Congruency effects on object recognition persist when objects are placed in the wild: An AR and mobile EEG study

Victoria I. Nicholls[1,2], Alexandra Krugliak[1], Benjamin Alsbury-Nealy[3,4], Klaus Gramann[5], Alex Clarke[1]

1. Department of Psychology, University of Cambridge, Cambridge, CB2 3EB, UK
2. Department of Psychology and Sports Sciences, Goethe University Frankfurt, Frankfurt am Main, 60323, Germany
3. Silicolabs, Toronto, Canada
4. Department of Psychology, University of Toronto, Toronto, M5S 1A1, Canada
5. Department of Biological Psychology and Neuroergonomics, Technische Universität Berlin, Berlin, 10623, Germany

Abstract

Objects in expected locations are recognised faster and more accurately than objects in incongruent environments. This congruency effect has a neural component, with increased activity for objects in incongruent environments. Studies have increasingly shown differences between neural processes in realistic environments and tasks, and neural processes in the laboratory. To what extent do findings obtained from a laboratory setting translate to neural processes elicited in real-world environments? We investigated how object recognition is modulated when objects are placed in real environments using augmented reality while recording mobile EEG. Participants approached, viewed, and rated how congruent they found the objects with the environment. We found significantly higher theta-band power for objects in incongruent contexts than objects in congruent contexts. This demonstrates that real-world contexts impact on how we recognize objects, and that mobile brain imaging and augmented reality are effective tools to study cognition in the wild.

Teaser

Combining augmented reality with mobile brain imaging to show that real-world contexts modulate object recognition processes.

Introduction

Every day we explore our environment, look at, and recognise objects. Being able to recognise objects quickly is important, for example, when crossing a road it is important to recognise if a vehicle is approaching to prevent a collision. Specific objects are generally experienced in certain locations and environments, knowledge about which we learn through experience in the form of schemas. This knowledge can aid recognition, such that objects which appear in expected (or congruent) environments are typically recognised faster and more accurately than objects which appear in unexpected (or incongruent) environments (1, 2, 3, 4, 5, 6, 7). This congruency effect also has a neural component, with studies showing a larger neural response for incongruent objects between around 250 and 500ms after the object appears, present in ERPs (e.g. the N400) and in theta oscillations (8, 9, 10, 11, 12, 13, 14). These findings have been linked to the integration of new semantic information into memory when the item was more unexpected (11, 12, 14, 15, 16, 17). However, the majority of studies investigating the congruency effect have been conducted in well-controlled laboratory

environments, where objects and scenes are presented together on a computer monitor before being replaced by an unrelated object and scene, breaking the temporal and spatial coherence that structure real-life experiences. Moreover, the participant is typically separated from the environment the objects are in, instead experiencing the 'environment' while being in a testing cubicle. This is different from how the real world is experienced, where there is a clear spatiotemporal coherence in the environment. As a consequence, we have little information about whether the tasks participants perform in laboratory settings are a good approximation of behaviours and associated neural processes in the real world. Indeed, recent findings have increasingly shown differences between neural processes in realistic environments and tasks, and neural processes in the laboratory (18, 19, 20, 21, 22, 23). This raises the question to what extent do research findings translate to real environments, and stress the importance of pursuing neurocognitive issues within their true dynamic functional context.

There is increasing acknowledgment that neuroscience must seek out more naturalistic ways to study cognition (24, 25, 26, 27, 28), however significant challenges remain for taking our experimental techniques out of the lab. With the developments in mobile electroencephalography (mEEG) technology over the last ten to fifteen years, such as the development of the MoBI framework (29) and improved signal processing tools (30), mEEG has become a reliable method which provides the opportunity to perform neuroimaging in the real world (25, 29, 31, 32, 33, 34). However, in experiments, we typically want to control or manipulate the stimuli we study, which poses technical barriers when experiments are situated in the real world. The development of mixed reality methods, such as head-mounted augmented reality (AR), allows researchers to control when and where stimuli are presented to participants while they freely move around the real world. While AR has been utilized to study assembly systems (35), pedagogy (36) and treatment of psychological disorders (37), a relatively novel approach is to use mobile EEG and head-mounted AR together, to study the neural processes underpinning cognition in real settings (38).

Here we study how object recognition processes are modulated by the relationship between the object and the environment, by presenting virtual objects in real locations using AR, while neural activity is recorded with mobile EEG. Participants walked around a large indoor and outdoor area seeing virtual objects that were either more or less congruent with their settings. Concurrent EEG then allows us to ask how the processing of objects is modulated by the congruency between the objects and the environment, thereby allowing us to test if our knowledge about the world modulates the processing of visual objects.

Results

The initial challenges were first to create a mixed reality experiment where objects were embedded into the real environment, set to appear at specific locations when the participant was in proximity to an augmented marker, and second to have sufficient synchronisation between the AR and EEG systems for millisecond precision in the analyses regarding the onset of augmented objects. We used a large indoor and outdoor space near the Department of Psychology at the University of Cambridge as our testing

arena, with the 80 virtual objects being assigned a physical location (Fig. 1C). The objects and locations were chosen such that approximately 40 objects would be more congruent and 40 less congruent with the environment. The study followed the procedures described in our previous protocol article (39).

