Predicting faces and houses: Category-specific visual action-effect prediction modulates late stages of sensory processing.

Gethin Hughes[1] and Florian Waszak[2,3].

1 Department of Psychology, University of Essex, UK.

2 Université Paris Descartes, Sorbonne Paris Cité, Paris, France

3 CNRS (Laboratoire Psychologie de la Perception, UMR 8158), Paris, France

Address for correspondence:

Gethin Hughes
Department of Psychology
University of Essex
Wivenhoe Park
Colchester, CO4 3SQ, UK
E-mail: hughes.gethin@gmail.com

Abstract

Our perception is fundamentally influenced by the way that we interact with the world. In particular, sensory events that are consistent with our planned actions are attenuated, both in terms of their phenomenology, and their neural response. Previous research in this domain has focused on simple-featured stimuli such as Gabor patches or sine wave tones, with attenuation normally occurring at early stages of sensory processing. In the current study we investigated this phenomenon using more ecologically valid stimuli that would likely involve higher-level visual predictions. More specifically, we trained participants to associate different actions with the presentation of a face or a house. By recording ERPs we could utilise the modularity of face processing to determine the locus of sensory attenuation for these high-level stimuli, as well as identify content-specific brain activity related to the prediction itself. In contrast to previous studies using low-level stimuli, we observed attenuation at later stages of visual processing, suggesting that higher-level predictions result in high-level prediction errors. We additionally observed significant differences over visual brain regions during action preparation dependent on whether participants were predicting to see a house or a face, perhaps reflecting preactivation of the predicted action effects. Furthermore, the degree to which participants showed evidence of preactivation, was correlated with the magnitude of their P2 attenuation. Taken together, these findings provide new insight into motor prediction and its influence on perception.

1 - Introduction

As part of everyday behaviour, humans learn to associate particular actions with specific sensory consequences. For example, one learns to associate different buttons on a remote control with different functions of the television set. Such ideomotor learning is important not only for action selection (Harless, 1861; James, 1890; Lotze, 1852; Shin, Proctor, & Capaldi, 2010) but also influences how we perceive the world around us (Waszak, Cardoso-Leite, & Hughes, 2012). Previous research has found that sensory events stemming from our actions are subject to sensory attenuation, investigated both using psychophysical (Cardoso-Leite, Mamassian, Schutz-Bosbach, & Waszak, 2010; Roussel, Hughes, & Waszak, 2013; Sato, 2008) and physiological methodologies (Baess, Widmann, Roye, Schroger, & Jacobsen, 2009; Blakemore, Wolpert, & Frith, 1998; Hughes, Desantis, & Waszak, 2013a; Hughes & Waszak, 2011). Sensory attenuation refers to the phenomenon whereby sensory action effects are attenuated as compared to the same stimuli presented independent of action (for recent reviews see Hughes, Desantis, & Waszak, 2013b; Waszak et al., 2012). A number of recent experiments have extended this methodology to investigate the extent to which the prediction of the identity of the stimulus influences sensory processing (Hughes et al., 2013a; Jones, Hughes, & Waszak, 2013; Roussel et al., 2013). These experiments revealed that action-effects that are consistent with the prediction generated by the chosen action are attenuated. This has been observed both using psychophysical approaches (Cardoso-Leite et al., 2010; Roussel et al., 2013) and ERP measures (Hughes et al., 2013a; Roussel, Hughes, & Waszak, 2014). For example, Cardoso-Leite et al. (2012) showed that detection sensitivity is reduced for congruent stimuli, while Roussel et al. (2013) showed that contrast sensitivity is reduced for predicted stimuli.

Importantly, in all these previous studies action effects were stimuli similar to those typically used in psychophysical experiments; sine wave tones, Gabor stimuli, or simple tactile stimuli. Given that these stimuli are very simple-featured input basically void of any semantic content, the internal prediction of the action effects in these experiments could also only concern low-level features. As such, any differences related to action-effect prediction would likely be present in low-level perceptual stages. Indeed, studies using EEG — most important in the present context — revealed that relatively early components of perceptual processing are attenuated (Baess et al., 2009; Hughes et al., 2013a). To give one example, Hughes et al. (2013a) showed that tones whose frequency was congruent with hand-specific action selection, were associated with a smaller auditory N1 response.

The current EEG study aimed to broaden the scope of the search for the locus of neurophysiological sensory attenuation. It does so by comparing ecologically valid, complex visual stimuli that have a well characterised differences in ERP scalp potentials – namely faces (Bentin, Allison, Puce, Perez, & McCarthy, 1996) and houses. As will be outlined below in more detail; this allowed us to assess sensory neurophysiological attenuation on a wide range of perceptual processing, from lower-level perceptual analysis to, more importantly, the creation of a higher-level perceptual representation. As with almost all previous studies on sensory attenuation, we did not include a behavioural index, but rather posited attenuation based on a reduced neurophysiological response. Nonetheless, it is important to maintain the distinction between behavioural and EEG indicies of this phenomenon. As such we refer to our measure as one of neurophysiological attenuation, to distinguish it from behavioural sensory attenuation (see also Roussel et al., 2014). A secondary aim of the study was to extract ERP differences during the preparation of a motor act predictive of a house or a face.

A wealth of experimental data exists on the different neural responses to faces as compared to other stimuli (for a recent review seeKanwisher & Yovel, 2006), using both brain imaging (Kanwisher, McDermott, & Chun, 1997) and ERP methodologies (Bentin et al., 1996). Most relevant to the current study, is the N170 ERP component (Bentin et al., 1996). This component is typically observed as a difference between face stimuli, and a vast array of different stimuli such as houses, hands or inverted faces (Eimer, 2000). This component, typically observed at bilateral parietal electrodes (with a polarity reversal over central electrodes), is thought to reflect some aspect of configural processing of faces, such that it is associated with the process of integrating different features of a face (Eimer, 2000; Eimer, Gosling, Nicholas, & Kiss, 2011). In addition to the N170, early visual ERP responses are also sometimes seen to differ between faces and other stimuli (Itier & Taylor, 2004) as well as between emotional and non-emotional faces (Eimer & Holmes, 2007; Kanwisher & Yovel, 2006) . Later modulations are often also reported associated with more reflexive aspect of face processing such as face familiarity (Gosling & Eimer, 2011), or conscious processing of emotional content (Eimer & Holmes, 2002). The current study will investigate how motor prediction might modulate these various components.

