were modest (54 ms and 70 ms, respectively), there were no differences in cruise-phase slopes, and search patterns were consistently organized. In Experiment 3, there were only an additional two items collected during synchronous trials. In general, then, foraging in *synchronous* conditions appeared to remain slow and effortful, albeit slightly faster than in the *asynchronous* condition.

While this lack of synchrony “pop out” may seem surprising – given evidence from the object recognition (Alais et al., 1998; Lee & Blake, 1999) and texture segregation (Blake & Lee, 2005; Fahle, 1993; Farid, 2002; Guttman et al., 2007; Leonards et al., 1996) literature discussed in

the *Introduction*, and in particular the claims made from the “pip-and-pop” phenomenon, it is nevertheless consistent with previous findings from the visual search literature (e.g., Cavanagh et al., 2001; Li et al., 2014; Thornton, Vuong, et al., 2021b). For example, Cavanagh et al. (2001) had participants find a synchronously moving point-light walker (Johansson, 1973) among scrambled distractors, and found highly inefficient search. Similarly, Li et al. (2014) found that a target element moving with a unique visual rhythm failed to pop out among asynchronously bouncing distractor elements. Most recently, Thornton et al. (2021b) examined search for horizontal or vertical target elements defined by groups of five dots with a unique temporal phase within a large field of randomly flickering dots. They again found highly inefficient search, although horizontal targets were found more quickly than vertical targets.

But if synchrony does not lead to automatic segregation of targets from distractors, how can we explain the performance advantages observed in the current work? Opting for the second answer to question 2 above, we suggest that our results are more consistent with the idea that synchrony helps via local grouping of targets. As we argued in the *Introduction*, this mechanism would still require participants to serially scan the display, but once an orientation target is selected, synchronized items in the local vicinity may be noticed, and attention could be guided towards those items. This pattern would repeat, leading to a connected series of responses. Our findings of positive cruise-phase ITT slopes and relatively organized search are both consistent with such locally guided foraging.

Our third question was whether cross-modal synchrony through a non-spatial sound might further modulate performance. Van der Burg et al. (2008) found large benefits for single-target visual search from a non-spatial sound that was synchronous with color changes in the target. They called this “pip and pop,” implying that sound can change the task domain from top-down to bottom-up processing – that the “pip” produces “pop out.”

But there were clearly no such benefits here. Firstly, the synchronous sound did not decrease total completion times (TCTs). We initially speculated that this was because observers foraged for multiple targets, rather than searching for only one target as in Van der Burg et al. (2008). But a problem with this explanation is that no benefit was seen even for the *first* selected target (RT1), which might be considered a single target search. Note that in a later study, Van der Burg, Awh, and Olivers (2013) argued that the binding of audition and vision could only occur for single items. In their study, 24 disks flashed asynchronously, while every 150 ms a randomly selected subset of one to eight discs changed polarity simultaneously in synchrony with a spatially uninformative sound. The task was to remember the target discs and determine whether a probe at trial end

indicated a target or distractor. But performance indicated that observers saw only one synchronized disk. Van der Burg et al. (2013) argued that this made ecological sense, since within natural settings, individual, object-related sounds come from a single source. But the question remains: why did we fail to see this benefit for a single item in the current experiments? One explanation could be that such binding of single items is prevented by grouping. But our results also mean that this binding does *not* occur for groups. The surprising hypothesis therefore emerges that grouping *eliminates* benefits from auditory synchrony.

We should also note that while the basic pip-and-pop effect has been replicated, there has been considerable discussion about underlying mechanisms. For example, some authors have argued that the pip-and-pop effect reflects improved oddball detection rather than temporal integration (Gao et al., 2021). Additionally, the effect is seemingly affected by top-down manipulations (Klapetek et al., 2012), so whether it counts as “bottom-up” attentional capture (or “pop-out”) is perhaps questionable.

