Expectation-Based Gist Facilitation: Rapid Scene Understanding and the Role of Top-Down Information

Dominic McLean (1), Antje Nuthmann (2), Louis Renoult (1), George L. Malcolm (1)

1) School of Psychology, University of East Anglia, Norwich, United Kingdom

2) Institute of Psychology, University of Kiel, Kiel, Germany

Word count: 19,138

Author Note

We have no known conflict of interest to disclose

Correspondence concerning this article should be addressed to Dr Dominic McLean, School of Psychology, Lawrence Stenhouse Building, University of East Anglia, Norwich Research Park, Norwich, NR4 7TJ, UK. Email: d.mclean@uea.ac.uk

Data available at: OSF | "Expectation-Based Gist Facilitation" - Data (https://osf.io/km6yg/)

Data from this study has been presented at the 2020 annual conferences of the Vision Sciences Society and the Experimental Psychology Society

Abstract

Scene meaning is processed rapidly, with 'gist' extracted even when presentation duration spans a few dozen milliseconds. This has led some to suggest a primacy of bottom-up information. However, gist research has typically relied on showing successions of unrelated scene images, contrary to our everyday experience in which the world unfolds around us in a predictable manner. Thus, we investigated whether top-down information – in the form of observers' predictions of an upcoming scene – facilitates gist processing. Within each trial, participants (*N* = 370) experienced a series of images, organised to represent an approach to a destination (e.g., walking down a sidewalk), followed by a target scene either congruous or incongruous with the expected destination (e.g., a store interior or a bedroom). A series of behavioural experiments revealed that: appropriate expectations facilitated gist processing; inappropriate expectations interfered with gist processing; sequentially-arranged scene images benefitted gist processing when semantically related to the target scene; expectation-based facilitation was most apparent when duration was most curtailed, and; findings were not simply the result of response bias. We then investigated the neural correlates of predictability on scene processing using ERP (*N*=24). Congruency-related differences were found in a putative scene-selective ERP component, related to integrating visual properties (P2), and in later components related to contextual integration including semantic and syntactic coherence (N400 and P600, respectively). Together, results suggest that in real-world situations, top-down predictions of an upcoming scene influence even the earliest stages of its processing, affecting both the integration of visual properties and meaning.

Keywords: scene processing, gist, top-down information, event-related potentials, semantic integration

Expectation-Based Gist Facilitation: Rapid Scene Understanding and the Role of Top-Down Information

Apart from at waking, every environment we encounter is part of a progression of scenes unfolding around us as we move through our surroundings. Any single scene is not confronted in isolation but instead is simply the most recently perceived environment within the continuous experiential flow of our passage through the world. However, scene perception research has largely ignored this point, focusing more on the mechanisms responsible for processing segregated, individual scene images. In a traditional experiment, a participant may be faced with a rapidly presented sequence of images, depicting a mountain, followed by a church, a kitchen, and so forth, a scenario clearly divergent from the progressive and structured environments one inhabits within the course of daily life.

We have learned a great deal from investigating the processing of isolated scene images, and such paradigms have been highly effective in identifying the mechanisms and visual features that facilitate processing of the initial meaning, or conceptual 'gist' (see Oliva, 2005), of a scene. This form of gist – i.e., the ability to derive the semantic information contained within a perceptual landscape – can be extracted even under conditions with viewing times spanning less than a tenth of a second (e.g., Bacon-Macé et al., 2005; Fei-Fei et al., 2007; Greene & Oliva, 2009; Joubert et al., 2007; Potter, 1975; Potter et al., 2014). Such limited durations have led many to infer the primacy of bottom-up visual factors in rapid scene perception (Fabre-Thorpe et al., 2001; Itti et al., 1998; Potter et al., 2014; Rumelhart, 1970), with the conviction that top-down information can have only a limited role under such brief time frames. This is a fair assessment if one contends that initial scene processing takes place in a classic hierarchical fashion. In such a scenario – of progressive activation through a linear pathway of anatomical areas divergent in terms of functional specificity – it is unlikely top-down feedback would be received prior to such rapid scene categorisation taking place. Therefore, while such models do not deny the role of feedback or re-entrant connectivity as processing continues through time, they propose feature-extraction mechanisms as sufficient for distinguishing conceptual information and meaning within complex natural scenes, with a single 'forward sweep' of neural activity through the ventral stream (e.g., Fabre-Thorpe et al., 2001; Potter et al., 2014).

However, the traditional view of the serial processing of visual input has been questioned for some time (Engel et al., 2001; Ullman, 1995) and, while artificial models of the visual system have demonstrated high recognition performance when reliant on bottom-up input (e.g., Cichy et al., 2016; Greene & Hansen, 2018; Serre et al., 2007), the latest recurrent models can better explain human visual recognition when compared to feedforward neural networks (e.g., Spoerer et al., 2020). Likewise, the past decade has seen great advances in our understanding of the broad extent of reciprocal connections within the neural architecture (e.g., Groen et al., 2017; Kauffman et al., 2015; Kravitz et al., 2013). For instance, research methodologies spanning MEG, EEG and TMS have all provided evidence for rapid local recurrent processes within early visual cortex (Boehler et al., 2008; Camprodon et al., 2010; de Graaf et al., 2014; Foxe & Simpson, 2002), with the proposal that these processes might start only a few tens of milliseconds after the arrival of the visual input (de Graaf et al., 2012). Furthermore, multiple feedforward-feedback loops have been hypothesised as taking place within the first 100 ms of stimulus onset (Bullier, 2001; Juan & Walsh, 2003), a proposition strengthened by the finding of activation in intermediate visual areas prior to the completed contribution of early visual cortex (Koivisto et al., 2011). So, while the separate contributions of top-down and bottom-up processing in visual recognition remain robustly debated (see, for example, Firestone & Scholl, 2016), there is reason to suggest that feedforward accounts cannot fully explain typical processing.

Similarly, from the object processing literature, evidence reveals that top-down processes are initiated prior to completion of target recognition, with the suggestion that early activation of higher-order brain regions facilitates the systematic analysis of bottom-up information (Bar et al., 2006). In other words, low spatial frequency information is passed rapidly to higher areas and is then used to form predictions as to the identity of the object being viewed. Consequently, this allows for the pre-activation of a limited set of object representations which are subsequently matched against the continuing flow of bottom-up information (Bar et al., 2006). It seems reasonable to infer that some equivalence may exist within the manner of operation for scene processing, whereby an initial 'sketch' (Marr, 1982; Rensink, 2000) of the environment may allow for the pre-activation of scene representations in higher-order areas. Indeed, parallel co-activity within higher regions has been observed even while perceptual coding of visual scenes is actively proceeding (Catherwood et al., 2014). While concurrent activation cannot be taken as direct evidence for interaction between regions, it provides the opportunity for such interactions to a far greater extent than models which assume somewhat step-by-step activation, whereby higher-order processing occurs only as perception subsides.

The above research shows that the selection and processing of those elements within even the precursory stages of the feed-forward wave of activity may be open to facilitation. Moreover, if top-down information can rapidly influence bottom-up processing in scenarios such as these – where no indication as to what will be displayed is provided prior to stimulus onset – then it seems appropriate to suggest that top-down influence might be even more rapid when pre-target cues allow for a subsequent visual image to be predicted. This suggestion is reinforced when considering growing evidence which demonstrates that activity within the visual cortex, including early striate cortex, can be affected by expectations alone (e.g., Aitken et al., 2020; Grill-Spector & Malach, 2004; Kok et al., 2012), that the shape-selectivity of neurons in area V1 is altered depending on what geometric shape is expected (McManus et al., 2011), and that *a priori* expectations generated by scenic context can lead to increased activation in higher-order areas during subsequent visual processing (Caplette et al., 2020).

The influence of context-based expectations on cognitive processing has been extensively investigated in object-scene relationships, which have repeatedly shown that target objects are found more quickly (Biederman et al., 1973; Võ & Henderson, 2011), and with higher accuracy (Antes et al., 1981; Davenport & Potter, 2004), when situated within 'appropriate' scenes (i.e., where the scene category and target object are semantically congruous). These context effects have been found not only during the simultaneous presentation of a scene and target object, but also when a scene image is presented prior to (Demiral et al., 2012; Ganis & Kutas, 2003; Võ & Wolfe, 2013), and independent of (Palmer, 1975), object presentation. Due to the speed with which objects can be detected and identified (e.g., Crouzet & Serre, 2011; Kirchner & Thorpe, 2006; Thorpe et al., 1996), these studies demonstrate that semantic information can rapidly influence visual processing, and that increased processing ability related to congruency is evident even when natural scene images are used to induce expectations. If scenes can provide semantic information capable of altering subsequent object processing, it would seem intuitive that such influence similarly extends to subsequent scene processing.

Indeed, experimental evidence has demonstrated that a scene can be primed by a preceding scene-image, termed the 'scene priming' effect, although this has tended to focus on priming at the perceptual – rather than conceptual – level. For instance, increased performance regarding spatial layout judgements have been elicited when target scenes are primed using an identical scene image (Sanocki, 2013) or with images of the target scene from different viewpoints (Sanocki & Epstein, 1997; but see Epstein et al., 2005), while image detection ability is improved if primed across scenes more closely matched in terms of spectral information (Caddigan et al., 2017). Furthermore, when primes and targets are adjacent segments of the same complete landscape – thus intrinsically different while being similar in general composition – biases to cortical responses, alongside improved feature detection performance, have been shown (Blondin & Lepage, 2005).

However, recent research has begun to suggest a potential influence of top-down factors on conceptual gist processing. Greene et al. (2015), for instance, briefly presented atypical scenes – such as a living room with a boulder in the centre, or a pillow-fight in a town square – were found to be more difficult to both process and understand compared to frequently encountered scene types (e.g., a car in a driveway). This indicates that an observer's prior semantic knowledge can influence the rapid processing of complex natural scenes, even over highly curtailed presentation durations.

However, the design of that study still involved the presentation of single, unrelated images on each trial, and so cannot apprise us of the interaction between immediately preceding information and predictability. So, while such research highlights the cost of violating the expectations held in long-term memory, it speaks less to the violation of expectations built upon the ‘on-line’ flow of information as it is received.

Here we extend previous findings to investigate if an observer’s expectations of an upcoming scene category have a direct effect on the initial stages of processing, i.e., the extraction of conceptual gist, and the timeframe involved, using a combination of behavioural and electrophysiological (EEG) measures. In so doing, we attempt to better replicate how scenes are processed outside the laboratory, namely as predictable settings preceded by contextually relevant visual information, and hence proffer that models based on a progression of activation across successive regions cannot provide an exhaustive account of functionality.

Recent research has started to address this directly, by pointing towards the influence of predictions on conceptual gist processing through the use of pre-target narrative sequences (Smith & Loschky, 2019). Here participants were presented with either a series of 10 sequentially-organised scenes (designed to simulate a journey between separate locations on, or nearby, their campus) or the same scenes but randomly ordered. When images were presented in a sequentially-organised manner, categorisation performance for – and predictability of – target scenes was significantly increased. Though this work concerned the effect of the ordering of pre-target images (i.e., their spatiotemporal coherence) from a familiar environment, rather than their congruency with an upcoming, novel target-scene, it does reveal that expectations can be informed by what has gone before and, moreover, that these expectations may have a functional role in terms of facilitating conceptual scene-gist processing. The authors suggest that narrative sequences help construct a current event model, which then in turn influences the extraction of gist information. As a consequence, an iterative process is created whereby ‘front end’ information extraction (such as scene gist derived from attentional selection mechanisms) informs ‘back end’ model construction (initially stored in working memory), which in turn influences front end processes, and so forth (Loschky et al., 2020).

To tease apart the role of on-line expectations within processing, the current study investigated the influence of visual information received immediately prior to target-scene onset. Across all experiments we employed a fundamental change to the traditional methodologies, which either position targets within a rapid serial visual presentation (RSVP) sequence of unrelated images (e.g., Potter, 1975) or present only a single image per trial (e.g., Greene et al., 2015). This was achieved by providing contextual information through presentation of antecedent 'lead-up' images, similar to Smith & Loschky (2019), allowing us to investigate the influence of just-prior experience on the understanding of scenes. These leading images provided a flow of movement through an environment and towards a scene, and so represented an approach to a destination. This creates a more naturalistic means by which to generate predictions based on lifelong experience, and as a result is somewhat removed from research investigating the effect of predictions on perception using simplistic pre-target cues (e.g., Summerfield & Koechlin, 2008), virtual reality experiments which maintain a flow of movement but in a limited physical space (e.g., Kit et al, 2014), or where predictability is manipulated by synthetic means such as the learning of arbitrary contingencies prior to task commencement (e.g., Hindy et al., 2016). A key aim of the current study, therefore, was to provide a more ecologically valid reflection of scene perception. While only an approximation of this can be achieved with a sedentary participant viewing static images on a monitor, careful construction of image-series was considered sufficient in affording an impression of progress through a landscape.

Then, by manipulating whether the target scene was congruous with these leading images, i.e., the 'approach-destination' congruency, we hoped to demonstrate whether there is indeed an influence of predictability on scene categorisation ability. In addition, across the separate behavioural experiments we manipulated the presentation duration of destination images, the spatiotemporal coherence of approach-image sequences, and the provision of pre-destination scenic context to investigate the mechanisms underlying the effect of expectations on gist processing, while also examining participants' discriminability of target scenes to ensure performance was not simply driven by response bias. Finally, we turned to electroencephalography to map changes in brain activity relating to the manipulation of approach-destination congruency, with the aim of identifying the forms of cognitive processing most readily affected by the violation of expectations.

Such an investigation helps inform a number of wide-ranging debates. Much discussion continues within scene research as to feed-forward vs. feed-back mechanisms (e.g., Greene et al., 2015; Potter et al., 2014; Smith & Loschky, 2019), but strongly contested disputes as to the separate contributions of top-down and bottom-up information are felt across visual recognition paradigms more generally (e.g., Firestone & Scholl, 2016, Kveraga et al., 2011). Even amongst those prescribing to a view that expectations can influence visual processing, extensive debate continues as to what stages of processing are accessible to predictions (e.g., Bar, 2004; Biederman et al., 1982). In short, therefore, the study of expectation-based effects on gist processing can help inform our understanding of cognitive processes in a broad and extensive manner.

General Methods

All experiments were approved by the ethics committee at the University of East Anglia's School of Psychology. The experiments were programmed using PsychoPy (version 1.85.3, Peirce et al., 2019) unless otherwise stated. The data are available at https://osf.io/km6yg/.

Participants

Participants in Experiments 1a and 1b were undergraduate Psychology students, recruited through the University of East Anglia's research pool, who received course credits for participating. The sample tested in Experiment 2 comprised both students and staff of the University, receiving either course credits or a small payment (£4) for taking part. For Experiments 3 and 4, participants were recruited online through Prolific (Palan & Schitter, 2018) and received a small payment (£3). Experiment 5 included Psychology students, again recruited through the research pool and given course-related credits for taking part. All participants provided informed consent prior to taking part in the study.

Stimuli

Each experiment used the same collection of images, comprised of photographs taken by the researchers and high-definition images of sceneries and video-stills freely available on the internet. The intention when constructing the series was to create progressions which mimicked movement through an environment (i.e., an 'approach') towards a target scene (i.e., a 'destination'), while reducing instances of over-similarity across viewpoints and avoiding sudden 'jumps' in the progression. Accordingly, variations in geographical distances between approach images needed to be considered across series, mainly due to the differing constraints imposed by the superordinate categories. For example, the distance between points during a progression through a house to, say, a bedroom would be inherently shorter when compared to the points of progression towards a beach. Therefore, we attempted to instil a 'semantic flow' within each series, with each of the transitional points of the approach represented in a manner which maintained the sense of a progression throughout.

There are four other important points to note relating to the construction of series. Firstly, the destination scene could not be determined from the earliest approach images. This was due to there being similar progressions across many series, in both interior-destination series (for instance, 'bathroom' and 'bedroom' targets would have similar approaches, involving stairways, hallways, etc.), and exterior-destination series (where many progressions shared similarities, such as traversing pavements, pathways and carparks). Furthermore, the eventual superordinate category of the target could not be anticipated at the start of the series: the approach images might represent a journey out in the open but with an indoor destination scene, or vice versa, such as walking across a garden before entering an outbuilding. Additionally, approaches frequently passed through other target categories. For example, images of a high street – a target category on some trials – might be passed through within the approach images of a series with a ‘shop’ target. This potential interplay across trials was at the category level, not the exemplar level, as no scenery (whether approach image or destination) was repeated at any point during the task.

Secondly, a balance was struck in terms of the final approach image representing a viewpoint geographically close enough to heighten expectations as to the destination, while trying to minimise the amount of similarity in low-level features across these two images. This was to ensure performance was based on semantic prediction rather than simply the repetition of low-level visual information. Therefore, while some features of a destination might be visible within the later approach images (such as the corner of a table and chair seen through a doorway prior to reaching a ‘dining room’ target) care was taken to maintain substantial differences in both the viewpoint and available visual features between the approach images and the destination scene. While it is inherent within this design that there are likely to be, on the whole, greater similarities in visual properties across Congruous than Incongruous series, the above safeguards aimed to minimise this contribution of low-level perceptual features.

Thirdly, the inclusion of people within images was kept to a minimum. It was not considered necessary to exclude pedestrians, shoppers, etc. from the sequences, but no images included individuals positioned close in the foreground or looking directly at the observer. Finally, all images (with the exception of multi-storey carparks) were of sceneries outside the county of the University’s location, in an attempt to limit any potential confounds due to familiarity with the specific exemplars used.

In total, 756 images were used as stimuli, of which 720 appeared in the experimental trials (600 for the online experiments). No images were repeated. Each trial consisted of sequences of spatiotemporally coherent approach images, followed by a target scene, resulting in 120 individual series. There were four series for each of the 30 scene categories (see Appendix A for list). Destination scenes that followed their respective approach series of images were considered Congruous trials. Depending on the experiment, a certain number of series were chosen at random to serve as Incongruous trials (30 for Experiment 1a; 90 for Experiment 1b; 60 for all other Experiments). The distribution of Incongruous trials was always evenly spread across scene categories. The target scenes of each of these selected series were then randomly reallocated amongst each other. This redistribution was conducted in adherence to two principles. Firstly, a target could not replace another target of the same scene category and, secondly, that a target could only replace another target of the same superordinate category (in terms of interior / exterior distinction). This division was maintained due to suggestions that discriminating between superordinate categories is not analogous to discriminating between basic-level categories (see, for example, Banno & Saiki, 2015).

For the behavioural experiments, each target image was followed by a set of five masks, presented rapidly in sequence. This technique of dynamic pattern masking has been shown to be more effective at masking visual features, due to the minimising of potential correspondences between the target and the mask (see, for example, Bacon-Macé et al., 2005; Greene & Oliva, 2009; Greene et al., 2015). A different set of masks was used after each target. To achieve this, 600 masks were generated from the approach images by using Portilla and Simoncelli’s (2000) texture synthesis algorithm in Matlab, in line with previous research showing this to be an effective method for placing temporal constraints on bottom-up processing of scene images (Loschky et al., 2010). For the online experiments, target images were followed by a single mask (due to software limitations), while the ERP experiment did not use visual masking.

Performance was judged through participants selecting the category that best described each destination scene from a list of six options (with the exception of Experiment 4 where the choice was binary). The available category options on each response screen were allocated randomly and were also randomised in terms of item position. All options were of the same superordinate category (indoor / outdoor) as the target. For Incongruous trials, the category the approach images would be expected to lead to was also included.

For the lab-based behavioural experiments, all images and masks were displayed with an image resolution of 800 × 600. All images were presented in colour, on a 24” monitor (HP EliteDisplay E240c; screen width × height: 21” × 12”; resolution: 1920 × 1080) with a typical grey-to-grey response time of 7 ms and a refresh rate of 60Hz. The experiment was run on an HP EliteDesk personal computer (processor: Intel® Core™ i5-6500T CPU@2.50GHz, 64-bit operating system, Intel® HD Graphics 530, Windows 10 Enterprise version 1607).

The ERP experiment was run on two Viglen Genie DQ77MK PCs (processor: intel core i7-3770 CPU@3.50GHz, 64-bit operating system, GeForce GT610 graphics card, Windows 7). Scene images were presented on a 24” computer monitor, resolution 1920 × 1080 (BenQ XL2411T).

Statistical Analysis using Mixed Models

The data from the five behavioural experiments were analysed with generalised linear mixed-effects models (GLMMs; Jaeger, 2008). The analyses were conducted using the `glmer` function from the R package *lme4* (version 1.1-27.1; Bates et al., 2015). To resolve convergence warnings, we used the BOBYQA optimizer and set the control parameter `calc.derivs` to `FALSE` (Brown, 2021); the default settings were used otherwise. For each experimental trial, the subject’s response was coded as correct (1) or incorrect (0). The binary variable categorisation success was analysed using binomial GLMMs with a logit link function (the default). Using the logit link function means that parameter estimates are obtained on the log-odds or logit scale. This scale is symmetric around zero and ranges from negative to positive infinity, with a logit of 0 corresponding to a probability of .5 (Jaeger, 2008).

To select an optimal random-effects structure for the GLMMs reported in this article, we pursued a data-driven approach using backward model selection. To minimise the risk of Type I error, we started with the maximal random-effects structure justified by the design (Barr et al., 2013). The maximal structure was stepwise backwards-reduced to arrive at a mixed model that was justified by the data (Matuschek et al., 2017). To select this parsimonious model (Bates et al., 2018), we conducted model comparisons using likelihood-ratio tests and information criteria (Matuschek et al., 2017).

The general advantages of mixed models in experimental research are well documented (Baayen et al., 2008; Jaeger, 2008; Kliegl et al., 2011). Concerning the data from Experiments 1a and 1b, a specific advantage of using GLMMs is that they consider the design-inherent imbalance of congruous and incongruous trials in these experiments (see Kliegl et al., 2011, for discussion). Moreover, mixed models account for the combined variance of subjects and items in a single model (Judd et al., 2012; Nuthmann et al., 2017). Note that our stimulus items consisted of target scenes that were preceded by a set of approach images.

In the GLMM analyses, data from individual trials (subject–item combinations) were considered. For figures showing categorisation accuracy, we calculated the proportion of correct answers for each participant in each experimental condition and plotted the means across participants' means. The data figures were created with the *ggplot2* R package (version 3.3.5; Wickham, 2016), with the error bars representing 95% confidence intervals (CIs) calculated with the *superb* R package (version 0.95.0; Cousineau et al., 2021).

Experiment 1a

The ability to categorise scenes even within very brief presentation durations has led many to argue that such rapidity of processing must take place largely outside the involvement of top-down influence (Fabre-Thorpe et al., 2001; Itti et al., 1998; Potter et al., 2014; Rumelhart, 1970). On the other hand, more recent research has found that semantic information can influence scene processing within shorter timeframes than previously thought (Greene et al., 2015; Võ & Wolfe, 2013). However, this research has largely focused on the semantic congruity of objects within a scene, rather than congruity between scenes. We here addressed this gap by manipulating the congruency between approach scenes and a destination (i.e., a target scene), allowing us to investigate whether semantic predictability of an upcoming scene influences processing.

