Time-resolved hierarchical frequency-tagging reveals markers of predictive processing in the action-perception loop

Roger Koenig-Robert[1], Thomas Pace[1], Joel Pearson[1], and Jakob Hohwy[2]

[1]University of New South Wales
[2]Monash University School of Psychological Sciences

January 10, 2023

Abstract

In everyday life, we use perception to guide our behaviour. While much effort has been devoted to neurophysiologically study perception and behaviour in isolation from each other, studies that conjoin perception and behaviour are rarer. Here, we devised a novel paradigm to dynamically study the action-perception loop, framed in terms of predictive processing as a guiding framework for brain function. We tracked the electrophysiological markers of predictive processing by using hierarchical frequency-tagging in an active foraging and recognition task. Participants had to forage a two-dimensional landscape to find three target images. They freely selected their foraging paths and when to finish and move to the next landscape. Temporally resolved analyses of hierarchical frequency-tagging signals revealed that putative prediction error signals triggered a cascade of neural signalling events leading to recognition. In addition, our results show that the accumulation of uncertainty is correlated with the decision to abort foraging and start a new search. For the first time, we tracked temporally-resolved frequency-tagged signals in an action-perception paradigm; this is consistent with contemporary iterations of predictive processing that increasingly focus on action (active inference). Our paradigm and findings open new ways to study such signals during the action-perception cycle beyond passive settings.

INTRODUCTION

Sensory information is used to inform our behaviour. A number of proposals have posited that brains evolved to predict interactions between behaving agents and their environment (Llinás, 2001; Pezzulo et al., 2022). Predictive processing theories of brain function (Friston, 2010; Hohwy, 2013) are useful frameworks to understand perception and behaviour in conjunction as they formally articulate how behaviour is updated to consider new sensory information. According to predictive processing theories, the brain produces internal generative models to predict outcomes of behaviour and perception, which in turn are continually updated by current behaviour and perception, thus providing an explanatory framework to link both perception and action. Under this explanatory framework, updating of the internal models is indicated through the different signalling channels used to pass sensory information, predictions generated by the model and the mismatch between them. These signalling channels are known as *sensory evidence,predictions* and *prediction errors* , respectively (Figure 1). In neurophysiological terms, sensory evidence corresponds to ascending neural signals emerging from sensory areas (Ahissar & Hochstein, 2004), predictions correspond to descending feedback signals from higher hierarchical areas (Bastos et al., 2012), and prediction error corresponds to signals generated by the interaction and comparison between these signals at each hierarchical level (Wacongne et al., 2012). Thus, when the internal model generates accurate predictions of the sensory input, prediction error signals should be low, and, inversely, when sensory input is incongruent with the model, prediction error signals should be high. Low prediction error then determines accurate perceptual inference of the states of the world (Friston et al., 2017; Rao & Ballard, 1999).

Key elements of predictive processing have received empirical support from many independent studies (Aitken

et al., 2020; Aitken & Kok, 2022; Alamia & VanRullen, 2019; Johnston et al., 2017; Kok et al., 2017). In particular, it has been shown that the gain of sensory signals depends on their degree of compatibility with inferred internal models (Kumar et al., 2021; Robinson et al., 2018; Tang et al., 2018). Experimental results have shown that sensory information that has not been considered by the inferred internal models (in other words, *surprising sensory information*) is enhanced compared to expected information. These *prediction error* signals, as hypothesized by sensory inference (Parr & Friston, 2019), can thus be used as a proxy to tell how well internal models predict perceptual information. To gain a better understanding of these signals beyond prediction error alone, we have developed a technique that tracks the main signalling components of predictive processing: sensory evidence, predictions and prediction error (Gordon et al., 2017). In short, this hierarchical frequency-tagging (HFT) technique uses SWIFT (Semantic Wavelet Induced Frequency Tagging), to track high-level visual representations (Koenig-Robert & VanRullen, 2013), which, under HFT, are considered to be encoding stimuli-specific predictions (Gordon et al., 2017, 2019). HFT also tracks lower visual representations, or sensory evidence, with SSVEP (Steady State Visual Evoked Potentials) (Regan, 1977). To track prediction error signals, HFT uses the non-linear interactions between the tagging frequencies of SWIFT and SSVEP (Gordon et al., 2017). Importantly, this method is constructed in such a way that low-level image attributes are conserved across time (Gordon et al., 2017; Koenig-Robert & VanRullen, 2013), thus ensuring that the effects found are not the result of changes in low-level visual input. Using HFT, studies have shown that prediction error related signals are modulated by expectation (Gordon et al., 2017), attention (Gordon et al., 2019) and autistic traits (Coll et al., 2020), thus experimentally testing this technique as well as the behaviour of putative key signalling elements of the predicting coding theory within passive perceptual settings (see Figure 1).

However, these studies as well as many others, have focused on studying perception in isolation. Despite the ecological importance of studying perception and action concurrently, as they naturally occur in everyday life, much of what we know about perception and perceptual inferences has come from studies using passive perceptual paradigms. Therefore, less is known about how predictive processing signals are modulated when participants are allowed to freely close the action-perception loop. The predictive processing framework itself is rapidly transitioning to a main emphasis on action and decision-making, including for perception, framed in terms of active inference (Parr et al., 2022); this development implies that prediction error message passing arises as agents close the action-perception loop. As mentioned, prior neurophysiological studies have been conducted in passive tasks, so there is a gap in our knowledge between predictive processing in the active inference framing and empirical studies of the conditions under which prediction error signals occur.

In this study, we contribute to closing this gap in knowledge about predictive processing by testing hypotheses, in a more ecologically valid setting, about the interplay among sensory signals during active foraging and error minimisation (Friston et al., 2017; Schwartenbeck et al., 2019). We were interested in tracking predictive processing signals while participants freely foraged a landscape to find targets, expecting to identify neural signatures of predictions, sensory evidence and prediction errors. Specifically, we reasoned that participants' motivation for foraging should be the guided by the imperative to minimize uncertainty about the location of the targets. Thus, the working assumption of the participants is that the more they forage for the images (i.e., explore the landscape), the more certainty they should have about the target identities and their location, as they find them. A violation of these assumptions would consist in the inability of finding targets after a period of foraging, which should be associated with curtailing foraging. In this circumstance, the hypothesis is that it would be attractive for the participant to decide to shift to a new foraging patch. On the other hand, decreased prediction error should be associated with extended foraging activity, as expectations match behaviour and perception (Mirza et al., 2018; Schwartenbeck et al., 2013).

In our paradigm, participants foraged a two-dimensional landscape (Figure 2) with the aim of finding three target images (represented by 3 orange blobs on Figure 2) embedded in dynamic visual noise (black background). We had specific hypotheses about the responses tracked by HFT (Figure 1). Steady state visual evoked potentials (SSVEP) have been shown to represent incoming sensory signals as their sources have been identified predominantly in early visual cortex (Tsoneva et al., 2021). Incoming sensory signals have been shown to be modulated by its precision or, in other words, their reliability (Den Ouden et al., 2012;

Kok et al., 2012), therefore, we expect SSVEP signals to be modulated positively by reliability of sensory information in our paradigm (Figure 1). Semantic wavelet induced frequency tagging (SWIFT), on the other hand, have been identified as correlates of high-level object representations (Koenig-Robert et al., 2015; Koenig-Robert & VanRullen, 2013). Within HFT, part of the signals tagged with SWIFT are considered as predictions feeding back from higher visual areas to lower ones (especially those integrated with SSVEP signals to produce intermodulation products). While there is not direct evidence testing this assumption, previous studies have given indirect experimental support for their role as predictions (Gordon et al., 2017, 2019). Putative prediction signals tracked with SWIFT should then increase after recognition, as object representations are only available once the image is found and recognized (Koenig-Robert & VanRullen, 2013). Finally, intermodulation frequencies (IM), as putatively indicating prediction error signals, should be modulated by the mismatch between expectations and sensory evidence or, in other words, *surprise* . The intensity of IM have been shown be *inversely correlated* with prediction errors (Coll et al., 2020; Gordon et al., 2017, 2019). That is, IM intensity is stronger when prediction errors are low and weaker when prediction errors are high.

We found that HFT signals corresponding to putative prediction errors triggered a cascade of neural signalling events leading to recognition. In addition, our results show that the decision to stop foraging and start a new search is correlated with an increase in uncertainty in the form of higher prediction error signals. For the first time, we tracked HFT signals in a time resolved manner, thus uncovering the temporal dynamics of perceptual inference during the action-perception loop. These results shed light into the neural signal dynamics during perception and action and are consistent with the predictive processing framework as it moves to an active inference framing.