We might expect neurophysiological sensory attenuation to be evident only in the very early stages of visual processing. In this case, we would expect to observe modulation of the visual P1 component as a function of stimulus predictability, but no modulation of the N170 component. Another possibility is that prediction influences both the early (P1), and content specific (N170) components. Finally, it might also be the case that we find no modulation of P1 or N170 components, but instead find a modulation of later components, suggesting that motor prediction of houses and faces influences rather later stages of stimulus processing, such as conscious appraisal of

the stimulus. Of course, it is also possible that we find neurophysiological attenuation in a combination or all of these stages.

In additional to modulating these sensory components, we should expect incongruent stimuli to activate mismatch components such as the action-effect negativity, a frontocentral component occurring around 200ms after an unexpected or incongruent action-effect (Band, van Steenbergen, Ridderinkhof, Falkenstein, & Hommel, 2009).

Additionally, by utilising visual stimuli that are known to produce distinctive ERP responses, we could also look for evidence that participants pre-activate the representation of the specific stimulus, while participants were preparing their action (in line with Waszak et al., 2012). Kuhn, Seurinck, Fias, and Waszak (2010) previously used a similar paradigm with fMRI and found that once participants had associated left- and right-hand actions with faces and houses, respectively, they showed activation of the fusiform face area and parahippocampal place area while performing these actions, even when no stimuli were presented. This was taken as evidence to suggest that the preparation of an action automatically involves a representation of the effects of that action, in line with ideomotor theory (see Shin et al., 2010). Nontheless, given the poor temporal resolution of the fMRI signal, this result tells us little about whether the house/face representation was truly pre-activated, or whether it was activated after the response was made. That is, did participants activate content specific brain regions already during action preparation, or did this activity rather occur at the moment when the predicted stimulus was expected to appear. This experiment, thus also aimed to clarify the time course of content-specific motor prediction.

2 - Methods

2.1 - *Participants*

Data was collected from 24 paid volunteers. All participants were right-handed, had normal or corrected-to-normal vision, and had no history of epilepsy. Individuals who were taking psychoactive drugs were excluded from participating. Two participants were excluded due to technical problems with the EEG recording, leaving 17 female and 5 male participants, with a mean age of 22 years and 7 months (range 19 to 29 yrs). All participants signed an informed consent prior to the experimental session, and were free to withdraw at any point.

2.2 - *Experimental Procedure*

The experiment was conducted in a dimly lit sound attenuated room. The experiment was conducted using Matlab (MathWorks) with the psychophysics toolbox (Brainard, 1997), with stimuli presented on a 21 inch monitor (60 cm from the participant). The experimental session began with a series of practice tasks, which were designed to familiarize participants with the task as well as to allow participants to build up the action-effect contingencies. In the first practice task (association phase) participants were asked to press either the z, or the m key on a keyboard, with their left or right hand, respectively. They were free to choose which button to press on each trial, and when to press the button, with the exception that they should not perform the action until at least 500 ms after the start of the trial (the onset of the fixation cross). If participants pressed too quickly then a red fixation cross would appear for 200 ms, before the trial was restarted (following a random inter trial interval and then a white fixation cross). Each valid button press triggered the presentation of a house or a face, with faces being presented following actions with one hand and houses following actions with the other (counterbalanced across participants). The stimulus was presented with a delay of 200 ms. 10 different house

and 10 different face stimuli were used (see Figure 1 for example images), with each stimulus being presented an equal number of times in each block. The stimuli were presented in the centre of the screen and measured 6.5 x 4.5 cm. Participants completed 30 trials of this practice task. On 10% of trials, the stimulus was presented upside down. Participants were required to press the space bar as quickly as possible in these instances. These trials were included to encourage participants to pay attention to the visual stimuli. They were not included in the analysis of the EEG data. At the end of each block the experimenter checked to confirm that participants were performing well (above 90%) on this task. Each trial was separated by a inter stimulus interval randomly selected between 1 and 1.5 seconds.

In the second practice task (memory phase) participants were presented with a sequence of houses and faces and asked to memorise the sequence, and recreate it by pressing the appropriate buttons on the keyboard. The first sequence contained 4 stimuli (2 houses and 2 faces), presented in a random order. Following a correct sequence 1 stimulus was added to the newly randomised sequence (up to a maximum of 7 stimuli). If participants failed to successfully replicate the sequence, 1 stimulus was removed (down to a minimum of 3 stimuli). Participants received feedback after each response sequence to inform them of their accuracy. This practice task contained a total of 10 sequences. If participants failed to replicate at least 6 out of the 10 sequences, they were asked to repeat the task to ensure that they had adequately learnt the action-stimulus contingencies.

Following these practice tasks, the participants were fitted with the EEG recording apparatus. They then completed 6 sequences of the association phase, the memory phase, and the test phase. The association phase consisted of 20 trials, and the memory phase of 7 trials. The test phase (100 trials) was identical to the association phase, except that on an average of 30 trials, the action-effect relationship was violated. For instance, if a participant had learnt that left hand actions always led

8

to faces, then on 30% of trials in the test phase left hand actions would result in the presentation of a house rather than a face. These will henceforth be referred to as incongruent trials, with consistent mappings referred to as congruent trials. As with the association block, 10% of trials were catch trials, requiring a response with the space bar. Catch trials fell randomly on any of the congruent trials. Including the association phase and memory phase between each test phase block was meant to ensure that participants maintained their learnt action-effect association. See Figure 1 for the trial timeline during the test phase.