Experiment 1a also manipulated target presentation duration, to investigate whether the influence of contextual information remained consistent across the different stages of scene processing. Specifically, models assuming primacy of bottom-up factors during gist processing would not expect differences in categorisation performance as a function of congruency at target durations below 100 ms. Under such models (e.g., Potter et al., 2014), the category of the lead-up scenes would be expected to have minimal influence during the gist processing of the subsequent target image. Conversely, if performance differences were found at such brief durations this would lend support to the proposition for top-down influences on gist processing.

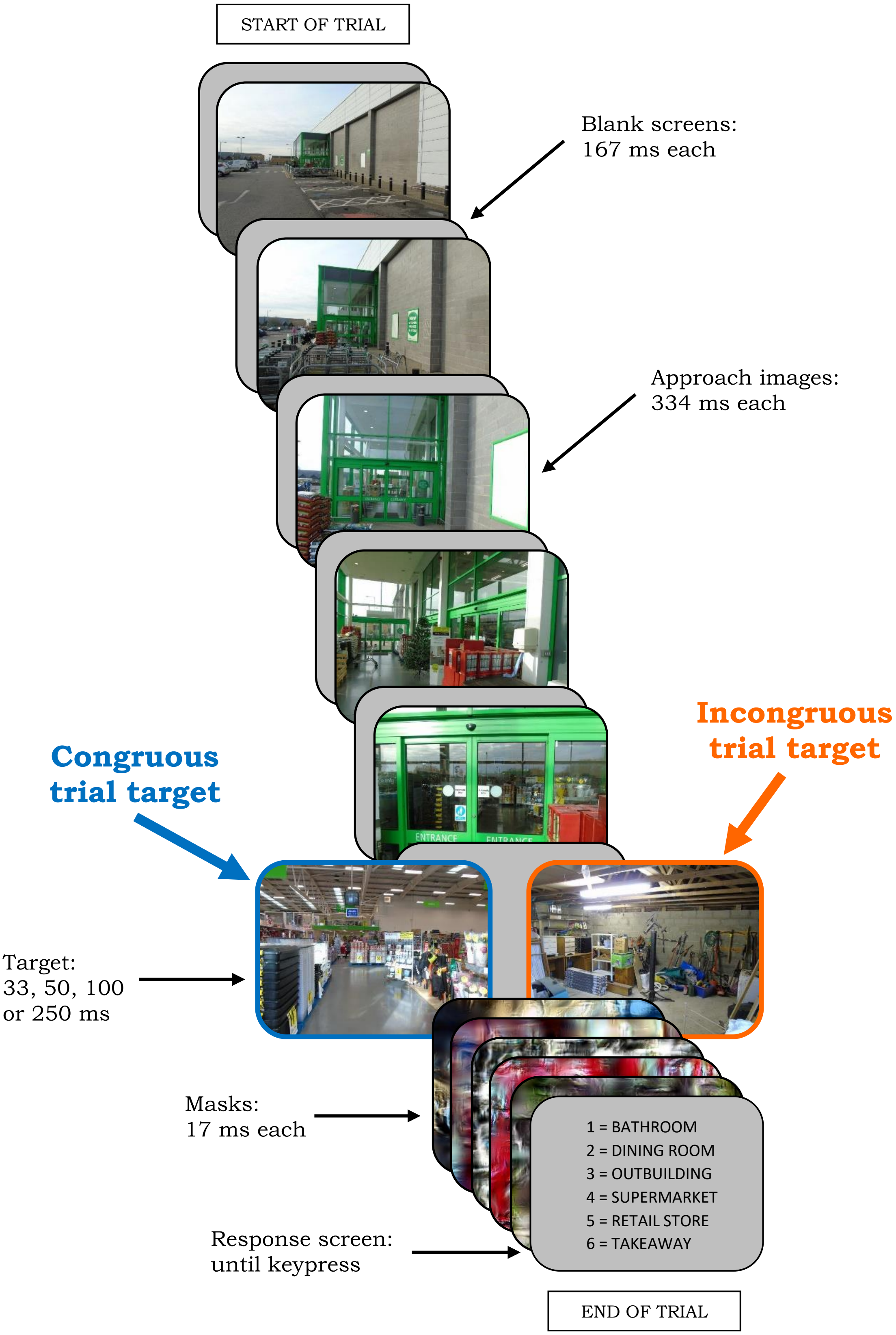
We hypothesised that destination scenes preceded by congruous approach images would be more accurately categorised, compared to those with incongruous approaches. Additionally, we predicted that this benefit would be most apparent at briefer presentation durations where the ability to extract visual information would be most curtailed.

Participants

Data from 129 participants were collected (*M*$_{\text{age}}$ = 20.09, *SD*$_{\text{age}}$ = 3.68; 103 women, 26 men; 113 right-handed, 15 left-handed, 1 ambidextrous).

Figure 1

Schematic of the Protocol for Experiments 1a and 1b


START OF TRIAL
Blank screens:
167 ms each
Approach images:
334 ms each
Congruous trial target
Incongruous trial target
Target:
33, 50, 100
or 250 ms
Masks:
17 ms each
1 = BATHROOM
2 = DINING ROOM
3 = OUTBUILDING
4 = SUPERMARKET
5 = RETAIL STORE
6 = TAKEAWAY
Response screen:
until keypress
END OF TRIAL

Procedure

Prior to starting the experiment, a series of instruction screens were displayed explaining the task and prompting participants to imagine travelling through the environments that were presented. This was followed by six practice trials, with the opportunity to ask any questions of the researcher on their completion. The same set of practice trials, in the same order, was experienced by each participant.

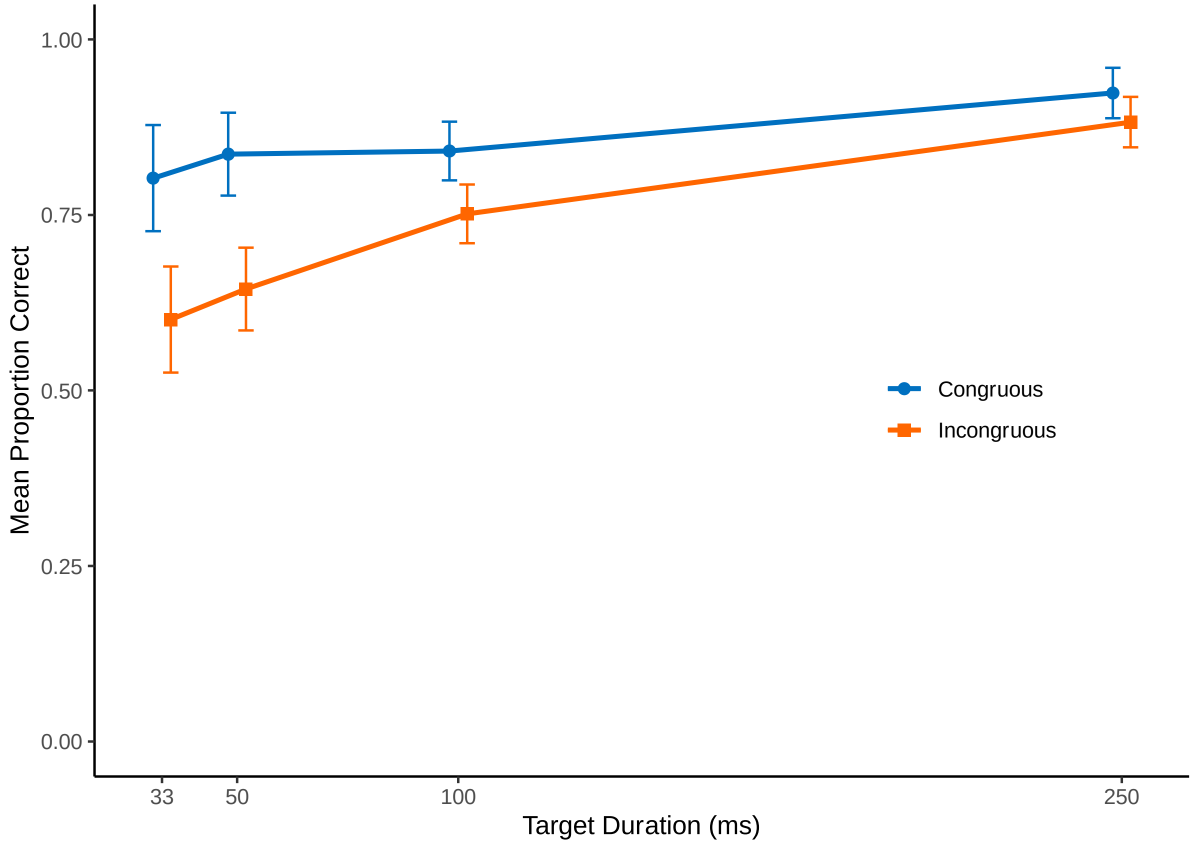
The 120 trials were presented in a randomised order for each participant. Each trial included five sequential approach images, separated by blank screens, followed by a destination image (see Figure 1). To ensure participants attended to the leading images, the destination scene was congruous with the approach scenes on 75% of the trials. Target scenes could be from one of 30 separate categories, split equally between interior and exterior sceneries, and could be presented for 33, 50, 100 or 250 ms (2, 3, 6 or 15 frames on a 60Hz monitor). A series of five masks began at target-offset, prior to a 6-AFC response screen. Once a response had been given, by pressing the number on the keyboard corresponding to the chosen category, the next trial began. At three equally spaced points within the task an ‘optional break’ screen was displayed, where participants could choose to pause if they wished and recommence once any key was pressed.

Results and Discussion

One participant was removed from the analysis due to a zero score for the Incongruous condition, suggestive of a misunderstanding of the task. Figure 2 displays the observed data for the included participants (*n* = 128).

Figure 2

Experiment 1a. Scene Categorisation Accuracy for Congruous and Incongruous Trials as a Function of Target Duration



Note. Error bars represent 95% CIs for mixed designs (Cousineau et al., 2021).

For inferential statistics, binomial logit mixed models were used. To test the effect of Congruency on categorisation accuracy, simple coding (–0.5 / +0.5) was used, with the Congruous condition serving as the reference level. With this contrast coding, the fixed-effect estimate for Congruency represents the mean difference in accuracy between Incongruous and Congruous conditions. To test the effect of Target duration, backward difference coding was used. To this end, the four-level factor target duration was ordered from shortest to longest duration. The three resulting contrasts describe the differences in mean categorisation accuracy between conditions 2 and 1 (50 ms – 33 ms), conditions 3 and 2 (100 ms – 50 ms), and conditions 4 and 3 (250 ms – 100 ms), respectively. Three interaction coefficients additionally describe whether the Congruency effect differs between adjacent target-duration conditions. The GLMM included altogether nine fixed effects and a parsimonious random-effects structure supported by the data (Table 1).

There was a significant main effect of approach-destination congruency on scene categorisation accuracy, with lower performance on Incongruous than Congruous trials, *b* = -0.85, *SE* = 0.26, *z* = -3.32, *p* = .001. Categorisation accuracy did not differ significantly between the two shortest target scene presentation durations, condition 2 – condition 1: estimate: *b* = 0.3, SE = 0.24, *z* = 1.21, *p* = .225. Further, categorisation accuracy was not significantly different at a target duration of 100 ms compared with 50 ms, condition 3 – condition 2: *b* = 0.44, *SE* = 0.24, *z* = 1.78, *p* = .075. However, categorisation accuracy was significantly higher for the 250 ms than for the 100 ms condition, condition 4 – condition 3: *b* = 1.13, *SE* = 0.26, *z* = 4.39, *p* < .001. None of the interactions between congruency and target duration were significant (Table 1).

To further explore how the congruency effect developed across target durations, we ran a post hoc GLMM with dummy coding for the factors Congruency (reference level: Congruous) and Target duration (reference level: 33 ms). With this contrast coding, the simple effect of Congruency represents the difference in categorisation accuracy for Incongruous compared with Congruous trials at a target duration of 33 ms. This difference score was significant, indicating lower accuracy for Incongruous compared with Congruous trials, *b* = -1.28, *SE* = 0.34, *z* = -3.82, *p* < .001. For Congruous trials, there was no significant difference in categorisation accuracy between the 33 ms condition and the 50 ms condition (*b* = 0.30, *SE* = 0.22, *z* = 1.38, *p* = .169) and between the 33 ms and the 100 ms condition (*b* = 0.40, *SE* = 0.22, *z* = 1.84, *p* = .066). However, categorisation accuracy for Congruous trials was significantly higher in the 250 ms condition than in the 33 ms condition, *b* = 1.33, *SE* = 0.23, *z* = 5.72, *p* < .001. The interaction terms describe the difference in performance for Incongruous compared with Congruous trials at each target duration, relative to the 33 ms baseline condition. Mirroring the results of the main analysis, the congruency effect was not significantly different for the 50 ms compared with the 33 ms condition, Congruency × Target duration 50 ms – 33 ms: *b* = -0.01, *SE* = 0.36, *z* = -0.02, *p* = .981. Moreover, the congruency effect was not significantly reduced when the target scenes were presented for 100 ms compared with 33 ms, Congruency × Target duration 100 ms – 33 ms: *b* = 0.66, *SE* = 0.36, *z* = 1.83, *p* = .067. For the 250 ms condition, however, the difference between Incongruous and Congruous trials was significantly smaller than for the 33 ms condition, Congruency × Target duration 250 ms – 33 ms: *b* = 1.08, *SE* = 0.39, *z* = 2.79, *p* = .005.

Table 1

Generalized linear mixed model fitting the effects of Congruency and Target duration on scene categorisation accuracy in Experiment 1a: estimates of coefficients (*b*), standard errors (*SE*), *z* values, and *p* values for fixed effects; variances, standard deviations (*SD*), and correlations for random effects

Fixed Effects	*b*	*SE*	*z*	*p*
Intercept	1.8851	0.1416	13.315	< .001
Congruency	-0.8512	0.2567	-3.316	.001
Target duration 2 – 1 (50 ms – 33 ms)	0.2973	0.245	1.214	.225
Target duration 3 – 2 (100 ms – 50 ms)	0.4362	0.2447	1.783	.075
Target duration 4 – 3 (250 ms – 100 ms)	1.1319	0.2578	4.391	< .001
Congruency × Target duration 2 – 1	-0.0087	0.3578	-0.024	.98
Congruency × Target duration 3 – 2	0.6674	0.3592	1.858	.063
Congruency × Target duration 4 – 3	0.4184	0.3867	1.082	.279
Random Effects				
Groups	Name	Variance	*SD*	Correlation
Subject ID	Intercept	0.8889	0.9428	Intercept
	Congruency	1.7372	1.3180	0.52
Target scene ID	Intercept	1.0793	1.0389	Intercept

As predicted, in Experiment 1a we found a significant benefit to categorisation performance when a target scene was preceded by semantically congruous approach images, revealing that participants' expectations were influencing scene processing. Furthermore, the largest performance difference across congruency conditions was found at shorter target durations (33-100ms) while the smallest was at the longer duration (250ms), indicative of gist extraction being modulated by top-down information.

These findings sit in agreement with a growing list of studies showing that expectations influence subsequent processing of scenes (Smith & Loschky, 2019), as well as research showing that object processing can be facilitated if situated within semantically compatible sceneries (e.g., Antes et al., 1981; Davenport & Potter, 2004). While facilitation of processing across scene-images has previously been observed in relation to the priming of visual features (Brady et al., 2017), we suggest that the benefit of congruency seen in Experiment 1a was due to the provision of semantically relevant context, and so more akin to the semantic priming of a scene when preceded by a relevant written word (Reinitz et al., 1989), or to work finding a disruption to gist processing when an observer views improbable sceneries (Greene et al., 2015). Therefore, these results demonstrate that top-down information – in the form of expectations generated prior to target scene appearance – was able to influence gist processing, a proposition at odds with models assuming minimal higher-order modulation of gist processing (e.g., Itti et al., 1998; Rumelhart, 1970).

Experiment 1b

The results from Experiment 1a demonstrated an advantage in categorisation performance for Congruous trials, apparent at target durations where the opportunity to process visual information was more limited. However, it was important to confirm that these findings were due to the congruency manipulation as opposed to unintended residual effects based on the experimental design. Specifically, 75% of trials were Congruous in Experiment 1a, and so higher performance on these trials was feasibly based on their increased frequency compared to Incongruous trials. We addressed this possibility in Experiment 1b, by switching the relative presentation frequencies of the congruency conditions.

The reduction in the number of Congruous trials in Experiment 1b also served a further purpose: in a task where most trials are incongruous it is not beneficial for participants to take account of the contexts provided by the approach images, as these are more often than not unrelated to the destination. If a pattern of results similar to those from Experiment 1a emerged, therefore, in terms of higher performance for Congruous compared to Incongruous trials, this would suggest that predictions as to an upcoming scene category were being generated automatically.

Design

We again employed a 3:1 split across trial congruency, but reversed the ratio of Congruous to Incongruous trials compared to Experiment 1a, such that now 75% of trials had destinations incongruous to the approach images. The decision was also taken to limit Experiment 1b to three target-duration conditions. This was due to the preceding iteration (Experiment 1a) showing very similar levels of performance, in terms of both congruency conditions, across the 33 and 50 ms target durations. There was also a noticeable amount of variation in performance across participants at 33 ms, with some failing to achieve scores above chance level. As a result, the 50 ms condition provided the most reliable reflection of general performance under circumstances of limited availability of visual stimulation.

Although the selection of Incongruous trials in Experiment 1a had been achieved by random assignment, it was prudent to ensure this had not led to any bias through unintentional systematic differences across the two congruency conditions. As such, Experiment 1b introduced a Latin Square design. Four separate versions of the protocol were programmed, each with a different set of 30 Congruous trials (one from each scene category). This meant that, over the course of the experiment as a whole, all series were presented in both congruous and incongruous fashion, with the specific makeup of conditions determined by which version a participant sat. Versions were cycled through for each new participant, separated by target-duration condition.

Participants and Stimuli

Our second experiment included 90 participants (M_{age} = 20.89, SD_{age} = 4.98; 68 women, 22 men; 81 right-handed, 9 left-handed). Experiment 1b used the same image set and masks as Experiment 1a. The response screens were redrawn, using the same randomisation procedures as the first experiment.

Procedure

The procedure for Experiment 1b mirrored that of 1a, with one alteration. A handful of participants had asked for clarification of certain category words during the previous iteration, most notably ‘Quay’. To eliminate this issue, prior to beginning Experiment 1b participants were shown a list of the 30 scene categories and were provided with explanations by the researcher where needed. Participants were assured that the list did not need to be memorised.

Results and Discussion

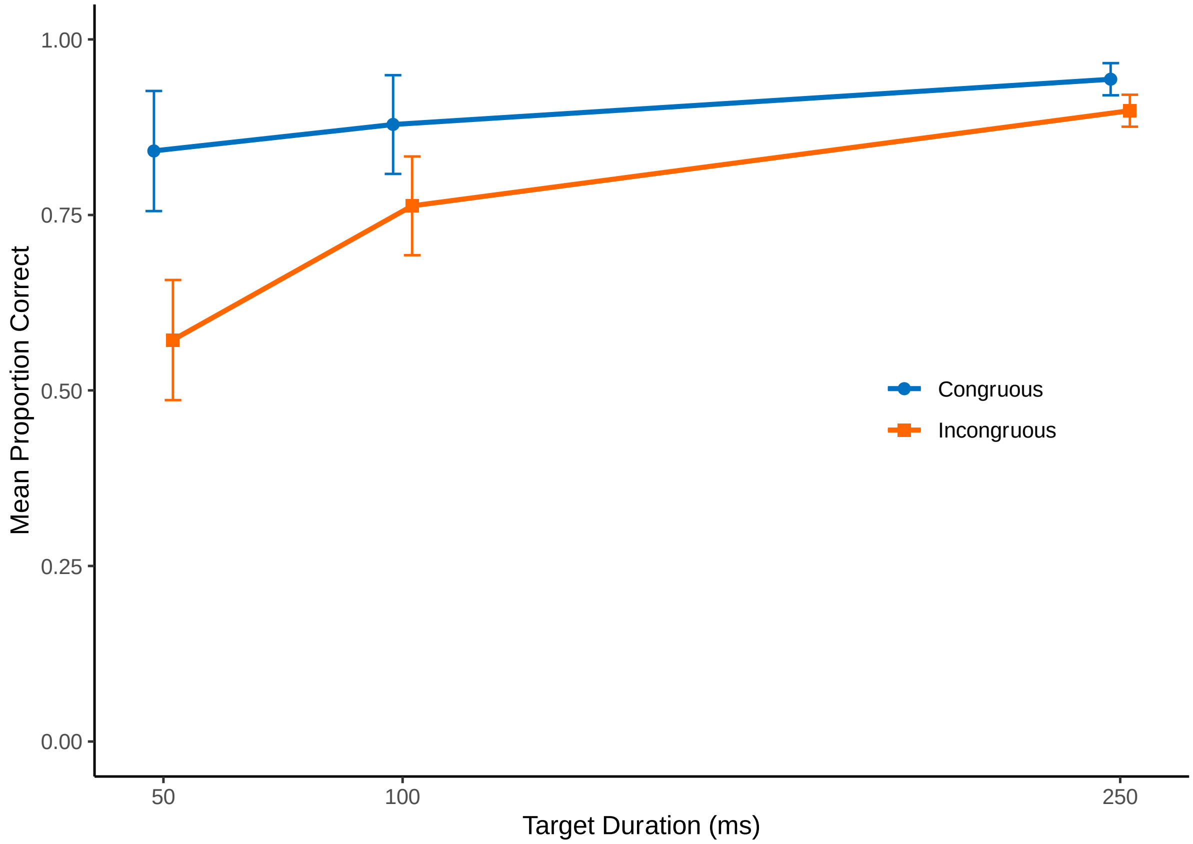
As in Experiment 1a, when the opportunity to extract visual information was limited due to a brief target duration (50 ms), performance on Incongruous trials was some distance below that on Congruous trials (see Figure 3). As target duration increased, the disparity across congruency conditions narrowed (100 and 250 ms).

The analysis strategy for Experiment 1b was analogous to that for Experiment 1a. The three target durations were represented by two sliding contrasts describing the differences in mean categorisation accuracy between conditions 2 and 1 (100 ms – 50 ms) and between conditions 3 and 2 (250 ms – 100 ms). As in Experiment 1a, we manipulated approach-destination congruency within subjects; however, here it was also manipulated within target scenes. Inclusion of both random slopes for Congruency improved the model fit. The GLMM results are summarized in Table 2.