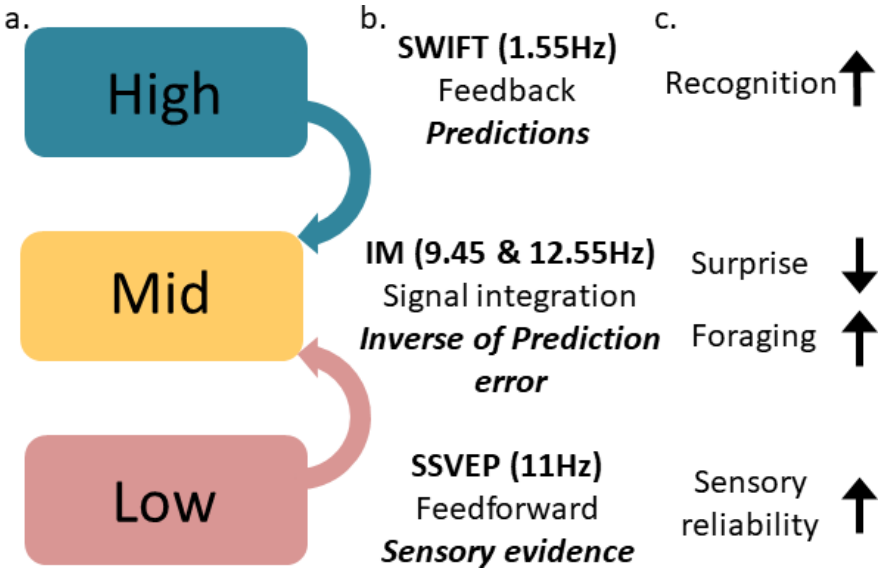


Figure 1. Brain signals tracked by hierarchical frequency-tagging. **A.** Schema of the different hierarchical visual levels. Note that while these hierarchies can be represented at different scales (such as cortical layers), here, they represent different vision-related brain areas. **B.** The different elements of hierarchical frequency-tagging (in bold) are aimed at tagging feedforward, feedback and the integration of both signals and the putative components according to predictive coding (in italics). The assumed role of these putative signals within the predictive processing framework have received experimental support (Gordon et al., 2017, 2019). The frequency tagged components are semantic wavelet-induced frequency tagging (SWIFT at 1.55Hz), steady-state visual evoked potentials (SSVEP at 11Hz) and the intermodulation frequencies (IM at 9.45 and 12.55Hz) produced by the non-linear interaction of these two tagging-frequencies. **C.** Cognitive factors modulating (increasing or decreasing) the strength of different signals tracked by hierarchical frequency-

tagging. While surprise should transiently increase prediction error, foraging should lead to its decrease in the long run. Note that the strength of the intermodulation frequencies represents the *inverse of the prediction error.* Thus, the strength of the IM increases when prediction error decreases and vice versa.

METHODS

Participants. Experimental procedures were approved by the University of New South Wales Human Research Ethics Committee (HREC#: HC12030). All methods in this study were performed in accordance with the guidelines and regulations from the Australian National Statement on Ethical Conduct in Human Research (https://www.nhmrc.gov.au/guidelines-publications/e72). All participants gave informed written consent to participate in the experiment. We tested a total of 20 participants. Three participants were discarded due to issues in the signal acquisition and two due to excessive artefacts in the EEG data. Finally, 15 participants, 4 women, aged 29.6±1.4 years old (mean±SEM) were retained for analyses. We choose the sample size based on previous studies using hierarchical frequency tagging (Coll et al., 2020; Gordon et al., 2017, 2019). Post-hoc power analyses revealed that the power achieved for our principal analysis (dynamic tracking of PC components, Figure 4) ranged between 1-β = 0.45 and 0.94, among the different frequencies of interest and timepoints.

Paradigm. The paradigm consisted of an interactive foraging game. Participants had to forage a two-dimensional landscape to find three images embedded in dynamic visual noise (Fig 2). In each trial of a total of 15, three unique images were positioned pseudo randomly in the landscape (see stimuli construction for details). The total amount of time that a participant could spend foraging across all 15 trials (divided in 5 blocks of 3 trials each) was 60min, therefore imposing a time trade-off between finding all the images in a trial and finishing all the trials in the experiment. We used a system of points to encourage participants to forage and in a more general way, reduce uncertainty (see details below). In each trial, the image triads, which composed the targets, were chosen such that they represented three exemplars of a common semantic category (see Table 1 for details). Participants pressed the arrows on a computer keyboard to forage for the target images within the landscape. Eight movement directions were allowed: four Cartesian directions plus diagonal orientations (achieved by pressing two contiguous arrows). While foraging, an arrowhead indicating the direction of movement was shown at fixation. Once the edge of the landscape was reached, a line indicated that further movement in that direction was not possible. Participants thus foraged the landscape in a freely chosen manner, which defined a foraging path for every trial that was recorded for further analysis (represented by a red line on Fig 2). As soon as participants found and recognized a target image, they pressed the space bar, indicating the discovery. Participants could abort the trial at any point, which could be before or after finding the 3 target images. In different trials, participants foraged the landscape to different extents. As an objective measure of foraging, we measured the path length (in pixels) in a given trial. Thus, the longer the path, the more the extent of the foraging. At the end of the trial, participants had to choose each image they found among a list of 4 distractors (this was repeated 3 times so all the target image labels were shown; see table 1 for details). Participants were instructed not to guess their answers. In order to encourage participants to find target images, we used a scoring system to reward (+10 points) correct recognition while also penalizing (-10 points) incorrect recognition to discourage guessing. If participants were not sure or had not seen a particular target image, they could choose option 6. "None", which led to no change in their score. Three prompts, one for each target image, were shown. After providing answers for the labels of the target images, participants had to choose the common concept shared by the triad. This was done to encourage participants to search for as many target images as possible, since the common concept becomes less uncertain as more images are found. After each target image and concept question (answered differently than "none"), participants were asked to rate their confidence from 1 to 5. Common concept labelling was rewarded or penalized with +/- 15 points. A "Current score" prompt showed the points earned across the entire experiment. After completing all questions, a prompt showing the remaining exploration time in the experiment was shown. The answering time (Figure 2, B to E) did not count toward the remaining exploration time, thus participants could have breaks during these prompts.

Stimuli construction. The key of our paradigm was to equalize the principal low-level image features

through the trial while tracking the different signal components of predictive coding using frequency-tagging. This allowed us to track genuine changes in the signalling strength rather than changes due to low-level image attributes. To do this, we employed hierarchical frequency-tagging (HFT) to track putative bottom-up, top-down and the interaction of both signals (Gordon et al., 2017). Full details of the HFT technique can be found elsewhere (Gordon et al., 2017; Koenig-Robert & VanRullen, 2013). In short, grey scale natural images were cyclically scrambled (at 1.55Hz) in the wavelet domain using semantic-wavelet image frequency-tagging (SWIFT) to track the higher-level representations of the images (putative predictions) using 3 individual scrambling cycles randomly selected during the presentation to avoid low-level responses at the tagging frequency (Koenig-Robert & VanRullen, 2013). On the other hand, to primarily track sensory input signals, we sinusoidally modulated (at 11Hz) the opacity (alpha channel) envelope of SWIFT sequences thus triggering steady-state visual evoked potentials (SSVEP) (Regan, 1982). The interaction of these signals was measured by tracking the second order intermodulation products (at 9.45 and 12.55Hz). While foraging the landscape, participants saw either dynamic visual noise (constructed using SWIFT) composed by an equal mix of the wavelet scrambled version of the three natural images, or one of the target images blended with the scrambled version of the other two (Figure 2, panels 1 to 4). Thus, the aim of participants was to move from areas containing only noise (plateau on Figure 2) to those containing the target images (peaks on Figure 2). The three target images were represented by 3 individual and non-connecting patches on the landscape. These patches were built by mapping each image into a portion of the landscape by randomly picking one pixel per image in a seed matrix filled with zeroes obeying two rules: the pixels cannot be contiguous and they cannot be part the edges. Zeroes were switched to ones at these 3 locations. After this, the seed matrix was resized into the landscape dimensions using spline interpolation thus producing smooth transitions from noise areas to target image areas.