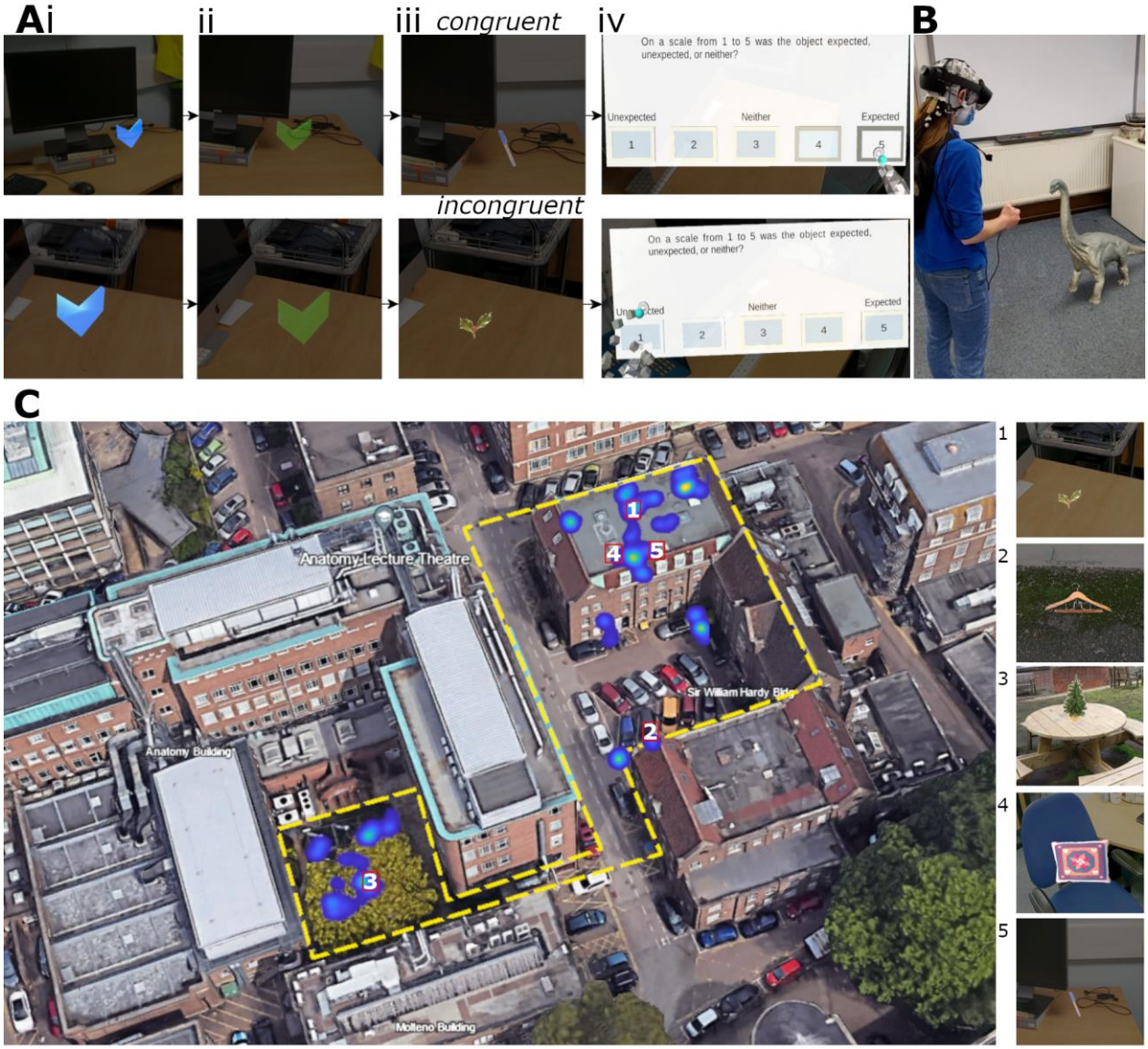

Fig. 1. *Experimental paradigm, example setup, and object locations. (Ai). Example of the blue virtual arrow indicators participants searched for, these indicated the object locations. (Aii) Example green arrow which would appear when participants moved close enough to the blue arrows. (Aiii) Example stimulus object which appeared for 5s after participants pressed the button on the button box once the arrows had turned green. (Aiv) Example congruency rating question participants were presented with once the stimulus object disappeared. For panel (A) the top row shows an example congruent trial, and the bottom row an example incongruent trial. (B) Example setup on one of the authors who consented to having their picture taken and shared. Author is wearing the Brain Products EEG system, with a shower cap, and then the Hololens 2 placed on top of the shower cap and EEG system. The author is holding the button box that was connected to the Hololens 2 and the LiveAmp. An example virtual object (Diplodocus)*

is presented to the author, the Diplodocus was not one of the objects presented to the participants. (C). A birds-eye view of the experiment environment taken from Google Earth v.10.55.01 (May 15th 2024), Cambridge, UK, and edited. The yellow dashed line indicates the full area that objects were placed in. The heatmap indicates locations a participant gazed at the objects, and example objects in locations 1-5 are presented on the right-hand side. Objects are ordered by congruency ratings made by one of the authors.

Once the participants were fitted with the mobile EEG system, the AR device was placed over the top of the electrodes. We used an initial calibration to align the real and virtual worlds, after which the experiment began (Fig. 1.B). On each trial, a virtual blue arrow appeared at the location of the object (Fig.1.Ai). To facilitate finding the object, the arrow could be seen through walls and surfaces. Once the participant was near the arrow, it turned green (Fig.1.Aii), signalling the participant could press a button after which the arrow would change into an object (Fig.1.Aiii). The object was presented for 5s, after which, a virtual display appeared to collect an expectancy rating on a 5-point scale, which is used to accommodate individual differences in how congruent each item was with its setting (Fig. 1.Aiv). The experiment had four blocks, two indoors and two outdoors, each containing 20 objects to find. The order of the blocks was counter-balanced between participants.

Following EEG data preprocessing pipelines developed for mobile EEG (30), we extracted EEG signals around the onset of the virtual objects. As a first step towards characterising the object recognition process in real world contexts, we examined the averaged EEG signals from the onset of the objects. Presenting participants with AR stimuli produced a clear visual evoked response over posterior electrodes, including a P100 followed by an N170, and P300 component (Fig. 2a), with ERPs declining after approximately 500ms. These peaks and characteristics were also present in the global field power of the EEG signals (Fig. 2b). A grand average time-frequency representation was also calculated (Fig. 2c), which showed a post-stimulus increase in low frequency power (lower-alpha/high-theta) between 100 and 400ms, followed by a decrease in power between 500 and 1000ms. There was a further increase in delta/theta power between 500ms and 1000ms, along with decreased gamma power throughout the post-stimulus period. These signals are in line with lab-based studies when presenting visual stimuli, suggesting that mobile EEG in combination with AR is a feasible approach to examining cognitive processes in real environments.

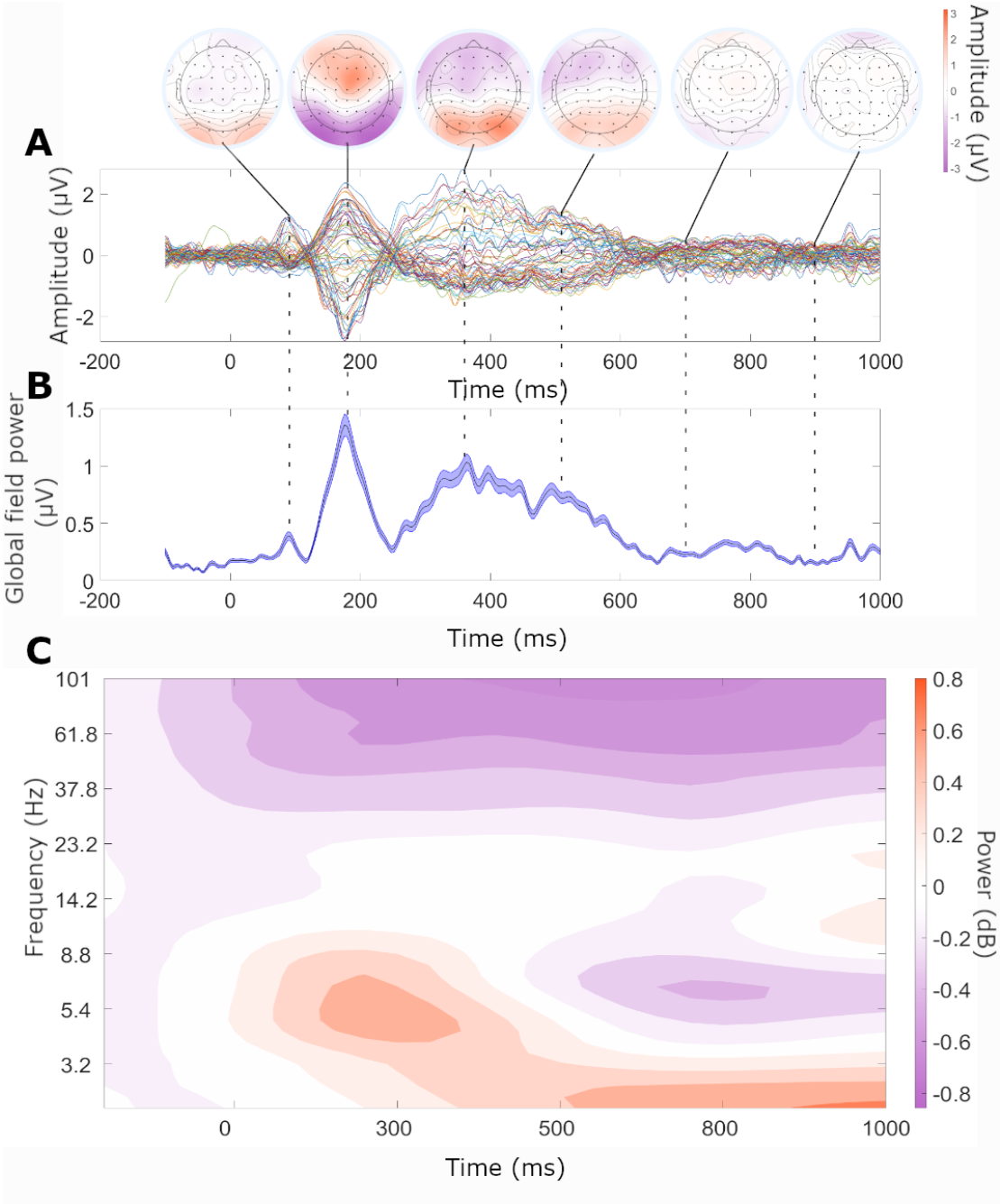

Fig. 2. Example of mEEG data quality *(A) Butterfly plot averaged across all participants and trials. Each line corresponds to an EEG channel. Above this are topographical plots showing the average amplitude at timepoints of: 100, 170, 350, 500, 700, and 900ms. (B) Global field power plot taken from the data used in A. The central line is the mean GFP and the lines above and below indicate the standard error. (C) Time-frequency representation of the EEG data from -100 to 1000ms and frequencies 2-101Hz. The frequencies are logarithmically scaled. The data was smoothed using a Gaussian kernel (3Hz, 40ms) across times and frequencies, and plotted using a filled 2D contour plot.*