2.3 - *EEG recording and data analysis*

Electroencephalography (EEG) data were recorded from 64 scalp locations using a Neuroscan Synamps system using the modified combinatorial nomenclature electrode placements. The EEG was digitized at 500 Hz with a 0.01 - 30 Hz band-pass filter, referenced to FCz. EEG analysis was conducted using EEGlab (Delorme & Makeig, 2004) and custom-built Matlab scripts. The data were re-sampled offline to a 250 Hz sample rate. Epochs were generated from -3000 to 1000 ms relative to the onset of the visual stimulus (face/house), with action onset at -200 ms. This long epoch was used to ensure that the epochs included the beginning of the trial to allow for baseline correction. Each epoch was baseline corrected by subtracting the average activity from the 100 ms prior to the start of the trial (the onset of the fixation cross). This baseline was used to ensure that the epoch was not contaminated by possible differences in motor potentials between our 2 conditions.

Initial artifact rejection was conducted in a semi-automatic manner (in EEGlab) by rejecting epochs where activity at any time point for any electrode was more that 5 standard deviations from the mean activity for that epoch. Any channels that contributed to the rejection of many epochs were considered for removal and later

interpolation. Ocular artifact correction was conducted in EEGlab in Matlab using independent component analysis (Delorme & Makeig, 2004). Following removal of eye blinks and eye movements, noisy channels were replaced by an interpolated weighted average from surrounding electrodes. A mean of 0.41 (maximum of 2) channels were interpolated for each participant. Data were then rereferenced to the common average. A final round of automatic artifact rejection with a threshold of +/- 250 μv was used to remove any remaining artifacts.

Data analysis focused on 4 regions of interest that were defined apriori. We included a frontocentral ROI (electrodes F1, Fz, F2, FC1, FCZ, FC2) to capture the Action-effect negativity (Band et al., 2009). The N170 was also investigated at left and right parietal electrodes (P7, P5, PO7, PO5 for left and P6, P8, PO6, PO8 for right). To ensure the reliability of results in the N170 period, we conducted additional analysis focusing only on electrodes P7 and P8 (equivalent to T5 and T6 on the 10-20 EEG system), where the N170 is known to peak (Bentin et al., 1996; Eimer, 2000). To capture the early visual components, and the later P2 component we also included a ROI over occipital leads (electrodes PO3, POz, PO3, O1, Oz, O2). Statistical anlaysis of post-stimulus ERP was conducted using ANOVA with 2 repeated measures factors: prediction (Face or House) and congruency (congruent or incongruent), with the average amplitude over each ROI entered as the dependent variable. Analysis was conducted in 3 time windows. An early time window focusing on the P1 (90 to 130 ms), a second time window focusing on the N170 (130 to 200 ms) and a third time window focusing on the P2 and AEN (200 to 260 ms).

Since we should expect pre-activation of the face specific brain regions to occur over similar electrodes, we included the same ROIs in our analysis of the pre-stimlus period. For this analysis, we generated averages of all trials where participants were predicting a face versus a house. Analysis was conducted using *t*-tests in 100 ms time windows from -1000 to -200 ms (800 ms prior to action onset).

3 - Results

   Initial analysis focused on the period prior to stimulus onset (see Figure 2). Analysis of the frontocentral ROI revealed no difference between predict face and predict house in any of the time windows. However, over occipital electrodes, 2 time windows revealed a significant effect of prediction (-500 – -400 ms; $t(21) = 2.18$; $p < .05$ and -400 – -300 ms; $t(21) = 2.18$; $p < .05$). In each of these time windows, trials where participants were performing actions associated with the presentation of a face showed more positive amplitude than trials where the action predicted a house (-500 – -400 ms: Face mean = .42 μv, std = 1.3; House mean = .086 μv, std = 1.3; -400 – -300 ms: Face mean = .42 μv, std = 1.3; House mean= .086 μv, std = 1.3). Similarly, over the left temporal parietal ROI, there was a significant effect of prediction from -600 ms to -200 ms; ($ts(1,21) = 2.39$; 2.21; 3.1; 2.12; $ps < .05$), with significantly more positive amplitude for face predicted than house predicted trials. The topography of this effect is shown in Figure 2. This topography reveals some lateralised differences, likely individual differences in motor potentials that do not completely cancel out from the counterbalancing of hand-stimulus pairings. However, a clear occipital difference is also observed, showing that occipital electrodes appear to pick up stimulus specific information regarding whether the participant is predicting a face or a house. In order to confirm that the differences observed at occipital electrodes did not differ between participants with different hand-stimulus mappings, we conducted an additional ANOVA where we included response mapping as an independent subjects factor. This analysis revealed no significant effects involving this variable (F < 1). This confirms that this effect was not related to motor potentials, but rather reflects differences in prediction of houses and faces over posterior electrode sites.

   Analysis of the post-stimulus ERP focused on the same 4 regions of interest in 3 different time windows. In the first time window (P1, 90-130 ms) we observed a significant main effect of stimulus type over occipital electrodes only ($F(1,21) =$

13.2; *p* < .01), with house trials (mean = 4.2 std = 2.29) being significantly more positive than face trials (mean = 3.5; std = 1.95). The topography of this P1 difference is in Figure 3, confirming an occipital difference between houses and faces. There was no effect of congruency in this time window, and no significant interactions.

In the second time window (N170, 130 – 200 ms) there was a significant main effect of stimulus type over frontocentral and parietal ROIs. Over frontocentral electrodes (*F* (1,21) = 8.17; *p* < .01), face stimuli were significantly more positive than house stimuli. This pattern was reversed over both left (*F* (1,21) = 31.7; *p* < .001) and right (*F* (1,21) = 88.3; *p* < .001) parietal regions where face stimuli (left = -1.71 µv; right = -1.44 µv) showed more negative amplitude than house stimuli (left = -0.12 µv; right = -0.36 µv). The topography of this difference can be observed in Figure 3, and reflects the N170 difference, as observed in numerous previous EEG studies comparing houses and faces (see Kanwisher and Yovel, 2006). As with the first time window, there was no effect of congruency in this time window, and no significant interactions.