EEG signal pre-processing. Electroencephalogram signals were recorded from 64 active scalp electrodes following the 10/20 placement using the Biosemi ActiveTwo system (Biosemi, The Netherlands). Signals were acquired at 1024 samples per second. We imported the data into MATLAB (R2015a, The Mathworks, Natick, USA) using the EEGLab toolbox v13.4.4b, (Delorme & Makeig, 2004) and the BIOSIG plugin. Noisy channels (> 400 μV) were replaced by a spline interpolation of neighbour channels using Andreas Widmann "repchan" function. To avoid slow drifts of the signal, we high-pass filtered the data at 0.5Hz, using two-way least-squares FIR filtering as implemented in EEGFILT in EEGLab. We then epoched the continuous data into individual trials. Note that trials had different lengths as participants were free to end trials at any time. We finally removed linear trends from the data before saving each trial in a file. From every trial, we then extracted epochs for every recognition report (from -10 to 10s). Additional detrending was then performed and data were saved for further analysis.

Tagging-frequencies analysis. We performed a fast Fourier transform (FFT) in the time domain data in order to extract the power at each temporal frequency. For each tagging-frequency and intermodulation products, we zero padded recognition epochs to the next integer multiple of their periods. We calculated the signal-to-noise ratio (SNR) in the frequency domain by dividing each frequency (signal) for the average of their neighbours (noise, 0.15Hz half bandwidth) and took the $\log_e$ of it. We selected a ROI for each of the four frequencies of interest consisting of six electrodes where the SNR at the frequency of interest was the greatest (see ROI section for details).

Time-resolved analysis of the frequency-tagged signals. In order to study the temporal dynamics of the frequencies of interest, we developed a time-resolved analysis of the strength of each signal. For each frequency of interest, we used a time-windowed FFT analysis elapsing three times the frequency period (1.936, 0.273, 0.318 and 0.239 seconds for SWIFT, SSVEP, IM1 and IM2 respectively). For each window and frequency, we calculated the signal-to-noise ratio by dividing the power at the frequency of interest by the power at neighbouring frequencies (the noise, ˜0.15Hz half bandwidth). To avoid discontinuity artefacts (leakage), we used a Hann tapering on each window. The centre of each window was taken as the time marker and windows were overlapped half length.

Stimuli strength analysis of frequency tagged signals. This analysis was analogous to the time

resolved analysis (previous section) with the main difference being that instead of calculating SNR for each time window, here we did it for each position along the path chosen by participants in each trial. Windowed FFT analysis parameters were the same as in the temporally resolved analysis (Hanning window, window length = 3 cycles, noise ~ 0.15Hz, 50% overlap). Each window was then binned into one of 5 bins representing the intensity of the target image (0 to 0.2, 0.2 to 0.4, 0.4 to 0.6, 0.6 to 0.8 and 0.8 to 1) and windows from within each bin were averaged together.

Frequency of interest normalization. To compare different frequencies of interest, we normalized them by taking their Z-score of their SNR across channels. This allowed us to visualize their amplitude changes within a comparable scale, since they had very dissimilar strengths (SSVEP >> SWIFT > IM1 and IM2).

Condition comparison analysis. In Figures 5 and 6, to compare each condition, we subtracted the averaged (across the channels of each ROI) z-scored SNR from the condition where the certainty is theoretically greater minus the condition where certainty should be lower (e.g., long foraging – short foraging) for each participant. The differences were then compared to zero (H0) using a two-tailed one-sample t-test against zero. Significant values thus represent differences in signal strength between the conditions that are significantly different from zero.

ROI selection. We selected the same ROIs of 6 channels for the analyses presented in Main Figures 3 and 4 and also Supplementary Figure S2. The selection criteria consisted in the selection of the channels showing the strongest $\log_e$(SNR) for each frequency of interest as showed on Figure 3. This yielded 4 ROI (each for each frequency of interest) as follows:

SWIFT: FCz, FC2, Cz, FT7, C1 and FC1; SSVEP: Oz, POz, PO4, O2, Iz and O1; IM1: C1, Iz, Fp2, FC1, P10 and CP2; IM2: F5, P2, POz, P1, T7 and Pz.

This procedure allowed optimizing the power of analyses. It is important to note that due to the nature of the paradigm, most of the data were recorded out of periods of recognition, thus hindering the amount of relevant data available for analyses.

For Figures 5 and 6, we selected ROI of 6 channels where the SNR was maximal for the condition that theoretically represents the highest certainty (long foraging for Figure 5, correctly recognized for Figure 6A and high confidence for Figure 6B). Note that the finding of significant differences is not trivial as the ROI definition (SNR maxima in the highest certainty condition) is orthogonal to the research question (is there a difference on these ROI between the conditions representing different certainty). These criteria yielded the following ROI:

Figure 5. SWIFT: CP6, T8, P8, PO3, P6 and Afz; SSVEP: Oz, POz, PO4, Iz, O2 and O1; IM1: CP3, AF4, C1, Afz, PO8 and Fpz and IM2: CPz, Cz, C4, TP8, CP2 and F6.

Figure 6A. SWIFT: POz, FCz, FC2, PO4, FC1 and F1; SSVEP: Oz, POz, O1, O2, Iz and PO4; IM1: CP2, C2, C4, P2, FC4 and P7; IM2: P9, PO7, P7, FC2, T7 and C2.

Figure 6B. SWIFT: C6, FC3, FC4, CPz, PO4 and O1; SSVEP: O2, Oz, O1, POz, PO4 and Iz; IM1: FC2, O1, Fp1, P9, FC4 and P5; IM2: P10, FC3, TP7, Fpz, FC4 and TP8.

RESULTS

Active exploration while tracking the signalling components of predictive coding using EEG. We developed an active paradigm where participants had to forage a virtual two-dimensional landscape to find three target images (Figure 2, see Methods for details). Briefly, target images were double frequency tagged using SWIFT (1.55Hz, tagging high level visual representations) and SSVEP (11Hz, tagging low level visual representations). Due to non-linear processing in the brain, intermodulation products (9.45 and 12.55Hz) indicate integration of these two streams (putative top-down and bottom up signals). When away from the target image, participants saw a blend of wavelet scrambled versions of the 3 target images (Figure 2, panel 4), which would morph into one of the images when sufficiently approaching the position of the orange blobs. Importantly, the low-level features of the 3 images (represented by their wavelet decomposition

products) were always presented, thus ensuring that any changes in the frequency tagging signals was due to finding and recognizing target images and not due to changes in low-level image features presented to the participants. Participants were rewarded via a point system, as they correctly identified the images and punished for incorrect identifications, thus encouraging foraging while hampering guessing. Participants could abort trials at any time (i.e., before finding the three targets) to reach a new landscape, and we limited the total of foraging time across all trials (N=15) to one hour. This imposed a trade-off between the exploration time spent on every trial versus the total of trials a participant would be able to complete during the whole session.

Analysis of participants' behaviour revealed that the task was demanding (Misses = 49.62%, False alarms = 14.76%), yet participants were able to find and identify the target images above chance (D'= 1.096, $p=1.57x10^{-6}$, one sample t-test against 0). For detailed analyses on behavioural performance refer to Supplementary Figure S1. We defined the extent of foraging within a trial by taking the length of the path (in pixels) traced by the participants while moving in the landscape (Figure 2). Supplementary Figure S1, panel H shows that participants had idiosyncratic average levels of foraging paths from 8.1 x 10^3 to 2.9 x 10^4pixels, mean = $1.46x10^4$ pixels across participants. Interestingly, the average time participants spent searching for targets per trial was quite short as shown in Supplementary Figure S1 panel I, ranging from 18s to 1min10s, mean = 37s, for an average trial length of 4 minutes by design (60 minutes over 15 trials). This shows that participants spent most of their time foraging (84.58%) as opposed to passive observation of the target images (15.42%). It is important to note that the short time spent on the target images puts constrains on the power of the EEG analysis presented below.