Our results do, on the other hand, fall in line with those of Fujisaki et al. (2006), who found that detection of targets that changed in synchrony with auditory stimuli (either flashes that were presented synchronously with amplitude modulated “pips” or visual rotations that were in synch with frequency modulated auditory stimulation) became gradually worse as the number of unsynchronized visual distractors increased. They explained this in terms of separate sequential processing stages in perception (see Kristjánsson & Egeth, 2020; Sternberg, 1969), speculating that spatiotemporal synchrony benefits reflect serial processing, meaning that there is no “pop-out” from a “pip.” RTs in the “pip-and-pop” paradigm do in fact increase with set-size (the slopes of RT and set size ranged from ~30 to 60 ms per item with the sound cue; Collignon et al., 2013; Van der Burg et al., 2008), making it doubtful that potential benefits from the auditory cue cause pop-out.

More generally, it is known that the temporal zone of integration for sight and sound can be broad and may depend on various temporal and spatial variables (Arnold et al., 2005; Driver, 1996; Reisberg, 1978), making such integration potentially quite sensitive to the precise perceptual and attentional parameters of a given task (Talsma et al., 2010). Subtle temporal differences could, for example, determine whether such across-modality benefits are seen, which is a topic we will address in future experiments.

Another notable observation is that there were 20 targets among 80 distractors in our study. Multiple target foraging is not the sum of searches for one target (T. Kristjánsson et al., 2020; Wolfe et al., 2016), and effects like “pip and pop” may not *by necessity* transfer from one to the other. Also, given the very robust nature of the visual synchrony effects reported in all of the current experiments, it is possible that an additional

benefit from the introduction of sound is present, but is too small to be detected given the current sample sizes. While we find this somewhat unlikely – given the absence of any clear cross-modal trends in the data – our power analysis was based on pilot testing of the visual synchrony effect, so a small, but potentially detectable, cross-modal effect does remain a theoretical possibility.

Overall, then, the current experiments appear to reveal clear limits on the benefits of cross-modally synchronous presentation within the context of multiple-target search. While this may not, by itself, be any reason to pass strong judgement on the “pip-and-pop” effect, the lack of a cross-modal foraging benefit – particularly on the time it takes to find the first target – is nevertheless quite surprising. Note that we find it unlikely that our results reflect a simple ceiling effect, in that there was no *additional* benefit from sound even when task difficulty was substantially increased in Experiment 3. However, rather than varying target-distractor similarity, it might be more informative to manipulate difficulty in other ways. As vision is undoubtedly the dominant sense in our foraging tasks, potential benefits from sound may be more readily observed if the overall reliability of the visual input were to be reduced (Ernst & Banks, 2002; Gori et al., 2011). Further experiments involving manipulations of parameters that influence cue combination, such as with degraded visual information (making visual synchrony harder to detect), could certainly be of interest.

Conclusion

We asked whether synchrony in stimulation can guide attention and aid multiple target foraging. Additionally, we tested whether cross-modal synchrony can guide attention and aid foraging. Overall, we found a clear benefit of visual synchrony while there were no additional benefits of a synchronous sound in contrast with the predictions of the so-called “pip-and-pop” effect (Van der Burg et al., 2008). This pattern was seen in all three experiments. While the benefits from visual synchrony add another dimension to various spatial and temporal grouping effects seen before in the literature on vision, the lack of benefits for visual selection from auditory synchronous sounds are surprising, and may reveal severe limits on the performance benefits from such synchrony.

Acknowledgement The work of the first author was funded by Technology Development Fund of the Icelandic Research fund (project no. 1910271).