There was a significant main effect of Congruency on categorisation accuracy, with lower performance on Incongruous than Congruous trials, *b* = -1.21, *SE* = 0.21, *z* = -5.74, *p* < .001. Categorisation accuracy was significantly higher at a target duration of 100 ms compared with 50 ms, condition 2 – condition 1: *b* = 0.93, *SE* = 0.27, *z* = 3.41, *p* = .001. Moreover, categorisation accuracy was significantly higher for the 250 ms than for the 100 ms condition, condition 3 – condition 2: *b* = 1.11, *SE* = 0.28, *z* = 3.91, *p* < .001. Regarding the interaction between Congruency and Target duration, the congruency effect was significantly smaller in the 100 ms than in the 50 ms condition, Congruency × Target duration 2 – 1: *b* = 1.06, *SE* = 0.46, *z* = 2.32, *p* = .021. However, the congruency effect at a target duration of 250 ms was not significantly different to that at 100 ms, Congruency × Target duration 3 – 2: *b* = -0.05, *SE* = 0.48, *z* = -0.10, *p* = .920.

Figure 3

Experiment 1b. Scene Categorisation Accuracy for Congruous and Incongruous Trials as a Function of Target Duration



Note. Error bars represent 95% CIs for mixed designs (Cousineau et al., 2021).

Table 2

Generalized linear mixed model fitting the effects of Congruency and Target duration on scene categorisation accuracy in Experiment 1b: estimates of coefficients (*b*), standard errors (*SE*), *z* values, and *p* values for fixed effects; variances, standard deviations (*SD*), and correlations for random effects

Fixed Effects	*b*	*SE*	*z*	*p*
Intercept	2.1615	0.1391	15.544	< .001
Congruency	-1.2084	0.2104	-5.745	< .001
Target duration 2 – 1 (100 ms – 50 ms)	0.9305	0.2732	3.406	.001
Target duration 3 – 2 (250 ms – 100 ms)	1.1076	0.2834	3.908	< .001
Congruency × Target duration 2 – 1	1.0645	0.4595	2.317	.021
Congruency × Target duration 3 – 2	-0.0486	0.4835	-0.101	.92

Random Effects				
Groups	Name	Variance	*SD*	Correlation
Subject ID	Intercept	0.9982	0.9991	Intercept
	Congruency	2.6830	1.6380	0.58
Target scene ID	Intercept	0.7250	0.8514	Intercept
	Congruency	0.7034	0.8387	-0.13

As predicted, the results from Experiment 1b mirrored those from Experiment 1a. Again, categorisation ability was significantly higher when target scenes were preceded by congruous approach images, and this differential in performance was greater when target presentation duration was at a briefer duration (50 ms). Consequently, Experiment 1b confirmed our findings were due to the congruency manipulation, as opposed to simply being based on the presentation frequency of experimental trials. In addition, these results show that context-based predictions were being generated automatically by participants as they viewed the approach images, in line with work demonstrating that pre-target natural scene images lead to the automatic generation of expectations as to the identity of an upcoming target object (Caplette et al., 2020).

Taken together, the findings from across these two experiments revealed that approach images influenced subsequent scene processing, and so suggest a role for top-down information in rapid gist processing. They do not, therefore, support proposals that suggest the extraction of scene-gist is exclusively based on feedforward processes.

Experiment 2

While Experiments 1a-b found an influence of trial congruity, the specific mechanisms still need to be determined. On one hand, the presentation order of approach images may have comparatively little bearing on performance, whereby the collective group of images simply provide a semantic context which increases the predictability of the eventual destination scene. For instance, observing an approach image which depicts surroundings commonly associated with the countryside may be sufficient for expectations to be formed as to the most likely eventual destination (e.g., a field, woods, etc.). In this scenario, there would be no cost to performance if approach images were not arranged in a meaningful sequence, as participants would still be provided with the same contextual information prior to target presentation. On the other hand, there may be an additional benefit, beyond the collective semantic context, from the spatiotemporally progressive nature of the series. If this were true, then we would expect to see lower performance on trials where the order of approach images is disrupted.

Previous research has highlighted the importance of narrative coherence for efficient processing. For instance, Cohn et al. (2012) disrupted the order and content of comic strips, finding that both semantic relatedness and narrative structure were advantageous, whereby the processing of a subsequent image was influenced by both the structure and meaning of the series that preceded it. In terms of scene processing, Smith and Loschky (2019) manipulated series of 10 familiar scene images to either be sequentially or randomly organised, finding that sequentially-organised trials yielded higher recognition performance. Coherent scene sequences could allow for the generation of a ‘perceived flow’ of movement through an environment, potentially facilitating processing by allowing for the extraction of more information, such as that derived from the semblance of optical flow (Gibson, 1966) or through aiding the transformation of the viewer-centred 2½D sketch into a three-dimensional representation (Marr, 1982). Additionally, the further away in space an approach image is from its eventual destination, the potentially weaker its predictive power. As an observer progresses through a series, each new leading image may further ‘fine-tune’ expectations, which could be a more additive process compared to that occurring from experiencing the same images in random order.

To investigate whether the sequence of the approach series plays a role in gist processing, in Experiment 2 we manipulated the presentation order of approach images while also continuing to manipulate their congruency to destination scenes. First, we predicted that performance on

Congruous trials would be better than Incongruous trials, regardless of sequentiality, as the approach images on Congruous trials provide semantically relevant context. Additionally, we predicted a categorisation advantage for sequentially coherent trials due to both the assumption that sequentiality would create a flow of information that more closely mirrored typical functioning in everyday environments, as well as previous research identifying an important role of narrative sequences for processing (e.g., Cohn et al., 2012; Smith & Loschky, 2019).

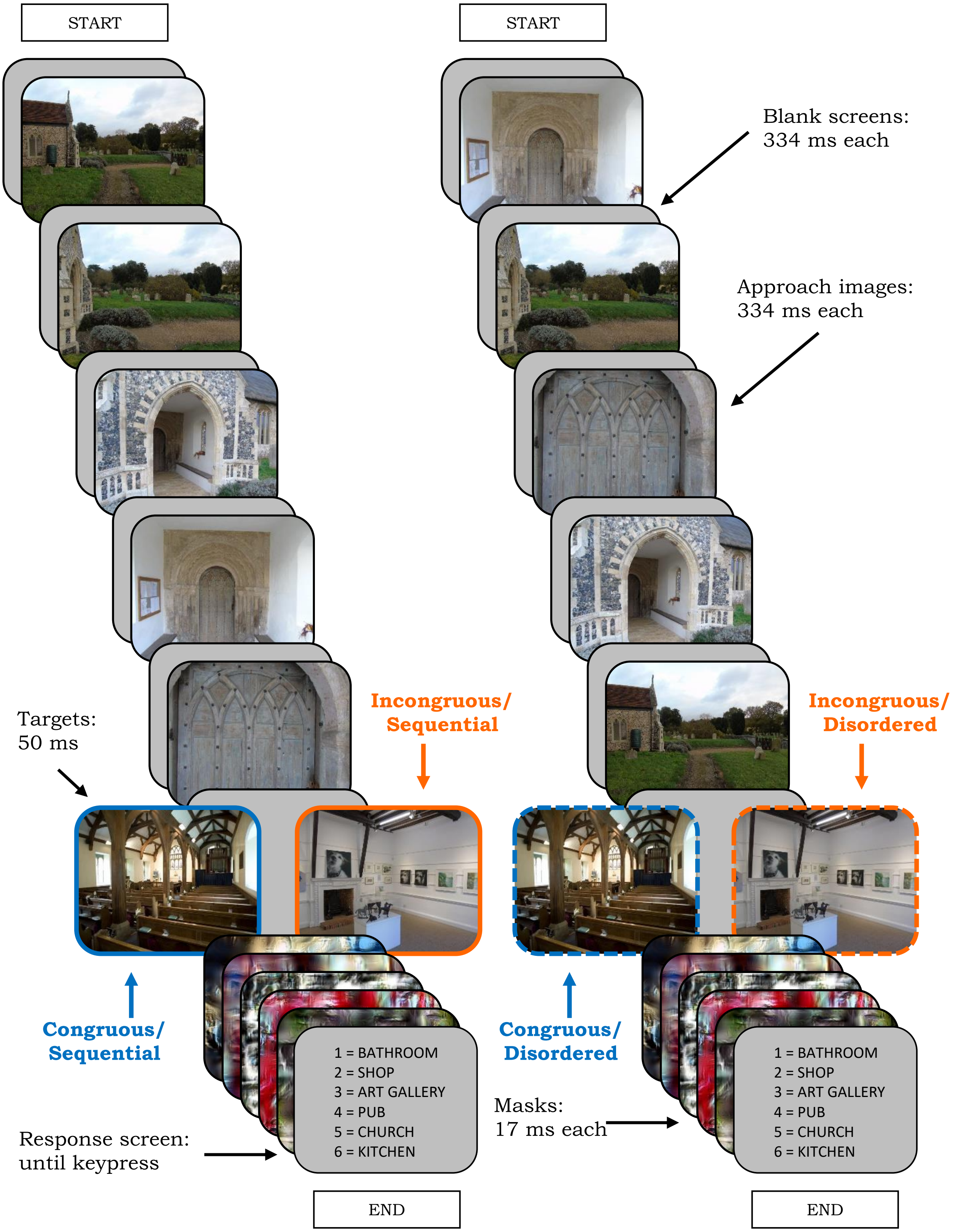
Design

While maintaining the approach-destination congruency manipulation of Experiments 1a-b, Experiment 2 departed from the previous iterations by also manipulating the sequentiality of approach images. Therefore, trials included approach images displayed either in a sequential or randomised order (disordered). This led to four within-participant conditions: Congruous-Sequential; Congruous-Disordered; Incongruous-Sequential; and Incongruous-Disordered. Each condition consisted of 30 trials and included one series for each of the scene categories. A Latin Square design was employed, so that each series alternated across all conditions within the four versions of the experiment. The presentation order of approach images within Disordered trials was randomly selected, but with two important constraints. First, the approach image in the closest geographical location to the destination scene could not be the final pre-target image. This parameter was to ensure that congruous targets were not simply being primed by the nearest geographical approach image. Secondly, such trials could not contain more than two approach images displayed in their original order. This was to safeguard the non-sequentiality of Disordered trials.

The presentation order of trials was randomised independently for each participant. Target duration was not manipulated in Experiment 2, with targets being presented for 50 ms. This was due to our previous experiments demonstrating that the effect of congruity was most apparent at brief target durations, diminishing as presentation length increased. See Figure 4 for a schematic of the experimental protocol.

Figure 4

Schematic of the Protocol for Experiment 2



Participants and Stimuli

Seventy-two participants took part in Experiment 2 (M_{age} = 23.58, SD_{age} = 11.22; 54 women, 18 men; 61 right-handed, 10 left-handed, 1 ambidextrous). Experiment 2 used the same image set and masks as Experiment 1a and 1b. The response screens were again redrawn, using the same randomisation procedures as the preceding experiments.

Procedure

The procedure followed the same routine as Experiment 1b, except that the display duration of blank screens was increased from 167 ms to 334 ms (10 to 20 frames on a 60Hz monitor). This was judged to provide a more comfortable viewing experience for participants, which better mimicked the sense of traversing an environment.

Results and Discussion

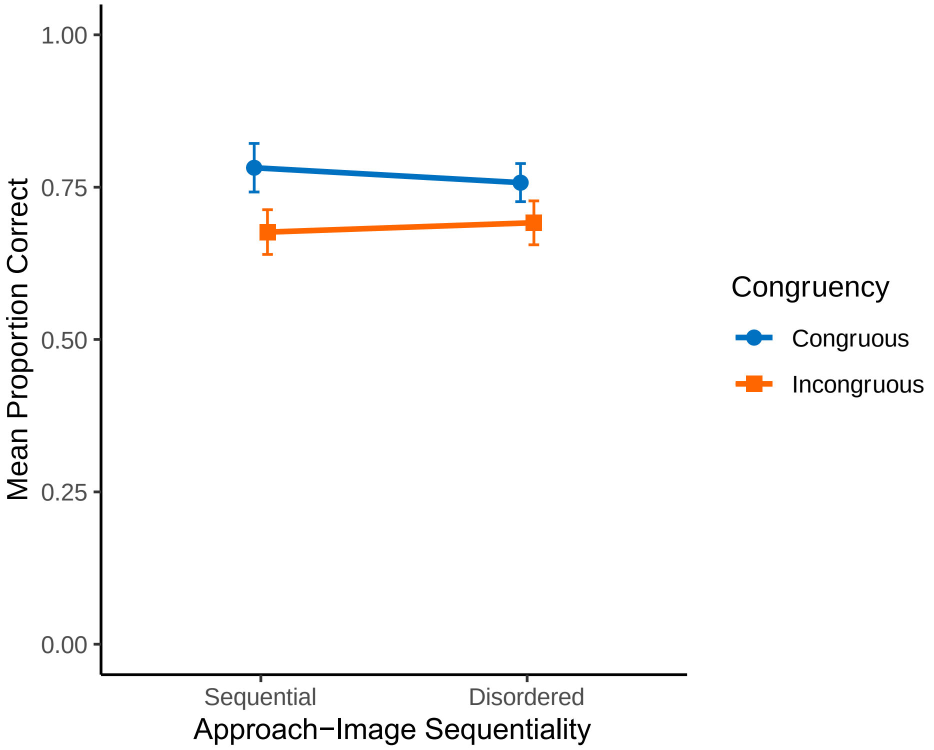
To specify the contrasts in the mixed-effects analysis, simple coding (–0.5 / +0.5) was used for the two-level factors Congruency (reference level: Congruent) and Sequentiality (reference level: Sequential). Because simple coding yields centred contrasts, the model intercept reflects the grand mean of the dependent variable. A summary of the GLMM results is provided in Table 3. There was a significant main effect of approach-destination Congruency on scene categorisation accuracy, with lower performance on Incongruous than Congruous trials, b = -0.43, SE = 0.19, z = -2.25, p = .024. There was no significant main effect of approach-image Sequentiality, b = -0.02, SE = 0.06, z = -0.34, p = .737. There was, however, a significant Congruency × Sequentiality interaction, b = 0.3, SE = 0.11, z = 2.64, p = .008. This indicated that the sequentiality of approach images had different effects on categorisation performance depending on whether the approach images were congruous with the target image (see Figure 5).

To investigate this interaction further, we ran two post hoc GLMMs using dummy coding for the factors Sequentiality (reference level: Sequential) and Congruency (reference level: Congruous in model 1, Incongruous in model 2). In model 1, the simple effect of Sequentiality represents the Sequentiality effect for Congruous trials. This analysis yielded a significant effect of Sequentiality, *b* = -0.17, *SE* = 0.08, *z* = -2.13, *p* = .033, which means that categorisation performance was reduced when approach images were presented in random order. In model 2, where the simple effect of Sequentiality represented the Sequentiality effect for Incongruous trials, the critical effect was not significant, *b* = 0.13, *SE* = 0.08, *z* = 1.61, *p* = .107.

As with the Experiments 1a-b, we again found a benefit to categorisation performance related to trial congruency. Sequentiality of the approach images, alone, did not affect performance. However, sequentiality did affect performance in the Congruous condition, replicating an effect of sequentiality in Smith & Loschky (2019) whose scenes were exclusively congruous. We failed to find evidence that sequentiality affects performance in the Incongruous condition. Sequentiality of the approach images therefore appears to build on their overall semantic relation to the target: when the semantic context provided by these images helps predict an upcoming target, having them in an appropriate order seems to further refine gist processing (possibly by narrowing the number of reasonable options that could appear next). Therefore, sequentially-arranged scene images seem to benefit gist processing when they are already semantically related to the target scene. However, we do not find evidence that simply observing a set of scenes in sequence facilitates scene categorisation.

Figure 5

Experiment 2. Scene Categorisation Accuracy for Sequential and Disordered Trials as a Function of Approach-Destination Congruency



Note. Error bars represent 95% within-subjects CIs, which were calculated using the Cousineau-Morey method (Cousineau, 2005; Morey, 2008).

Table 3

Generalized linear mixed model fitting the effects of Congruency and Sequentiality on scene categorisation accuracy in Experiment 2: estimates of coefficients (*b*), standard errors (*SE*), *z* values, and *p* values for fixed effects; variances, standard deviations, and correlations for random effects

Fixed Effects	*b*	*SE*	*z*	*p*
Intercept	1.3487	0.1579	8.539	< .001
Congruency	-0.4323	0.1919	-2.252	.024
Sequentiality	-0.0193	0.0575	-0.335	.737
Congruency × Sequentiality	0.3039	0.1149	2.644	.008
Random Effects				
Groups	Name	Variance	*SD*	Correlation
Subject ID	Intercept	1.2476	1.1170	Intercept
	Congruency	2.0749	1.4405	0.56
Target scene ID	Intercept	0.7861	0.8867	Intercept
	Congruency	0.4745	0.6888	0.18

Experiment 3

The findings from Experiment 2 suggested the influence of approach images on subsequent scene processing was due to participants being provided a semantic context, in sequential order, prior to target onset. However, up to this point the assumption had been made that this effect was driven by congruous approaches facilitating the subsequent processing of destinations. A possible alternative explanation was that the difference across experimental conditions was the result of incongruous approaches interfering with the processing of destinations.

To answer this question, in Experiment 3 we introduced a third, neutral condition whereby approach images were replaced by images of coloured patterns. As such, provision of semantic context was absent within the trials of this condition, meaning participants were unable to generate expectations as to the identity of the upcoming destination. By comparing categorisation performance in the neutral (No-context) condition to both the Congruous and Incongruous conditions, we hoped to uncover more fully the role of the congruency manipulation on gist processing. We expected better performance on Congruous trials, compared to No-context and Incongruous trials; due to a lack of direct evidence from previous research, we made no predictions as to whether performance on Incongruous trials would be significantly lower than that of the No-context condition.

Design

The design was very similar to that of Experiments 1a-b, with two changes. First, we introduced a third congruency condition in which the approach images provided no semantic context to participants prior to destination-onset. This led to three within-participant conditions: Congruous; Incongruous; and No-context. Each condition consisted of 40 trials, including at least one, and no more than two, series for each of the scene categories. Secondly, target duration was set at 50 ms for all participants. See Figure 6 for a schematic of the experimental protocol.

Participants

Participants were recruited through Prolific (https://www.prolific.co). Demographic screeners were used to ensure all participants were adults who lived in the UK, US, Canada, Australia or New Zealand, and were fluent speakers of English. This filtering was to ensure that all participants would both be able to fully understand the task instructions and would be familiar with the types of sceneries used in the experiment. Forty-five participants took part in Experiment 3 (M_{age} = 33.53, SD_{age} = 11.62; 30 women, 15 men; 37 right-handed, 7 left-handed, 1 ambidextrous).

Stimuli

Experiment 3 used the same image set and masks as the previous experiments, although in this iteration some of the mask-images were repurposed to act as leading images in the No-context condition (as set out below). The response screens were again reconfigured, using the same randomisation procedures as before.

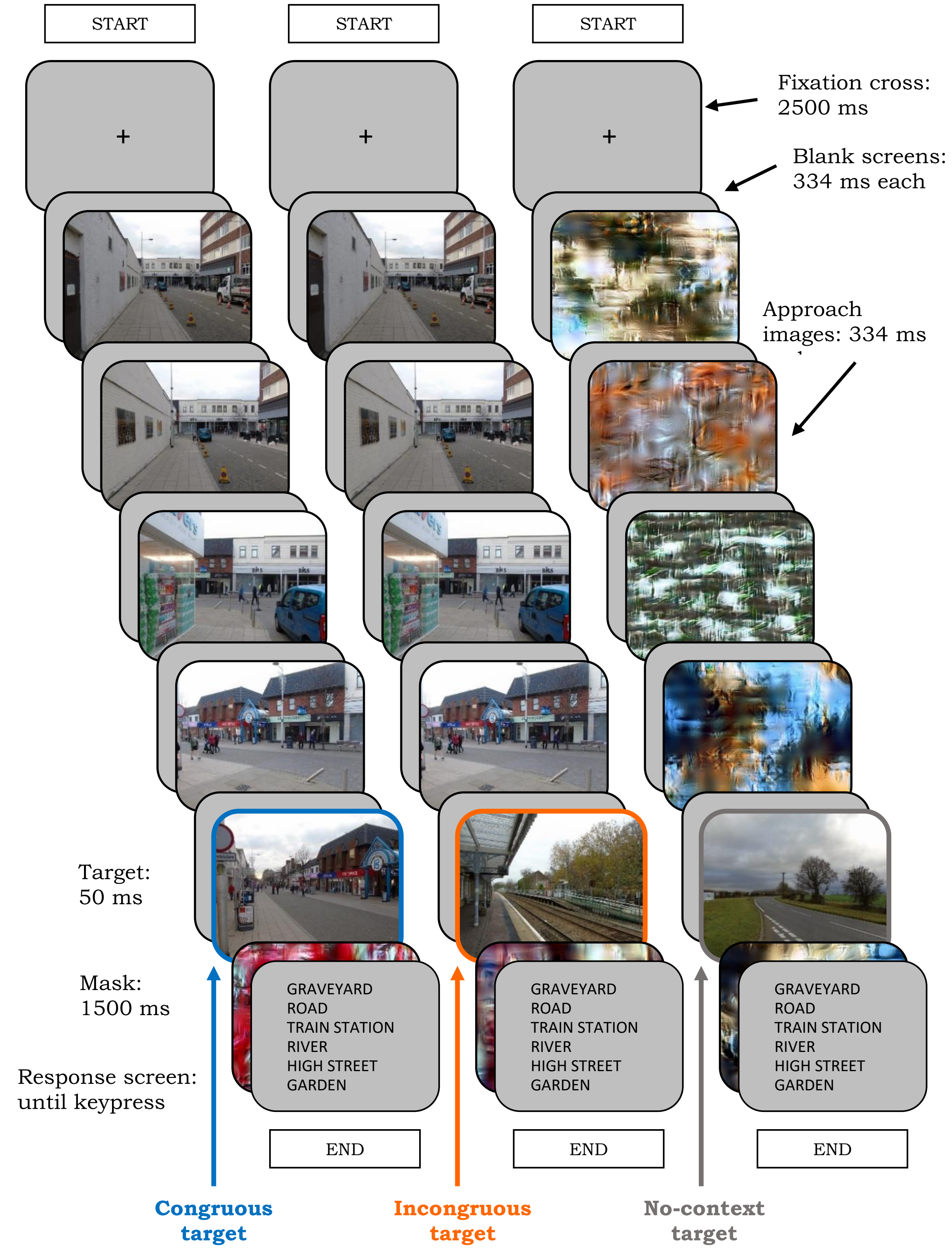
Procedure

The procedure followed a similar routine as Experiment 2, with some minor alterations necessary for the experiment to be run online. The experiment was programmed using Testable (www.Testable.org) and, due to constraints imposed by the software, the number of images displayed per trial needed to be reduced. This was achieved in two ways. Firstly, only four approach images were presented per trial, as we removed the first approach image from each series (i.e., the image most geographically distant from the destination). Secondly, destination images were followed by a single mask rather than a set of five dynamic masks. The duration of these individual masks was extended to 1500 ms to ensure a suitable disruption to processing from target-offset was maintained. The previous experiments each used 600 mask-images, and so 120 of these were randomly selected to again be used as masks in Experiment 3. A further 160 were then randomly selected to serve as leading images in the No-context condition. The order of presentation of these images, both within series and across trials, was also randomised. These randomisation procedures were followed for each of the three Latin Square versions, with the proviso that a mask-image could not be used as a leading image and a mask within a single version, and that all 600 mask-images were used across the experiment as a whole.