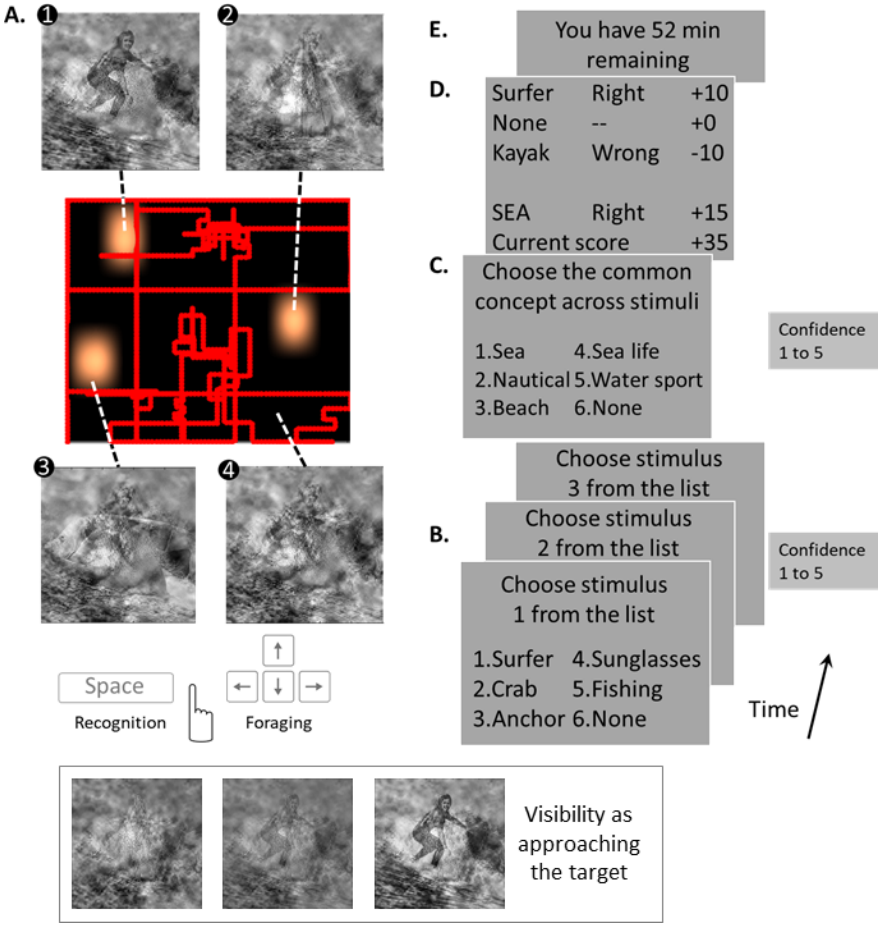


Figure 2. Paradigm. A. Participants (N=18) freely explored a 2D space, represented by the black rectangle, in order to find 3 target images (panels 1 to 3), represented by the 3 orange blobs (the more intense the colour, the more visible the target was), by pressing arrows on a computer keyboard. Images were double frequency tagged using SWIFT (1.55Hz, tagging high level visual representations) and SSVEP (11Hz, tagging low level visual representations). Thanks to the non-linear processing in the brain, intermodulation products (9.45 and 12.55Hz) indicate integration of the two information streams (putative sensory evidence and predictions). When not on the blobs, participants saw a blend of wavelet scrambled versions of the 3 target images (panel 4), which would morph into one of the images as they approached the orange blobs (see inset at the bottom). Importantly, the low-level features of the 3 images (represented by their wavelet decomposition products) were always presented, thus ensuring that any changes in the frequency tagging signals was due to recognizing the target image and not due to the presentation of its low-level features. In this particular trial, the 3 images were: surfer, boat and fish. Participants explored the 2D space in a freely chosen path, represented by the red line, to find the images. Participants could not see the layout of the 2D space nor the positions of the target images during the experiment but only the result of approaching to or moving away from the target images (1 to 4). As soon as participants found and recognized the target images, they pressed space bar. Participants were free to end a trial whenever they pleased, which could be before or after finding the 3 target images. Participants could thus choose to forage more or less of a given 2D space. The path length in a given trial was taken as a proxy for the amount of foraging in such trial. The total amount of time that a participant could spend foraging across all trials was 60min, so the amount of foraging in a given trial was a compromise considering the time left to forage other trials. The image triads were chosen such that they represented 3 exemplars of a common semantic category. In order to encourage participants to find target images, we used a scoring system to reward correct recognition (B to E). **B.** After ending the trial, participants had to choose the right target image among 4 distractors (see Table 1 for details). If not sure or not seen that particular target image, they chose option 6. "None". Participants were instructed to not guess. Three such prompts, one for each target image, was shown. **C.** After providing answers for the labels of the target images, participants had to choose the common concept shared by the triad. This was done to encourage participants to search for as many of target images as possible, since the common concept becomes less uncertain as more images are seen. After each target image and concept question (answered differently than "none"), participants were asked to rate their confidence from 1 to 5. **D.** Scoring. Correctly identified target images were rewarded with 10 points, whereas incorrect labelling was penalized by -10 points. When "None" was selected, no points were added. Analogously, common concept labelling was rewarded or penalized with +/- 15 points. "Current score" showed the points earned across the entire experiment. **E.** After completing the questions, a prompt showing the remaining exploration time in the experiment, consisting in a maximum of 15 trials, was shown. The answering time (B to E) did not count toward the remaining exploration time, thus participants could have breaks during these prompts.

Hierarchical frequency tagging of predictive coding signals under active foraging. We first verified that the tagging frequencies could be found in the EEG signal. For this, we selected four frequency-specific ROI comprised of the 6 electrodes where the signal was the strongest (Figure 3, white dots on the topographies). We found significant ($p<0.05$, FDR-corrected for multi-comparisons, right-tailed t-test against baseline = 0) peaks at the tagging frequencies (SWIFT, 1.55Hz and SSVEP, 11Hz) as well as their intermodulation products (IM1, 9.45Hz and IM2, 12.55Hz). The topographies of the frequencies are largely consistent with previous reports despite the different paradigms (Coll et al., 2020; Gordon et al., 2017, 2019), pointing to the robustness of the loci of these generators.

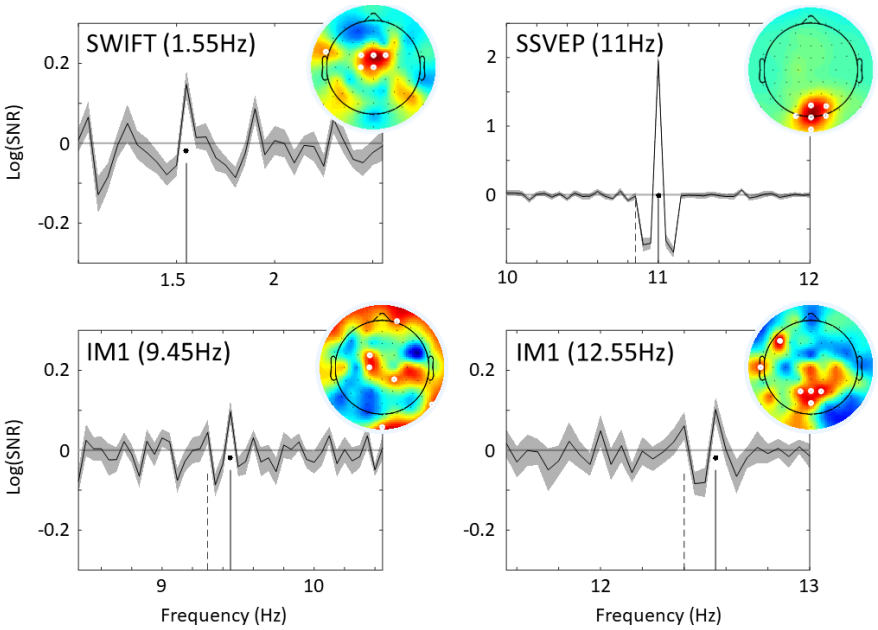


Figure 3. Frequencies of interest spectra and topographies. Tagging frequencies elicited significant spectral responses ($p<0.05$, FDR-corrected for multicomparisons, $q=0.05$) albeit reaching relatively low signal-to-noise ratios, which is likely the result of the dynamic nature of the paradigm (i.e., participants saw the target images momentarily while exploring the landscape). Tagging frequencies show distinctive topographies: central-frontal for SWIFT, occipital for SSVEP, central, frontal and occipital-temporal for IM1 and centro-occipital and parietal for IM2. Spectra were calculated around recognition time (-10 to +10 from recognition) on the 6 channels showing strongest responses for a given frequency (white dots). Full vertical lines represent tagging frequencies, dotted lines represent tagging frequency harmonics.

Time-resolved putative error prediction signals trigger a cascade of events leading to recognition. We then focused on the time of recognition as we assumed it to be a key moment of updating perceptual inferences. We devised a time resolved analysis of the frequency-tagged signals (see Methods for details) with the aim of answering two questions. First, how are perceptual signals modulated at the time of recognition? Second, what is the temporal order of appearance of these signals?