To further examine the potential modulation of the N170, we compared peak amplitude in the N170 time range (100 to 240 ms) across our conditions at electrodes P7 and P8. This revealed a significant main effect of condition for both electrodes (*F* (1,21) = 46.6; *p* < .001; and *F* (1,21) = 416.6; *p* < .001), and a trend towards more negative amplitude for congruent trials (*F* (1,21) = 3.19; *p* < .1; and *F* (1,21) = 3.60; *p* < .001). We further investigated (at P7 and P8) the average amplitude in the 32 ms around the peak of the N170 (164 to 196 ms), with additional analysis focusing on the rising aspect of the peak (164 to 180 ms) and the falling aspect (180 ms to 196 ms). These analyses all revealed significant main effects of condition at both electrodes (at *p* < .001), confirming a strong N170 difference between faces and houses. While analysis on the rising peak showed no trends for the congruency factor (*F* (1,21) = 1.65; *p* = .21; and *F* (1,21) < 1), the falling bank showed a very near significant effect

for electrode P7 (F (1,21) = 4.07; p = .057). These analyses confirm that there were no clear effects of congruency in the N170 time range, and that the trends that were observed were more towards the falling bank of the N170 and therefore most likely reflect a modulation of the subsequent ERP component (described below).

In the third and final time window (200 to 260 ms) we observed a significant main effect of congruency at all four regions of interest, but no significant main effects of stimulus type and no interactions. The topography of this P2 difference between congruent and incongruent trials can be seen in Figure 3. Over the frontocentral ROI (F (1,21) = 8.17; p < .01), incongruent trials (mean = -1.95 μv; std = 1.69 μv) showed more negative amplitude compared to congruent trials (mean = -1.6 μv; std = 1.65 μv). This pattern was reversed over the occipital ROI (F (1,21) = 6.26; p < .05), with more positive amplitude for incongruent (mean = 5.1 μv; std = 2.66 μv) trials, compared to congruent (mean = 4.7 μv; std = 2.67 μv) trials. A similar pattern was also observed over left (F (1,21) = 6.5; p < .05) and right (F (1,21) = 5.71; p < .01) temporal parietal ROIs, with incongruent trials (left = 1.94 μv; right = 3.14 μv) showing more positive amplitude than congruent trials (left = 1.64 μv; right = 2.85 μv).

Our final analysis focused on the relationship between the effects prior to stimulus onset and the congruency effects observed in response to the stimulus presentation. If the prestimulus effects reflect prediction of the sensory consequences of the action then they should correlate with the subsequent ERP differences related to prediction congruency. To investigate this we correlated the magnitude of the pre-stimulus effects over occipital electrodes, with the magnitude of the congruency effects across participants. We calculated the magnitude of congruency effects by taking the combined difference between congruent and incongruent trials for both faces and houses in the P2 time range at each ROI. Positive correlations would suggest that participants who make stronger predictions (as indexed by the difference

13

between faces and houses prior to stimulus onset) show greater congruency effects. In line with this hypothesis we found that the prediction effect over occipital electrodes correlated with congruency effects at occipital and left parietal electrodes. More specifically the prediction effect over occipital electrodes from -500 to -400 ms was seen to significantly correlate with the congruency effect over left parietal electrodes (r (20) = .53, p < .05), while the prediction effect from -400 to -300 ms was significantly correlated with the congruency effect at both occipital (r (20) = .45, p < .05) and left parietal (r (20) = .51, p < .05) ROIs. To ensure that these correlations were specific to the congruency effect and did not simply reflect the fact that some participants had larger ERP effects, we also correlated the prediction effect with the N170 face/house difference, revealing no significant correlations (all p values > .3). Thus, these analyses point towards an association between our pre-stimulus prediction effects, and our post-stimulus neurophysiological attenuation effects.

4 - Discussion

The current study aimed to address two important questions in relation to motor prediction and action-effect processing. Firstly, we were interested in determining whether neurophysiological sensory attenuation was manifested in early and/or late ERP components associated with the different steps in the processing of ecologically valid, semantically meaningful effect stimuli. Secondly, we were interested in investigating the time course of the predictive processes themselves. To test these predictions, we trained participants to associate left- and right-hand button presses differentially with face and house stimuli. We found no modulation of the P1, or N170 components as a function of motor prediction, but rather in later stimulus processing from 200ms after stimulus onset. We further observed significant differences over posterior electrodes while participants prepared actions that were associated with faces and houses, which could reflect the preacitvation of content

specific brain regions associated with the stimulus. This possibility is supported by the observation that the magnitude of this preactivation effect was correlated with the size of the congruency effect observed following stimulus presentation.

Our analysis of the post-stimulus ERPs revealed significant attenuation in the P2 time range over occipital electrodes, for congruent as compared to incongruent stimuli. We also observed differences in earlier P1 and N170 peaks. However, these components only varied as a function of stimulus type, but not as a function of congruency. This suggests that, in the current study, neurophysiological attenuation occurred on later, more general visual processes. One might argue that in fact our congruency modulation began already in the N170 time window, but only reached statistical significance in the P2 time window. Our detailed analysis of the N170 suggests that any modulation in the N170 component appeared only in the falling bank of the N170. Furthermore, while the N170 topography showed a temporal parietal topography, the congruency modulation included significant differences observed over occipital leads. Thus, the most parsimonious explanation for our data seems to be that congruency modulates the P2 component, not the N170 component.