Two further minor alterations were included to ensure the smooth running of the experiment, due to the inevitable reduction in researcher oversight during an online study. Firstly, a

Figure 6

Schematic of the Protocol for Experiment 3



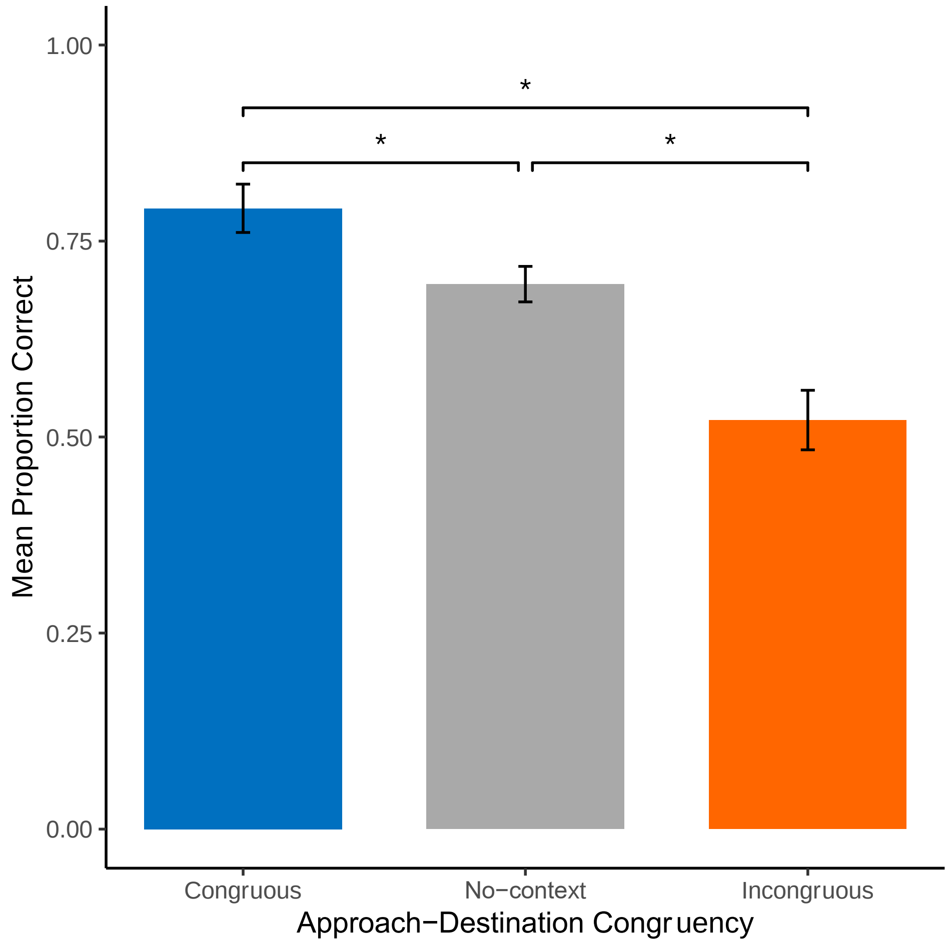
fixation cross was displayed in the centre of the screen prior to the start of each series. Secondly, selection of a response was made by navigating a cursor to the chosen textbox, rather than by pressing a number on a keypad.

Results and Discussion

As can be seen in Figure 7, the proportion of correct responses was greatest in the Congruous condition (*M* = 0.79, *SE* = 0.02), followed by the No-context condition (*M* = 0.70, *SE* = 0.01), and with weakest performance in the Incongruous condition (*M* = 0.52, *SE* = 0.02). For the mixed-effects analysis, backward difference coding was used for the 3-level factor approach-destination congruency (ordering: Congruous, No-context, Incongruous). The GLMM with the maximal random-effects structure justified by the design (Barr et al., 2013) provided the best fit to the data (Table 4). Scene categorisation accuracy was significantly lower in the No-Context condition than in the Congruous condition, *b* = -0.5, *SE* = 0.14, *z* = -3.45, *p* = .001. Moreover, categorisation accuracy was significantly reduced in the Incongruous condition compared with the No-context condition, *b* = -1.11, *SE* = 0.15, *z* = -7.57, *p* < .001. A post hoc GLMM using dummy coding for the factor approach-destination congruency (reference level: Congruous) confirmed that accuracy was significantly lower in the Incongruous than in the Congruous condition, *b* = -1.61, *SE* = 0.18, *z* = -9.18, *p* < .001).

Figure 7

Experiment 3. Scene Categorisation Accuracy for Congruous, Incongruous and No-context Trials



Note. Error bars represent 95% within-subjects CIs, which were calculated using the Cousineau-Morey method (Cousineau, 2005; Morey, 2008). * denotes *p* < .05.

Table 4

Generalized linear mixed model fitting the effect of Congruency on scene categorisation accuracy in Experiment 3: estimates of coefficients (*b*), standard errors (*SE*), *z* values, and *p* values for fixed effects and variances and correlations for random effects

Fixed Effects	*b*	*SE*	*z*	*p*
Intercept	1.0081	0.1573	6.409	< .001
No-context – Congruous	-0.5002	0.145	-3.449	.001
Incongruous – No-context	-1.1074	0.1463	-7.57	< .001
Random Effects				
Groups	Name	Variance	Correlation	
Subject ID	Intercept	0.5371	Intercept	
	No-context – Congruous	0.0836	0.82	No-context – Congr.
	Incongruous – No-context	0.3998	0.45	0.39
Target scene ID	Intercept	1.3662	Intercept	

	No-context – Congruous	1.2419	0.43	No-context – Congr.
	Incongruous − No-context	0.6034	-0.48	-0.45

A clear pattern of results emerged in Experiment 3, with significantly contrasting levels of categorisation performance seen across each of the three conditions. As with our previous experiments, categorisation performance was greater in Congruous rather than Incongruous trials. Further to this, performance was also higher on Congruous trials compared to No-context trials, thus supporting our contention that semantic congruity facilitates gist processing.

In addition, we found that performance in the Incongruous conditions was significantly below that of the No-context condition, suggesting that the ability to categorise a scene is inhibited if expecting a different scene category. This appears to agree with previous research showing scenes containing unexpected features are more difficult to extract meaning from (Greene et al., 2015), as well as work demonstrating that object-scene contextual violations interfere with object recognition (Biederman et al., 1982; Davenport & Potter, 2004; Lauer et al., 2020; Palmer, 1975). In sum, this pattern of results clearly shows that approach images were eliciting expectations as to the likely identity of an upcoming target scene, resulting in a benefit to gist processing if expectations were realised but, alternatively, resulting in a cost to processing if violated.

Experiment 4

The findings across the behavioural experiments consistently revealed higher categorisation scores on Congruous trials as compared to Incongruous trials. Furthermore, Experiment 3 showed that this equated to improved performance compared to baseline when the approach images were congruous with the target scene, and inhibited performance when these were incongruous. However, the possibility remained that a participant may achieve such a pattern of results even if they paid no attention to the target scene, and simply always chose the option on the response screen that corresponded to the category of the approach images. In other words, it could be argued that higher scores on Congruous compared to Incongruous trials were not due to participants being sensitive to the visual signal associated with the target scene, and instead were the result of a response bias towards selecting the approach category in situations of uncertainty (Hollingworth & Henderson, 1998; but see Auckland et al., 2007).

There are several reasons to suggest that this is not the case. First, the pattern of results across Experiments 1a-b remained strikingly similar despite the ratio of Congruous to Incongruous trials varying significantly between these two experiments (3:1 and 1:3 target to distractor ratios, respectively). If the results were due to response strategies, then there should have been a marked difference in the findings across these experiments. Secondly, the inclusion of a No-context condition in Experiment 3 demonstrates that participants' ability to categorise a briefly presented target scene was substantial, with a mean performance score of around 70% when scenes were displayed for 50 ms and no context was provided. Since chance level performance would be near 17%, this shows a particularly high level of sensitivity to the target images.

Furthermore, while it is possible that a response strategy bias could lead to higher scores in the Congruous condition as compared to No-context, as it would substantially increase the number of correct responses to Congruous trials that the participant was not sensitive to, this is not supported by the pattern of findings in Experiment 3. Instead, we found the performance gain for Congruous trials compared to baseline as relatively modest (~9%) when compared to the performance loss for Incongruous trials compared to baseline (~18%). This cannot simply be attributed to ceiling effects adding a downward pressure to the Congruous condition, as the mean score for this condition did not rise above 80% correct.

However, to better quantify the separate contribution of participants' sensitivity towards target scenes and their response strategies to the behavioural results, we ran a further experiment that was more suited to the investigation of response bias. This experiment retained the same manipulation of trial congruency as previous iterations, but with response scenes now including only two options. Unlike the previous 6-AFC design, a 2-choice discrimination task is suitable for investigation through signal detection measures, thus allowing for participants' response strategies and sensitivity to target scene images to be quantified. While we would predict some bias towards responding that any given trial was congruous in nature, we would expect this to be relatively small and, conversely, that participants would score highly on their ability to discriminate between target scene images.

Design

Experiment 4 maintained the approach-destination congruency manipulation of previous experiments but did not include the manipulation of approach sequentiality seen in Experiment 2, nor the baseline condition introduced in Experiment 3. There were two within-participant conditions: Congruous; Incongruous. Each condition consisted of 60 trials, including two series for each of the scene categories. A Latin Square design was employed, so that each series alternated across each condition within the two versions of the experiment. For each version, the destination images for those series selected for Incongruous trials were randomly reallocated amongst each other, in line with the principles of previous iterations. Target duration was set at 50 ms for all participants. Unlike previous experiments, response screens only included two category options.

Participants

Participants were recruited through Prolific and received payment for taking part. All participants were from the UK. Thirty-eight participants took part (M_{age} = 27.89, SD_{age} = 9.09; 20 women, 17 men, 1 Other; 35 right-handed, 3 left-handed).

Procedure

The procedure followed a similar routine as the original iterations, while incorporating the same necessary alterations to make it suitable to be run remotely as seen in our previous online experiment (Experiment 3). The experiment was again programmed using Testable (www.Testable.org), so only four approach images were presented per trial, with the destination image followed by a single mask. We randomly selected 240 mask images from our pool of 600, and these were randomly distributed across the two Latin Square versions of the experiment. Finally, as with Experiment 3, we included a fixation cross at the centre of the screen at the start of each series, and responses were made by moving a cursor and clicking on the chosen textbox.

Results and Discussion

Three participants were removed as their mean accuracy across both conditions was at chance level (<51%), indicating that they did not engage with the task properly. The remaining 35 participants were included in the analysis (M_{age} = 27.17, SD_{age} = 8.67; 19 women, 15 men, 1 other; 32 right-handed, 3 left-handed). As expected, higher categorisation ability was demonstrated on Congruous trials (*M* = 0.93, *SE* = 0.03) compared to Incongruous trials (*M* = 0.63, *SE* = 0.03). For the mixed-effects modelling, simple coding was used for the factor Congruency (see Experiment 2). The model with by-subject and by-item slopes for Congruency along with random intercepts for subjects and items provided the best fit to the data. Categorisation accuracy was significantly lower for Incongruous compared with Congruous trials, *b* = -2.39, *SE* = 0.30, *z* = -8.05, *p* < .001.

Next, we invoked signal detection theory (SDT) to estimate sensitivity and response bias within a mixed-effects modelling approach (Wright et al., 2009). The goal of a mixed-model signal-detection analysis is to predict what the participant responds from the true nature of the trial. For a given trial, the outcome variable captured whether the participant made a Congruous response. The outcome variable was coded with 1, if the participant responded that the target scene came from the same category as the approach scenes. If the participant responded that the target scene came from a different category than the approach scenes, the outcome variable was coded with 0. The predicting variable Congruency represented the true relationship between the approach images and the destination image (1 congruous, 0 incongruous).

The data were analysed with a binomial GLMM with a probit link function. Using the probit (i.e., inverse normal) link allows for interpretations in line with the traditional SDT model based on the normal distribution (DeCarlo, 1998). Simple coding was used to centre the Congruency predictor. With this contrast coding, the model intercept represents the criterion location (*c*) as a measure of response bias. Specifically, the fixed effect for the intercept reflects the overall probability of making a Congruous response (i.e., across both congruency conditions). Given that the parameters estimates are obtained in probit space, an intercept of 0 corresponds to a probability of .5 and therefore indicates that there is no response bias. The fixed effect of Congruency represents the mean difference between Congruous and Incongruous trials. This difference can be viewed as the difference between hits (correctly responding 'congruous' on Congruous trials) and false alarms (incorrectly responding 'congruous' on Incongruous trials), which is why the main effect of Congruency provides an estimate of sensitivity (*d*'). Higher values of *d*' represent greater sensitivity, with *d*' = 0 representing chance performance.

The maximal GLMM with by-subject and by-item slopes for Congruency provided the best fit to the data. The model intercept was significantly larger than zero (*b* = 0.65, *SE* = 0.08, *z* = 7.93, *p* < .001), which means that there was a general bias towards choosing the response option that corresponded to the approach images. Importantly, the fixed effect estimate for Congruency was also significantly larger than zero (*b* = 2.27, *SE* = 0.20, *z* = 11.28, *p* < .001), indicating high sensitivity towards the target scene images. Taken together, the results suggest that participants were able to effectively discriminate target scene categories while showing a small-to-medium liberal response bias.

Experiment 5

The behavioural experiments revealed that providing semantically informative scenes, preferentially sequenced as if walking through the environment, leads to increased performance on subsequent scene gist processing. Building on these results, we turned to an investigation of the neural signature. In Experiment 5 we investigated the event-related potential (ERP) correlates of scene processing and the role of expectations on gist extraction. This investigation was exploratory in nature as, to date, we are unaware of any prior use of this methodology for the examination of the role of sequential, naturalistic leading images on subsequent scene-gist processing. However, previous research suggested which ERP components were most likely to be correlated with the effect of congruency on scene processing. Specifically, we focused on both an early component commonly associated with the processing of a scene's global visual features (P2), as well as later components associated with higher order processing, such as semantic and syntactic integration (N400 and P600, respectively). Determining the pattern of amplitude changes across these separate components can help us better understand whether the approach-destination congruency manipulation was influencing feature extraction mechanisms, semantic processing, syntactic processing, or indeed all of these.

The P2 component arises rapidly within Parieto-occipital regions, around 200 ms after target onset; this component has been proposed as the earliest known marker for scene-specific processing (Harel et al., 2016), and is affected by changes in global scene properties but not top-down observer-based goals (Hansen et al., 2018). However, the exact influence of top-down information on the P2 remains unclear. While evidence indicates early components such as this are sensitive to low-level visual information such as salience (Straube & Fahle, 2010), as well as object identification (Viggiano & Kutas, 2000), the influence of higher-level processes is less well determined. For example, differences in ERPs at around 200 ms have been found when identifying the presence of objects within briefly presented natural images, potentially reflecting decision-related activation (Thorpe et al., 1996; VanRullen & Thorpe, 2001). As a result, a lack of agreement exists, both in terms of whether the P2 is altered by top-down processing at all and, if so, what form of top-down processing might hold influence. Furthermore, it has been suggested the P2 may index an intermediary processing stage, somewhat bridging perceptual and higher-order processes, such as segmentation and categorisation, respectively (De Cesarei et al., 2013). In terms of the current study, the above implies predictions relating to the P2 must be tentative. We can contend, however, that if congruency-based differences in activation were shown to exist within the earliest indicant of scene-specific processing (Harel et al., 2016), this would be representative of expectations influencing early perceptual processing. More broadly, finding activation differences would signify that the P2 component is open to influence from top-down information.

An ERP analysis can also help elucidate the mechanisms underlying expectation-related performance changes as cognitive processing continues. Previous scene processing research has shown two later components susceptible to experimental manipulation. The first of these – the N400 – has long been associated with semantic processing, with its amplitude observed as being inversely proportional to semantic expectancy (e.g., Kutas & Hillyard, 1984) and more generally to the ease with which conceptual information can be retrieved (Van Petten & Luka, 2006). For this component a certain level of consensus has been reached: across both central and anterior sites, increased negativity within the N400 time window has been related to scene-object semantic violations in static images (Ganis & Kutas, 2003; Mudrik et al., 2010; Võ & Wolfe, 2013) and within video clips (Sitnikova et al., 2008; Sitnikova et al., 2003). N400 effects have also been found to be sensitive to the semantic association between pairs of sequential pictures (Barrett & Rugg, 1990), and to violations of semantic expectation in language comprehension studies (Holcomb, 1993; Van Petten, 1995). A similar pattern of N400 changes across conditions within the current study would, therefore, indicate that differential behavioural performance derived from congruency-based manipulations in the behavioural experiments was due to semantic violations, rather than simply violations of expected low-level visual information.

The second of these later components is the P600. Like the N400, this component was initially described in language comprehension studies, where syntactic errors leading to sentence reanalysis are associated with increased posterior positivity at ~600 ms (e.g., Hagoort et al., 1993). Perhaps the most reproducible findings from this domain are through the use of 'garden path' sentences, whereby violation of the expected structure of a sentence creates the need for reanalysis of the preceding sequence of words (e.g., Frazier & Fodor, 1978; Osterhout & Holcomb, 1992). Within scene processing research, increased positivity at the P600 has been reported as reflecting reanalysis prompted by mis-located objects (Võ & Wolfe, 2013). There, increased late positivity was found when appropriate objects were positioned in inappropriate places within a scene (such as a dishtowel on a kitchen floor), irregularities proposed by the authors as reflecting syntactic – rather than semantic – violations. However, a lack of agreement should be noted regarding the functional role of the P600. It has also been suggested that this increased late positivity may not exclusively represent syntactic violations, as its sensitivity to semantic information has also been demonstrated (Gunter et al., 2000; Gunter et al., 1997; Kuperberg, 2007; Sitnikova et al., 2003). Furthermore, such changes to late positive components have not always been observed when objects break syntactic rules within scenes (e.g., Demiral et al., 2012). To confuse matters further, while Võ and Wolfe (2013) did not find alterations to the P600 when inappropriate objects were placed in appropriate locations – taken by the authors as evidence of the dissociation between the effects of semantic and syntactic violations – this form of semantic violation was shown to elicit a reduction in P600 amplitude in previous work (Mudrik et al., 2010). It is possible this inconsistency across studies is rooted in contrasting methodological choices, with one allowing for expectations to be generated due to the context-scene appearing prior to the target object (Võ & Wolfe, 2013), and the other avoiding this through simultaneous presentation of targets and their associated scenes (Mudrik et al., 2010).

Based on the above, we predict that there would be greater N400 amplitudes across central and anterior regions for Incongruous trials as compared to Congruous trials and increased P600 amplitude across posterior sites. Finally, due to debate remaining as to the influence of top-down factors on the P2 component, we do not make predictions as to whether Incongruous trials will elicit increased positivity in posterior regions during this time-window. However, if such changes were observed, we would take this as signalling the violation of expectations was able to influence the earliest stages of scene processing, including the integration of visual properties.

Design

To maintain consistency throughout the study the experimental protocol mirrored previous iterations closely, although with certain alterations necessary to improve the suitability of the trial routine for use with electroencephalography. The 120 experimental trials were split equally across two conditions of approach-destination congruity, and there was no manipulation of approach sequentiality. Sixty image-series were randomly selected to serve as Incongruous trials, with their destination images randomly redistributed amongst themselves.

Participants

Twenty-six participants took part in Experiment 5 (M_{age} = 20.31, SD_{age} = 2.77; 20 women, 6 men; 19 right-handed, 7 left-handed). All were Psychology students at the University of East Anglia, and received course credits for taking part. This is slightly below the number suggested from our initial power calculation (~30 participants) but is standard for ERP research (see, for example, Cohn & Foulsham, 2020; Harel et al., 2016; Mudrik et al., 2010; Võ & Wolfe, 2013). Furthermore, the relatively large congruency effects seen in the previous iterations allowed for confidence in the experiment remaining sufficiently powered despite this reduction. One participant was removed as their comprehension of the task could not be assured (incorrectly responding to 77% of Incongruous trials), and another removed due to excessive high-frequency noise across multiple channels. Analyses were conducted on the remaining 24 participants (M_{age} = 20.42, SD_{age} = 2.86; 18 women, 6 men; 18 right-handed, 6 left-handed). All participants reported to have no history of neurological disorders.

Stimuli

The same image set and response screens were used, although masks were removed.

Procedure

Each trial began with a ‘blink’ screen, followed by a blank screen including a jitter (duration: 2.5, 3, 3.5 or 4 seconds) to protect the ERPs from the potential systematic influence of slow baseline drifts coinciding with the routine. The jitter was pseudo-randomised to ensure a different blank screen duration prior to each of the four approach-series per scene category. A second, shorter jitter (duration: 350, 367, 383 or 400 ms) was also introduced to the last blank screen prior to target presentation, to shield against artefacts caused by participants being able to predict the exact onset time of the target. This jitter was pseudo-randomised in the same manner as before, and was evenly distributed across the two congruency conditions. There was no manipulation of target duration, with presentation length set at 1 second. This extended duration served two purposes: firstly, as only correctly answered trials were used in the analysis it sustained a high level of categorisation performance and, secondly, it protected against noise within the ERP caused by the offset of the stimulus or the onset of the response screen. No masking was used, with the target followed by a blank screen prior to a 6-AFC response screen.

Data Acquisition

The EEG was recorded using a Brain Vision 64-channel active electrode system, embedded within a nylon cap (10/20 system). Electrode FT9 was removed from the cap and placed under the left eye to monitor blinks and eye movements. The signal was acquired at a 1000 Hz sampling rate with FCz used as the online reference.