We had specific hypotheses regarding the sign of the modulation of each signal at the time of recognition (see Figure 1). As participants found and recognized target images, we expected that putative predictions tracked with SWIFT would increase. This was based on results from previous studies showing that SWIFT selectively tracks higher-level visual representations of recognized images, while being largely insensitive to unrecognized images and meaningless visual textures (Koenig-Robert et al., 2015; Koenig-Robert & VanRullen, 2013). Thus, when a target image was found and recognized, target image representations and putative image-specific prediction signals should increase. On the other hand, we expected that putative sensory evidence tracked with SSVEP should decrease briefly at the moment of recognition. This, while perhaps counterintuitive, is a result of the inherent sensory ambiguity of the stimuli. As target images are embedded in a large amount of visual "noise"(scrambled versions of the other two target images, Figure 2 and Methods for details), signals tagged with SSVEP are mostly composed of irrelevant information. A decrease in the SSVEP signal would then be the result of selectively filtering irrelevant information (visual noise) from the stimuli. The ambiguity of the stimuli should thus turn the balance towards predictions (signals tagged with SWIFT) at the moment of recognition while weighting down the sensory evidence (SSVEP), which is noisy and unreliable (Kanai et al., 2015; Weilnhammer et al., 2020). Finally, we expected a short-lived*decrease* in the intermodulation frequencies as a result of surprise due to the finding and recognition

of the target image. As shown in previous studies, the intensity of intermodulation products is *inversely proportional* to the prediction error or surprise (Coll et al., 2020; Gordon et al., 2017, 2019). Since the discovery and recognition of the target images is inherently surprising (participants do not know a priori what exact image they will find), prediction error signals are expected to transiently increase at the moment of recognition thus decreasing the amplitude of the intermodulation frequencies, with respect to the local baseline. This effect is of course expected to be transient as the discovery of the target ultimately leads to a decrease of prediction error over time, as certainty about the identity of the other images increases, in the long term, since target images are semantically related (the more target images found, the least surprising a new one should be), as we show later.

Figure 4 shows amplitude modulation of the tagging frequencies at the time of recognition. As expected, SWIFT-tracked signals showed an increase in amplitude (Figure 4, blue line); while SSVEP signals showed a decrease (Figure 4, green line). As for the intermodulation frequencies, IM1 and IM2 both showed a decrease in amplitude, consistent with our predictions (Figure 1) and previous studies (Coll et al., 2020; Gordon et al., 2017, 2019).

As for the second aim, namely investigating the temporal dynamics of the signals at the moment of recognition, we identified the time at which each of the signals was significantly different from baseline ($p<0.05$, two-tailed, FDR-corrected, dashed lines). We found that the first signal to cross this threshold was IM1 at –1.73s from recognition followed by SSVEP at 0.16s and IM2 at 0.21s. The last signal to cross the threshold was SWIFT after 0.64s from recognition. This finding is interesting as it suggests that prediction error, as tracked by IM1, precedes by more than a second the recognition of the image. In other words, these results indicate that a transient increase of prediction error triggers the dampening ambiguous sensory evidence while enhancing image-specific predictions. See the Discussion section for further discussion.

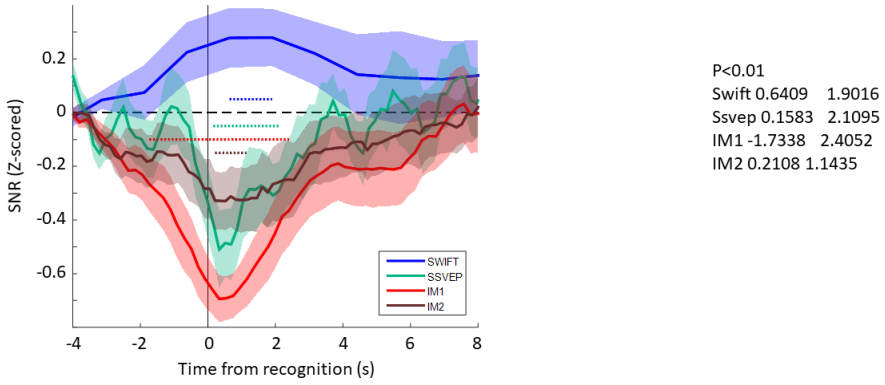


Figure 4. Dynamic modulation of tagging frequencies around recognition. Tagging frequencies were significantly modulated around behaviourally indicated recognition (time-windowed FFT at the tagging frequencies, dotted lines, $p<0.01$). Signals tagged with SWIFT (putative predictions carried by high level representations, blue line) were positively modulated around recognition time, which is consistent with high-level representations appearing at the moment of recognition. SSVEP signals (putative low-level representations or sensory evidence, green line) were negatively modulated around recognition, likely as a result of the filtering out of the noisy background in which the target images were embedded. Intermodulation frequencies (putative inverse of the prediction errors, red and brown lines) were also negatively modulated around recognition, consistent with the notion that prediction errors should diminish when the target image is found and recognized. Interestingly, the onset differed among different signals shedding light into the cascade of events around recognition. Intermodulation signals represented by IM1 reached significance the earliest (-1.73s from recognition) suggesting that the decrease of prediction errors precedes recognition. Both SSVEP and IM2 reached significance after recognition (0.16 and 0.21s from recognition). While SWIFT signals seem to raise from earlier than recognition, the relatively low SNR and the variability (shaded area) of

the signal resulted in its late reach of significance (0.6s from recognition). Topographies represent signals before (-4 to -2s) shortly after (1 to 3s) and long after recognition (4 to 6s). Signal-to-noise ratios were averaged across ROI channels shown on Figure 3.

Perceptual signalling streams are modulated by the scrambling level of the images in the landscape. Next, we verified whether the target image visibility or intensity (the inverse of the distance to the target image) modulated signals similarly as recognition did. We performed this analysis as a sanity check in which, instead of time from recognition, we analysed the signals as a function of the intensity of the target image (Supplementary Figure S2). In our design, the lower-level image features (wavelet scrambled versions) of all 3 target images are always presented to the participants (see Methods for details). These ensures that changes in the tagging frequencies strength are due to the recognition of the images rather than due to changes in low level image features as shown in previous studies (Gordon et al., 2017; Koenig-Robert et al., 2015; Koenig-Robert & VanRullen, 2013). Importantly, the results from this control analysis are smeared over time as participants could hover around the maximum intensity of the target image for different lengths at each instantiation. Also, this analysis does not distinguish between the periods when participants were getting closer and periods when they were getting further away from the target image; neither does it discriminate instances where participants did not recognize the target image but got closer to it. Despite of the above mentioned, stimulus intensity (target image visibility) and time from recognition time modulated frequency tagged signals in a similar way (Figure 4 vs Supplementary Figure S2). As previously, SSVEP, IM1 and IM2 showed a decrease in amplitude as participants got closer to the target image, becoming significantly different from baseline ($p<0.05$, FDR-corrected, coloured dots) from the 40% target image intensity bin (see Methods for binning and analysis details). On the other hand, SWIFT signals became stronger as participants got closer to the target image, becoming significant at the 100% intensity bin. This means that the behaviour of the signals seen in Figure 4 is not the result of an artefact from the selection of the epochs around the time of recognition but the result of active foraging, discovery and closeness to the target images. Again, it is important to note that these changes cannot be attributed to changes in the low-level features of the visual stimulation as visual features of all the three images are presented continuously with the same strength in the form of their wavelet-scrambled versions (see Methods for details).

The extent of foraging modulates the amplitude of the intermodulation products (or inverse of prediction error). We then sought to test whether foraging would translate into a modulation of putative prediction error signals. In our paradigm, foraging is instrumental to the discovery of target images. As the three target images are semantically related (e.g., surfer, boat, fish correspond to the common concept *sea*), the discovery of target images should decrease uncertainty about the identity of the remaining target images. Thus, prediction error should decrease in the long term as exploration increases, because the uncertainty (about the targets) should be reduced by gathering more information (foraging from the landscape).

We thus partitioned trials with more or less foraging (path length mean split, refer to the Methods sections for details). Figure 5 shows the difference in signal strength between long and short foraging trials. While all signals showed a trend of increased amplitude for long foraging trials compared to short (0.04, 0.44, 0.30 and 0.19 SNR z-score for SWIFT, SSVEP, IM1 and IM2, Figure 5), only the IM1 signals showed a modest but significant difference from 0 ($p=0.022$, two-tailed t-test against 0, uncorrected).

These results indicate that longer foraging paths lead to a reduction in prediction error (more IM strength) in the long run, as expected from improving internal models of the environment (Mirza et al., 2018). This result also suggests that the decision to keep foraging is correlated with the overall level of prediction error. In other words, short foraging trials (trials where the decision of aborting the trial came earlier) are characterized by higher prediction error (less intermodulation frequency amplitude) than longer foraging trials, where prediction error is lower (more IM signal strength).