References

- Alais, D., Blake, R., & Lee, S.-H. (1998). Visual features that vary together over time group together over space. *Nature Neuroscience*, 1(2), 160–164. <https://doi.org/10.1038/414>

- Arnold, D. H., Johnston, A., & Nishida, S. (2005). Timing sight and sound. *Vision Research*, 45(10), 1275–1284. <https://doi.org/10.1016/j.visres.2004.11.014>
- Bella-Fernández, M., Suero Suárez, M., & Gil-Gómez De Liaño, B. (2023). The time course of visual foraging in the lifespan: Spatial scanning, organization search, and target processing. *Psychonomic Bulletin & Review*. <https://doi.org/10.3758/s13423-023-02345-8>
- Blake, R., & Lee, S.-H. (2005). The role of temporal structure in human vision. *Behavioral and Cognitive Neuroscience Reviews*, 4(1), 21–42. <https://doi.org/10.1177/1534582305276839>
- Bond, A. B. (1983). Visual search and selection of natural stimuli in the pigeon: The attention threshold hypothesis. *Journal of Experimental Psychology: Animal Behavior Processes*, 9(3), 292–306.
- Bond, A. B., & Kamil, A. C. (2006). Spatial heterogeneity, predator cognition, and the evolution of color polymorphism in virtual prey. *Proceedings of the National Academy of Sciences*, 103(9), 3214–3219. <https://doi.org/10.1073/pnas.0509963103>
- Bridges, D., Pitiot, A., MacAskill, M. R., & Peirce, J. W. (2020). The timing mega-study: Comparing a range of experiment generators, both lab-based and online. *PeerJ*, 8, e9414. <https://doi.org/10.7717/peerj.9414>
- Cavanagh, P., Labianca, A. T., & Thornton, I. M. (2001). Attention-based visual routines: Sprites. *Cognition*, 80(1–2), 47–60. [https://doi.org/10.1016/S0010-0277\(00\)00153-0](https://doi.org/10.1016/S0010-0277(00)00153-0)
- Chetverikov, A., Campana, G., & Kristjánsson, Á. (2016). Building ensemble representations: How the shape of preceding distractor distributions affects visual search. *Cognition*, 153, 196–210. <https://doi.org/10.1016/j.cognition.2016.04.018>
- Chetverikov, A., & Upravitelev, P. (2016). Online versus offline: The Web as a medium for response time data collection. *Behavior Research Methods*, 48(3), 1086–1099. <https://doi.org/10.3758/s13428-015-0632-x>
- Collignon, O., Dormal, G., Albouy, G., Vandewalle, G., Voss, P., Phillips, C., & Lepore, F. (2013). Impact of blindness onset on the functional organization and the connectivity of the occipital cortex. *Brain*, 136(9), 2769–2783. <https://doi.org/10.1093/brain/awt176>
- Dawkins, M. S. (1983). Battery hens name their price: Consumer demand theory and the measurement of ethological ‘needs.’ *Animal Behaviour*, 31(4), 1195–1205. [https://doi.org/10.1016/S0003-3472\(83\)80026-8](https://doi.org/10.1016/S0003-3472(83)80026-8)
- Driver, J. (1996). Enhancement of selective listening by illusory mislocation of speech sounds due to lip-reading. *Nature*, 381(6577), 66–68. <https://doi.org/10.1038/381066a0>
- Driver, J., & Spence, C. (1998). Attention and the crossmodal construction of space. *Trends in Cognitive Sciences*, 2(7), 254–262. [https://doi.org/10.1016/S1364-6613\(98\)01188-7](https://doi.org/10.1016/S1364-6613(98)01188-7)
- Dukas, R., & Elner, S. (1993). Information processing and prey detection. *Ecology*, 74(5), 1337–1346. <https://doi.org/10.2307/1940064>
- Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, 96(3), 433–458. <https://doi.org/10.1037/0033-295x.96.3.433>
- Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415(6870), 429–433. <https://doi.org/10.1038/415429a>
- Fahle, M. (1993). Figure–ground discrimination from temporal information. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 254(1341), 199–203. <https://doi.org/10.1098/rspb.1993.0146>
- Farid, H. (2002). Temporal synchrony in perceptual grouping: A critique. *Trends in Cognitive Sciences*, 6(7), 284–288. [https://doi.org/10.1016/S1364-6613\(02\)01927-7](https://doi.org/10.1016/S1364-6613(02)01927-7)
- 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), 1149–1160. <https://doi.org/10.3758/BRM.41.4.1149>