A further possibility is that the absence of P1 and N170 modulations as a function of congruency in the current experiment could reflect the fact that participants unlearnt the particular association during the test phase, as this includes some incongruent trials. However, we have strong reasons to believe this not to be the case. Firstly, in previous studies using a very similar paradigm we have observed significant behavioural (Cardoso-Leite et al., 2010; Roussel et al., 2013) and electrophysiological (Hughes et al., 2013a) markers of sensory attenuation. Indeed, in each of those experiments, sensory attenuation was observed where participants were presented with completely unpredictable stimuli (50% congruent) in the test-phase. In the current experiment, actions were somewhat predictive of the stimulus (70% congruent) even during the test phase, making unlearning even less likely in the

current experiment. Furthermore, it is important to note that other than the learnt associations between the actions and the stimulus class, congruent and incongruent trials in the current experiment were essentially identical. If participants had unlearnt the associations, then we would have been unlikely to observe any congruency-related modulations in our ERPs.

Previous ERP research has suggested that activity in the time regions where we observed our congruency effect appears to be modulated by characteristics such as conscious evaluation of emotional aspects of face stimuli (Eimer & Holmes, 2002), or face recognition processes (Gosling & Eimer, 2011). The neurophysiological attenuation effect observed in the current study, thus, occurs at a more reflective level of face processing, rather than the early recognition of a face. As such, in the current study, while later processing is reduced for predicted stimuli, this does not reduce the differential ERP signatures of faces and houses, and therefore does not reduce the brain's ability to distinguish a face from a house. This is also consistent with the observation that while participants appear to have reduced stimulus sensitivity for predicted stimuli (Cardoso-Leite et al., 2010), they do not show decreased stimulus discrimination (Kok, Jehee, & de Lange, 2012).

Our finding seems to be in contrast to neurophysiological sensory attenuation in the auditory domain, which is commonly reported over the primary auditory N1 response (Baess, Horvath, Jacobsen, & Schroger, 2011; Baess, Jacobsen, & Schroger, 2008; Hughes et al., 2013a, 2013b). They are at a first glance also in contrast to a recent study which showed visual sensory attenuation over the same component that was modulated by stimulus contrast (Roussel et al., 2014). These studies suggest a rather early locus of modulation. We see two possibilities to reconcile the different patterns of result, both related to the type of stimuli used as action effects in the different studies. First, the fact that fast processing of faces is evolutionarily important for humans, as highly social animals, may mean that such early aspects of face

processing are spared from modulation from motor predictions. Second, one might speculate that neurophysiological sensory attenuation depends on the level on which effect stimuli are represented. In the previous experiments, effect stimuli involved only very simple features. In the current experiment, by contrast, participants' actions triggered presentation of complex, semantically meaningful stimuli. It might be that action effect anticipation takes place on the highest possible level of representation. Accordingly, novel, semantically meaningless stimuli are anticipated on a rather low–level of representation, whereas semantically meaningful stimuli are anticipated on a higher level. Future research should attempt to investigate this issue further by comparing different types of stimuli and different paradigms, to further determine the stage of processing involved in visual sensory attenuation.

In the present study, since participants were presented with 10 different possible faces, precise low-level predictions might not be possible (although note that faces share many low-level features). A recent study by SanMiguel, Saupe, and Schroger (2013) suggests that in situations where a general prediction is made (i.e. that a sound will be presented), but not a specific prediction (the identity of the sound), ERPs to the omission of the tone do not result in an auditory omission response (cf.SanMiguel, Widmann, Bendixen, Trujillo-Barreto, & Schroger, 2013). They suggest that the absence of an omission response in this condition reflects the fact that precise physical characteristics of the sound are unknown to the participants. This finding further supports our conclusion that the late neurophysiological attenuation effects observed in the present study result from the fact that participants predictions involved higher-level stimulus features than those reported in previous studies using this paradigm (Cardoso-Leite et al., 2010; Hughes et al., 2013a; Roussel et al., 2013, 2014).

Another intriguing possibility is that the differences observed in the P2 time range do not reflect a process of prediction matching (i.e. sensory attenuation), but

rather that they reflect a mismatch negativity (MMN) for unpredicted visual stimuli. While the majority of research on MMN has focused on auditory stimuli, a number of recent studies have identified a posterior component (vMMN) somewhere between 100 and 300 ms after the onset of a visual oddball (for a recent review see Winkler & Czigler, 2012). As such, the modulations observed in the current experiment could provide further evidence of mismatch-related activity in the visual domain. Distinguishing between sensory attenuation and sensory mismatch processes, in both visual (including the current study) and auditory experiments in this field is difficult, since comparing expected and unexpected stimuli will naturally entail both these processes. Indeed, predictive coding accounts consider mismatch (or prediction error) to be a building block of perception, thus blurring the distinction between reduced early sensory processing, and later mismatch related activity (see Clark, 2013).

In addition to this posterior modulation, we observed a significant negativity over frontocentral electrodes for incongruent compared to congruent trials. This modulation is consistent with the action-effect negativity (AEN), which is thought to reflect performance monitoring processes coming from anterior cingulate cortex (Band et al., 2009). Similar modulations have also been previously observed in sensory attenuation paradigms (Hughes & Waszak, 2011; Jones et al., 2013).

Turning our attention now to the pre-stimulus ERP results, the observation that actions that lead to the prediction of a face were associated with a different ERP pattern to actions that lead to a house is consistent with the idea that the selection of an action includes a representation of the effects of that action (Hommel, Müsseler, Aschersleben, & Prinz, 2001). This finding builds on previous research (Kuhn et al., 2010) showing that even when action-effects are no longer presented participants nonetheless activate face and house specific brain regions predicted by their chosen action. The current data extends that finding, by showing that these content specific regions are activated already several hundred milliseconds before participants perform

their voluntary action. The presence of differences between predicted houses and faces in advance of the presentation of the action-effect itself, is also consistent with a recent account of the possible mechanism of sensory attenuation, whereby attenuation results from a preactivation of prediction sensory action-effects (Waszak et al., 2012). Under such an account, attenuation of predicted action-effects comes about due to an increase in baseline activation driven by the prediction. The resulting stimulus-driven increase in signal is therefore lower, compared to a situation where no preactivation occurs. As such, sensory attenuation may rely on similar neural mechanisms to those involved repetition suppression (for a review see Grill-Spector, Henson, & Martin, 2006), whereby sensory neurons are fatigued followed prediction-related activation, or preactivation (see Waszak et al., 2012).