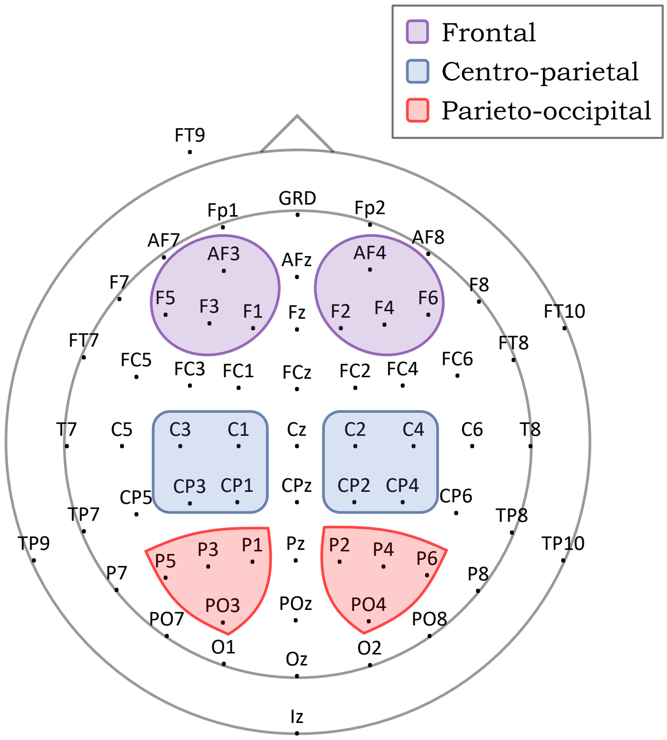
Processing

Offline processing and analyses were conducted using EEGLAB (Delorme & Makeig, 2004) and ERPLAB (Lopez-Calderon & Luck, 2014), running under Matlab 9.2.0 (R2017a, Mathworks, Natick, MA). Trials with incorrect responses were removed from the continuous EEG (3.89% of Congruous trials and 5.01% of Incongruous trials across participants). Ocular artefact correction took place through Independent Component Analysis (ICA) to identify blinks and lateral eye movements. These artefacts are located at anterior electrodes and can be identified based on their characteristic shapes (frequent clear spikes or step-like functions, respectively). Therefore, removal of these components was conducted manually by simultaneously comparing the continuous EEG to the time-course of the Independent Components. This led to removal of 41 Independent Components across the sample as a whole, with no more than two components removed for any single participant.

Re-referencing to the average of the TP9 and TP10 electrodes (which approximate to the location of the mastoids) was computed offline. Using such a reference is in line with recent work investigating the processing of sequences of images (e.g., Cohn & Foulsham, 2020; Cohn & Kutas, 2015, 2017), and re-referencing to the mastoids has been used to good effect in designs involving complex natural scenes (e.g., Demiral et al., 2012) and, in the literature more widely, for investigating the N400 (e.g., Henderson et al., 2011; Martín-Loeches et al., 2017), P600 (e.g., Frisch et al., 2002; Tanner et al., 2017) and posterior P2 components (e.g., Antal et al., 2000; Handy et al., 2001). An important benefit of using such a technique is that it helps avoid some of the inherent limitations of using an average across all sites (see, for example, Dien, 1998; Junghöfer et al., 1999; Yao et al., 2007), especially that analyses based on the average reference make comparison of waveforms and scalp distributions across studies difficult (Luck, 2014). Any channels suffering from persistent high-frequency noise were interpolated using the mean signal from the surrounding electrodes (mean percentage of channels interpolated across participants: < 1%). After removal of DC trends, an IIR Butterworth filter was applied for high- and low-pass filtering the data with half-amplitude cut off values of 0.01 Hz and 80 Hz, respectively (12

Figure 8

Map of Electrode Placement Including the ROIs



Note. FT9 was removed from the cap and placed on the left cheekbone to monitor blinks.

dB/oct; 40 dB/dec). The EEG was segmented into epochs of 1 second, from 200 ms before to 800 ms after target-scene onset. The length of the baseline used to correct epochs was the 200 ms immediately preceding target onset. Epochs contaminated with excessive artefacts were identified, and rejected, by setting a peak-to-peak voltage threshold of 100 µV across a moving window of 200 ms with a window step of 50 ms. This resulted in the rejection of 6.94% of Congruous trials and 7.10% of Incongruous trials across participants.

The amplitudes of the P2, N400 and P600 were measured as the mean of all data points between 175-250 ms, 300-500 ms and 500-700 ms, respectively. These specific components were chosen as the P2 has previously been suggested as the earliest indicator of scene selectivity (Harel et al., 2016), while the N400 and P600 have been associated with semantic and syntactic integration, respectively (e.g., Friederici et al., 1993; Hagoort & Brown, 2000; Holcomb, 1993; Mudrik et al., 2010; Van Petten, 1995; Võ & Wolfe, 2013).

The time windows chosen are commonly used as boundaries for investigating the N400 (e.g., Ganis & Kutas, 2003; Guillaume et al., 2018; Mudrik et al., 2010) and P600 (Angrilli et al., 2002; Cohn et al., 2014; De Vincenzi et al., 2003). Less standardisation exists regarding the P2, however, with previous research involving the processing of scenes employing time windows ranging anywhere between 140 to 320 ms post-stimulus onset (see, for example, De Cesarei et al., 2013; Ferrari et al., 2017; Harel et al., 2020; Yuan et al., 2007). We, therefore, determined our window of interest based on visual inspection of the grand average ERP. As a result, a window of 175-250 ms was selected as it covered the 220 ms timepoint previously identified as showing maximal amplitude for scene processing (Harel et al., 2016), while offering as large a span as was achievable without incorporating elements of the proximal P1 and P3 components.

Key electrode sites were grouped into three regions of interest (ROIs), each incorporating eight electrodes (split equally across hemispheres). A Centro-parietal ROI included electrodes C1/C2, C3/C4, CP1/CP2 and CP3/CP4, a Parieto-occipital ROI comprised electrodes P1/P2, P3/P4, P5/P6 and

PO3/PO4, and a Frontal ROI contained electrodes F1/F2, F3/F4, F5/F6, and AF3/AF4 (see Figure 8). The posterior ROI was selected as Parieto-occipital regions are associated with maximal amplitude of the P600 (e.g., Gouvea et al., 2010) and the P2 (e.g., Hansen et al., 2018). The more central and anterior ROIs were chosen as the amplitude of the N400 has previously been found to be maximal at Centro-parietal regions (e.g., Ganis & Kutas, 2003), while the processing of semantic information related to images, as compared to text, has often been shown to elicit a Frontal negativity during the 300-500 ms temporal window (e.g., Ganis et al., 1996; Holcomb & McPherson, 1994; Mudrik et al., 2014).

Results

Previous scene processing studies investigating either of these three key ERP components (P2, N400, P600) have primarily used ANOVAs. To facilitate direct comparisons between the current experiment and those studies, analysis was conducted on the mean amplitudes for each time-period of interest using 2 (Hemisphere: Left; Right) × 3 (Region: Centro-parietal; Parieto-occipital; Frontal) × 2 (Congruency: Congruous; Incongruous) repeated-measures ANOVAs. Hemisphere was included as a factor within the analysis due to this being the common approach in recent studies of visual narrative sequences (Cohn & Foulsham, 2020; Cohn & Kutas, 2015, 2017), as well as suggestions that scene-related amplitude changes to the P2 component may be greater within the right hemisphere than the left (Hansen et al., 2018; Harel et al., 2020; although see Harel et al., 2016). Where Mauchly's test revealed possible violations of the sphericity assumption Greenhouse-Geisser corrected values are reported (Greenhouse & Geisser, 1959). Significant interactions were followed up with paired t-tests where appropriate. See Appendix B for a summary of the statistical analyses conducted.

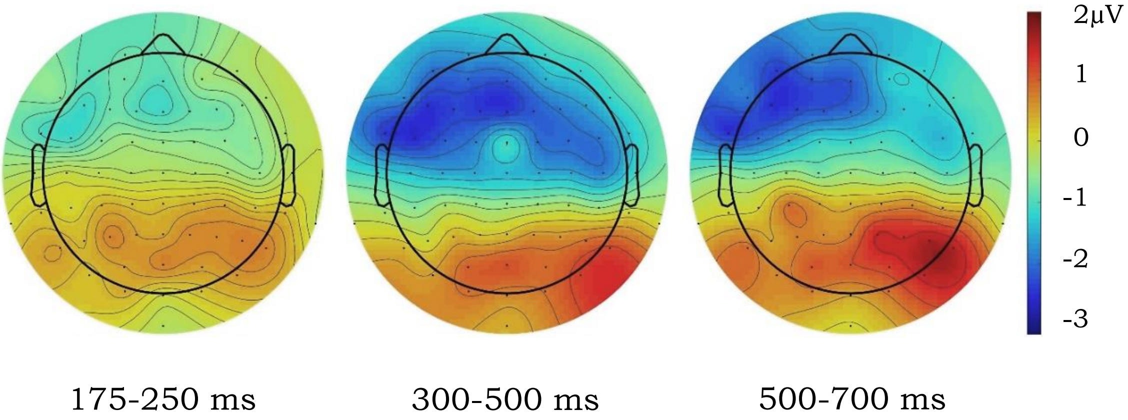
175-250 ms Window

A three-way ANOVA revealed no main effect of Congruency (*p* = .842). There was also no three-way interaction (*p* = .552), nor a Hemisphere × Region interaction (*p* = .396), nor a Hemisphere

× Congruency interaction (*p* = .424). There was, however, a significant Region × Congruency interaction *F*(2, 46) = 15.68, *p* < .001, ηp^2 = .41. See Figure 9 for scalp maps of voltage differences across conditions. In terms of this interaction, follow-up paired t-tests revealed no significant effect of congruency within the Centro-parietal ROI (*p* = .893). There was a significant effect within the Frontal region, *t*(23) = 2.54, *p* = .018, *r* = .47, due to there being a significantly more negative mean amplitude for Incongruous trials (*M* = -3.06 µV) than Congruous trials (*M* = -2.27 µV). See Figure 10 for the grand-averaged Frontal ERPs. This represents a medium-to-large effect. There was also a significant effect within the Parieto-occipital region, *t*(23) = -2.08, *p* = .048, *r* = -.40, representing a medium-sized effect. This was due to there being a significantly more positive mean amplitude for Incongruous trials (*M* = 4.90 µV) than Congruous trials (*M* = 4.33 µV) within Parieto-occipital areas

Figure 9

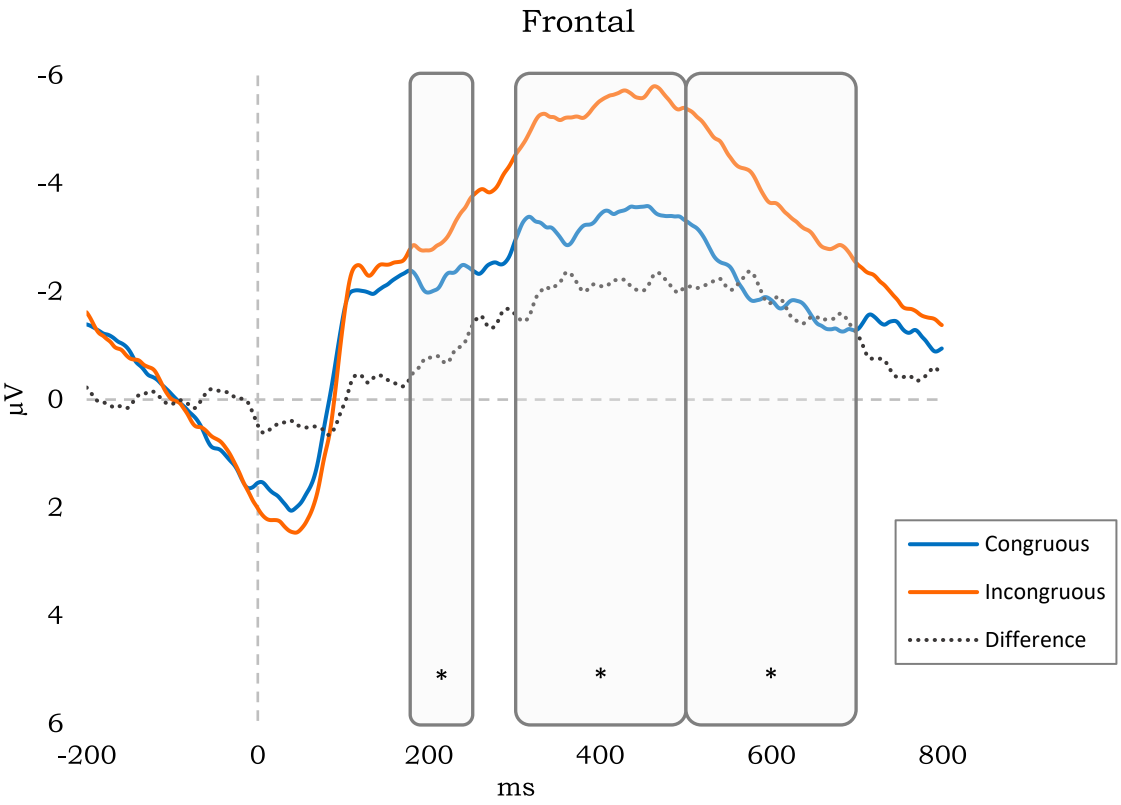
Scalp Maps of the Mean Voltage Difference Between the Congruency Conditions for Each of the Time Windows Under Investigation



Note. Scalp maps represent Incongruous minus Congruous amplitudes. Blue colours indicate the difference is negative, while red colours indicate the difference is positive.

Figure 10

Grand-averaged ERPs for the Frontal Region, Collapsed Across Hemispheres



Note. Blue lines represent amplitudes for Congruous trials and orange lines represent amplitudes for Incongruous trials. Dotted line represents the difference wave (Incongruous minus Congruous). Waveforms low-pass filtered at 30Hz for display purposes (*n* = 24). Grey boxes represent the three time-windows of interest. * denotes *p* < .05.

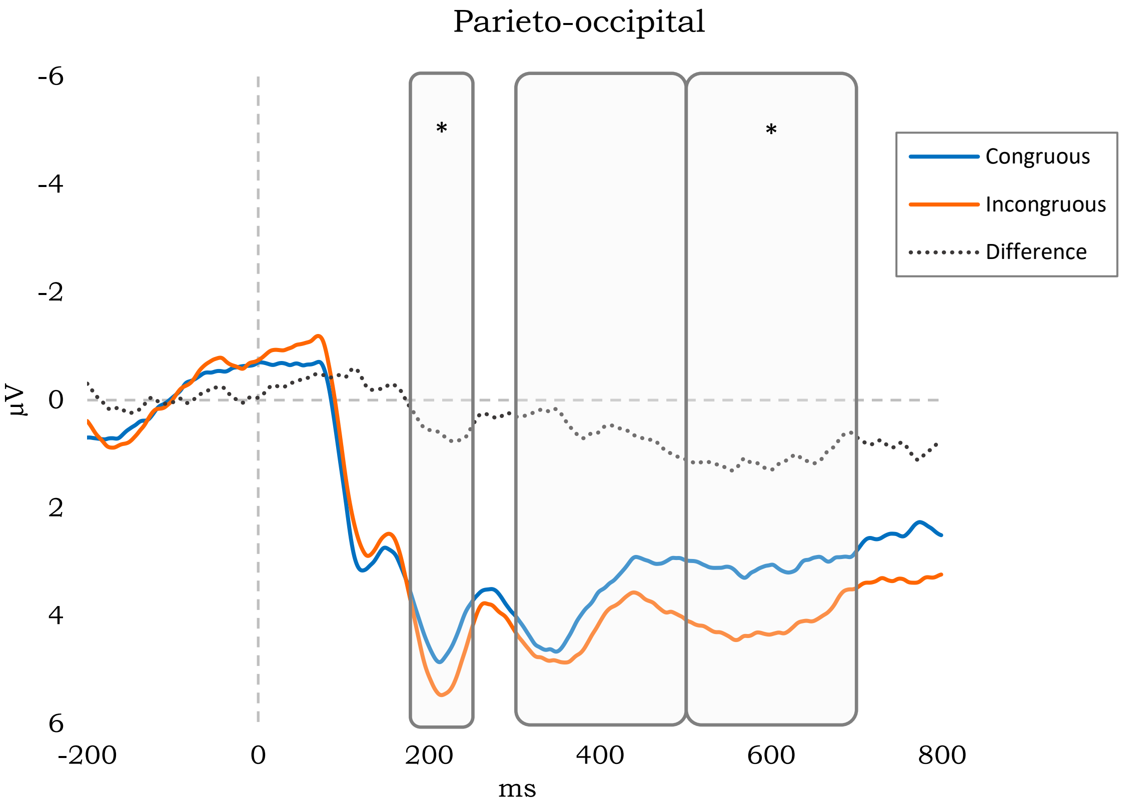
(see Figure 11 for the grand-averaged Parieto-occipital ERPs). Additionally, we re-ran our analysis at slightly more lateral posterior sites, in regions where maximal P2 changes have previously been shown (e.g., Harel et al., 2016; Harel et al., 2020; Hansen et al., 2018). This confirmed our finding of congruency-related changes to the P2 component (see Appendix C for further details).

300-500 ms Window

A three-way ANOVA revealed a main effect of Congruency, *F*(1, 23) = 6.16, *p* = .021, ηp^2 =

Figure 11

Grand-averaged ERPs for the Parieto-Occipital Region, Collapsed Across Hemispheres



Note. Blue lines represent amplitudes for Congruous trials and orange lines represent amplitudes for Incongruous trials. Dotted line represents the difference wave (Incongruous minus Congruous). Waveforms low-pass filtered at 30Hz for display purposes (*n* = 24). Grey boxes represent the three time-windows of interest. * denotes *p* < .05.

.21, due to there being significantly more negative mean amplitudes for Incongruous trials (*M* = -0.85 µV) than Congruous trials (*M* = -0.02 µV) during this time-window. There was no three-way interaction (*p* = .136), nor a Hemisphere × Region interaction (*p* = .274), nor a Hemisphere × Congruency interaction (*p* = .117). There was, however, a significant Region × Congruency interaction *F*(1.46, 33.65) = 32.92, *p* < .001, ηp^2 = .59. In terms of this interaction, follow-up paired t-tests revealed no significant effect of congruency within the Parieto-occipital region (*p* = .129). However, there was a significant effect within the Centro-parietal region, *t*(23) = 2.40, *p* = .025, *r* = .45. This represents a medium-to-large effect. This was due to there being a significantly more negative mean amplitude for Incongruous trials (*M* = -1.39 µV) than Congruous trials (*M* = -0.41 µV) within the Centro-parietal region during this time-window (see Figure 12 for the grand-averaged Centro-parietal ERPs). There was also a significant effect within the Frontal region, $t(23) = 5.37$, $p < .001$, $r = .75$, representing a large effect. This was due to there being a significantly more negative mean amplitude for Incongruous (M = -5.40 µV) than Congruous trials (M = -3.33 µV).

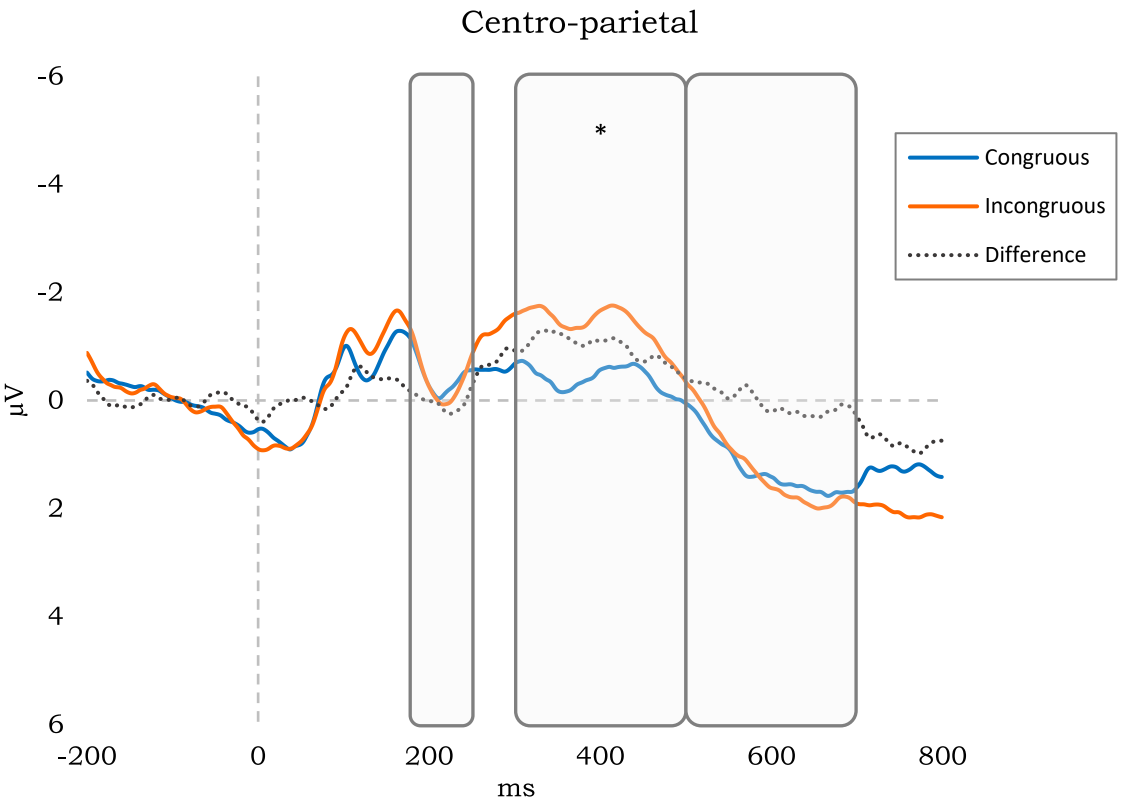
500-700 ms Window

A three-way ANOVA revealed no main effect of Congruency ($p = .553$). There was also no three-way interaction ($p = .056$), nor a Hemisphere × Region interaction ($p = .656$). There was, however, a significant Hemisphere × Congruency interaction, $F(1, 23) = 5.72$, $p = .025$, $\eta p^2 = .20$.

Follow up paired t-tests for this interaction, collapsed across region, revealed no significant congruency-related difference in amplitude in either the left *(p* = .185) or right hemisphere (p = .958). There was also a significant Region × Congruency interaction $F(2, 46) = 34.05$, $p < .001$, $\eta p^2 = .60$. In terms of this interaction, follow-up paired t-tests revealed no significant effect of congruency within the Centro-parietal region ($p = .972$), but did find a significant effect within the Parieto-occipital region, $t(23) = -2.41$, $p = .025$, $r = -.45$. This represents a medium-to-large effect. This was due to there being a significantly more positive mean amplitude for Incongruous trials (M = 4.15 µV) than Congruous trials (M = 3.06 µV) within the Parieto-occipital region during this time-window. There was also a significant effect of congruency within the Frontal region, $t(23) = 4.57$, $p < .001$, $r = .69$, representing a large effect. This was due to there being significantly more negative mean amplitudes for Incongruous (M = -3.83 µV) than Congruous trials (M = -1.99 µV).

Figure 12

Grand-averaged ERPs for the Centro-Parietal Region, Collapsed Across Hemispheres



Note. Blue lines represent amplitudes for Congruous trials and orange lines represent amplitudes for Incongruous trials. Dotted line represents the difference wave (Incongruous minus Congruous). Waveforms low-pass filtered at 30Hz for display purposes (*n* = 24). Grey boxes represent the three time-windows of interest. * denotes *p* < .05.

ERP Results Summary and Discussion

Compared to Congruous trials, Incongruous trials displayed a significantly more positive mean amplitude for the P2 across the Parieto-occipital region (Figure 11), a significantly more negative mean amplitude for the N400 across the Centro-parietal (Figure 12) and Frontal regions (Figure 10), and a significantly more positive mean amplitude for the P600 within the Parieto-occipital region. We also found significantly more negative amplitudes for Incongruous trials within the Frontal region across the early and late windows of interest.