- Fujisaki, W., Koene, A., Arnold, D., Johnston, A., & Nishida, S. (2006). Visual search for a target changing in synchrony with an auditory signal. *Proceedings of the Royal Society B: Biological Sciences*, 273(1588), 865–874. <https://doi.org/10.1098/rspb.2005.3327>
- Fujisaki, W., & Nishida, S. (2007). Feature-based processing of audio-visual synchrony perception revealed by random pulse trains. *Vision Research*, 47(8), 1075–1093. <https://doi.org/10.1016/j.visres.2007.01.021>
- Gao, M., Chang, R., Wang, A., Zhang, M., Cheng, Z., Li, Q., & Tang, X. (2021). Which can explain the pip-and-pop effect during a visual search: Multisensory integration or the oddball effect? *Journal of Experimental Psychology: Human Perception and Performance*, 47(5), 689–703. <https://doi.org/10.1037/xhp0000905>
- Gil-Gómez De Liaño, B., & Wolfe, J. M. (2022). The FORAGEKID Game: Hybrid-Foraging as a new way to study aspects of executive function in development. *Cognitive Development*, 64, 101233. <https://doi.org/10.1016/j.cogdev.2022.101233>
- Gilliam, J. F., & Fraser, D. F. (1987). Habitat selection under predation hazard: Test of a model with foraging minnows. *Ecology*, 68(6), 1856–1862. <https://doi.org/10.2307/1939877>
- Gori, M., Mazzilli, G., Sandini, G., & Burr, D. (2011). Cross-sensory facilitation reveals neural interactions between visual and tactile motion in humans. *Frontiers in Psychology*, 2. <https://doi.org/10.3389/fpsyg.2011.00055>
- Guttman, S. E., Gilroy, L. A., & Blake, R. (2007). Spatial grouping in human vision: Temporal structure trumps temporal synchrony. *Vision Research*, 47(2), 219–230. <https://doi.org/10.1016/j.visres.2006.09.012>
- Ho, C., & Spence, C. (2005). Assessing the effectiveness of various auditory cues in capturing a driver’s visual attention. *Journal of Experimental Psychology: Applied*, 11(3), 157–174. <https://doi.org/10.1037/1076-898X.11.3.157>
- Ho, C., Spence, C., & Gray, R. (2013). Looming auditory and vibrotactile collision warning for safe driving. *Driving Assessment Conference*, 7. <https://doi.org/10.17077/drivingassessment.1540>
- Horowitz, T. S., & Thornton, I. M. (2008). Objects or locations in vision for action? Evidence from the MILO task. *Visual Cognition*, 16(4), 486–513. <https://doi.org/10.1080/13506280601087356>
- Jeffreys, H. (1998). *Theory of Probability*. Clarendon Press. https://books.google.is/books?id=_PuRmAEACAAJ
- Jóhannesson, Ó. I., Kristjánsson, Á., & Thornton, I. M. (2017). Are foraging patterns in humans related to working memory and inhibitory control? Foraging patterns, working memory, and inhibitory control. *Japanese Psychological Research*, 59(2), 152–166. <https://doi.org/10.1111/jpr.12152>
- Jóhannesson, Ó. I., Thornton, I. M., Smith, I. J., Chetverikov, A., & Kristjánsson, Á. (2016). Visual foraging with fingers and eye gaze. *I-Perception*, 7(2), 204166951663727. <https://doi.org/10.1177/2041669516637279>
- Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics*, 14(2), 201–211. <https://doi.org/10.3758/BF03212378>
- Kassambara A. (2021). *rstatix: Pipe-Friendly Framework for Basic Statistical Tests* [Computer software]. <https://cran.r-project.org/web/packages/rstatix/index.html>
- Klapetek, A., Ngo, M. K., & Spence, C. (2012). Does crossmodal correspondence modulate the facilitatory effect of auditory cues on visual search? *Attention, Perception, & Psychophysics*, 74(6), 1154–1167. <https://doi.org/10.3758/s13414-012-0317-9>
- Kong, G., Alais, D., & Van der Burg, E. (2017). Orientation categories used in guidance of attention in visual search can differ in strength. *Attention, Perception, & Psychophysics*, 79(8), 2246–2256. <https://doi.org/10.3758/s13414-017-1387-5>
- Kristjánsson, Á., Björnsson, A. S., & Kristjánsson, T. (2020). Foraging with Anne Treisman: Features versus conjunctions, patch leaving and memory for foraged locations. *Attention, Perception,*