Some previous studies have begun to investigate similar prediction related EEG activity in a self-triggered tone task (SanMiguel, Saupe, et al., 2013; SanMiguel, Widmann, et al., 2013). However, these studies only observed differences related to motor prediction either at, or soon after action onset. SanMiguel, Widmann, et al. (2013) showed that even when expected tones are omitted, an auditory N1 component is observed. The authors suggest that this reflects preactivation of the expected action effect, despite its non-appearance. In another study (SanMiguel, Saupe et al., 2013) the authors observed ERP differences from -20 to 40 ms relative to action onset in a similar sound omission paradigm. The authors suggest that this activation might reflect the generation of a motor efference copy, rather than content-specific auditory predictions. The current study extends these findings to show differences related to the content of the motor prediction, several hundred milliseconds before action onset.

In a previous study (Hughes & Waszak, 2011), we investigated the difference between trials where participants predicted a stimulus to occur, and trials where no stimulus was predicted. Unlike in the present study, we found no significant differences over occipital electrodes in the action preparation period. The fact that we

observed significant differences between the two classes of stimuli in the present study suggests that asking participants to prepare actions associated with distinct stimuli is a more effective approach for localising ERP signatures of sensory predictions. This approach may serve to highlight the prediction of a particular type of stimulus, while in a context where participants predict the presence or absence of a single class of stimulus, both actions might elicit a representation of the relevant stimulus (Hughes et al., 2011).

The finding that in the current study participants who showed greater pre-stimulus ERP differences also showed greater congruency modulations, provides further evidence that neurophysiological sensory attenuation results from stimulus-specific predictions made during action preparation, that likely involve preactivation of predicted action effects. Note that since several different stimuli were used from each category, participants would not be able to predict, or preactivate the precise identity of the stimulus but rather they would predict the specific category (face or house). Importantly, the fact that these correlations were specific to the congruency modulation strongly suggests that this effect reflects prediction related activity, rather than simply differences in visual attention, which would likely result in larger ERP effects on all aspects of stimulus processing.

In summary, the current study utilised the modularity of face processing to investigate the time course of action-effect prediction, and content specific sensory processing of predicted and unpredicted action effects. We observed significant differences over posterior electrodes during action preparation that differed dependent on whether participants were predicting to see a house or a face. This difference could reflect an early correlate of the preactivation of the predicted sensory consequences of one's action (Waszak et al., 2012). Later processing of predicted and unpredicted faces differed from around 200ms after stimulus onset, and was independent of content-specific ERP modulations, suggesting that when motor predictions are

associated with higher level visual features, it is later visual processing stages that are modulated by these predictions.

Author Notes

References

Baess, P., Horvath, J., Jacobsen, T., & Schroger, E. (2011). Selective suppression of self-initiated sounds in an auditory stream: An ERP study. *Psychophysiology, 48*(9), 1276-1283.

Baess, P., Jacobsen, T., & Schroger, E. (2008). Suppression of the auditory N1 event-related potential component with unpredictable self-initiated tones: evidence for internal forward models with dynamic stimulation. *Int J Psychophysiol, 70*(2), 137-143.

Baess, P., Widmann, A., Roye, A., Schroger, E., & Jacobsen, T. (2009). Attenuated human auditory middle latency response and evoked 40-Hz response to self-initiated sounds. *Eur J Neurosci, 29*(7), 1514-1521.

Band, G. P., van Steenbergen, H., Ridderinkhof, K. R., Falkenstein, M., & Hommel, B. (2009). Action-effect negativity: irrelevant action effects are monitored like relevant feedback. *Biol Psychol, 82*(3), 211-218.

Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological Studies of Face Perception in Humans. *J Cogn Neurosci, 8*(6), 551-565.

Blakemore, S. J., Wolpert, D. M., & Frith, C. D. (1998). Central cancellation of self-produced tickle sensation. *Nat Neurosci, 1*(7), 635-640.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*(4), 433-436.

Cardoso-Leite, P., Mamassian, P., Schutz-Bosbach, S., & Waszak, F. (2010). A new look at sensory attenuation: action-effect anticipation affects sensitivity, not response bias. *Psychol Sci, 21*(12), 1740-1745.

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

Eimer, M. (2000). The face-specific N170 component reflects late stages in the structural encoding of faces. *Neuroreport, 11*(10), 2319-2324.

Eimer, M., Gosling, A., Nicholas, S., & Kiss, M. (2011). The N170 component and its links to configural face processing: a rapid neural adaptation study. *Brain Res, 1376*, 76-87.

Eimer, M., & Holmes, A. (2002). An ERP study on the time course of emotional face processing. *Neuroreport, 13*(4), 427-431.

Eimer, M., & Holmes, A. (2007). Event-related brain potential correlates of emotional face processing. *Neuropsychologia, 45*(1), 15-31.

Gosling, A., & Eimer, M. (2011). An event-related brain potential study of explicit face recognition. *Neuropsychologia, 49*(9), 2736-2745.

Harless, E. (1861). Der Apparat des Willens. *Zeitschrift für Philosophie und Philosophische Kritik, 38*, 50-73.

Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The Theory of Event Coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences, 24*(5), 849-+.

Hughes, G., Desantis, A., & Waszak, F. (2013a). Attenuation of auditory N1 results from identity-specific action-effect prediction. *Eur J Neurosci, 37*(7), 1152-1158.