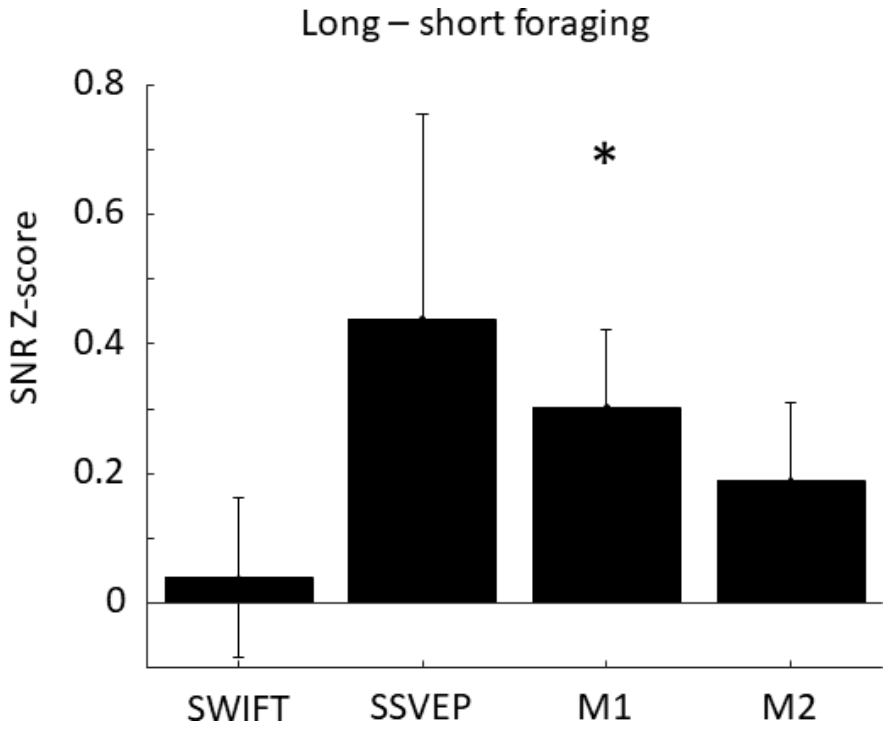


Figure 5. The amount of foraging (path length) modulates intermodulation frequency strength. Trials with longer path distances (more foraging) showed more intermodulation frequencies (IM1) amplitude (less prediction error) than did shorter ones ($p<0.05$, uncorrected). This can be interpreted as that the decision to stop foraging and pass to the next trial was correlated with an overall increase in prediction error (i.e., violation of internal model predictions). Signals were averaged over a ROI of 6 channels defined as the maxima in long foraging trials. Error bars represent SEM across participants.

Correct target image recognition and confidence, as behavioural markers of error minimization, modulate putative prediction error signals. We then focused on how correct recognition of the target images and confidence in their recognition were reflected in the strength of the signals. Both correct recognition and confidence are evidence of error minimization, although in different ways: correct recognition is an objective measure of error minimization while confidence is a subjective measure (how sure participants were of their own answers). Thus, correct recognition should be associated with a reduction in prediction error (more IM strength) when compared to instances of incorrect recognition, with a similar effect of confidence. We thus partitioned data into correct and incorrect target recognition trials (Figure 6A), and high and low confidence (Figure 6B, mean split). Correct recognition of target images was associated with stronger SWIFT and intermodulation products (IM1 and IM2) compared to incorrect ones (Figure 6A, $p<0.05$, one-sample t-test against 0, uncorrected). Increase in SWIFT strength in correctly recognized images is explained by stronger high-level representations and thus stronger image-specific priors (Gordon et al., 2017; Koenig-Robert & VanRullen, 2013). The increase in the amplitude of intermodulation products is expected as prediction error is minimized. Confidence modulated signals similarly (Figure 6B), however, only the difference on IM1 signals was significantly different from zero ($p<0.05$, one-sample t-test against 0, uncorrected). This could be a result of noisier subjective reports (confidence) compared to objective accounts (objective recognition performance). Again, these results are consistent with the idea of error minimization, in the form of increased certainty and decreasing prediction errors.

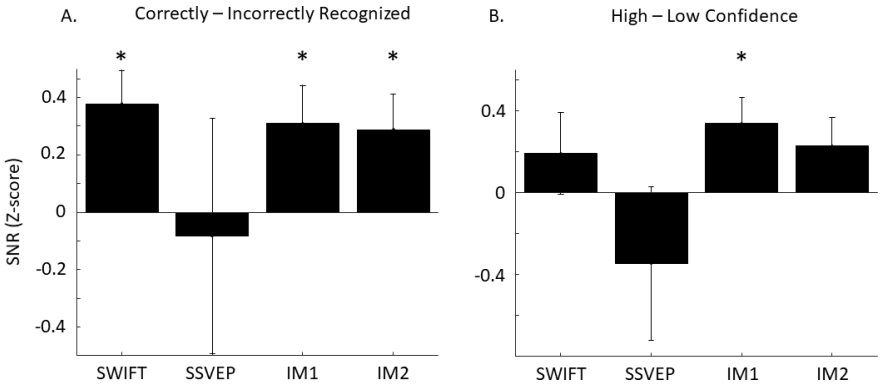


Figure 6 Recognition and confidence modulate signals similarly. A. Correctly vs incorrectly recognized target images. Signals for SWIFT, IM1 and IM2 were significantly higher for images successfully recognized than for unsuccessful ones ($p<0.05$, uncorrected).**B. High vs low confidence recognized images.** Signals for IM1 were significantly higher for high confidence recognized images than for low confidence ones ($p<0.05$, uncorrected). Signals were averaged over a ROI of 6 channels defined as the maxima in correctly recognized targets (A) and high confidence targets (B).

DISCUSSION

Discovery and recognition of target images trigger a cascade of signalling events during active foraging . We showed that the discovery and recognition of target images lead to a transient decrease in the lower intermodulation product IM1 at 9.45Hz (i.e., an increase in putative prediction errors). This decrease precedes the recognition of the images by 1.73s, which can be interpreted as surprise (prediction error) related to the discovery of the target image.

Shortly after recognition (0.12s), SSVEP signals (or sensory evidence) decreased significantly below baseline. This result can be interpreted as the correlate of decreasing the weighting on the noisy incoming sensory evidence (the mixture of the target image and the scrambled version of the other targets) in favour of predictions (Kanai et al., 2015; Weilnhammer et al., 2020).

Following closely afterward (0.21s), we observed a significant decrease in the IM2 (12.55Hz). It is unclear whether differences in the temporal responses between both intermodulation frequencies IM1 and IM2 are due to their specific temporal frequencies, that is, some frequency bands could be more sensitive for detecting subtle effects due to their closeness to resonant modes of the visual system (Herrmann, 2001). Another possibility is that different intermodulation products represent different channels of information as they could originate from different types of physical integration of ascending and descending signals (i.e., different neural circuits). Recent data supports the notion that different intermodulation frequencies represent signals from different hierarchical levels (Gordon et al., 2019). Future research should expand on this.

Finally, at 0.6s from recognition, SWIFT signals (recognized object representations and putative predictions) increased significantly above baseline. This is expected as SWIFT tracks complete object representations, but not unrecognized images or complex textures (Koenig-Robert et al., 2015; Koenig-Robert & VanRullen, 2013). Part of the signals tracked with SWIFT could also be interpreted as image-specific predictions being implemented and fed back to lower visual areas after recognition (Ahissar & Hochstein, 2004), although only indirect experimental support for this is currently available (Gordon et al., 2017, 2019).

Altogether, we can tentatively interpret these results as follows. Surprise (prediction error) marks the discovery of the target image. Prediction error signals then transiently increase above baseline leading on to a decrease in the weighting of noisy sensory evidence in favour of the enhancement of target-specific predictions once the target image is recognized. After this perceptual inference updating, the signals return to baseline. These results could thus provide support for the hypothesis that predictive processing occurs in

active foraging, for recognition of targets during the action-perception loop and consistent with contemporary formulations of predictive processing in terms of active inference.

Tracking active error minimization with hierarchical frequency-tagging. The use of a paradigm that includes perception and action allowed us to explore error minimization as a function of behaviour (exploration in the shape of foraging in a landscape). Our analyses suggest that more foraging is correlated with less prediction error. This is expected as foraging is instrumental to the discovery of target images and that the discovery of each target image reduces uncertainty of the identity of the to-be-discovered targets as they are related semantically (e.g., surfer, boat, fish). By separating long and short foraging trials, we observed that the amplitude of the lower second-order intermodulation product (IM1, 9.45Hz) was significantly higher in long foraging trials (Figure 6). This can be interpreted as that the decision to stick with the foraging policy is driven by a reduction in prediction error, while the decision to adopt a policy to abort foraging (i.e., having a shorter foraging path) would be triggered by higher prediction error (i.e., violation of internal model predictions). This result is thus consistent with predictions from theoretical accounts of active inference (Mirza et al., 2018; Schwartenbeck et al., 2013). Other signals (SWIFT, SSVEP and IM2) failed to show a significant difference between trials with more foraging compared to those with less foraging, and further research is needed to resolve if this is a result of the limited data available (refer to the next section, Limitations, for details).