- & Psychophysics, 82(2), 818–831. <https://doi.org/10.3758/s13414-019-01941-y>
- Kristjánsson, Á., & Draschkow, D. (2021). Keeping it real: Looking beyond capacity limits in visual cognition. *Attention, Perception, & Psychophysics*, 83(4), 1375–1390. <https://doi.org/10.3758/s13414-021-02256-7>
- Kristjánsson, Á., & Egeth, H. (2020). How feature integration theory integrated cognitive psychology, neurophysiology, and psychophysics. *Attention, Perception, & Psychophysics*, 82(1), 7–23. <https://doi.org/10.3758/s13414-019-01803-7>
- Kristjánsson, Á., Jóhannesson, Ó. I., & Thornton, I. M. (2014). Common attentional constraints in visual foraging. *PLoS ONE*, 9(6), e100752. <https://doi.org/10.1371/journal.pone.0100752>
- Kristjánsson, Á., Ólafsdóttir, I. M., & Kristjánsson, T. (2019). Visual foraging tasks provide new insights into the orienting of visual attention: Methodological considerations. In S. Pollmann (Ed.), *Spatial Learning and Attention Guidance* (Vol. 151, pp. 3–21). Springer US. https://doi.org/10.1007/7657_2019_21
- Kristjánsson, T., Draschkow, D., Pálsson, Á., Haraldsson, D., Jónsson, P. Ö., & Kristjánsson, Á. (2022). Moving foraging into three dimensions: Feature- versus conjunction-based foraging in virtual reality. *Quarterly Journal of Experimental Psychology*, 75(2), 313–327. <https://doi.org/10.1177/1747021820937020>
- Kristjánsson, T., Páll Thorvaldsson, T., & Kristjánsson, Á. (2014). Divided multimodal attention: Sensory trace and context coding strategies in spatially congruent auditory and visual presentation. *Multisensory Research*, 27(2), 91–110. <https://doi.org/10.1163/22134808-00002448>
- Kristjánsson, T., Thornton, I. M., & Kristjánsson, Á. (2018). Time limits during visual foraging reveal flexible working memory templates. *Journal of Experimental Psychology: Human Perception and Performance*, 44(6), 827–835. <https://doi.org/10.1037/xhp0000517>
- Lawrence MA. (2016). *ez: Easy Analysis and Visualization of Factorial Experiments*. R package version 4.4-0 [Computer software]. <https://CRAN.R-project.org/package=ez>
- Lee, S.-H., & Blake, R. (1999). Visual form created solely from temporal structure. *Science*, 284(5417), 1165–1168. <https://doi.org/10.1126/science.284.5417.1165>
- Leonards, U., Singer, W., & Fahle, M. (1996). The influence of temporal phase differences on texture segmentation. *Vision Research*, 36(17), 2689–2697. [https://doi.org/10.1016/0042-6989\(96\)86829-5](https://doi.org/10.1016/0042-6989(96)86829-5)
- Li, H., Bao, Y., Pöppel, E., & Su, Y.-H. (2014). A unique visual rhythm does not pop out. *Cognitive Processing*, 15(1), 93–97. <https://doi.org/10.1007/s10339-013-0581-1>
- Mark, V. W., Woods, A. J., Ball, K. K., Roth, D. L., & Mennemeier, M. (2004). Disorganized search on cancellation is not a consequence of neglect. *Neurology*, 63(1), 78–84. <https://doi.org/10.1212/01.WNL.0000131947.08670.D4>
- Masson, M. E. J. (2011). A tutorial on a practical Bayesian alternative to null-hypothesis significance testing. *Behavior Research Methods*, 43(3), 679–690. <https://doi.org/10.3758/s13428-010-0049-5>
- Ngo, M. K., & Spence, C. (2010). Crossmodal facilitation of masked visual target identification. *Attention, Perception & Psychophysics*, 72(7), 1938–1947. <https://doi.org/10.3758/APP.72.7.1938>
- Ólafsdóttir, I. M., Gestsdóttir, S., & Kristjánsson, Á. (2020). Age differences in foraging and executive functions: A cross-sectional study. *Journal of Experimental Child Psychology*, 198, 104910. <https://doi.org/10.1016/j.jecp.2020.104910>
- Ólafsdóttir, I. M., Gestsdóttir, S., & Kristjánsson, Á. (2021). The development of foraging organization. *Attention, Perception, & Psychophysics*, 83(7), 2891–2904. <https://doi.org/10.3758/s13414-021-02328-8>