Hughes, G., Desantis, A., & Waszak, F. (2013b). Mechanisms of intentional binding and sensory attenuation: the role of temporal prediction, temporal control, identity prediction, and motor prediction. *Psychol Bull, 139*(1), 133-151.

Hughes, G., & Waszak, F. (2011). ERP correlates of action effect prediction and visual sensory attenuation in voluntary action. *Neuroimage, 56*(3), 1632-1640.

Itier, R. J., & Taylor, M. J. (2004). N170 or N1? Spatiotemporal Differences between Object and Face Processing Using ERPs. *Cerebral Cortex, 14*(2), 132-142.

James, W. (1890). *The principles of psychology* (Vol. 2). New York, NY: Holt.

Jones, A., Hughes, G., & Waszak, F. (2013). The interaction between attention and motor prediction. An ERP study. *Neuroimage, 83C*, 533-541.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neurosci, 17*(11), 4302-4311.

Kanwisher, N., & Yovel, G. (2006). The fusiform face area: a cortical region specialized for the perception of faces. *Philos Trans R Soc Lond B Biol Sci, 361*(1476), 2109-2128.

Kok, P., Jehee, J. F., & de Lange, F. P. (2012). Less is more: expectation sharpens representations in the primary visual cortex. *Neuron, 75*(2), 265-270.

Kuhn, S., Seurinck, R., Fias, W., & Waszak, F. (2010). The Internal Anticipation of Sensory Action Effects: When Action Induces FFA and PPA Activity. *Front Hum Neurosci, 4*, 54.

Lotze, R. H. (1852). *Medicinische Psychologie oder Physiologie der Seele [Medical psychology or physiology of the soul].* Leipzig, Germany: Weidmann.

Roussel, C., Hughes, G., & Waszak, F. (2013). A preactivation account of sensory attenuation. *Neuropsychologia*.

Roussel, C., Hughes, G., & Waszak, F. (2014). Action prediction modulates both neurophysiological and psychophysical indices of visual sensory attenuation. *Frontiers in Human Neuroscience, 8*(115).

SanMiguel, I., Saupe, K., & Schroger, E. (2013). I know what is missing here: electrophysiological prediction error signals elicited by omissions of predicted "what" but not "when". *Front Hum Neurosci, 7*, 407.

SanMiguel, I., Widmann, A., Bendixen, A., Trujillo-Barreto, N., & Schroger, E. (2013). Hearing silences: human auditory processing relies on preactivation of sound-specific brain activity patterns. *J Neurosci, 33*(20), 8633-8639.

Sato, A. (2008). Action observation modulates auditory perception of the consequence of others' actions. *Conscious Cogn, 17*(4), 1219-1227.

Shin, Y. K., Proctor, R. W., & Capaldi, E. J. (2010). A review of contemporary ideomotor theory. *Psychol Bull, 136*(6), 943-974.

Waszak, F., Cardoso-Leite, P., & Hughes, G. (2012). Action effect anticipation: neurophysiological basis and functional consequences. *Neurosci Biobehav Rev, 36*(2), 943-959.
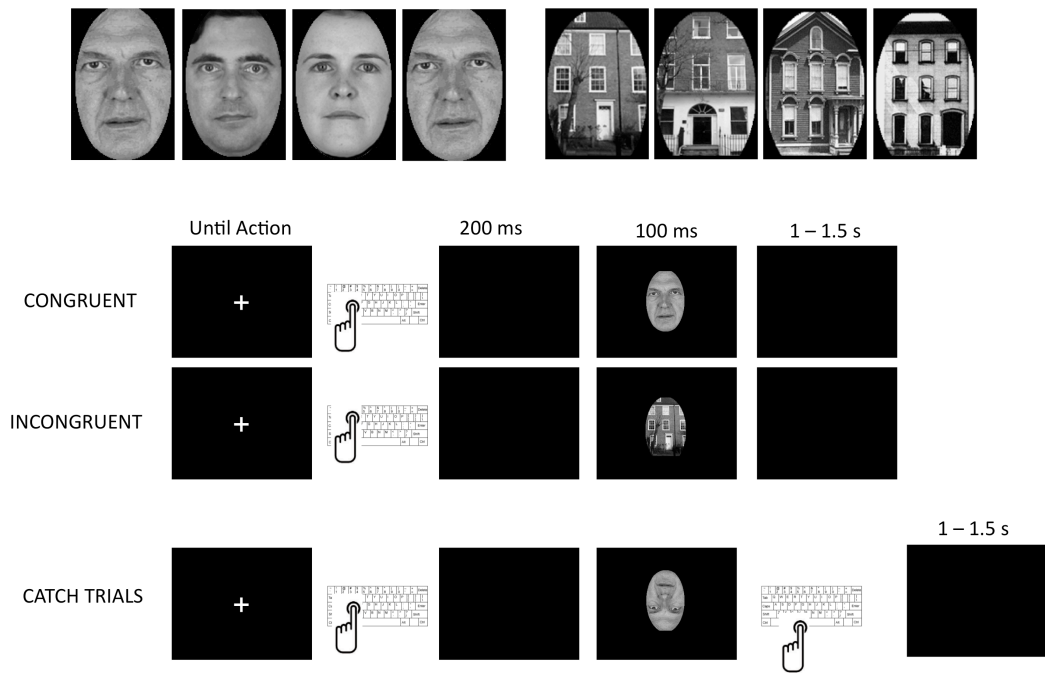
Figures:

Figure 1: Examples of the Face and House stimuli used in the current experiment. Timeline of the experimental procedure for the test phase for congruent trials, incongruent trials and catch trials.
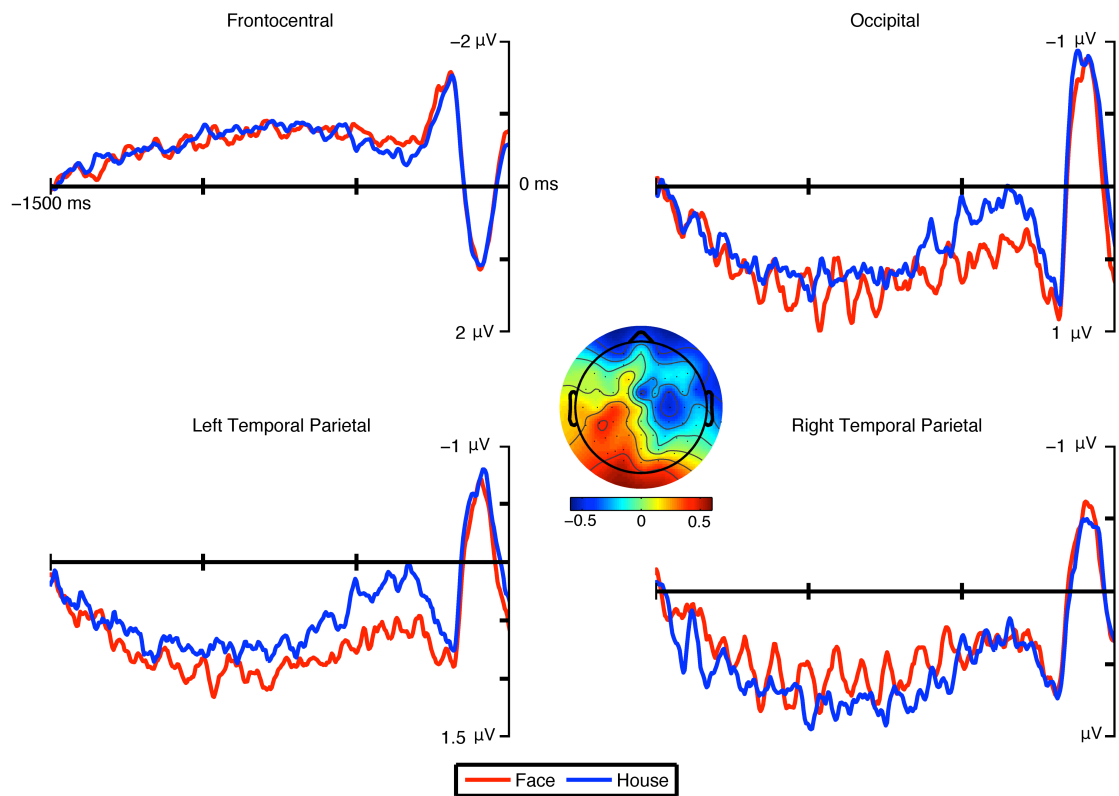
Figure 2: ERP for the four regions of interest during the action preparation period (with action onset at -200 ms). Topographic map shows average amplitude of predict Face minus predict House conditions from -600 ms to -200 ms.
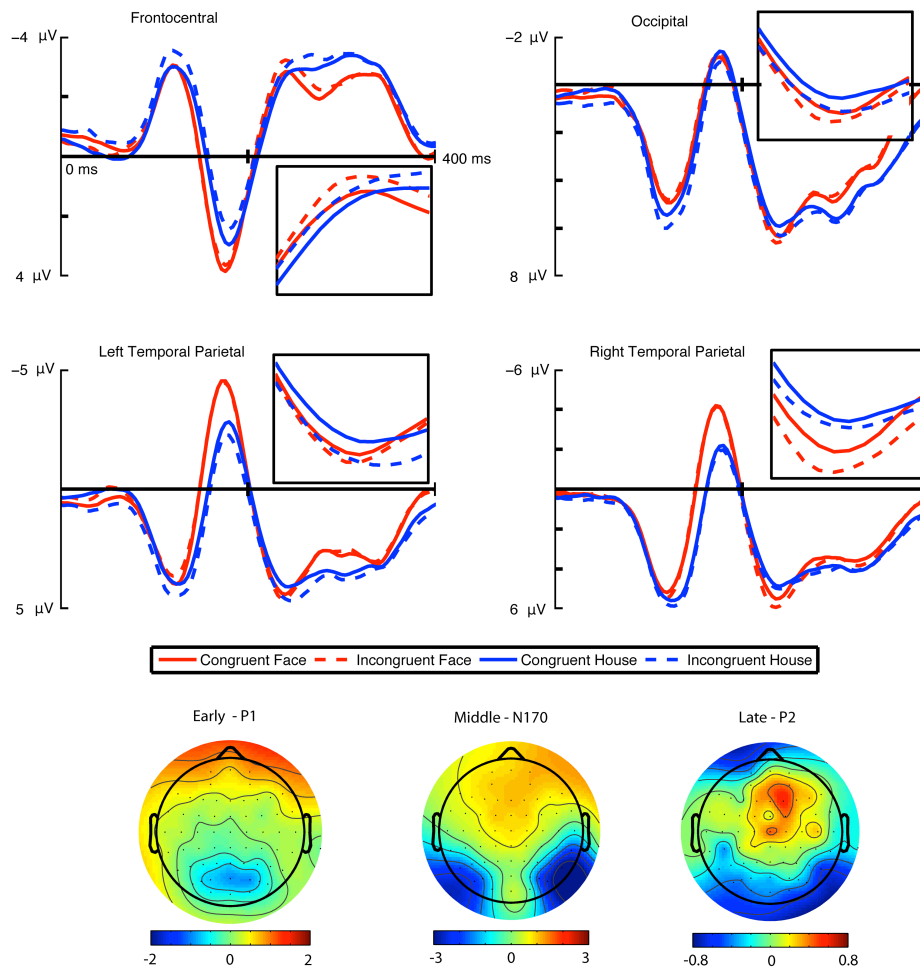
Figure 3: ERPs in the 400 ms following stimulus onset for the four regions of interest. Inlays show ERPs for the middle (P2) time window. Topographic plots for Face minus House conditions in the early and middle time windows, and for Congruent minus Incongruent stimuli in the late time window.