The congruency-related amplitude changes seen in the P2 component support our contention that observer expectations were affecting gist processing, as this suggests top-down information was having an influence while perceptual processing was still ongoing. The sensitivity of the P2 to top-down information is still debated (e.g., Hansen et al., 2018), although previous work has relied on the presentation of individual images. It may be that changes to this early component seen here result from participants being able to generate expectations prior to target onset, thus providing the opportunity for more immediate top-down influence once the destination image is presented.

Amplitude changes across conditions were also apparent in the N400, supporting our prediction that expectations were influencing higher-level cognitive processes rather than simply the extraction of low-level visual information. Furthermore, changes to the N400 have repeatedly been demonstrated for violations to semantic expectations within language (e.g., Van Petten, 1995) and across object-scene pairs (e.g., Demiral et al., 2012), and so the current findings suggest that an equivalent effect exists for violations between separate scenes. We found these N400 amplitude changes within the Centro-parietal as well as the Frontal region, in line with research showing the semantic processing of pictorial stimuli elicits a more anteriorly located negativity during this temporal window (e.g., Ganis et al., 1996). Further to this, we saw morphological dissimilarities within the ERPs across these two regions. In particular, the congruency-related amplitude differences in the anterior region spanned all three time-windows investigated, meaning the effect of approach-destination congruency was apparent in Frontal sites within 175-250 ms from target onset.

As with the N400, changes to the P600 again suggest an influence of higher-level processes on gist extraction. There is still debate, within both scene and language research, as to whether alterations to the P600 are a reflection of difficulties with semantic (e.g., Mudrik et al., 2010; Sitnikova et al., 2003) or syntactic (Hagoort & Brown, 2000; Võ & Wolfe, 2013) processing, or indeed an integration of both (Friederici & Weissenborn, 2007). Additionally, different aspects of syntactic processing have been proposed as being reflected in the P600 (e.g., Friederici et al., 2002; Gouvea et al., 2010; Kaan et al., 2000). For example, increased positivity at P600 has been elicited during ‘garden path’ sentences, which require a re-interpretation of expectations while reading a sentence due to an atypical grammatical format (Osterhout & Holcomb, 1992). Some equivalence to the current study is apparent, whereby expectations are built during a progression of approach images only to require re-evaluation once violated by the appearance of an incongruous destination.

In sum, congruency-based differences were found in a putative scene-selective ERP component, related to integrating visual properties (P2), as well as later components related to contextual integration including semantic and syntactic coherence (N400 and P600, respectively).

General Discussion

As predicted, across experiments we found a benefit for categorising scenes when semantically congruous with lead-up images. Experiment 1a-b revealed an advantage that was greatest at shorter target durations, where the opportunity to process visual information was most limited. Experiment 2 revealed that the performance advantage seen for Congruous trials was based on approach images providing a semantic context for upcoming targets, and were benefitted when they were presented in sequential order; however, no sequence effect was found in the Incongruous condition. Experiment 3 then confirmed that Congruous approach images facilitated gist processing, while Incongrous approach images inhibited performance compared to baseline. Experiment 4 demonstrated that these effects were not solely the product of response bias, but rather the approach images changing the sensitivity to the target scene.

Experiment 5 then swapped to ERP analysis to investigate the neural correlates of predictability on rapid scene processing, showing an effect across all tested ERP components. For Incongruous trials, the P2 and P600 showed significantly greater mean amplitudes within the Parieto-occipital region, while a significantly more negative mean amplitude for the N400 was seen within the Centro-parietal and Frontal regions. Furthermore, Incongruous trials were also associated with a significantly more negative amplitude across the early and late time-windows within Frontal sites. Taken together, this meant we found congruency-related changes within the earliest known indicant of scene-specific processing (P2), within the component classically proposed as an index of semantic expectancy as well as the retrieval of conceptual information (N400), and within the component associated with both semantic and syntactic processing (P600).

The collective results demonstrate that expectations can alter scene gist processing. Importantly, the most substantial differences were found at target durations of 50 milliseconds and below, indicative of expectations influencing the earliest stages of processing. Top-down information therefore appears to modulate the extraction of scene gist.

Our findings from the initial behavioural experiments stand in agreement with recent findings that suggest pre-target narrative sequences can affect subsequent scene processing (Smith & Loschky, 2019). That work demonstrated an influence of sequential predictions on gist recognition through manipulating the spatiotemporal coherence of connected routes in a familiar environment and asking participants to identify a target scene embedded within a rapidly presented sequences of associated scene images (24 ms scene presentation with 276ms interstimulus interval). It was found that categorisation accuracy was greater for targets in sequentially-coherent series, that targets in coherent sequences were more predictable, that this predictability contributed to categorisation performance independently to the visual similarity between the prime and target, and that the advantage of coherent sequences was related to greater perceptual sensitivity for scenes rather than response biases (Smith & Loschky, 2019). Therefore, despite having different approaches – namely in terms of whether the primary manipulation was one of congruency or sequentiality, whether the presentation duration of target scenes was manipulated or not, or whether scenes were from familiar environments or not – both studies support one another in pointing towards visually-evoked expectations as able to influence a scene's subsequent processing, above and beyond that based on the perceptual priming of low-level image features.

The relationship between expectations and gist processing also builds on complimentary work concerning improbable scenes (Greene et al., 2015). That research uncovered increased difficulty in understanding the meaning of atypical scenes, pointing to a disruption in gist processing when scenes diverge from what an observer expects. Such findings strongly point to a role of top-down information in rapid scene understanding, although there is an important distinction to our work. The violation of expectations within single scenes – such as a boulder inside a room (Greene et al., 2015) – would potentially result from inconsistencies between the bottom-up signal and a template stored in long-term memory. On the other hand, our study showed the effect of violating predictions based on the on-line flow of information: the introduction of approach images meant predictions could be formed prior to target onset, potentially resulting in the pre-activation of templates expected to be required for matching against the stimulus. Such pre-activation provides the opportunity for a stored representation to be available prior to the appearance of the target-derived signal, conceivably resulting in more rapid matching or, through predictive coding mechanisms, allowing for the detection of inconsistencies at an earlier processing level due to pre-emptive changes in error thresholds (Rauss et al., 2011).

These results stand as a challenge to ‘forward sweep’ models, which assume minimal top-down modulation of gist processing (Fabre-Thorpe et al., 2001; Itti et al., 1998; Potter et al., 2014; Rumelhart, 1970). However, we do not suggest our results reject the primacy of bottom-up visual factors in scene perception: across each of our behavioural experiments the accuracy with which scenes were categorised far exceeded chance level, even when approach images were incongruous with destinations. Some degree of gist processing was still possible when no relevant semantic information was provided prior to destination-scene onset. Therefore, we propose that feature extraction mechanisms may well be capable of rapidly distinguishing a great deal of information within complex natural scenes (e.g., Potter et al., 2014), but that these mechanisms are susceptible to influence from higher-level processing. This might particularly be the case when, as in our design, antecedent information is provided before gist processing begins, thereby allowing for the formation of expectations prior to a scene being encountered. It may be argued that a design eliciting *a priori* predictions allows for the generation of a pre-emptive attentional set or cognitive state (e.g., Gilbert & Li, 2013), and that visual information then proceeds through this in a feedforward manner once received, but it is still the case that such frameworks do not currently address this sufficiently due to their focus on immediate visual stimulation as determining gist processing.

While the preceding behavioural experiments provided clear support for an effect of expectations on gist processing, it was important to investigate the manner of such influence. The 'No-context' condition in Experiment 3 served as a measure of baseline performance, in terms of gist processing ability in the absence of antecedent contextual information, to which the congruency conditions could be compared. As predicted, we saw significantly increased categorisation ability on Congruous trials when compared to baseline performance, confirming that contextual information facilitated subsequent gist processing. Such a finding was expected due to the well-understood mechanisms of visual processing, where increased efficiency is achieved through utilisation of learned regularities to generate expectations as to the current environment (Chaumon et al., 2008; Fiser et al., 2016; Gregory, 1997; Li et al., 2004; Rock, 1997; Ullman, 1980). Experiment 3 also demonstrated interference to gist extraction when participants were provided with inappropriate contextual information. This appears to agree with previous gist processing research, where improbable scenes – i.e., those which contain unexpected features – were found by participants to be more difficult to extract meaning from, as compared to typical scenes (Greene et al., 2015). It is similarly in line with work demonstrating that object recognition is inhibited when there is a contextual violation with the surrounding scene (Biederman et al., 1982; Davenport & Potter, 2004; Lauer et al., 2020; Palmer, 1975).

The exact mechanisms governing such interference remain open to interpretation, although it seems reasonable to suggest that the deficit in performance results from an attempt to match an unexpected bottom-up signal to an inappropriate, internally generated representation. This may be in the form of predictive coding mechanisms, whereby a significant disparity between expectations and ascending signal leads to prediction errors substantial enough to force reanalysis of the sensory input (e.g., Barrett & Simmons, 2015; Macpherson, 2017; Talsma, 2015). Alternatively, the Scene Perception and Event Comprehension Theory (SPECT; e.g., Loschky et al., 2020) proposes that an observer creates an internal current event model while progressing through a narrative, which represents their understanding of what is happening in that moment. Within this framework, significant changes in situational continuity initiate an automatic cognitive shifting towards creation of a new event model, and this operation is associated with distinct processing costs (Loschky et al., 2020). In the current study, therefore, reduced performance may have resulted from the disruption to processing due to the break in contextual continuity within Incongruous series. On the other hand, the case could be made that participants continued to search for an associative link when confronted with the lack of coherence within Incongruous trials, resulting in a protracted cognitive load that affected low-level perceptual processes (Afiki & Bar, 2020), or even that violations to predictions invoked increased encoding of the current scene-image while actively suppressing retrieval mechanisms (Sherman & Turk-Browne, 2020).

The electrophysiological analysis of the underlying neural signatures help elucidate both the time-course and means by which the violation of expectations affects processing. The congruency-related amplitude differences in the P2 component, appearing so soon after target onset, suggest an influence of top-down information while perceptual processing was still ongoing, similar to that proposed for object recognition (Bar, 2003; Fenske et al., 2006). While the P2 has previously been suggested to be a marker for scene processing (Harel et al., 2016), there is debate as to whether this component is sensitive to top-down influence. For example, recent research found no top-down modulatory effect (Hansen et al., 2018), at least in relation to observer-based goals. Conversely, some forms of early higher-order influence have been implied, as changes to amplitude at ~200 ms post-stimulus have been observed with tasks involving the detection of objects within natural scenes, potentially reflecting decision-related activation (Thorpe et al., 1996; VanRullen & Thorpe, 2001), and tasks that manipulated the emotional nature of scene-images, argued as being driven by motivational systems (Schupp et al., 2006). Relatedly, recent work investigating scene-object congruity has found context effects arising ~170 ms post stimulus onset (Guillaume et al., 2018).

It is possible that different forms of top-down information are integrated at different temporal points, or simply that modulations to such early ERP components are more apparent under certain experimental designs than others. It may be that changes to the P2, found here, result from the use of antecedent information. Our use of approach images allowed for an expectation of the upcoming target category to be formed prior to its onset, meaning that this top-down information was available to facilitate processing from the moment the destination scene was presented. This is a clear departure from a task that involves a single image, whereby bottom-up input, perhaps in terms of low spatial frequency information (e.g., Bar et al., 2006), must first be employed at scene onset to form expectations and only then is available as a tool for the ongoing evaluation of the incoming signal. As a result, it appears reasonable that a design eliciting expectations prior to target-onset would be able to more swiftly affect early ERP components such as the P2, as compared to single-image designs.

Our display of increased negativity at the N400 mirrors previous work related to the semantic violation of object-scene pairs. Such effects have been observed both when a scene is presented prior to target-object presentation, thereby allowing for *a priori* expectations as to the identity of the upcoming object to be formed (e.g., Demiral et al., 2012; Ganis & Kutas, 2003; Võ & Wolfe, 2013), as well as during simultaneous presentation of objects and scenes, where expectations as to object appropriateness cannot be formed prior to onset (e.g., Mudrik at al., 2010). However, while this previous work investigated violations to the semantic relationship between single scenes and their objects, our results show a comparable neural signature resulting from semantic violations between scenes.

The processing of semantic information related to images, as opposed to text, has often been shown to elicit a more anterior negativity during this temporal window (e.g., Ganis et al., 1996;

Holcomb & McPherson, 1994; Kutas et al., 2006), and our results reflect this. However, while both Frontal and Centro-parietal sites here displayed typical N400 effects, in terms of increased negativity for Incongruous trials, the pattern of amplitude changes are morphologically dissimilar across regions. Notably, the congruency-based amplitude changes in anterior sites began to emerge earlier (~200 ms) and were sustained for a far greater period of time (until at least 750 ms after target onset), with significantly more negative amplitudes for Incongruous trials across all three time-windows. There is minimal research regarding similar late effects at anterior sites, although it has previously been attributed to late processes of semantic evaluation (Mudrik et al., 2014).

On the other hand, investigations of pre-N400 negativity across frontal regions have been more frequent. In particular, the earlier emergence of effects at anterior compared to central sites has repeatedly been observed in object-scene research, leading to the proposition that this reflects a separate component, namely the N300 (Barrett & Rugg, 1990; Demiral et al., 2012; McPherson & Holcomb, 1999; Truman & Mudrik, 2018). This has been offered as reflecting context effects at a perceptual level (e.g., Schendan & Kutas, 2002; Mudrik et al., 2010), immediately prior to the semantic processing indicated by the subsequent N400. Furthermore, the N300 appears to be sensitive to alterations in global stimulus features rather than to low-level visual elements (e.g., Schendan & Kutas, 2007), and recent work has suggested it may be an index of perceptual hypothesis testing at a scale of whole scenes and objects, such as template matching routines based on perceptual structure (Kumar et al., 2021). It has also been put forward that components prior to the N300 may reflect predictive coding mechanisms in relation to expected low-level visual features (Kumar et al., 2021). However, distinguishable N300 effects have often not been forthcoming (e.g., Demiral et al., 2012; Ganis & Kutas, 2003) and this dissociation between the N300 and N400 is still debated (see, for example, Draschkow et al., 2018; Willems et al., 2008).

It is important to note that our early window of interest (175-250 ms) preceded the window typically used for investigating the N300 (e.g., Kumar et al., 2021; Lauer et al., 2020), and so our intention is not to comment directly on the debate surrounding that particular component. What we do assert, however, is that – if the N300 is taken as indexing perceptual, rather than higher-order, processing – then our early effects across anterior regions should be similarly categorised. In other words, due the early amplitude changes within Frontal sites as well as the alterations to the P2 discussed above, we suggest that expectations generated prior to target presentation were able to influence the extraction of scene gist at the level of perceptual processing. Predictions as to the category of an upcoming scene are likely to contain predictions not just of its identity, but also its expected perceptual features. At one level an observer may expect to see a beach, but on another level they may be expecting a certain spatial layout (Sanocki & Epstein, 1997) or specific form of spatial envelope (Oliva & Torralba, 2001), or a certain array of colours (Castelhano & Henderson, 2008; Gegenfurtner & Rieger, 2000), textures (Renninger & Malik, 2004), edge-based information (Walther & Shen, 2014) or other low-level features (Shafer-Skelton & Brady, 2019). However, whether the expectation-based violations to processing seen here were related to global properties or to lower-level information remains open to debate.

In terms of the P600, the changes observed here help further elucidate the potential mechanisms underlying the effect of expectations on scene processing. As with the N400, previous scene-related studies investigating this component have focused on object-scene pairs, but findings have proved inconsistent. For instance, Mudrik and colleagues (2010) found that positioning inappropriate objects in appropriate places (a semantic violation, such as a chessboard – rather than a baking tray – being placed into an oven) led to a more negative amplitude at 600 ms, compared to scenes containing appropriate objects. Võ and Wolfe (2013), alternatively, found no alterations to the P600 with similar object-scene semantic violations, but did find an increased P600 when appropriate objects were presented in a position considered to be atypical (such as a dishtowel on the floor, as opposed to hanging on a nearby towel rail). The authors proposed that these images created syntactic – rather than semantic – violations, as the objects contravened structural rules while remaining semantically congruous with their scenes. Thus, they reported the P600 as reflecting syntactic violations to scene processing (Võ & Wolfe, 2013), and so there appears to be a lack of consensus regarding the types of context-based violation that lead to changes in this component. However, it may be the case that these differing results reflect sensitivity to different methodological choices across studies, such as whether the scene is presented prior to the object or simultaneously with it, and whether the object is in a position of stable rest or being acted upon by agents within the image.

The current study, on the other hand, found alterations to the P600 without such violations to object location or appropriateness. It may be, therefore, that these similar ERP patterns are reflecting different phenomena, as research has shown the P600 to be associated with different forms of syntactic anomaly (Gouvea et al., 2010). Increased positivity at the P600 for inconsistent syntax between scenes and objects may be akin to grammatical errors in sentences (e.g., Hagoort et al., 1993), whereas the increased positivity seen here might be more similar to that elicited by 'garden path' sentences (e.g., Osterhout & Holcomb, 1992). Although containing no grammatical errors, progression through such sentences reaches a point where re-interpretation of expectations is necessary, through parsing the word-sequence in a different way. A similar form of violation may be responsible for our P600 pattern. The progression of sequential approach images built an expectation in the observer until the final, incongruous destination disrupted the assumed end-point and resulted in an attempted re-evaluation of meaning. The P600 has similarly been linked to target word inferences in a noisy-channel, such as when the meaning of a semantically incorrect, but orthographically and phonologically similar, word in a sentence can be recovered (e.g. '"The storyteller could turn any incident into an amusing *antidote*" rather than *anecdote*; Ryskin et al., 2021). So, it may not be the case that the P600 is exclusively within the purview of violations to syntax, as it could also be a marker of the sudden need for reanalysis elicited by the disruption to an expected sequence. Such an explanation remains speculative, and further work surrounding the similarities in neural signatures across scene processing and language comprehension is certainly warranted. Both the N400 and P600 in scene processing appear somewhat analogous to those from language comprehensions studies and, while the specific forms of 'grammar' involved in these differing tasks likely diverge, a strong case can be made for the existence of commonalities (e.g., Cohn, 2020).

The congruency-based alterations to the ERP across all time-windows of interest suggest the existence of a singular temporal or cortical point at which top-down predictions affect processing is unlikely. Indeed, the concept of having a specific point of effect is perhaps only valid if a linear hierarchy of visual processing is accepted, as opposed to a cognitive network displaying abundant re-entrant connections (e.g., Boehler et al., 2008; Bullier, 2001; Kauffman et al., 2015; Koivisto et al., 2011). It may be, therefore, more germane to think of predictions of an upcoming scene as influencing manifold areas within the hierarchy simultaneously, whereby expectations set a cortical 'state' deemed appropriate for processing the predicted upcoming signal across the whole network (Gilbert & Li, 2013). Such an account could be considered as fitting within predictive coding frameworks (e.g., Friston, 2010; Friston & Kiebel, 2009; Rao & Ballard, 1999). Internal representations, activated through expectations as to the upcoming scene category, could allow for top-down predictions to propagate across processing areas (e.g., Lewis & Bastiaansen, 2015). As such, regions are informed by predictions based on the approach images, where reanalysis becomes necessary if the bottom-up signal is fundamentally at odds with what was expected (e.g., Talsma, 2015). Importantly, under such a model, a priori expectations may alter prediction error thresholds not only in early visual areas but also within higher-order processing regions (e.g., Hindy et al., 2016; Huang & Rao, 2011; Lewis & Bastiaansen, 2015; Summerfield et al., 2006), thus potentially resulting in a situation where difficulties in matching become apparent across separate levels of abstraction, such as at a perceptual and conceptual level.

This is a nascent area of research and the current study opens several important lines for further investigation. Perhaps most fundamentally, the consistent findings suggest that an investigation of the required level of context to facilitate gist processing would be valuable. For example, it may be the case that presenting visually degraded approach images (e.g., containing only high or low-spatial frequency information) still provide enough information to enhance processing.

Additionally, the findings raise the question as to what other sources of expectations might influence scene processing. These could range from differing forms of top-down communication, such as an observer's goals, to the role of other sensory information, such as potential cross-modal facilitation through the parallel presentation of visual scenes and their related sounds. Finally, there appears to be clear similarities with how expectations affect the processing of meaning across both scenes and language. While the precise mechanisms by which expectations affect the processing of scenes are still to be discovered, we argue that semantically relevant antecedent information may allow for pre-activation of scene-category templates across the visual hierarchy.