- Prpic, V., Kniesed, I., Camilleri, E., Maureira, M. G., Kristjánsson, Á., & Thornton, I. M. (2019). A serious game to explore human foraging in a 3D environment. *PLoS ONE*, 14(7), e0219827. <https://doi.org/10.1371/journal.pone.0219827>
- Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, 25, 111. <https://doi.org/10.2307/271063>
- Reisberg, D. (1978). Looking where you listen: Visual cues and auditory attention. *Acta Psychologica*, 42(4), 331–341. [https://doi.org/10.1016/0001-6918\(78\)90007-0](https://doi.org/10.1016/0001-6918(78)90007-0)
- RStudio Team. (2020). *RStudio: Integrated Development for R* [Computer software]. <http://www.rstudio.com/>
- Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica*, 30, 276–315. [https://doi.org/10.1016/0001-6918\(69\)90055-9](https://doi.org/10.1016/0001-6918(69)90055-9)
- Tagu, J., & Kristjánsson, Á. (2022). Dynamics of attentional and oculomotor orienting in visual foraging tasks. *Quarterly Journal of Experimental Psychology*, 75(2), 260–276. <https://doi.org/10.1177/1747021820919351>
- Talsma, D., Senkowski, D., Soto-Faraco, S., & Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration. *Trends in Cognitive Sciences*, 14(9), 400–410. <https://doi.org/10.1016/j.tics.2010.06.008>
- Thornton, I. M., de'Sperati, C., & Kristjánsson, Á. (2019). The influence of selection modality, display dynamics and error feedback on patterns of human foraging. *Visual Cognition*, 27(5–8), 626–648. <https://doi.org/10.1080/13506285.2019.1658001>
- Thornton, I. M., & Horowitz, T. S. (2004). The multi-item localization (MILO) task: Measuring the spatiotemporal context of vision for action. *Perception & Psychophysics*, 66(1), 38–50. <https://doi.org/10.3758/BF03194859>
- Thornton, I. M., & Horowitz, T. S. (2020a). MILO Mobile: An iPad app to measure search performance in multi-target sequences. *I-Perception*, 11(3), 204166952093258. <https://doi.org/10.1177/2041669520932587>
- Thornton, I. M., & Horowitz, T. S. (2020b). Searching through alternating sequences: Working memory and inhibitory tagging mechanisms revealed using the MILO task. *I-Perception*, 11(5), 204166952095801. <https://doi.org/10.1177/2041669520958018>
- Thornton, I. M., Nguyen, T. T., & Kristjánsson, Á. (2020). Foraging tempo: Human run patterns in multiple-target search are constrained by the rate of successive responses. *Quarterly Journal of Experimental Psychology*, 75(2), 297–312. <https://doi.org/10.1177/1747021820961640>
- Thornton, I. M., Nguyen, T. T., & Kristjánsson, Á. (2022). Foraging tempo: Human run patterns in multiple-target search are constrained by the rate of successive responses. *Quarterly Journal of Experimental Psychology*, 75(2), 297–312. <https://doi.org/10.1177/1747021820961640>
- Thornton, I. M., Tagu, J., Zdravković, S., & Kristjánsson, Á. (2021a). The Predation Game: Does dividing attention affect patterns of human foraging? *Cognitive Research: Principles and Implications*, 6(1), 35. <https://doi.org/10.1186/s41235-021-00299-w>
- Thornton, I. M., Vuong, Q. C., & Pilz, K. S. (2021b). A search advantage for horizontal targets in dynamic displays. *I-Perception*, 12(2), 204166952110046. <https://doi.org/10.1177/20416695211004616>
- Treisman, A. (1999). Solutions to the Binding Problem. *Neuron*, 24(1), 105–125. [https://doi.org/10.1016/s0896-6273\(00\)80826-0](https://doi.org/10.1016/s0896-6273(00)80826-0)
- Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12(1), 97–136. [https://doi.org/10.1016/0010-0285\(80\)90005-5](https://doi.org/10.1016/0010-0285(80)90005-5)
- Van der Burg, E., Awh, E., & Olivers, C. N. L. (2013). The Capacity of Audiovisual Integration Is Limited to One Item. *Psychological Science*, 24(3), 345–351. <https://doi.org/10.1177/0956797612452865>