We next sought to determine if EEG activity was modulated based on the congruency between the objects and the environment. Previous lab-based EEG studies have typically shown a congruency effect in the N300 and N400 time range (250 to 500 ms) across central electrodes, with a larger negativity for objects that are incongruent with the environment than for objects that are congruent with the environment. We attempted to replicate this finding in our data using the same data analysis approach as that in (10). We computed the mean EEG amplitude in two response windows: between 250 and 350ms (N300), and between 350 and 500ms (N400), across the central electrodes FC1, CP1, CP2, Cz, FC2, C1, CPz, and C2. As we had five levels of congruency ratings we averaged 1 and 2 into one incongruent level, and 4 and 5 into one congruent level. We computed t-tests in the N300 and N400 time-windows between the congruent and incongruent levels. We found no significant congruency effect (Fig. 3; N300: t(33)=0.63, SE=0.15, p=0.536, Cohen's *d*=0.096; N400: t(33)=0.20, SE=0.13, p=0.841, Cohen's *d*=0.024).

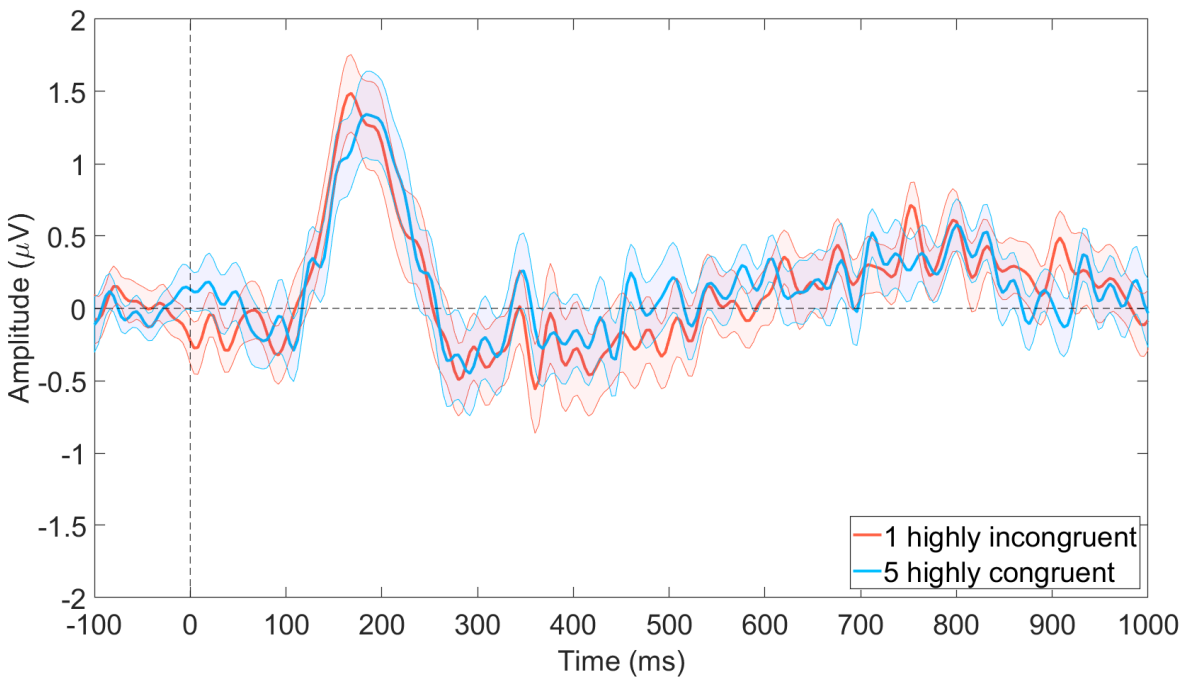

Fig. 3. ERPs examining the congruency effect. *Mean ERPs across subjects from -100 to 1000ms post-stimulus presentation. The bold red line indicates the mean ERP for trials where the object was rated as congruent with the environment (4 or 5 rating), and the bold blue line indicates the mean ERP for trials where the object was rated as incongruent (1 or 2 rating). The fill surrounding the mean line indicate the standard error. The dashed horizontal and vertical black lines indicate an amplitude of 0μV and the stimulus presentation point (0ms).*

We next performed a more exploratory analysis to test for congruency effects across all electrodes and time points, whilst also utilising the full range of the 5-point congruency scores provided for each trial. A linear regression was performed at each electrode and time point to determine how EEG signals varied with congruency ratings alongside a cluster-based permutation test. Like for our more restrictive analysis, we found no evidence for a modulation of ERPs by the congruency between objects and the real-world setting.

Given the more fluid dynamics of real-world experiences, and that the context was ever present, it may be that analyses less dependent on strict time-locking across trials might be more sensitive. As such, we explored whether the congruency between the object and the environment modulated different neural oscillations (Fig. 4). EEG power for each frequency and time-point was calculated for each trial, and averaged within each of the 5 congruency ratings. We then ran a cluster-based permutation linear regression analysis across all electrodes, timepoints, and frequencies. This revealed a significant cluster showing a positive relationship between incongruency (congruency rating 1) and power between approximately 100 and 600ms post-stimulus onset, centred in the theta range (Fig. 4a; 4-7Hz; cluster $p = 0.014$). The cluster was primarily located over posterior electrodes, both centrally and on the right. This positive relationship indicates that there was significantly more power in the theta range between 100-600ms for objects that were more incongruent with the environment compared to those that were more congruent. This shows that when more unexpected events occur, low-frequency neural responses increase over sensory and parietal electrodes in proportion to how unexpected the item was (Fig. 4a-d).