Our results also suggest that objective and subjective behavioural measures of error minimization correlate with an increase in the amplitude of intermodulation products (a decrease of prediction error). Accurate recognition of target images, as an objective measure of error minimization, was correlated with higher amplitude of both intermodulation products, IM1 and IM2. Also, SWIFT signals were higher for accurately recognized target images when compared to incorrectly recognized ones. Since SWIFT signals have been shown to represent recognized objects (Koenig-Robert & VanRullen, 2013), this result was indeed expected. From a predictive coding perspective, the increase of SWIFT can be interpreted in this case as stronger target-specific predictions for correctly recognized target images. Similarly, subjective behavioural accounts of error minimization in the form of confidence for the recognition of target images correlated with an increase of the lower intermodulation product IM1, representing a decrease of prediction error.

Limitations. There are some limitations inherent to our paradigm. First, the precise timing of the signals around recognition can be smeared out by the windowed analysis (refer to the Methods for details). Specifically, the lower the frequency, the more time is necessary to get an amplitude estimate at that frequency. Thus, time dynamics of lower frequencies (SWIFT) are more smeared out than higher frequencies (IM2). Another limitation is the amount of informative data through the experiment. Only about 15% of the time were participants near target images and only a fraction of this time was employed recognizing the target images. It is important to stress that the choice of the paradigm design was made to promote participants' freedom at performing the task (and thus maximizing the changes of measuring genuine predictive processing signals under action and perception) over maximal relevant data collection, as a passive paradigm would. This design nonetheless puts constraints on the power of the analyses, potentially precluding identifying more subtle effects.

Implications for predictive coding. These results suggest that putative indicators of message passing in predictive coding, previously observed in passive perception studies, also arise when participants actively forage to gather information and diminish uncertainty. This addresses a gap in knowledge that arises as the predictive processing framework is moving towards a focus on action (Parr et al., 2022), showing that indeed predictive coding appear to occur as agents close the action-perception loop. The results also add to previous studies using hierarchical frequency tagging by demonstrating a time-resolved profile of the indicators of predictions, evidence and prediction errors that is consistent with the hypothesised cascade of signals under predictive coding. Finally, the results suggest that prediction error indeed scales with learning, and provide some confirmation for the hypothesis that foraging decisions relate to accumulated uncertainty, which is again consistent with an active inference framing of predictive processing.

ACKNOWLEDGEMENTS

This work was supported by Australian Research Council Discovery Grant DP160102770, DP190101805 awarded to JH. Australian NHMRC grants APP1046198 and APP1085404 and an ARC discovery project DP140101560 awarded to JP; Career Development Fellowship APP1049596 awarded to JP.

REFERENCES

Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in Cognitive Sciences* ,*8* (10), 457–464. https://doi.org/10.1016/j.tics.2004.08.011

Aitken, F., & Kok, P. (2022). Hippocampal representations switch from errors to predictions during acquisition of predictive associations.*Nature Communications 2022 13:1* , *13* (1), 1–13. https://doi.org/10.1038/s41467-022-31040-w

Aitken, F., Menelaou, G., Warrington, O., Koolschijn, R. S., Corbin, N., Callaghan, M. F., & Kok, P. (2020). Prior expectations evoke stimulus-specific activity in the deep layers of the primary visual cortex. *PLoS Biology* , *18* (12), 1–19. https://doi.org/10.1371/journal.pbio.3001023

Alamia, A., & VanRullen, R. (2019). Alpha oscillations and traveling waves: Signatures of predictive coding? *PLOS Biology* ,*17* (10), e3000487. https://doi.org/10.1371/journal.pbio.3000487

Bastos, Usrey, W. M., Adams, R. a, Mangun, G. R., Fries, P., & Friston, K. J. (2012). Canonical microcircuits for predictive coding.*Neuron* , *76* (4), 695–711. https://doi.org/10.1016/j.neuron.2012.10.038

Coll, M. P., Whelan, E., Catmur, C., & Bird, G. (2020). Autistic traits are associated with atypical precision-weighted integration of top-down and bottom-up neural signals. *Cognition* , *199* , 104236. https://doi.org/10.1016/j.cognition.2020.104236

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods* , *134* (1), 9–21. http://www.ncbi.nlm.nih.gov/pubmed/15102499

Den Ouden, H. E. M., Kok, P., & de Lange, F. P. (2012). How prediction errors shape perception, attention, and motivation. *Frontiers in Psychology* , *3* (DEC), 1–12. https://doi.org/10.3389/fpsyg.2012.00548

Friston, K. (2010). The free-energy principle: a unified brain theory?*Nature Reviews Neuroscience 2010 11:2* , *11* (2), 127–138. https://doi.org/10.1038/nrn2787

Friston, K., Fitzgerald, T., Rigoli, F., Schwartenbeck, P., & Pezzulo, G. (2017). Active Inference: A Process Theory. *Neural Computation* , *29* , 1–49. https://doi.org/10.1162/NECO_a_00912

Gordon, N., Koenig-Robert, R., Tsuchiya, N., van Boxtel, J. J., & Hohwy, J. (2017). Neural markers of predictive coding under perceptual uncertainty revealed with Hierarchical Frequency Tagging. *ELife* ,*6* . https://doi.org/10.7554/eLife.22749

Gordon, N., Tsuchiya, N., Koenig-Robert, R., & Hohwy, J. (2019). Expectation and attention increase the integration of top-down and bottom-up signals in perception through different pathways. *PLOS Biology* , *17* (4), e3000233. https://doi.org/10.1371/journal.pbio.3000233

Herrmann, C. S. (2001). Human EEG responses to 1-100 Hz flicker: Resonance phenomena in visual cortex and their potential correlation to cognitive phenomena. *Experimental Brain Research* ,*137* (3–4), 346–353. https://doi.org/10.1007/s002210100682

Hohwy, J. (2013). *The Predictive Mind* . Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199682737.001.0001

Johnston, P., Robinson, J., Kokkinakis, A., Ridgeway, S., Simpson, M., Johnson, S., Kaufman, J., & Young, A. W. (2017). Temporal and spatial localization of prediction-error signals in the visual brain.*Biological Psychology* , *125* , 45–57. https://doi.org/10.1016/J.BIOPSYCHO.2017.02.004

Kanai, R., Komura, Y., Shipp, S., & Friston, K. (2015). Cerebral hierarchies: Predictive processing, precision and the pulvinar.*Philosophical Transactions of the Royal Society B: Biological Sciences* , *370* (1668), 20140169. https://doi.org/10.1098/rstb.2014.0169

Koenig-Robert, R., & VanRullen, R. (2013). SWIFT: a novel method to track the neural correlates of recognition. *NeuroImage* ,*81* , 273–282. https://doi.org/10.1016/j.neuroimage.2013.04.116

Koenig-Robert, R., VanRullen, R., & Tsuchiya, N. (2015). Semantic Wavelet-Induced Frequency-Tagging (SWIFT) Periodically Activates Category Selective Areas While Steadily Activating Early Visual Areas.*PloS One* , *10* (12), e0144858. https://doi.org/10.1371/journal.pone.0144858

Kok, P., Mostert, P., & De Lange, F. P. (2017). Prior expectations induce prestimulus sensory templates. *Proceedings of the National Academy of Sciences of the United States of America* , *114* (39), 10473–10478. https://doi.org/10.1073/PNAS.1705652114/SUPPL_FILE/PNAS.201705652SI.PDF

Kok, P., Rahnev, D., Jehee, J. F. M. M., Lau, H. C., & De Lange, F. P. (2012). Attention reverses the effect of prediction in silencing sensory signals. *Cerebral Cortex (New York, N.Y. : 1991)* , *22* (9), 2197–2206. https://doi.org/10.1093/cercor/bhr310

Kumar, M., Federmeier, K. D., & Beck, D. M. (2021). The N300: An Index for Predictive Coding of Complex Visual Objects and Scenes.*Cerebral Cortex Communications* , *2* (2), 1–14. https://doi.org/10.1093/TEXCOM/TGAB030