References

Afiki, Y., & Bar, M. (2020). Our need for associative coherence. *Humanities & Social Sciences Communications*, *7*(1), Article 80. https://doi.org/10.1057/s41599-020-00577-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), Article e3001023. https://doi.org/10.1371/journal.pbio.3001023

Angrilli, A., Penolazzi, B., Vespignani, F., De Vincenzi, M., Job, R., Ciccarelli, L., Palomba, D., & Stegagno, L. (2002). Cortical brain responses to semantic incongruity and syntactic violation in Italian language: an event-related potential study. *Neuroscience Letters*, *322*(1), 5-8. https://doi.org/10.1016/s0304-3940(01)02528-9

Antal, A., Kéri, S., Kovács, G., Janka, Z., & Benedek, G. (2000). Early and late components of visual categorization: an event-related potential study. *Cognitive Brain Research*, *9*(1), 117-119. https://doi.org/10.1016/s0926-6410(99)00053-1

Antes, J. R., Penland, J. G., & Metzger, R. L. (1981). Processing global information in briefly presented pictures. *Psychological Research-Psychologische Forschung*, *43*(3), 277-292. https://doi.org/10.1007/bf00308452

Auckland, M. E., Cave, K. R., & Donnelly, N. (2007). Nontarget objects can influence perceptual processes during object recognition. *Psychonomic Bulletin & Review*, *14*(2), 332-337. https://doi.org/10.3758/bf03194073

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390-412. https://doi.org/10.1016/j.jml.2007.12.005

Bacon-Macé, N., Macé, M. J.-M., Fabre-Thorpe, M., & Thorpe, S. J. (2005). The time course of visual processing: Backward masking and natural scene categorisation. *Vision Research*, *45*(11), 1459-1469. https://doi.org/10.1016/j.visres.2005.01.004

Banno, H., & Saiki, J. (2015). The processing speed of scene categorization at multiple levels of description: The superordinate advantage revisited. *Perception*, *44*(3), 269-288. https://doi.org/10.1068/p7683

Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object recognition. *Journal of Cognitive Neuroscience*, *15*(4), 600-609. https://doi.org/10.1162/089892903321662976

Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience*, *5*(8), 617-629. https://doi.org/10.1038/nrn1476

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmidt, A. M., Dale, A. M., Hämäläinen, M. S., Marinkovic, K., Schacter, D. L., Rosen, B. R., & Halgren, E. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(2), 449-454. https://doi.org/10.1073/pnas.0507062103

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255-278. https://doi.org/10.1016/j.jml.2012.11.001

Barrett, L. F., & Simmons, W. K. (2015). Interoceptive predictions in the brain. *Nature Reviews Neuroscience*, *16*(7), 419-429. https://doi.org/10.1038/nrn3950

Barrett, S. E., & Rugg, M. D. (1990). Event-related potentials and the semantic matching of pictures. *Brain and Cognition*, *14*(2), 201-212. https://doi.org/10.1016/0278-2626(90)90029-n

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1-48. https://doi.org/10.18637/jss.v067.i01

Bates, D. M., Kliegl, R., Vasishth, S., & Baayen, R. H. (2018). Parsimonious mixed models. *arXiv preprint*. https://doi.org/10.48550/arXiv.1506.04967

Biederman, I., Glass, A. L., & Stacy, E. W. (1973). Searching for objects in real-world scenes. *Journal of Experimental Psychology*, *97*(1), 22-27. https://doi.org/10.1037/h0033776

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology*, *14*(2), 143-177. https://doi.org/10.1016/0010-0285(82)90007-X

Blondin, F., & Lepage, M. (2005). Decrease and increase in brain activity during visual perceptual priming: An fMRI study on similar but perceptually different complex visual scenes. *Neuropsychologia*, *43*(13), 1887-1900. https://doi.org/10.1016/j.neuropsychologia.2005.03.021

Boehler, C. N., Schoenfeld, M. A., Heinze, H. J., & Hopf, J. M. (2008). Rapid recurrent processing. gates awareness in primary visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(25), 8742-8747. https://doi.org/10.1073/pnas.0801999105

Brady, T. F., Shafer-Skelton, A., & Alvarez, G. A. (2017). Global ensemble texture representations are critical to rapid scene perception. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(6), 1160-1176. https://doi.org/10.1037/xhp0000399

Brown, V. A. (2021). An introduction to linear mixed-effects modeling in R. *Advances in Methods and Practices in Psychological Science*, *4*(1), 1-19. https://doi.org/10.1177/2515245920960351

Bullier, J. (2001). Integrated model of visual processing. *Brain Research Reviews*, *36*(2-3), 96-107. https://doi.org/10.1016/s0165-0173(01)00085-6

Caddigan, E., Choo, H., Fei-Fei, L., & Beck, D. M. (2017). Categorization influences detection: A perceptual advantage for representative exemplars of natural scene categories. *Journal of Vision*, *17*(1), Article 21. https://doi.org/10.1167/17.1.21

Camprodon, J. A., Zohary, E., Brodbeck, V., & Pascual-Leone, A. (2010). Two phases of V1 activity for visual recognition of natural images. *Journal of Cognitive Neuroscience*, *22*(6), 1262-1269. https://doi.org/10.1162/jocn.2009.21253

Caplette, L., Gosselin, F., Mermillod, M., & Wicker, B. (2020). Real-world expectations and their affective value modulate object processing. *Neuroimage*, *213*, Article 116736. https://doi.org/10.1016/j.neuroimage.2020.116736

Castelhano, M. S., & Henderson, J. M. (2008). The influence of color on the perception of scene gist. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(3), 660-675. https://doi.org/10.1037/0096-1523.34.3.660

Catherwood, D., Edgar, G. K., Nikolla, D., Alford, C., Brookes, D., Baker, S., & White, S. (2014). Mapping brain activity during loss of situation awareness: an EEG investigation of a basis for top-down influence on perception. *Human Factors*, *56*(8), 1428-1452. https://doi.org/10.1177/0018720814537070

Chaumon, M., Drouet, V., & Tallon-Baudry, C. (2008). Unconscious associative memory affects visual processing before 100 ms. *Journal of Vision*, *8*(3), Article 10. https://doi.org/10.1167/8.3.10

Cichy, R. M., Khosla, A., Pantazis, D., Torralba, A., & Oliva, A. (2016). Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence. *Scientific Reports*, *6*, Article 27755. https://doi.org/10.1038/srep27755

Cohn, N. (2020). Your brain on comics: A cognitive model of visual narrative comprehension. *Topics in Cognitive Science*, *12*(1), 352-386. https://doi.org/10.1111/tops.12421

Cohn, N., & Foulsham, T. (2020). Zooming in on the cognitive neuroscience of visual narrative. *Brain and Cognition*, *146*, Article 105634. https://doi.org/10.1016/j.bandc.2020.105634

Cohn, N., Jackendoff, R., Holcomb, P. J., & Kuperberg, G. R. (2014). The grammar of visual narrative: Neural evidence for constituent structure in sequential image comprehension. *Neuropsychologia*, *64*, 63-70. https://doi.org/10.1016/j.neuropsychologia.2014.09.018

Cohn, N., & Kutas, M. (2015). Getting a cue before getting a clue: Event-related potentials to inference in visual narrative comprehension. *Neuropsychologia*, *77*, 267-278. https://doi.org/10.1016/j.neuropsychologia.2015.08.026

Cohn, N., & Kutas, M. (2017). What's your neural function, visual narrative conjunction? Grammar, meaning, and fluency in sequential image processing. *Cognitive Research-Principles and Implications*, *2*, Article 27. https://doi.org/10.1186/s41235-017-0064-5

Cohn, N., Paczynski, M., Jackendoff, R., Holcomb, P. J., & Kuperberg, G. R. (2012). (Pea)nuts and bolts of visual narrative: Structure and meaning in sequential image comprehension. *Cognitive Psychology*, *65*(1), 1-38. https://doi.org/10.1016/j.cogpsych.2012.01.003

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorial in Quantitative Methods for Psychology*, *1*(1), 42-45. https://doi.org/10.20982/tqmp.01.1.p042

Cousineau, D., Goulet, M. A., & Harding, B. (2021). Summary plots with adjusted error bars: The superb framework with an implementation in R. *Advances in Methods and Practices in Psychological Science*, *4*(3), 1-18, Article 25152459211035109. https://doi.org/10.1177/25152459211035109

Crouzet, S. M., & Serre, T. (2011). What are the visual features underlying rapid object recognition? *Frontiers in Psychology*, *2*, Article 326. https://doi.org/10.3389/fpsyg.2011.00326

Davenport, J. L., & Potter, M. C. (2004). Scene consistency in object and background perception. *Psychological Science*, *15*(8), 559-564. https://doi.org/10.1111/j.0956-7976.2004.00719.x

De Cesarei, A., Mastria, S., & Codispoti, M. (2013). Early spatial frequency processing of natural images: an ERP study. *PLOS ONE*, *8*(5), Article e65103. https://doi.org/10.1371/journal.pone.0065103

de Graaf, T. A., Goebel, R., & Sack, A. T. (2012). Feedforward and quick recurrent processes in early visual cortex revealed by TMS? *Neuroimage*, *61*(3), 651-659. https://doi.org/10.1016/j.neuroimage.2011.10.020

de Graaf, T. A., Koivisto, M., Jacobs, C., & Sack, A. T. (2014). The chronometry of visual perception: Review of occipital TMS masking studies. *Neuroscience and Biobehavioral Reviews*, *45*, 295-304. https://doi.org/10.1016/j.neubiorev.2014.06.017

De Vincenzi, M., Job, R., Di Matteo, R., Angrilli, A., Penolazzi, B., Ciccarelli, L., & Vespignani, F. (2003). Differences in the perception and time course of syntactic and semantic violations. *Brain and Language*, *85*(2), 280-296. https://doi.org/10.1016/s0093-934x(03)00055-5

DeCarlo, L. T. (1998). Signal detection theory and generalized linear models. *Psychological Methods*, *3*(2), 186-205. https://doi.org/10.1037//1082-989x.3.2.186

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. https://doi.org/10.1016/j.jneumeth.2003.10.009

Demiral, Ş. B., Malcolm, G. L., & Henderson, J. M. (2012). ERP correlates of spatially incongruent object identification during scene viewing: contextual expectancy versus simultaneous processing. *Neuropsychologia*, *50*(7), 1271-1285. https://doi.org/10.1016/j.neuropsychologia.2012.02.011

Dien, J. (1998). Issues in the application of the average reference: Review, critiques, and recommendations. *Behavior Research Methods Instruments & Computers*, *30*(1), 34-43. https://doi.org/10.3758/BF03209414

Draschkow, D., Heikel, E., Võ, M. L.-H., Fiebach, C. J., & Sassenhagen, J. (2018). No evidence from MVPA for different processes underlying the N300 and N400 incongruity effects in object-scene processing. *Neuropsychologia*, *120*, 9-17. https://doi.org/10.1016/j.neuropsychologia.2018.09.016

Engel, A. K., Fries, P., & Singer, W. (2001). Dynamic predictions: Oscillations and synchrony in top-down processing. *Nature Reviews Neuroscience*, *2*(10), 704-716. https://doi.org/10.1038/35094565

Epstein, R. A., Higgins, J. S., & Thompson-Schill, S. L. (2005). Learning places from views: Variation in scene processing as a function of experience and navigational ability. *Journal of Cognitive Neuroscience*, *17*(1), 73-83. https://doi.org/10.1162/0898929052879987

Fabre-Thorpe, M., Delorme, A., Marlot, C., & Thorpe, S. (2001). A limit to the speed of processing in ultra-rapid visual categorization of novel natural scenes. *Journal of Cognitive Neuroscience*, *13*(2), 171-180. https://doi.org/10.1162/089892901564234

Fei-Fei, L., Iyer, A., Koch, C., & Perona, P. (2007). What do we perceive in a glance of a real-world scene? *Journal of Vision*, *7*(1), Article 10. https://doi.org/10.1167/7.1.10

Fenske, M. J., Aminoff, E., Gronau, N., & Bar, M. (2006). Top-down facilitation of visual object recognition: object-based and context-based contributions. *Progress in Brain Research*, *155*, 3-21. https://doi.org/10.1016/S0079-6123(06)55001-0

Ferrari, V., Codispoti, M., & Bradley, M. M. (2017). Repetition and ERPs during emotional scene processing: A selective review. *International Journal of Psychophysiology*, *111*, 170-177. https://doi.org/10.1016/j.ijpsycho.2016.07.496

Firestone, C., & Scholl, B. J. (2016). Cognition does not affect perception: Evaluating the evidence for "top-down" effects. *Behavioral and Brain Sciences*, *39*, Article e229. https://doi.org/10.1017/s0140525x15000965

Fiser, A., Mahringer, D., Oyibo, H. K., Petersen, A. V., Leinweber, M., & Keller, G. B. (2016). Experience-dependent spatial expectations in mouse visual cortex. *Nature Neuroscience*, *19*(12), 1658-1664. https://doi.org/10.1038/nn.4385

Foxe, J. J., & Simpson, G. V. (2002). Flow of activation from V1 to frontal cortex in humans: A framework for defining “early” visual processing. *Experimental Brain Research*, *142*(1), 139–150. https://doi.org/10.1007/s00221-001-0906-7

Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition*, *6*, 291-325. https://doi.org/10.1016/0010-0277(78)90002-1

Friederici, A. D., Hahne, A., & Saddy, D. (2002). Distinct neurophysiological patterns reflecting aspects of syntactic complexity and syntactic repair. *Journal of Psycholinguistic Research*, *31*(1), 45-63. https://doi.org/10.1023/a:1014376204525

Friederici, A. D., Pfeifer, E., & Hahne, A. (1993). Event-related brain potentials during natural speech processing: effects of semantic, morphological and syntactic violations. *Cognitive Brain Research*, *1*(3), 183-192. https://doi.org/10.1016/0926-6410(93)90026-2

Friederici, A. D., & Weissenborn, J. (2007). Mapping sentence form onto meaning: The syntax-semantic interface. *Brain Research*, *1146*, 50-58. https://doi.org/10.1016/j.brainres.2006.08.038

Frisch, S., Schlesewsky, M., Saddy, D., & Alpermann, A. (2002). The P600 as an indicator of syntactic ambiguity. *Cognition*, *85*(3), B83-B92. https://doi.org/10.1016/s0010-0277(02)00126-9

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

Friston, K. J., & Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society B-Biological Sciences*, *364*(1521), 1211-1221. https://doi.org/10.1098/rstb.2008.0300

Ganis, G., & Kutas, M. (2003). An electrophysiological study of scene effects on object identification. *Cognitive Brain Research*, *16*(2), 123-144. https://doi.org/10.1016/S0926-6410(02)00244-6

Ganis, G., Kutas, M., & Sereno, M. I. (1996). The search for ''common sense'': An electrophysiological study of the comprehension of words and pictures in reading. *Journal of Cognitive Neuroscience*, *8*(2), 89-106. https://doi.org/10.1162/jocn.1996.8.2.89

Gegenfurtner, K. R., & Rieger, J. (2000). Sensory and cognitive contributions of color to the recognition of natural scenes. *Current Biology*, *10*(13), 805-808. https://doi.org/10.1016/s0960-9822(00)00563-7

Gibson, J. J. (1966). *The senses considered as perceptual systems*. Houghton Mifflin.

Gilbert, C. D., & Li, W. (2013). Top-down influences on visual processing. *Nature Reviews Neuroscience*, *14*(5), 350-363. https://doi.org/10.1038/nrn3476

Gouvea, A. C., Phillips, C., Kazanina, N., & Poeppel, D. (2010). The linguistic processes underlying the P600. *Language and Cognitive Processes*, *25*(2), 149-188. https://doi.org/10.1080/01690960902965951

Greene, M. R., Botros, A. P., Beck, D. M., & Fei-Fei, L. (2015). What you see is what you expect: rapid scene understanding benefits from prior experience. *Attention, Perception, & Psychophysics*, *77*(4), 1239-1251. https://doi.org/10.3758/s13414-015-0859-8

Greene, M. R., & Hansen, B. C. (2018). Shared spatiotemporal category representations in biological and artificial deep neural networks. *PLOS Computational Biology*, *14*(7), Article e1006327. https://doi.org/10.1371/journal.pcbi.1006327

Greene, M. R., & Oliva, A. (2009). The briefest of glances: The time course of natural scene understanding. *Psychological Science*, *20*(4), 464-472. https://doi.org/10.1111/j.1467-9280.2009.02316.x

Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, *24*(2), 95-112. https://doi.org/10.1007/bf02289823

Gregory, R. L. (1997). Knowledge in perception and illusion. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, *352*(1358), 1121-1127. https://doi.org/10.1098/rstb.1997.0095

Grill-Spector, K., & Malach, R. (2004). The human visual cortex. *Annual Review of Neuroscience*, *27*, 649-677. https://doi.org/10.1146/annurev.neuro.27.070203.144220

Groen, I. I. A., Silson, E. H., & Baker, C. I. (2017). Contributions of low- and high-level properties to neural processing of visual scenes in the human brain. *Philosophical Transactions of the Royal Society B-Biological Sciences*, *372*(1714), Article 20160102. https://doi.org/10.1098/rstb.2016.0102

Guillaume, F., Tinard, S., Baier, S., & Dufau, S. (2018). An ERP investigation of object-scene incongruity: The early meeting of knowledge and perception. *Journal of Psychophysiology*, *32*(1), 20-29. https://doi.org/10.1027/0269-8803/a000181

Gunter, T. C., Friederici, A. D., & Schriefers, H. (2000). Syntactic gender and semantic expectancy: ERPs reveal early autonomy and late interaction. *Journal of Cognitive Neuroscience*, *12*(4), 556-568. https://doi.org/10.1162/089892900562336

Gunter, T. C., Stowe, L. A., & Mulder, G. (1997). When syntax meets semantics. *Psychophysiology*, *34*(6), 660-676. https://doi.org/10.1111/j.1469-8986.1997.tb02142.x

Hagoort, P., Brown, C., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes*, *8*(4), 439-483. https://doi.org/10.1080/01690969308407585

Hagoort, P., & Brown, C. M. (2000). ERP effects of listening to speech compared to reading: the P600/SPS to syntactic violations in spoken sentences and rapid serial visual presentation. *Neuropsychologia*, *38*(11), 1531-1549. https://doi.org/10.1016/s0028-3932(00)00053-1

Handy, T. C., Green, V., Klein, R. M., & Mangun, G. R. (2001). Combined expectancies: Event-related potentials reveal the early benefits of spatial attention that are obscured by reaction time measures. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(2), 303-317. https://doi.org/10.1037/0096-1523.27.2.303

Hansen, N. E., Noesen, B. T., Nador, J. D., & Harel, A. (2018). The influence of behavioral relevance on the processing of global scene properties: An ERP study. *Neuropsychologia*, *114*, 168-180. https://doi.org/10.1016/j.neuropsychologia.2018.04.040

Harel, A., Groen, I. I. A., Kravitz, D. J., Deouell, L. Y., & Baker, C. I. (2016). The temporal dynamics of scene processing: A multifaceted EEG investigation. *eNeuro*, *3*(5), Article e0139. https://doi.org/10.1523/eneuro.0139-16.2016

Harel, A., Mzozoyana, M. W., Al Zoubi, H., Nador, J. D., Noesen, B. T., Lowe, M. X., & Cant, J. S. (2020). Artificially-generated scenes demonstrate the importance of global scene properties for scene perception. *Neuropsychologia*, *141*, Article 107434. https://doi.org/10.1016/j.neuropsychologia.2020.107434

Henderson, L. M., Baseler, H. A., Clarke, P. J., Watson, S., & Snowling, M. J. (2011). The N400 effect in children: Relationships with comprehension, vocabulary and decoding. *Brain and Language*, *117*(2), 88-99. https://doi.org/10.1016/j.bandl.2010.12.003

Hindy, N. C., Ng, F. Y., & Turk-Browne, N. B. (2016). Linking pattern completion in the hippocampus to predictive coding in visual cortex. *Nature Neuroscience*, *19*(5), 665-667. https://doi.org/10.1038/nn.4284

Holcomb, P. J. (1993). Semantic priming and stimulus degradation: Implications for the role of the N400 in language processing. *Psychophysiology*, *30*(1), 47-61. https://doi.org/10.1111/j.1469-8986.1993.tb03204.x

Holcomb, P. J., & McPherson, W. B. (1994). Event-related brain potentials reflect semantic priming in an object decision task. *Brain and Cognition*, *24*(2), 259-276. https://doi.org/10.1006/brcg.1994.1014

Hollingworth, A., & Henderson, J. M. (1998). Does consistent scene context facilitate object perception? *Journal of Experimental Psychology: General*, *127*(4), 398-415. https://doi.org/10.1037//0096-3445.127.4.398

Huang, Y. P., & Rao, R. P. N. (2011). Predictive coding. *Wiley Interdisciplinary Reviews-Cognitive Science*, *2*(5), 580-593. https://doi.org/10.1002/wcs.142

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *20*(11), 1254-1259. https://doi.org/10.1109/34.730558

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434-446. https://doi.org/10.1016/j.jml.2007.11.007

Joubert, O. R., Rousselet, G. A., Fize, D., & Fabre-Thorpe, M. (2007). Processing scene context: Fast categorization and object interference. *Vision Research*, *47*(26), 3286-3297. https://doi.org/10.1016/j.visres.2007.09.013

Juan, C. H., & Walsh, V. (2003). Feedback to V1: a reverse hierarchy in vision. *Experimental Brain Research*, *150*(2), 259-263. https://doi.org/10.1007/s00221-003-1478-5

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, *103*(1), 54-69. https://doi.org/10.1037/a0028347

Junghöfer, M., Elbert, T., Tucker, D. M., & Braun, C. (1999). The polar average reference effect: a bias in estimating the head surface integral in EEG recording. *Clinical Neurophysiology*, *110*(6), 1149-1155. https://doi.org/10.1016/s1388-2457(99)00044-9

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes*, *15*(2), 159-201. https://doi.org/10.1080/016909600386084

Kauffmann, L., Chauvin, A., Pichat, C., & Peyrin, C. (2015). Effective connectivity in the neural network underlying coarse-to-fine categorization of visual scenes. A dynamic causal modeling study. *Brain and Cognition*, *99*, 46-56. https://doi.org/10.1016/j.bandc.2015.07.004

Kirchner, H., & Thorpe, S. J. (2006). Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited. *Vision Research*, *46*(11), 1762-1776. https://doi.org/10.1016/j.visres.2005.10.002

Kit, D., Katz, L., Sullivan, B., Snyder, K., Ballard, D., & Hayhoe, M. (2014). Eye movements, visual search and scene memory, in an immersive virtual environment. *PLOS ONE*, *9*(4), Article e94362. https://doi.org/10.1371/journal.pone.0094362

Kliegl, R., Wei, P., Dambacher, M., Yan, M., & Zhou, X. (2011). Experimental effects and individual differences in linear mixed models: estimating the relationship between spatial, object, and attraction effects in visual attention. *Frontiers in Psychology*, *1*, Article 238. https://doi.org/10.3389/fpsyg.2010.00238

Koivisto, M., Railo, H., Revonsuo, A., Vanni, S., & Salminen-Vaparanta, N. (2011). Recurrent processing in V1/V2 contributes to categorization of natural scenes. *Journal of Neuroscience*, *31*(7), 2488-2492. https://doi.org/10.1523/jneurosci.3074-10.2011

Kok, P., Jehee, J. F. M., & de Lange, F. P. (2012). Less is more: Expectation sharpens representations in the primary visual cortex. *Neuron*, *75*(2), 265-270. https://doi.org/10.1016/j.neuron.2012.04.034

Kravitz, D. J., Saleem, K. S., Baker, C. I., Ungerleider, L. G., & Mishkin, M. (2013). The ventral visual pathway: an expanded neural framework for the processing of object quality. *Trends in Cognitive Sciences*, *17*(1), 26-49. https://doi.org/10.1016/j.tics.2012.10.011

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

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research*, *1146*, 23-49. https://doi.org/10.1016/j.brainres.2006.12.063

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*(5947), 161-163. https://doi.org/10.1038/307161a0

Kutas, M., Van Petten, C. K., & Kluender, R. (2006). Psycholinguistics electrified II (1994–2005). In M. J. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (2nd ed., pp. 659-724). Academic Press. https://doi.org/10.1016/B978-012369374-7/50018-3

Kveraga, K., Ghuman, A. S., Kassam, K. S., Aminoff, E. A., Hämäläinen, M. S., Chaumon, M., & Bar, M. (2011). Early onset of neural synchronization in the contextual associations network. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(8), 3389-3394. https://doi.org/10.1073/pnas.1013760108

Lauer, T., Willenbockel, V., Maffongelli, L., & Võ, M. L.-H. (2020). The influence of scene and object orientation on the scene consistency effect. *Behavioural Brain Research*, *394*, Article 112812. https://doi.org/10.1016/j.bbr.2020.112812

Lewis, A. G., & Bastiaansen, M. (2015). A predictive coding framework for rapid neural dynamics during sentence-level language comprehension. *Cortex*, *68*, 155-168. https://doi.org/10.1016/j.cortex.2015.02.014

Li, W., Piech, V., & Gilbert, C. D. (2004). Perceptual learning and top-down influences in primary visual cortex. *Nature Neuroscience*, *7*(6), 651-657. https://doi.org/10.1038/nn1255

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: an open-source toolbox for the analysis of event related potentials. *Frontiers in Human Neuroscience*, *8*, Article 213. https://doi.org/10.3389/fnhum.2014.00213

Loschky, L. C., Hansen, B. C., Sethi, A., & Pydimarri, T. N. (2010). The role of higher order image statistics in masking scene gist recognition. *Attention, Perception, & Psychophysics*, *72*(2), 427-444. https://doi.org/10.3758/app.72.2.427

Loschky, L. C., Larson, A. M., Smith, T. J., & Magliano, J. P. (2020). The Scene Perception & Event Comprehension Theory (SPECT) applied to visual narratives. *Topics in Cognitive Science*, *12*(1), 311-351. https://doi.org/10.1111/tops.12455

Luck, S. J. (2014). *An introduction to the event-related potential technique* (2nd ed.). MIT Press.