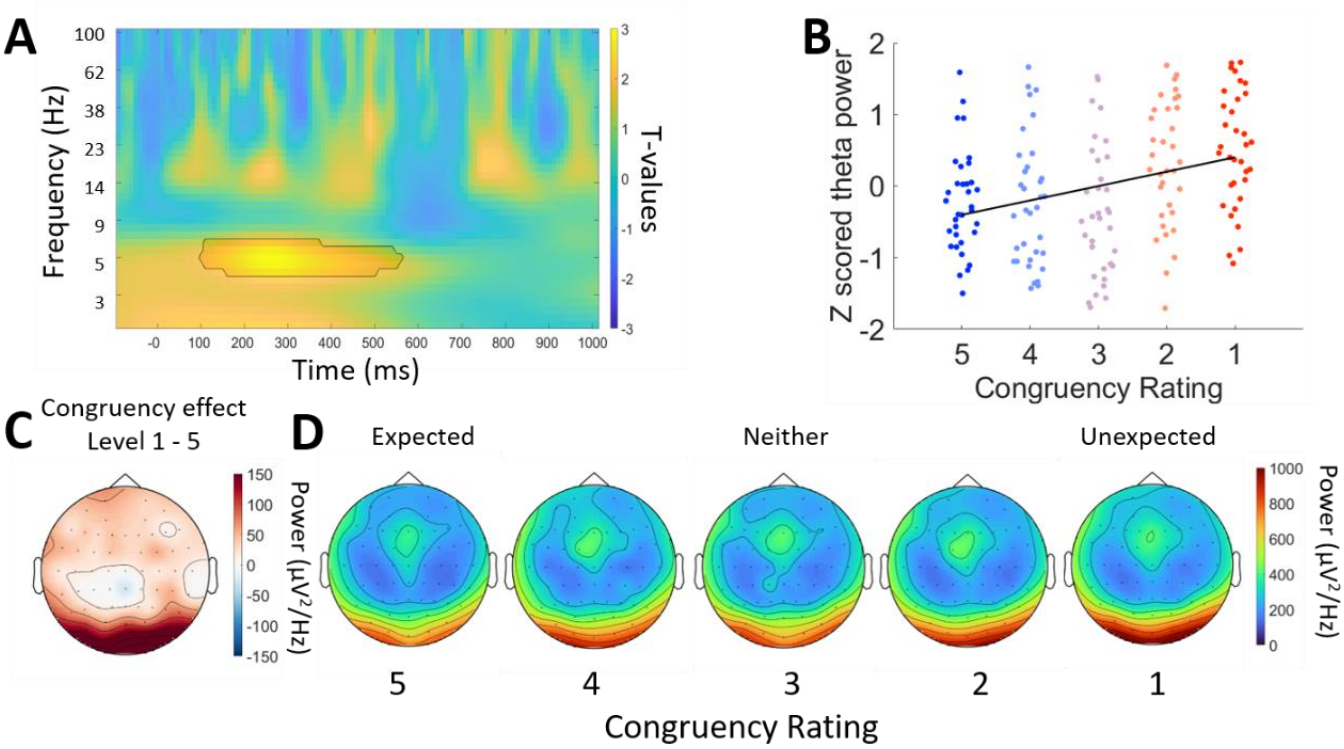

Fig. 4. Figures demonstrating the time-frequency analyses. *(a) T-values calculated across time and frequencies of the EEG data. The black line indicates the significant cluster for electrode O2. (b) The z-scored theta power (4-7Hz) averaged across each of the five congruency ratings for each participant. The black line is a first-degree polynomial fitted to the z-scored theta power. The colours indicate an increase in congruency (light to dark blue), an increase in incongruency (red to dark red), or neither (light purple). (c) The mean difference topography showing the difference in theta power between trials where objects were rated as highly incongruent (1) and highly congruent (5), between 200 and 500ms, averaged across participants. (d) Topographies for the mean theta power (5-7Hz) between 200 and 500ms at each congruency rating sorted so that objects rated as highly congruent (5) are on the left to objects rated as highly incongruent (1) are on the right.*

Discussion

Here we set out to determine whether object recognition is impacted by context, when the context is a real spatiotemporally coherent environment, the real world. Using mEEG and AR we examined whether neural object recognition processes that are impacted by scene context are also impacted by real-world outdoor and indoor contexts. We found that context modulated theta-band power, with significantly higher theta-band power for objects in incongruent contexts than objects in congruent contexts. This demonstrates that real-world contexts impact on how we recognise objects, and that mobile brain imaging and AR are effective tools to study cognition in the wild.

Our main finding concerned context effects in the theta band across posterior electrodes between 100 and 600ms, where theta-band power was higher for objects that were incongruent with the environment than for objects that were congruent with the environment. This is in line with previous research that has shown an increase in theta power when targets are incongruent, for example critical words not matching the preceding sentence (11, 12, 13, 14), which has been suggested to reflect memory processes relating to the integration of semantic information into the currently held representations of the semantic context (11, 12, 14, 15, 16, 17). Our results expand on these findings by showing that even in real spatiotemporally consistent environments, expectations about which objects appear in the environment influences how objects are recognised. These results give promising support to findings from laboratory studies applying to real world scenarios even though the laboratory studies are highly reduced compared to the real world.

Theta oscillations have also been linked to the ongoing semantic processing of objects and words in the ventral visual pathway, at a similar time range to our modulations of theta activity (40, 41, 42). This might suggest that the knowledge we have about the environment we are in, modulates how we access semantic information, which is primarily reflected through theta oscillations. This is also consistent with recent EEG work showing that semantic representations for objects that are incongruent with the scene are associated with enhanced and extended sematic processing compared to more congruent objects (43). This points towards a system whereby the semantic processing of objects that are incongruent with the environment show enhanced theta activity as a consequence of the object recognition process not benefiting from the schema-based predictions generated by the immersion in a particular environment. While our EEG analyses do not allow us to make claims about the neural regions involved, it is noteworthy that previous studies suggest that the precuneus might integrate object sensory information with expectations generated prior to the presentation of an object (3, 44), which would correspond with the location of our theta effects over occipital and parietal electrodes. The intriguing question of how activity in this posterior medial system interacts with the ventral visual pathway could be explored in future studies using more spatially specific methods such as MEG or fMRI.

In addition to our oscillatory analysis, we also examined ERPs. Previous ERP research examining the impact context has on the neural processes involved in object recognition has found an impact on the N300 and N400 components (8, 9, 10), with a greater amplitude being shown for objects that are incongruent with the environment than objects that are congruent with the environment. This amplitude increase has been associated with an increase in semantic processing. As a first step we attempted to replicate this effect using the same electrodes as (10), but found no evidence of a modulation. As our reference electrode was placed at FCz, while theirs was placed at the mastoids, any potential N400 effects in our study may be pushed to different electrodes than that seen by (10). Therefore, we performed a cluster-based permutation test across all electrodes and timepoints to determine whether a congruency effect existed in the ERPs. We found no impact of congruency on the amplitude across any electrodes or timepoints. It is unclear why we did not replicate the well-established N400 congruency effect, although a combination of three possibilities might go some

way to explaining this. First, we cautiously suggest there could be differences between the neural processes behind the congruency effect when participants are placed in a real environment compared to a laboratory environment, as has been shown for other neural effects recently (18, 19, 20, 21, 22, 23). For example, in laboratory experiments objects are often presented as pictures within a scene that appears for a short period of time before being replaced with another picture, removing the spatial and temporal coherence that exists when moving from one environment to the next in a real-world setting. Removing this spatial and temporal coherence may artificially increase the N400 effect. Indeed, previous studies have shown that an N400 is generated when an unexpected ending of an action occurs, such as an incorrect dance step (45, 46) or when spatial coherence is disrupted, such as a plane in a parking lot rather than the sky of the same scene (47). Therefore, the combined disruption of semantic, temporal, and spatial coherence of the scene context in controlled lab-based contexts may artificially inflate the N400 compared to real-world environments where the temporal and spatial coherence of the context is preserved.

A second possibility is that real world cognitive effects might be less time-locked across events compared to lab-based tasks – for instance, in the lab there is a relatively short, largely fixed, time between a scene and object appearing, meaning that the dynamics of expectations influencing object processing are relatively predictable, which might lead to more stable congruency effects in time over trials. In our experiment, whilst our objects have discrete onsets that we align our analysis to, the context is ever present and relatively stable over time. This might add variability to when expectations impact object processing, and reduce the ability of time-locked ERPs to reveal congruency effects. A third possibility is that ERP effects are partly produced by very low frequency signals, which are more problematic for mobile EEG. Our ERP analysis filtered out signals below 1 Hz due to artefacts caused by movement. The removal of these very low frequency signals is known to distort ERP morphologies (48, 49), which may in turn reduce our ability to see clear N300 or N400 effects. However, our paradigm and analyses clearly suggest that modulations of theta activity (4-7 Hz) are present, and that time-frequency analyses, with less reliance on time-locked activity (50), are a fruitful analytic approach for mobile EEG.

Recently there has been increasing interest in examining human neural processes where participants perform tasks in more naturalistic contexts, such as VR or even the real world, whilst performing increasingly natural behaviours (25, 29, 31, 51, 52, 53, 54). We make a significant expansion to these studies by demonstrating the use of AR in combination with mEEG to investigate fundamental neural processes such as those underpinning object recognition. This novel combination opens the door to a whole new direction for neuroscience research, where not only can one conduct highly controlled studies in the laboratory or in highly realistic VR, but also studies in real-world locations beyond the lab. This allows for the exploration of neuroscience while participants perform natural behaviours within real environments, giving us a way to understand our neural processes involved in going about our day to day lives. Moreover, AR allows to explore what can help us improve our environments and the efficiency of our neural processing. For example, one could examine whether we can improve our memory when viewing augmented information compared to information presented in a

text or static display (e.g. 55, 56). Mobile EEG and AR also allows us to explore these questions in diverse arenas, such as museums and galleries, where standard text displays and exhibits could be augmented with animations to enhance memory and visitor engagement.

One limitation of our paradigm, is that the objects were not permanently embedded into the environment, but rather appeared on a button press. We chose this as a first step towards testing object recognition in the real world with mobile EEG and AR, as it allowed us to have control over the objects and clear time periods to analyse. However, it enforces an artificial temporal constraint which would not normally be present in the real world when examining objects. As these methods and analytical approaches develop, it will become important to be able to track object recognition in a more ecologically valid way if we are to study natural cognitive processes. This may be facilitated through continued development of the integration of head-mounted AR devices, eye tracking, body sensors and mobile brain imaging, where the visual and semantic processing of our visual world can be modelled continuously depending on what we fixate upon and choose to interact with.