Llinás, R. R. (2001). I of the Vortex: From Neurons to Self. *I of the Vortex* . https://doi.org/10.7551/MITPRESS/3626.001.0001

Mirza, M. B., Adams, R. A., Mathys, C., & Friston, K. J. (2018). Human visual exploration reduces uncertainty about the sensed world.*PLOS ONE* , *13* (1), e0190429. https://doi.org/10.1371/journal.pone.0190429

Parr, T., & Friston, K. J. (2019). Generalised free energy and active inference. *Biological Cybernetics* , *113* (5–6), 495–513. https://doi.org/10.1007/s00422-019-00805-w

Parr, T., Pezzulo, G., & Friston, K. J. (2022). Active Inference. In*Active Inference* . The MIT Press. https://doi.org/10.7551/mitpress/12441.001.0001

Pezzulo, G., Parr, T., & Friston, K. (2022). The evolution of brain architectures for predictive coding and active inference.*Philosophical Transactions of the Royal Society B* ,*377* (1844), 20200531. https://doi.org/10.1098/RSTB.2020.0531

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience* , *2* (1), 79–87. https://doi.org/10.1038/4580

Regan, D. (1977). Steady-state evoked potentials. *Journal of the Optical Society of America* , *67* (11), 1475–1489. http://www.ncbi.nlm.nih.gov/pubmed/411904

Regan, D. (1982). Comparison of transient and steady-state methods.*Annals of the New York Academy of Sciences* , *388* (1 Evoked Potent), 45–71. https://doi.org/10.1111/j.1749-6632.1982.tb50784.x

Robinson, J. E., Breakspear, M., Young, A. W., & Johnston, P. J. (2018). *Dose-dependent modulation of the visually evoked N1/N170 by perceptual surprise: a clear demonstration of prediction-error signalling* . https://doi.org/10.1111/ejn.13920

Schwartenbeck, P., FitzGerald, T., Dolan, R. J., & Friston, K. (2013). Exploration, novelty, surprise, and free energy minimization.*Frontiers in Psychology* , *4* (OCT), 710. https://doi.org/10.3389/fpsyg.2013.00710

Schwartenbeck, P., Passecker, J., Hauser, T. U., FitzGerald, T. H., Kronbichler, M., & Friston, K. J. (2019). Computational mechanisms of curiosity and goal-directed exploration. *ELife* , *8* , 1–45. https://doi.org/10.7554/elife.41703

Tang, M. F., Smout, C. A., Arabzadeh, E., & Mattingley, J. B. (2018). Prediction error and repetition suppression have distinct effects on neural representations of visual information. *ELife* , *7* , 1–21. https://doi.org/10.7554/eLife.33123

Tsoneva, T., Garcia-Molina, G., & Desain, P. (2021). SSVEP phase synchronies and propagation during repetitive visual stimulation at high frequencies. *Scientific Reports 2021 11:1* , *11* (1), 1–13. https://doi.org/10.1038/s41598-021-83795-9

Wacongne, C., Changeux, J.-P., & Dehaene, S. (2012). A neuronal model of predictive coding accounting for the mismatch negativity. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience* , *32* (11), 3665–3678. https://doi.org/10.1523/JNEUROSCI.5003-11.2012

Weilnhammer, V., Rod, L., Eckert, A. L., Stuke, H., Heinz, A., & Sterzer, P. (2020). Psychotic experiences in schizophrenia and sensitivity to sensory evidence. *Schizophrenia Bulletin ,46* (4), 927–936. https://doi.org/10.1093/schbul/sbaa003

Time resolved hierarchical frequency-tagging reveals markers of predictive processing in the action-perception loop

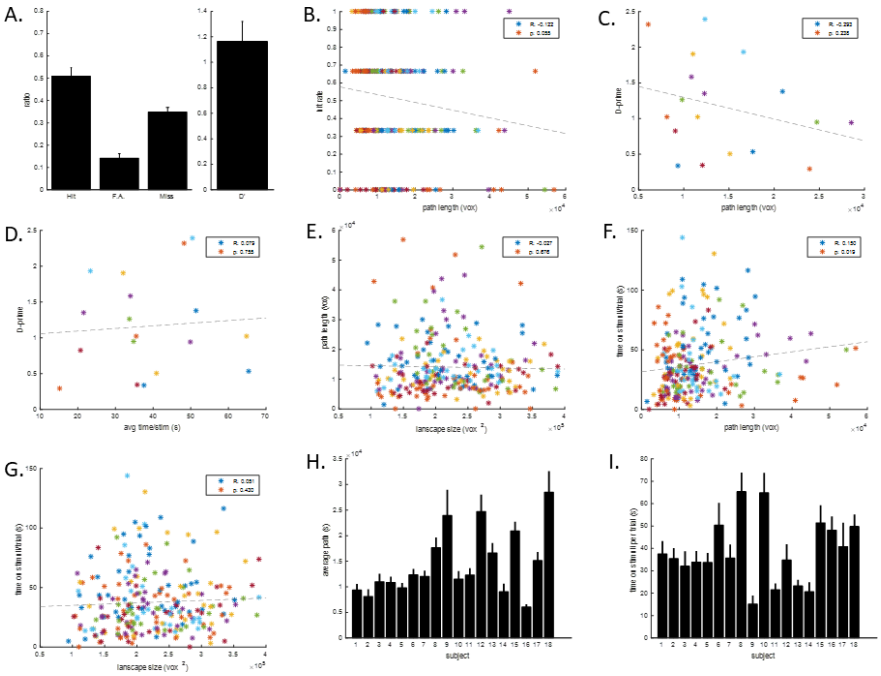
Roger Koenig-Robert, Thomas Pace, Joel Pearson and Jakob Hohwy.

SUPPLEMENTARY MATERIAL

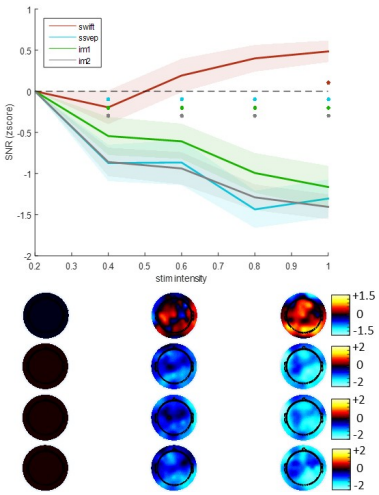
Hosted file

`image7.emf` available at https://authorea.com/users/574720/articles/618323-time-resolved-hierarchical-frequency-tagging-reveals-markers-of-predictive-processing-in-the-action-perception-loop

Supplementary Table 1 . Target image labels (light grey) and their common concept (dark grey). Distractors grouped by colour and their common concept shown at the bottom. For example, in the first column, the target common concept "sea" includes the target images "boat", "fish", and "surfer". The distractor common concept "beach" includes "sand castle", "sunglasses", and "beach towel".



Supplementary Figure S1. Behaviour . **A.** Performance on the recognition of target images. Participants were able to discriminate correct labelling of the target images above chance. However, the task was demanding as shown by the hit ratio. **B.** Hit rate as a function of path length. Hit rate was not correlated with the path length (R= -0.122, p= 0.055). **C.** D-prime as a function of path length. D-prime was not correlated with the path length (R= -0.293, p= 0.238). **D.** D-prime as a function of average time per stimulus. D-prime was not correlated with the time spent on stimuli (R= 0.079, p=0.755). **E.** Path length as a function of landscape size. Path length was not correlated with landscape size (R= -0.027, p= 0.676).**F.** Time on stimuli as a function of path length. Time spent on target images was correlated with path length (R= 0.150, p=0.019). This might be a result of the difficulty on recognizing certain target images: more difficult, more foraging and more time spent on a particular image to recognize it. **G.** Time on stimuli as a function of landscape size. Time spent on images was not correlated with landscape size (R= 0.051, p=0.43). **H.** Average path per trial for every subject. Different participants showed different levels of foraging (exploration) as shown by their average path lengths per trial.**I.** Average time spent on target images per trial for each participant.



Supplementary Figure S2. Signal modulation as a function of target image strength. Signals were modulated by stimulus intensity in a similar way than by recognition. Stimulus intensity (closeness to the target image in the landscape) was binned into 5 intensities. SWIFT signals significantly increased as a function of stimulus intensity ($p<0.05$, FDR-corrected, red dot), while SSVEP and IMs decreased as a function of stimulus intensity. Interestingly, while SSVEP and IM seemed to show a linear behaviour as a function of stimulus intensity, SWIFT on the other hand showed more of an all or none sigmoid behaviour.