- Van der Burg, E., Olivers, C. N. L., Bronkhorst, A. W., & Theeuwes, J. (2008). Pip and pop: Nonspatial auditory signals improve spatial visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 34(5), 1053–1065. <https://doi.org/10.1037/0096-1523.34.5.1053>
- Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., & von der Heydt, R. (2012). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure-ground organization. *Psychological Bulletin*, 138(6), 1172–1217. <https://doi.org/10.1037/a0029333>
- Wagenmakers, E.-J., Wetzels, R., Borsboom, D., & van der Maas, H. L. J. (2011). Why psychologists must change the way they analyze their data: The case of psi: Comment on Bem (2011). *Journal of Personality and Social Psychology*, 100(3), 426–432. <https://doi.org/10.1037/a0022790>
- Wertheimer, M. (1923). Untersuchungen zur Lehre von der Gestalt. II. *Psychologische Forschung*, 4(1), 301–350. <https://doi.org/10.1007/BF00410640>
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., ... Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, 4(43), 1686. <https://doi.org/10.21105/joss.01686>
- Wolfe, J. M., Aizenman, A. M., Boettcher, S. E. P., & Cain, M. S. (2016). Hybrid foraging search: Searching for multiple instances of multiple types of target. *Vision Research*, 119, 50–59. <https://doi.org/10.1016/j.visres.2015.12.006>
- Wolfe, J. M., Friedman-Hill, S. R., Stewart, M. I., & O'Connell, K. M. (1982). The role of categorization in visual search for orientation. *Journal of Experimental Psychology: Human Perception and Performance*, 18(1), 34–49. <https://doi.org/10.1037/0096-1523.18.1.34>
- Wolfe, J. M., & Horowitz, T. S. (2017). Five factors that guide attention in visual search. *Nature Human Behaviour*, 1(3), 0058. <https://doi.org/10.1038/s41562-017-0058>
- Woods, A. J., Göksun, T., Chatterjee, A., Zelonis, S., Mehta, A., & Smith, S. E. (2013). The development of organized visual search. *Acta Psychologica*, 143(2), 191–199. <https://doi.org/10.1016/j.actpsy.2013.03.008>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.