Conclusion

This study provides some of the first insights into how interactions between objects and real environments modulate object recognition. Previously, research has described in detail how object congruency modulates object recognition processes in well controlled laboratory environments. However, what has been missing is information on how this process occurs when participants are present in the same environment as the stimulus objects, the environment is relatively stable and long lasting, and while participants perform naturalistic behaviours. We find that even real environments impact on how we recognise objects. These results are a positive for laboratory studies as it would seem even with low ecological validity, some findings appear universal, or to be similar in highly ecologically valid conditions. Moreover, we suggest mobile brain imaging and AR are excellent tools to study real world cognition.

Materials and Methods

The methods follow the protocol described in (39), with the details pertinent to this study reported below. The RSA analyses of the EEG data and analyses of the eye tracking data will be reported separately. In addition to the ERP analyses described in (39), we have conducted additional exploratory analyses of the time-frequency data described in the EEG Analyses subsection.

Participants

To determine the number of participants required, a power analysis was conducted on data from (10) who examined EEG congruency effects of object-scene processing. The power analysis was conducted using the 'sampsizepwr()' function in MATLAB (57), based on the mean difference between the amplitude of the N400 for congruent and incongruent conditions, to obtain a power of 0.8, at an alpha of 0.05. The analysis showed that a minimum sample size of 34 participants was required.

Overall forty-two participants took part in the experiment (13 female, aged 18-35 years). Eight participants had exceptionally noisy data and were excluded from the analysis, leaving a total of 34 participants for the final analyses in line with the outcome of the power analysis. All participants had normal or corrected to normal vision and reported no history of neurological conditions. Ethical approval for the study was granted by the ethics committee of the Department of Psychology at the University of Cambridge. Participants signed informed consent prior to the start of the experiment. This study was performed in accordance with all appropriate institutional guidelines and international guidelines and regulations, in line with the principles of the Helsinki Declaration.

Apparatus

Participants were presented with virtual stimuli using a Hololens 2 device (https://www.microsoft.com/en-us/hololens/). The Hololens 2 is a mixed reality headset with a horizontal field of view of 42°, a vertical field of view of 29°, and presents images at 60Hz. EEG was recorded using the Brainvision LiveAmp 64 mobile system (Brain Products GmbH, Gilching, Germany), with an online low pass filter of 131Hz. In all sessions we recorded 64-channel EEG through ActiCap Slim active Ag/AgCl electrodes located in an elastic cap (actiCAP snap) positioned according to the international 10/20 system (58), with a reference electrode placed at FCz, ground electrode placed at Fpz and a sampling rate of 500Hz. The EEG electrodes were placed on the participant with the Hololens 2 placed on top of the electrodes, and the LiveAmp and electrode cables placed in a backpack that the participant wore while performing the experiment. A custom-built button box was plugged into the Hololens 2 USB-C port and the LiveAmp 1-bit trigger port, meaning that when the button was pressed, it simultaneously sent a signal to Hololens 2 and a 5V signal to the LiveAmp. The LiveAmp then converted that signal to a trigger that was marked in the EEG recording. The signal to the Hololens 2 triggered the appearance of an object during the experiment. An example of the setup can be seen in Fig. 1.B.

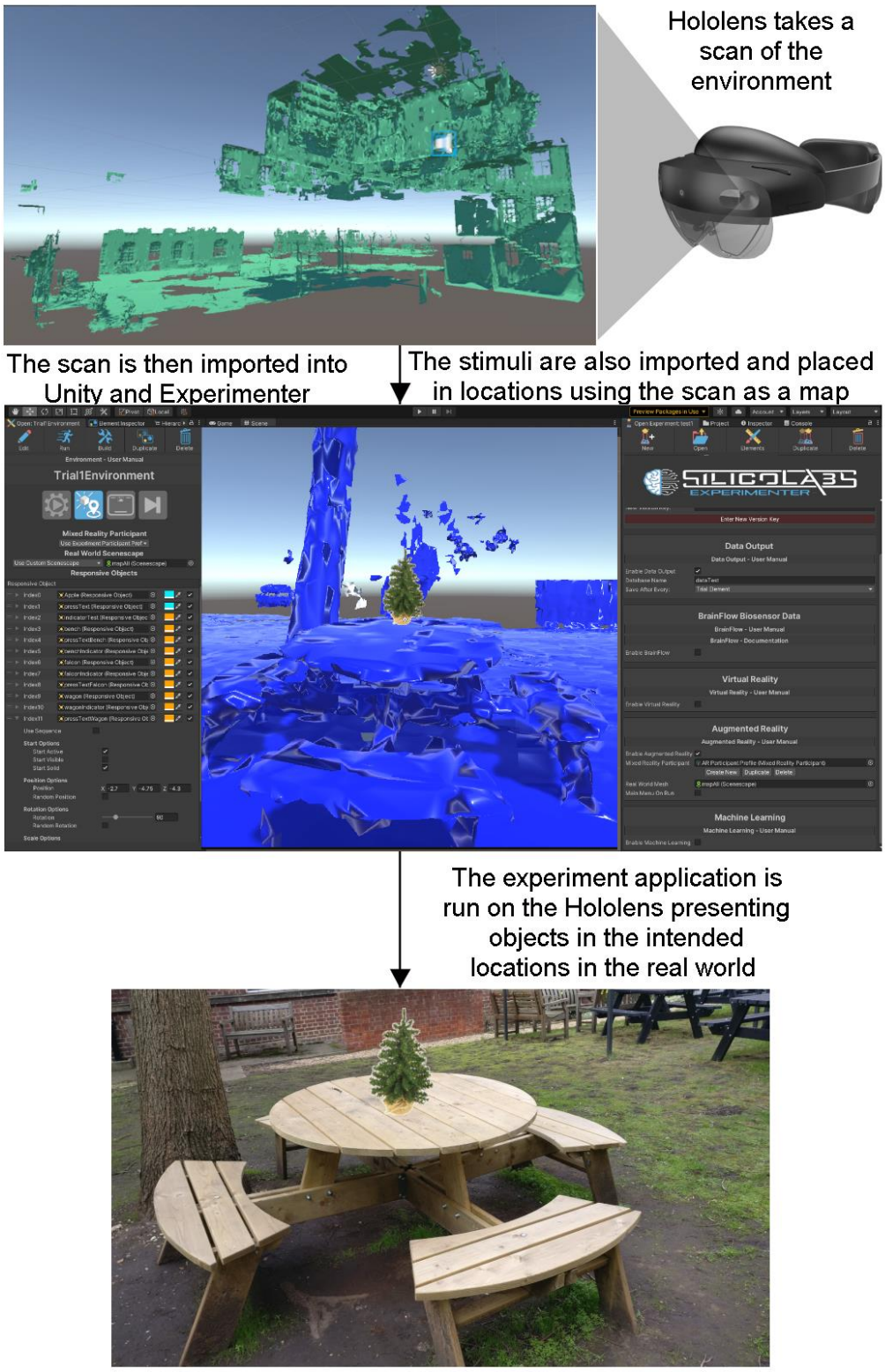

Fig. 5. A description of how the experiment was created. *First the Hololens 2 is used to take a scan of the real world. This is done by walking around the real world and looking at as many parts of the world as possible. For example, to scan a door one would first look at the top left corner of the door then move their head down slowly to the bottom left corner. Then look slightly to the right and move your head slowly upwards. This would be repeated until you had looked at all parts of the door. As you are looking and scanning the environment, the Hololens constructs 3D models or maps*

of the real world, an example of our environment is shown in A. This 3D map can be downloaded from the device and then imported into Unity, shown in panel B, and used in the same way as a virtual environment. The stimuli were then placed in the intended locations. As the 3D map is an accurate map of the real world, the locations the objects are placed in within Unity correspond to their real world locations. Therefore, when exporting the virtual environment, with the stimuli objects, back to the Hololens 2, the device will present the objects in the intended locations in the real world (As in panel C).