Macpherson, F. (2017). The relationship between cognitive penetration and predictive coding. *Consciousness and Cognition*, *47*, 6-16. https://doi.org/10.1016/j.concog.2016.04.001

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. W. H. Freeman and Company.

Martín-Loeches, M., Ouyang, G., Rausch, P., Stürmer, B., Palazova, M., Schacht, A., & Sommer, W. (2017). Test-retest reliability of the N400 component in a sentence-reading paradigm.

Language, Cognition and Neuroscience, *32*(10), 1261-1272. https://doi.org/10.1080/23273798.2017.1330485

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305-315. https://doi.org/10.1016/j.jml.2017.01.001

McManus, J. N. J., Li, W., & Gilbert, C. D. (2011). Adaptive shape processing in primary visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(24), 9739-9746. https://doi.org/10.1073/pnas.1105855108

McPherson, W. B., & Holcomb, P. J. (1999). An electrophysiological investigation of semantic priming with pictures of real objects. *Psychophysiology*, *36*(1), 53-65. https://doi.org/10.1017/S0048577299971196

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorial in Quantitative Methods for Psychology*, *4*(2), 61-64. https://doi.org/10.20982/tqmp.04.2.p061

Mudrik, L., Lamy, D., & Deouell, L. Y. (2010). ERP evidence for context congruity effects during simultaneous object-scene processing. *Neuropsychologia*, *48*(2), 507-517. https://doi.org/10.1016/j.neuropsychologia.2009.10.011

Mudrik, L., Shalgi, S., Lamy, D., & Deouell, L. Y. (2014). Synchronous contextual irregularities affect early scene processing: Replication and extension. *Neuropsychologia*, *56*, 447-458. https://doi.org/10.1016/j.neuropsychologia.2014.02.020

Nuthmann, A., Einhäuser, W., & Schütz, I. (2017). How well can saliency models predict fixation selection in scenes beyond central bias? A new approach to model evaluation using generalized linear mixed models. *Frontiers in Human Neuroscience*, *11*(10), Article 491. https://doi.org/10.3389/fnhum.2017.00491

Oliva, A. (2005). Gist of the scene. In L. Itti, G. Rees, & J. K. Tsotsos (Eds.), *Neurobiology of attention* (pp. 251-256). Elsevier. https://doi.org/10.1016/B978-012375731-9/50045-8

Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, *42*(3), 145-175. https://doi.org/10.1023/A:1011139631724

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, *31*(6), 785-806. https://doi.org/10.1016/0749-596x(92)90039-z

Palan, S., & Schitter, C. (2018). Prolific.ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, *17*, 22-27. https://doi.org/10.1016/j.jbef.2017.12.004

Palmer, S. E. (1975). Effects of contextual scenes on identification of objects. *Memory & Cognition*, *3*(5), 519-526. https://doi.org/10.3758/BF03197524

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195-203. https://doi.org/10.3758/s13428-018-01193-y

Portilla, J., & Simoncelli, E. P. (2000). A parametric texture model based on joint statistics of complex wavelet coefficients. *International Journal of Computer Vision*, *40*(1), 49-71. https://doi.org/10.1023/A:1026553619983

Potter, M. C. (1975). Meaning in visual search. *Science*, *187*(4180), 965-966. https://doi.org/10.1126/science.1145183

Potter, M. C., Wyble, B., Hagmann, C. E., & McCourt, E. S. (2014). Detecting meaning in RSVP at 13 ms per picture. *Attention, Perception, & Psychophysics*, *76*(2), 270-279. https://doi.org/10.3758/s13414-013-0605-z

Rao, R. P. N., & 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

Rauss, K., Schwartz, S., & Pourtois, G. (2011). Top-down effects on early visual processing in humans: A predictive coding framework. *Neuroscience and Biobehavioral Reviews*, *35*(5), 1237-1253. https://doi.org/10.1016/j.neubiorev.2010.12.011

Reinitz, M. T., Wright, E., & Loftus, G. R. (1989). Effects of semantic priming on visual encoding of pictures. *Journal of Experimental Psychology: General*, *118*(3), 280-297. https://doi.org/10.1037/0096-3445.118.3.280

Renninger, L. W., & Malik, J. (2004). When is scene identification just texture recognition? *Vision Research*, *44*(19), 2301-2311. https://doi.org/10.1016/j.visres.2004.04.006

Rensink, R. A. (2000). Scene perception. In A. E. Kazdin (Ed.), *Encyclopedia of psychology* (Vol. 7, pp. 151-155). Oxford University Press. https://doi.org/10.1037/10522-061

Rock, I. (1997). *Indirect perception*. MIT Press/Bradford Books.

Rumelhart, D. E. (1970). A multicomponent theory of the perception of briefly exposed visual displays. *Journal of Mathematical Psychology*, *7*(2), 191-218. https://doi.org/10.1016/0022-2496(70)90044-1

Ryskin, R., Stearns, L., Bergen, L., Eddy, M., Fedorenko, E., & Gibson, E. (2021). An ERP index of real-time error correction within a noisy-channel framework of human communication. *Neuropsychologia*, 158, 107855. https://doi.org/10.1016/j.neuropsychologia.2021.107855

Sanocki, T. (2013). Facilitatory priming of scene layout depends on experience with the scene. *Psychonomic Bulletin & Review*, *20*(2), 274-281. https://doi.org/10.3758/s13423-012-0332-9

Sanocki, T., & Epstein, W. (1997). Priming spatial layout of scenes. *Psychological Science*, *8*(5), 374-378. https://doi.org/10.1111/j.1467-9280.1997.tb00428.x

Schendan, H. E., & Kutas, M. (2002). Neurophysiological evidence for two processing times for visual object identification. *Neuropsychologia*, *40*(7), 931-945. https://doi.org/10.1016/s0028-3932(01)00176-2

Schendan, H. E., & Kutas, M. (2007). Neurophysiological evidence for the time course of activation of global shape, part, and local contour representations during visual object categorization and memory. *Journal of Cognitive Neuroscience*, *19*(5), 734-749. https://doi.org/10.1162/jocn.2007.19.5.734

Schupp, H. T., Flaisch, T., Stockburger, J., & Junghöfer, M. (2006). Emotion and attention: event-related brain potential studies. *Progress in Brain Research, 156*, 31-51. Elsevier. https://doi.org/10.1016/S0079-6123(06)56002-9

Serre, T., Oliva, A., & Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(15), 6424-6429. https://doi.org/10.1073/pnas.0700622104

Shafer-Skelton, A., & Brady, T. F. (2019). Scene layout priming relies primarily on low-level features rather than scene layout. *Journal of Vision*, *19*(1), Article 14. https://doi.org/10.1167/19.1.14

Sherman, B. E., & Turk-Browne, N. B. (2020). Statistical prediction of the future impairs episodic encoding of the present. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(37), 22760-22770. https://doi.org/10.1073/pnas.2013291117

Sitnikova, T., Holcomb, P. J., Kiyonaga, K. A., & Kuperberg, G. R. (2008). Two neurocognitive mechanisms of semantic integration during the comprehension of visual real-world events. *Journal of Cognitive Neuroscience*, *20*(11), 2037-2057. https://doi.org/10.1162/jocn.2008.20143

Sitnikova, T., Kuperberg, G., & Holcomb, P. J. (2003). Semantic integration in videos of real-world events: An electrophysiological investigation. *Psychophysiology*, *40*(1), 160-164. https://doi.org/10.1111/1469-8986.00016

Smith, M. E., & Loschky, L. C. (2019). The influence of sequential predictions on scene-gist recognition. *Journal of Vision*, *19*(12), Article 14. https://doi.org/10.1167/19.12.14

Spoerer, C. J., Kietzmann, T. C., Mehrer, J., Charest, I., & Kriegeskorte, N. (2020). Recurrent neural networks can explain flexible trading of speed and accuracy in biological vision. *PLOS Computational Biology*, *16*(10), Article e1008215. https://doi.org/10.1371/journal.pcbi.1008215

Straube, S., & Fahle, M. (2010). The electrophysiological correlate of saliency: Evidence from a figure-detection task. *Brain Research*, *1307*, 89-102. https://doi.org/10.1016/j.brainres.2009.10.043

Summerfield, C., Egner, T., Greene, M., Koechlin, E., Mangels, J., & Hirsch, J. (2006). Predictive codes for forthcoming perception in the frontal cortex. *Science*, *314*(5803), 1311-1314. https://doi.org/10.1126/science.1132028

Summerfield, C., & Koechlin, E. (2008). A neural representation of prior information during perceptual inference. *Neuron*, *59*(2), 336-347. https://doi.org/10.1016/j.neuron.2008.05.021

Talsma, D. (2015). Predictive coding and multisensory integration: an attentional account of the multisensory mind. *Frontiers in Integrative Neuroscience*, *9*, Article 19. https://doi.org/10.3389/fnint.2015.00019

Tanner, D., Grey, S., & van Hell, J. G. (2017). Dissociating retrieval interference and reanalysis in the P600 during sentence comprehension. *Psychophysiology*, *54*(2), 248-259. https://doi.org/10.1111/psyp.12788

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, *381*(6582), 520-522. https://doi.org/10.1038/381520a0

Truman, A., & Mudrik, L. (2018). Are incongruent objects harder to identify? The functional significance of the N300 component. *Neuropsychologia*, *117*, 222-232. https://doi.org/10.1016/j.neuropsychologia.2018.06.004

Ullman, S. (1980). Against direct perception. *Behavioral and Brain Sciences*, *3*(3), 373-381. https://doi.org/10.1017/s0140525x0000546x

Ullman, S. (1995). Sequence seeking and counter streams: a computational model for bidirectional information flow in the visual cortex. *Cerebral Cortex*, *5*(1), 1-11. https://doi.org/10.1093/cercor/5.1.1

Van Petten, C. (1995). Words and sentences: Event-related brain potential measures. *Psychophysiology*, *32*(6), 511-525. https://doi.org/10.1111/j.1469-8986.1995.tb01228.x

Van Petten, C., & Luka, B. J. (2006). Neural localization of semantic context effects in electromagnetic and hemodynamic studies. *Brain and Language*, *97*(3), 279-293. https://doi.org/10.1016/j.bandl.2005.11.003

VanRullen, R., & Thorpe, S. J. (2001). The time course of visual processing: From early perception to decision-making. *Journal of Cognitive Neuroscience*, *13*(4), 454-461. https://doi.org/10.1162/08989290152001880

Viggiano, M. P., & Kutas, M. (2000). Overt and covert identification of fragmented objects inferred from performance and electrophysiological measures. *Journal of Experimental Psychology: General*, *129*(1), 107-125. https://doi.org/10.1037/0096-3445.129.1.107

Võ, M. L.-H., & Henderson, J. M. (2011). Object-scene inconsistencies do not capture gaze: evidence from the flash-preview moving-window paradigm. *Attention, Perception, & Psychophysics*, *73*(6), 1742-1753. https://doi.org/10.3758/s13414-011-0150-6

Võ, M. L.-H., & Wolfe, J. M. (2013). Differential electrophysiological signatures of semantic and syntactic scene processing. *Psychological Science*, *24*(9), 1816-1823. https://doi.org/10.1177/0956797613476955

Walther, D. B., & Shen, D. (2014). Nonaccidental properties underlie human categorization of complex natural scenes. *Psychological Science*, *25*(4), 851-860. https://doi.org/10.1177/0956797613512662

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis* (2d ed.). Springer. https://doi.org/10.1007/978-3-319-24277-4

Willems, R. M., Özyürek, A., & Hagoort, P. (2008). Seeing and hearing meaning: ERP and fMRI evidence of word versus picture integration into a sentence context. *Journal of Cognitive Neuroscience*, *20*(7), 1235-1249. https://doi.org/10.1162/jocn.2008.20085

Wright, D. B., Horry, R., & Skagerberg, E. M. (2009). Functions for traditional and multilevel approaches to signal detection theory. *Behavior Research Methods*, *41*(2), 257-267. https://doi.org/10.3758/brm.41.2.257

Yao, D. Z., Wang, L., Arendt-Nielsen, L., & Chen, A. C. N. (2007). The effect of reference choices on the spatio-temporal analysis of brain evoked potentials: The use of infinite reference. *Computers in Biology and Medicine*, *37*(11), 1529-1538. https://doi.org/10.1016/j.compbiomed.2007.02.002

Yuan, J. J., Zhang, Q. L., Chen, A. T., Li, H., Wang, Q. H., Zhuang, Z. C. X., & Jia, S. W. (2007). Are we sensitive to valence differences in emotionally negative stimuli? Electrophysiological evidence from an ERP study. *Neuropsychologia*, *45*(12), 2764-2771. https://doi.org/10.1016/j.neuropsychologia.2007.04.018

Appendix A

List of Scene Categories

ART GALLERY; BATHROOM; BEACH; BEDROOM; CARPARK; CHURCH; DINING ROOM; ENTRANCE HALL; FIELD; GARDEN; GRAVEYARD; HIGH STREET; KITCHEN; LIVING ROOM; MULTISTOREY CARPARK; OUTBUILDING; PARK; PETROL STATION; PUB; QUAY; RECYCLING AREA; RETAIL STORE; RIVER; ROAD; SHOP; SPORTS PITCH; SUPERMARKET; TAKEAWAY; TRAIN STATION; WOODS

Appendix B

Statistical Analyses

Window	Factor	*df*	*F*	*t*	*p*	ηp^2	*r*
175-250 ms	Hemisphere	1, 23	10.21		.004*	.31	
	Region	1.17, 27.00	51.28		.000*	.69	
	Congruency	1, 23	0.04		.842	.00	
	Hemisphere*Region*Congruency	1.61, 37.04	0.54		.552	.02	
	Hemisphere*Region	2, 46	0.95		.396	.04	
	Hemisphere*Congruency	1, 23	0.66		.424	.03	
	Region*Congruency	2, 46	15.68		.000*	.41	
	Paired t-tests (for R*C interaction)						
	Frontal	23		2.54	.018*		.47
	Centro-parietal	23		-0.14	.893		.03
	Parieto-occipital	23		-2.08	.048*		.40
300-500 ms	Hemisphere	1, 23	6.39		.019*	.22	
	Region	1.21, 27.90	45.37		.000*	.66	
	Congruency	1, 23	6.16		.021*	.21	
	Hemisphere*Region*Congruency	2, 46	2.08		.136	.08	
	Hemisphere*Region	2, 46	1.33		.274	.06	
	Hemisphere*Congruency	1, 23	2.65		.117	.10	
	Region*Congruency	1.46, 33.65	32.92		.000*	.59	
	Paired t-tests (for R*C interaction)						
	Frontal	23		5.37	.000*		.75
	Centro-parietal	23		2.40	.025*		.45
	Parieto-occipital	23		-1.57	.129		.31
500-700 ms	Hemisphere	1, 23	6.87		.015*	.23	
	Region	1.52, 34.92	44.53		.000*	.66	
	Congruency	1, 23	0.36		.553	.02	
	Hemisphere*Region*Congruency	2, 46	3.07		.056	.12	
	Hemisphere*Region	2, 46	0.43		.656	.02	
	Hemisphere*Congruency	1, 23	5.72		.025*	.20	
	Paired t-tests (for H*C interaction)						
	Left hemisphere	23		1.37	.185		.32
	Right hemisphere	23		-0.05	.958		.01
	Region*Congruency	2, 46	34.05		.000*	.60	
	Paired t-tests (for R*C interaction)						
	Frontal	23		4.57	.000*		.69
	Centro-parietal	23		-0.04	.972		.01
	Parieto-occipital	23		-2.41	.025*		.45
175-250 ms	Hemisphere	1, 23	6.57		.017*	.22	
(Lateral P2)	Congruency	1, 23	5.81		.024*	.20	
	Hemisphere*Congruency	1, 23	1.26		.274	.05	

Note. The three windows of the main analysis were analysed with 2 × 3 × 2 ANOVAs. The additional analysis of the 175-250 ms window for the lateral Parieto-occipital region was analysed with a 2 × 2 ANOVA. * denotes *p* < .05

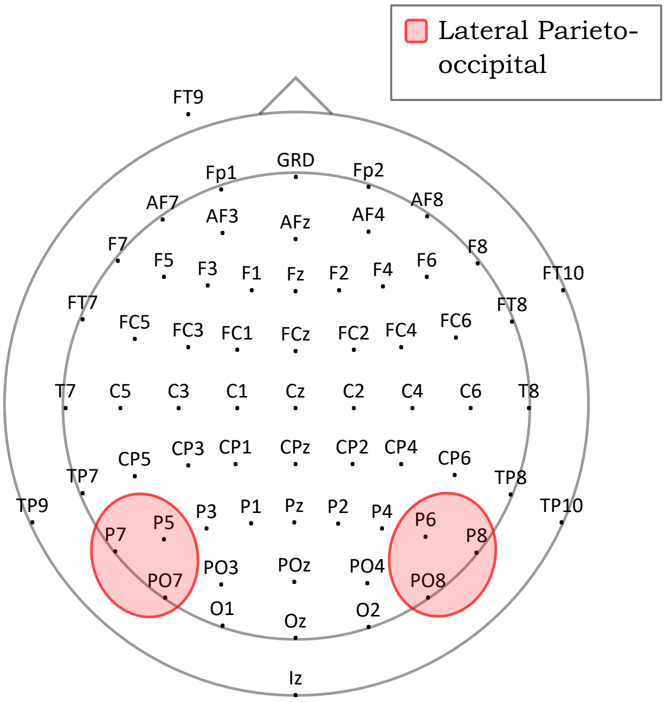
Appendix C

Additional Analysis: Lateral P2

In our initial analyses we found a significant effect of Congruency within the P2 time-window at posterior sites. However, in the interest of completeness we decided further investigation would be insightful. Previous scene processing research concerned with the P2 component has found effects to be maximal at sites more lateral than our initial ROIs (Hansen et al., 2018; Harel et al., 2016; Harel et al., 2020). Consequently, we created a Lateral Parieto-occipital ROI comprising six electrodes (split equally across hemispheres). The position of these regions was chosen to mirror previous work as closely as possible. Specifically, Harel and colleagues (2016; 2020) use a lateral region including eight electrodes across the two hemispheres (P5/P6, P7/P8, P9/P10 and PO7/PO8). Exact duplication of this setup was not possible, as instead of the electrode pair P9/P10 our array included TP9/TP10, which were located near the mastoids, and had been used as our re-referencing electrodes. Therefore, our lateral regions consisted of P5/P6, P7/P8 and PO7/PO8 (see Figure D1).

Figure C1

Map of Electrode Placement Including the Lateral ROIs



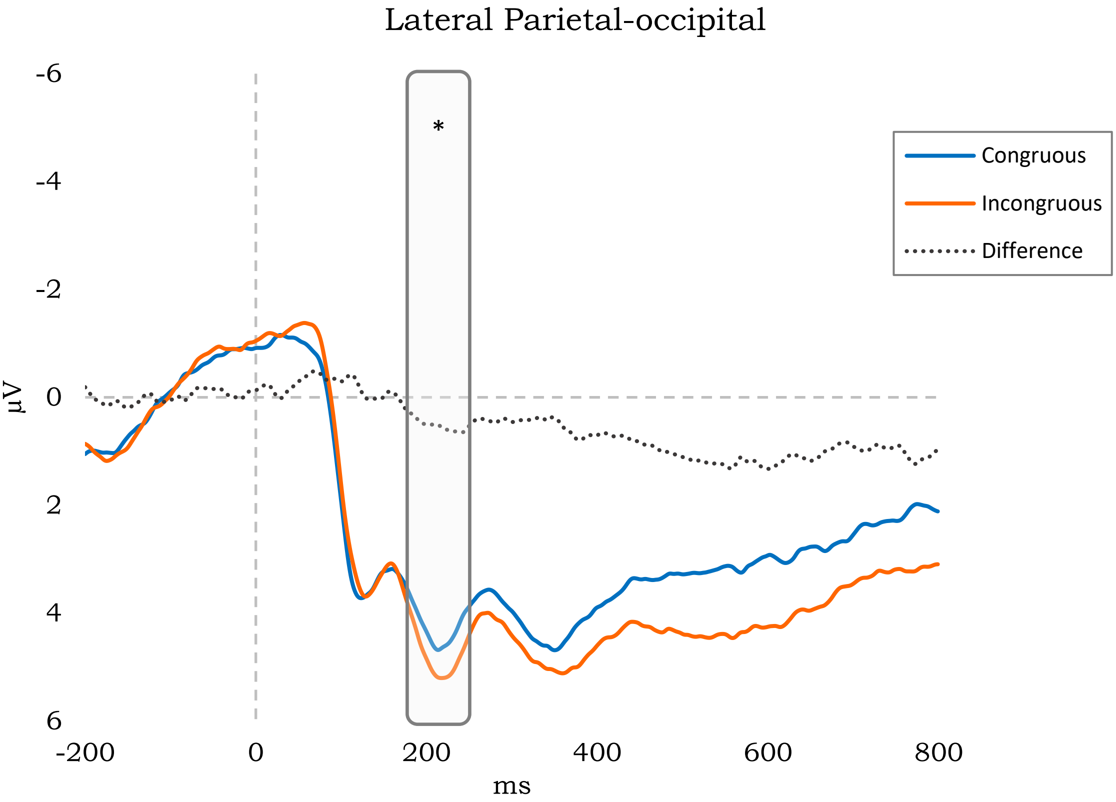
Note. FT9 was removed from the cap and placed on the left cheekbone to monitor blinks.

Analysis was conducted on the mean amplitudes for the same time-period as before (175-250 ms) using a 2 (Hemisphere: Left; Right) × 2 (Congruency: Congruous; Incongruous) repeated-measures ANOVA. This revealed a main effect of Congruency, *F*(1, 23) = 5.81, *p* = .024, ηp^2 = .20, with more positive amplitudes for Incongruous (*M* = 4.78 µV) than Congruous (*M* = 4.26 µV) trials.

The Hemisphere × Congruency interaction did not reach significance (*p* = .274). See Figure D2 for grand averaged ERPs.

Figure C2

Grand-averaged ERPs for the Lateral Parieto-occipital Region, Collapsed Across Hemispheres



Note. Blue lines represent amplitudes for Congruous trials and orange lines represent amplitudes for Incongruous trials. Dotted line represents the difference wave (Incongruous minus Congruous). Waveforms low-pass filtered at 30Hz for display purposes (*n* = 24). Grey box represents the time-window of interest. * denotes *p* < .05.