Stimuli and environments

During the experiment, participants explored a large real-world environment, including an indoor space (consisting of three offices, a corridor, and a meeting room) and an outdoor space (consisting of a parking lot and a picnic area) where they were presented with virtual objects in various locations (Fig. 1.C). A total of 80 images corresponding to concepts from a property norming study (59) were presented to the participants, and five additional images used as practice trials.

The 80 images used in the main experiment were selected from the one thousand images in the property norms set. The one thousand images were narrowed down to 400 by the lead author, based on which objects could plausibly appear in the experimental locations. These objects were then rated as either congruent or incongruent with the indoor and outdoor environments. From this we extracted 40 objects that were the most expected in the environments, and 40 objects that were the most unexpected in the environments. For the unexpected objects we made sure that they were highly unexpected, but not impossible. For example, a reindeer in a parking lot in the UK is highly unexpected but not impossible. However, if an object would be highly unexpected and impossible, we would not use that object (for example a dinosaur in the picnic area). The congruency ratings for the 80 objects were then verified by a second experimenter. All objects were placed in plausible locations within the environments (i.e. a bottle on a table, not floating in the air).

Constructing the paradigm

To place the stimuli in the correct locations in the real world, a virtual 3D map of the real-world testing area was created. This involved scanning the experimental environment through the Hololens 2, which uses spatial mapping to create detailed representations of real-world surfaces which are built into a map. This spatial map was then downloaded from the device, and imported into Unity for use as a virtual environment to place the objects within. The locations used in the virtual environment would correspond to the matching locations in the real world (an example is shown in Fig. 5). The experiment was coded using SilicoStudio (beta testing version, formerly Experimenter, https://www.silicolabs.ca/), an experimental design software package built for the Unity engine, and presented to participants with the Hololens 2. SilicoStudio was used to place the objects in the correct real-world locations and to enable interactions between the participant and the objects. The locations of the stimuli were the same for all participants. A list of the stimuli used and their locations in the

experiment can be found in the Supplementary Materials (Table S1, Figs. S1 and S2). The interactions included the changing of the object indicator arrow colour when participants stepped close enough to the indicator. The second interaction included changing the indicator into an object when the participants pressed the button on the button box. Furthermore, SilicoStudio was used to manage the order in which objects appeared, and to record behavioural data which included head, and eye tracking, and behavioural responses.

Procedure

The 80 images were presented to participants in four blocks of 20. The images in each block were matched on the following predictors: congruency, category, domain, hit rate on a visual task, hit rate on a lexical task, false alarm rate for a lexical task, false alarm rate for a visual task, image energy, image size, proportion of non-white space in the image, hue of the image, saturation of the image, frequency in the COCA database, proportion of participants identifying the modal name of the object, and the number of non-taxonomic features. Except for congruency and environment, the predictors for each image were taken from (59). The matching was performed using the Match software (version 2.09.02, 57).

To familiarise the participants with the Hololens and the appearance of AR stimuli they performed five practice trials in an indoor environment. The trials follow the same procedure as for the experiment described below. During the experiment participants walked through an environment, with indoor and outdoor sub-environments. In each sub-environment, participants completed two blocks of 20 trials (40 total per sub-environment). In each trial participants needed to find an arrow indicating the location of the next object (Fig. 1.Ai). When participants were close enough to the arrow, the arrow changed colour (Fig. 1.Aii) indicating that the participants could press the button on their button box (Fig. 1.A). Once the button was pressed, an object appeared for five seconds (Fig. 1.Aiii). At the same time a trigger was sent to the 1-bit TTL input of the LiveAmp by the button press. Participants were instructed to look at the object for the entire time it was visible, and to keep as still as possible. After the object disappeared, participants were presented with a virtual screen, asking them how expected or congruent they found the object within the environment on a scale from one to five (Fig. 1.Aiv). One was unexpected (incongruent), three was neither, and five was expected (congruent). Participants indicated their rating by pressing the appropriate button on the virtual screen. We included participant ratings as participants may have different ideas or experiences of which objects are congruent and incongruent in the environments. The next trial began once they had responded. This process was repeated until all 80 objects were found. During the whole session, the participants' EEG activity and eye movements were recorded. The four blocks took approximately 60 minutes to complete.

EEG Pre-processing

Pre-processing was performed using the EEGLAB toolbox (61) with the BeMoBIL pipeline (30), which allowed for the automated analyses of mobile brain imaging signals. The single-subject data were truncated by removing parts of the EEG recording that occurred before the start of the experiment, the inter-block intervals, and after the

end of the experiment. The data were then downsampled to 250Hz and power line noise was cleaned using the Zapline toolbox (62). Noisy channels were automatically identified using the EEGLAB clean_artifacts() function with the maximum tolerated fraction of the recording duration during which the channel was flagged as 'bad' was set to 0.5 and the minimum channel correlation was set to 0.75. Otherwise the template configuration settings from the BeMoBIL pipeline were used for the clean_artifacts() function. The identified bad channels were then interpolated (mean interpolated = 3, SD = 3) and an average reference was applied to the data.

At this point, the dataset was duplicated so that IC weights could be computed on a filtered version of the data and then applied to a non-filtered dataset. For ICA, the duplicated data were highpass filtered using a zero-phase Hamming window FIR filter with a cutoff frequency of 1.5 Hz (order 1650, transition width 0.5Hz). Independent components (ICs) were then identified using an adaptive mixture component analysis (AMICA; 63). The AMICA decomposition uses an automatic time domain cleaning where samples are removed based on the log likelihood that the samples do not correspond to the algorithm's estimate of model fit. Three standard deviations were used as the removal criterion. The ICs identified by AMICA were copied into the original unfiltered dataset, and classified as 'brain', 'eye', 'heart', 'channel noise', 'line noise', 'muscle', and 'other' using IClabel Lite (64). ICs that were labelled as 'eye', 'heart', 'channel noise', and 'line noise' were then excluded from the reconstructed dataset (mean excluded = 8, SD = 3). Another version of the data analysis results where ICs labelled as 'muscle' are shown in Fig. S3 in the Supplementary Materials. All effects reported in the Results remained when 'muscle' components were removed. However, as IClabel is trained on a very small portion of mobile EEG data we are not confident that IClabel accurately labels 'muscle' ICs over 'brain' ICs (30), therefore, we went with the preprocessing pipeline where we were confident no 'brain' ICs were removed.

The IC-cleaned data were then filtered with a one-pass zero-phase Blackman-windowed sinc FIR filter with cutoff frequencies of 1 and 40Hz, and a transition width of 0.1Hz for the ERP analyses, and cutoff frequencies of 1 and 100Hz for the time-frequency analyses.

The data were then epoched between -2 and 3 seconds around the onset of the images. Channel baseline means were removed between -100 and 0ms before the onset of the stimulus. Noisy epochs were automatically removed using the bemobil_reject_epochs() function, based on the mean of channel means (to detect general large shifts), the mean of channel standard deviations (SDs; to detect large fluctuations in many channels, e.g. large muscle activity), the SD of channel means (to detect phases where single channels had large fluctuations), and the SD of channel SDs (to detect large fluctuations in single channels). We used an automatic kneepoint with an offset of 5% to determine a threshold for epoch rejection. On average, 8 trials were rejected per participant (SD=4).

EEG Analyses

We analysed event-related potentials (ERPs) and time-frequency representations (TFRs) from channel data using cluster-based permutation tests from the FieldTrip toolbox (65).

For the calculation of the ERPs, we first calculated the mean ERPs for objects rated as congruent and incongruent. As participants gave congruency ratings between 1 and 5, we averaged 1 and 2 into one incongruent ERP, and 4 and 5 into one congruent ERP. The ERPs were calculated from 0 to 600ms post-image appearance. This was done for every channel individually giving two mean ERPs per channel per participant. For the ERP analysis we attempted to replicate the data analysis approach taken in (10). We computed the mean EEG amplitude in two response windows: between 250 and 350ms (N300), and between 350 and 500ms (N400), across the central electrodes FC1, CP1, CP2, Cz, FC2, C1, CPz, and C2. We computed t-tests in the N300 and N400 time-windows between the congruent and incongruent levels.

For the TFR calculation, we followed a similar process as for the ERPs. The TFRs were calculated for each trial, and averaged for each congruency rating the participants gave, giving five TFRs per participant. The TFRs were calculated using 5 cycle Morlet wavelets for 40 logarithmically scaled frequencies from 1 to 100Hz, a time window of interest of -2 to +3ms around the appearance of the image. The TFRs were calculated for each channel individually giving five mean TFs per channel per participant.

The cluster based permutation tests were then performed on the ERP and TFR data. For each channel, time point, and frequency (TFR only), a linear regression was performed determining the relationship between EEG amplitude and congruency rating, producing t-values for each channel-time (and frequency) sample. All samples with t-values below the 2.5th and 97.5th quantile were selected and clustered based on temporal, spatial, and spectral adjacency. The t-values in each cluster are then summed giving a cluster mass.

The significance of each cluster was determined by using the Monte Carlo method. For this the trials for each of the congruency rating levels were placed in a single set. Trials are randomly drawn out from this set that match the number of trials for each congruency rating. The same regression analysis as above is then performed on this randomised dataset. The random drawing and regression analysis on the randomised dataset was repeated 5000 times and a histogram of the test statistic was constructed. If the original cluster mass is below the 2.5th or above the 97.5th quantile of the constructed histogram, then they were considered to be significant at an alpha level of 0.025.

References

1. M. Bar, Visual objects in context. *Nat. Rev. Neurosci.* **5**(8), 617-629 (2004)

2. I. Biederman, R. J. Mezzanotte, J. C. Rabinowitz, Scene perception: Detecting and judging objects undergoing relational violations. *Cogn. Psychol.* **14**(2), 143-177 (1982)

3. T. Brandman, & M. V. Peelen, Interaction between scene and object processing revealed by human fMRI and MEG decoding. *J. Neurosci.* **37**(32), 7700-771 (2017)

4. T. H. Cornelissen, M.L.H. Võ, Stuck on semantics: Processing of irrelevant object-scene inconsistencies modulates ongoing gaze behaviour. *Atten., Percept., & Psychophys.* **79**, 154-168 (2017)

5. J. L. Davenport, M. C. Potter, Scene consistency in object and background perception. *Psychol. Sci.* **15**(8), 559-564 (2004)

6. J. M. Henderson, P. A. Weeks Jr, A. Hollingworth, The effects of semantic consistency on eye movements during complex scene viewing. *J. Exp. Psychol. Hum. Percept. Perform.* **25**(1), 210 (1999)

7. G. R. Loftus, N. H. Mackworth, Cognitive determinants of fixation location during picture viewing. *J. Exp. Psychol. Hum. Percept. Perform.* **4**(4), 565 (1978)

8. M. I. Coco, S. Araujo, K. M. Petersson, Disentangling stimulus plausibility and contextual congruency: Electro-physiological evidence for differential cognitive dynamics. *Neuropsychologia,* **96**, 150-163.0 (2017)

9. M. I. Coco, A. Nuthmann, O. Dimigen, Fixation-related brain potentials during semantic integration of object–scene information. *J. Cogn. Neurosci.* **32**(4), 571-589 (2020)

10. D. Draschkow, E. Heikel, M. L. H. Vo, C. J. Fiebach, J. Sassenhagen, No evidence from MVPA for different processes underlying the N300 and N400 incongruity effects in object-scene processing. *Neuropsychologia,* **120**, 9-17. (2018)

11. M. C. Bastiaansen, M. Van Der Linden, M. Ter Keurs, T. Dijkstra, P. Hagoort, Theta responses are involved in lexical—Semantic retrieval during language processing. *J. Cogn. Neurosci.* **17**(3), 530-541 (2005)

12. R. M. Willems, R. Oostenveld, P. Hagoort, Early decreases in alpha and gamma band power distinguish linguistic from visual information during spoken sentence comprehension. *Brain Res.* **1219**, 78-90 (2008)

13. D. J. Davidson, P. Indefrey, An inverse relation between event-related and time–frequency violation responses in sentence processing. *Brain Res.,* **1158***,* 81-92 (2007)

14. W. Klimesch, EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain Res. Rev.* **29**(2-3), 169-195 (1999)

15. J. Riddle, J. M. Scimeca, D. Cellier, S. Dhanani, M. D'Esposito, Causal evidence for a role of theta and alpha oscillations in the control of working memory. *Curr. Biol.* **30**(9), 1748-1754 (2020)

16. P. Metzner, T. von der Malsburg, S. Vasishth, F. Rösler, Brain responses to world knowledge violations: A comparison of stimulus-and fixation-triggered event-related potentials and neural oscillations. *J. Cogn. Neurosci.* **27**(5), 1017-1028 (2015)

17. L. A. Hald, C. M. Bastiaansen, P. Hagoort, EEG theta and gamma responses to semantic violations in online sentence processing. *Brain Lang.* **96**(1), 90-105 (2006)

18. S. V. David, W. E. Vinje, J. L. Gallant, Natural stimulus statistics alter the receptive field structure of v1 neurons. *J. Neurosci.* **24**(31), 6991-7006 (2004)

19. Z. M. Aghajan, L. Acharya, J. J. Moore, J. D. Cushman, C. Vuong, M. R. Mehta, Impaired spatial selectivity and intact phase precession in two-dimensional virtual reality. *Nat. Neurosci.* **18**(1), 121-128 (2015)

20. D. Hölle, M. G. Bleichner, Smartphone-based ear-EEG to study sound processing in everyday life. *BioRxiv*, 2023-01 (2023)

21. R. R. Hampton, B. M. Hampstead, E. A. Murray, Selective hippocampal damage in rhesus monkeys impairs spatial memory in an open-field test. *Hippocampus,* **14***(7*), 808-818 (2004)

22. E. Redcay, L. Schilbach, Using second-person neuroscience to elucidate the mechanisms of social interaction. *Nat. Rev. Neurosci.* **20**(8), 495-505 (2019)

23. G. Maimon, A. D. Straw, M. H. Dickinson, Active flight increases the gain of visual motion processing in Drosophila. *Nat. Neurosci.* **13**(3), 393-399 (2010)

24. S. A. Nastase, A. Goldstein, U. Hasson, Keep it real: rethinking the primacy of experimental control in cognitive neuroscience. *NeuroImage,* **222**, 117254 (2020)

25. K. Gramann, D. P. Ferris, J. Gwin, S. Makeig, Imaging natural cognition in action. *Int. J. Psychophysiol.* **91**(1), 22-29 (2014)

26. A. Grasso-Cladera, M. Bremer, S. Ladouce, F. J. Parada, A Systematic Review of Mobile Brain/Body Imaging Studies Using the P300 Event-Related Potentials to Investigate Cognition Beyond the Laboratory. *PsyArxiv* (2024)

27. S. G. Shamay-Tsoory, A. Mendelsohn, Real-life neuroscience: an ecological approach to brain and behavior research. *Perspect. Psychol. Sci.* **14**(5), 841-859 (2019)

28. M. Stangl, S. L. Maoz, N. Suthana, Mobile cognition: imaging the human brain in the 'real world'. *Nat. Rev. Neurosci.* **24**(6), 347-362 (2023)

29. S. Makeig, K. Gramann, T. P. Jung, T. J. Sejnowski, H. Poizner, Linking brain, mind and behavior. *Int. J. Psychophysiol.* **73**(2), 95-100 (2009)

30. M. Klug, S. Jeung, A. Wunderlich, L. Gehrke, J. Protzak, Z. Djebbara, Z., A. Argubi-Wollesen, B. Wollesen, K. Gramann, The BeMoBIL Pipeline for automated analyses of multimodal mobile brain and body imaging data. *BioRxiv*, 2022-09 (2022)

31. G. Niso, E. Romero, J. T. Moreau, A. Araujo, L. R. Krol, Wireless EEG: A survey of systems and studies. *NeuroImage,* **269**, 119774 (2023)

32. B. Griffiths, A. Mazaheri, S. Debener, S. Hanslmayr, Brain Oscillations track the formation of episodic memories in the real world. *NeuroImage*. **143**, 256-266 (2016)

33. M. Mustile, D. Kourtis, M. G. Edwards, S. Ladouce, D. Volpe, M. Pilleri, E. Pelosin, G. Learmonth, D. I. Donaldson, M. Ietswaart, Characterising neurocognitive impairments in Parkinson's disease with mobile EEG when walking and stepping over obstacles. *Brain Commun.* **5**(6), 326 (2023).

34. S. Ladouce, D. I. Donaldson, P. A. Dudchenko, M. Ietswaart, Mobile EEG identifies re-allocation of attention during real world activity. *Sci. Rep.* **9**(1), 15851 (2019)

35. X. Wang, S. K. Ong, A. Y. Nee, A comprehensive survey of augmented reality assembly research. *Adv. Manuf.* **4***,* 1-22 (2016)

36. N. F. Saidin, N. D. A. Halim, N. Yahaya, A review of research on augmented reality in education: Advantages and applications. *Int.l Educ. Stud.* **8**(13), 1-8 (2015)

37. I. A. C. Giglioli, F. Pallavicini, E. Pedroli, S. Serino, G. Riva, Augmented reality: A brand new challenge for the assessment and treatment of psychological disorders. *Comput. Math. Methods Med.* ***862942***, 1-13 (2015)

38. A. Krugliak, A. Clarke, Towards real-world neuroscience using mobile EEG and augmented reality. *Sci. Rep.* **12**(1), 2291 (2022)

39. V. I. Nicholls, B. Alsbury-Nealy, A. Krugliak, A. Clarke, Context effects on object recognition in real-world environments: A study protocol. *Wellcome Open Res.* **7** (2022)

40. A. Clarke, B. J. Devereux, L. K. Tyler, Oscillatory dynamics of perceptual to conceptual transformations in the ventral visual pathway. *J. Cogn. Neurosci.* **30**(11), 1590-1 (2018)

41. A. Clarke, Dynamic activity patterns in the anterior temporal lobe represents object semantics. *Cogn. Neurosci.* **11**(3), 111-121 (2020)

42. E. Halgren, E. Kaestner, K. Marinkovic, S. S. Cash, C. Wang, D. L. Schomer, J. R. Madsen, I. Ulbert, Laminar profile of spontaneous and evoked theta: Rhythmic modulation of cortical processing during word integration. *Neuropsychologia,* **76**, 108-124 (2015)

43. A. Krugliak, D. Draschkow, M. L. H. Võ, A. Clarke, Semantic object processing is modulated by prior scene context. *Lang. Cogn. Neurosci.* 1-10 (2023)

44. A. Clarke, J. Crivelli-Decker, C. Ranganath, Contextual expectations shape cortical reinstatement of sensory representations. *J. Neurosci.* **42**(30), 5956-5965 (2022)

45. L. Amoruso, L. Sedeño, D. Huepe, A. Tomio, J. Kamienkowski, E. Hurtado, J. F. Cardona, M. A. A. Gonzalez, A. Rieznik, M. Sigman, F. Manes, A. Ibáñez, Time to tango: Expertise and contextual anticipation during action observation. *Neuroimage,* **98**, 366-385 (2014)

46. Y. C. Wu, S. Coulson, Meaningful gestures: Electrophysiological indices of iconic gesture comprehension. *Psychophysiology,* **42**(6), 654-667 (2005)

47. Ş. B. Demiral, G. L. Malcolm, J. M. Henderson, ERP correlates of spatially incongruent object identification during scene viewing: Contextual expectancy versus simultaneous processing. *Neuropsychologia,* **50**(7), 1271-1285 (2012)

48. G. Zhang, D. R. Garrett, S. J. Luck, Optimal filters for ERP research II: Recommended settings for seven common ERP components. *Psychophysiology*, e14530 (2024)

49. D. Tanner, K. Morgan-Short, S. J. Luck, How inappropriate high-pass filters can produce artifactual effects and incorrect conclusions in ERP studies of language and cognition. *Psychophysiology,* **52**(8), 997-1009 (2015)

50. C. Tallon-Baudry, O. Bertrand, Oscillatory gamma activity in humans and its role in object representation. *Trends Cogn. Sci.,* **3**(4), 151-162 (1999)

51. D. Iggena, S. Jeung, P. M. Maier, C. J. Ploner, K. Gramann, C. Finke, Multisensory input modulates memory-guided spatial navigation in humans. *Commun. Biol.* **6**(1), 1167 (2023)

52. K. Gramann, F. U. Hohlefeld, L. Gehrke, M. Klug, Human cortical dynamics during full-body heading changes. *Sci. Rep.* **11**(1), 18186 (2021)

53. C. Hilton, A. Kapaj, S. I. Fabrikant, Fixation-related potentials during mobile map assisted navigation in the real world: The effect of landmark visualization style. *Atten., Percept., & Psychophys.* 1-16 (2024)

54. A. Delaux, J. B. de Saint Aubert, S. Ramanoël, M. Bécu, L. Gehrke, M. Klug, M., R. Chavarriaga, J-A. Sahel, A. Arleo, Mobile brain/body imaging of landmark-based navigation with high-density EEG. *Eur. J. Neurosci.* **54**(12), 8256-8282 (2021)

55. T. M. Mahmoudi, K. Badie, M. Valipour, Assessing the role of AR-based content in improving learning performance considering Felder-Silverman learning style. In 2015 International Conference on Interactive Collaborative Learning (ICL) *IEEE.* 838-843 (2015)

56. N. Capuano, A. Gaeta, G. Guarino, S. Miranda, S. Tomasiello, Enhancing augmented reality with cognitive and knowledge perspectives: A case study in museum exhibitions. *Behav. Inf. Tech.* **35**(11), 968-979 (2016)

57. MATLAB: version 9.11.0 (R2021b). Natick, Massachusetts: The MathWorks Inc.2021.

58. H. H. Jasper, Ten-twenty electrode system of the international federation. *Electroencephalogr. Clin. Neurophysiol.* **10**, 371-375 (1958)

59. M. Hovhannisyan, A. Clarke, B. R. Geib, R. Cicchinelli, Z. Monge, T. Worth, A. Szymanski, R. Cabeza, S. W. Davis, The visual and semantic features that predict object memory: Concept property norms for 1,000 object images. *Mem. Cognit.* **49**, 712-731 (2021)

60. M. van Casteren, M. H. Davis, Match: A program to assist in matching the conditions of factorial experiments. *Behav. Res. Methods.* **39**(4), 973-978 (2007)

61. A. Delorme, S. Makeig, EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods.* **134**(1), 9-21 (2004)

62. A. de Cheveigné, ZapLine: A simple and effective method to remove power line artifacts. *NeuroImage,* **207**, 116356 (2020)

63. J. A. Palmer, S. Makeig, K. Kreutz-Delgado, B. D. Rao, Newton method for the ICA mixture model.In: 2008 IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 1805–1808 (2008)

64. L. Pion-Tonachini, K. Kreutz-Delgado, S. Makeig, ICLabel: An automated electroencephalographic independent component classifier, dataset, and website. *NeuroImage,* **198**, 181-197 (2019)

65. R. Oostenveld, P. Fries, E. Maris, J. M. Schoffelen, FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.* **2011**, 1-9 (2011)

Acknowledgements

Funding: This research was funded in whole, or in part, by the Wellcome Trust [Grant number 211200/Z/18/Z to AC]. For the purpose of open access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

Author Contributions:

Conceptualisation: VIN, AK, AC

Methodology: VIN, AK, BA-N, KG, AC

Investigation: VIN, AK, AC

Visualisation: VIN, AC

Supervision: AC, KG

Writing – original draft: VIN, AC

Writing – review & editing: VIN, AK, BA-N, KG, AC

Competing interests: Benjamin Alsbury-Nealy is the CEO and founder of SilicoLabs, which provided the software for the experiment creation. Alex Clarke is a scientific advisor for SilicoLabs.

Data materials and availability: All data, code, and materials can be found on zenodo at: https://zenodo.org/doi/10.5281/zenodo.11394630