You *Must* See the Point: Automatic Processing of Cues to the Direction of Social Attention

Stephen R. H. Langton and Vicki Bruce

Department of Psychology, University of Stirling

Correspondence to: Dr. S. R. H. Langton, Dept. of Psychology, University of Stirling, Stirling, FK9 4LA, Scotland.

E-Mail: srhl1@stirling.ac.uk

Fax: +44 (0) 1786-467-641

Abstract

Four experiments are reported that explore the processing of pointing gestures comprising hand, head and gaze cues to direction. The cross-modal interference effect exerted by pointing hand gestures on the processing of spoken directional words, first noted by Langton et al. (1996), was found to be moderated by the orientation of the gesturer's head/gaze (Experiment 1). Hand and head cues also produced bi-directional interference effects in a within-modality version of the task (Experiment 2). These findings suggest that both head/gaze and hand cues to direction are processed automatically and in parallel up to a stage in processing where a directional decision is computed. In support of this model, head/gaze cues produced no influence on non-directional decisions to social emblematic gestures in Experiment 3, but exerted significant interference effects on directional responses to arrows in Experiment 4. It is suggested that the automatic analysis of head, gaze and pointing gestures occurs because these directional signals are processed as cues to the direction of another individual's social attention.

It has long been realised that spoken language is not the exclusive communication medium available to the human species. In addition to the voice, we make use of a variety of facial gestures in most social situations in which we find ourselves. For instance, facial expressions convey a variety of emotions and interpersonal attitudes (see Argyle, 1988), movements of the lips, teeth and tongue assist our interpretation of what is said to us (McGurk & MacDonald, 1976), and finally gaze appears to be useful in regulating turn-taking in conversation, in expressing certain attitudes such as liking or loving, and in exercising social control (see Kleinke, 1986 for a review).

Alongside facial expression, gaze and lip-configurations, spontaneous speech is also accompanied by a myriad of gestures. During speech, the hands are in a constant state of flux, either making rhythmical movements of varying amplitude, or describing particular shapes and patterns in the air. However, despite their ubiquity, there is little agreement as to the functions performed by these speech-related gestures, nor to the processes by which they are generated and understood (for reviews see Kendon, 1994; Rimé & Schiaratura, 1991). In particular, with the exception of a handful of recent studies (Langton, O'Malley & Bruce, 1996; Thompson & Massaro, 1986, 1994), gesture *comprehension* remains a "neglected field in cognitive psychology" (Feyereisen, 1991, p.57). Accordingly, this article focuses on the comprehension of a particular type of co-speech gesture, the deictic or pointing gesture.

The popular idea that gestures and other aspects of non-verbal behaviour form part of a separate "body language" carries with it the common assumption that a speaker's gestural behaviour plays a critical role in communication. However, this view is not universally accepted. Some argue that gestures are produced for the benefit of the speaker, serving an instrumental role in maintaining the fluency of speech by facilitating the retrieval of items from lexical memory (e.g., Morrel-Samuels & Krauss, 1992; Rimé & Schiaratura, 1991). It follows from this view that the information conveyed by gesture is largely redundant with speech. Accordingly, under normal circumstances, listeners are not expected to attend to, or to process a speaker's gestural

activity (Rimé & Schiaratura, 1991). On the other hand, researchers such as Kendon (1986, 1994) and McNeill (1985, 1987, 1989) are enthusiastic supporters of the idea that gestures play a critical role in the communicative process. They argue that gestural and verbal behaviours serve to represent different aspects of the underlying meaning that a speaker is striving to express. It follows that a listener must process both the gestural and the verbal components of the utterance, combining this information at some point in processing to provide an integrated representation of the speaker's intended meaning (Langton, O'Malley & Bruce, 1996; McNeill, Cassell & McCullough, 1994).

A number of studies have yielded results which are consistent with the suggestion that listeners do indeed process gestural information in comprehension (see Kendon, 1994 for a review). These studies have shown that listeners' understanding of verbal utterances are influenced by speakers' accompanying gestural performances. Langton et al. (1996) chose to explore this issue using a variant of the Stroop interference paradigm (Stroop, 1935).

The Stroop paradigm has proved useful to researchers in a variety of disciplines as a tool with which to demonstrate the automatic processing of a to-be-ignored or irrelevant aspect of a stimulus (see MacLeod, 1991 for a review). For instance, in the original version of the Stroop colour word task, participants were slower to name the colour of the ink in which an incongruent colour word was printed (e.g., *RED* in blue ink) relative to a control condition of naming colour squares. This finding clearly indicates that the to-be-ignored dimension (the colour word) nevertheless receives some form of analysis by the perceptual or cognitive system despite participants' intentions. One interpretation of this kind of "breakdown" in selective attention is that the component dimensions under scrutiny interact or exchange information in some way, and at some stage in information processing. From this perspective, the observation of an interference effect therefore entails two things; that the to-be-ignored dimension has been processed automatically, and that information from this dimension has interacted with information encoded from the target dimension.

Langton et al. (1996) placed pointing (deictic) gestures into conflict with spoken directional words and asked participants to respond to the direction of the gesture in one block of trials, and the identity of the spoken word in another block of trials. Langton et al. (1996) showed that vocal or manual responses to spoken words were influenced by to-be-ignored gestures, and reciprocally, that responses to pointing gestures were influenced by spoken words. Thus, participants' reaction times to the word "up" were slowed if they simultaneously saw a person pointing downwards compared with upwards. Similarly, when asked to respond to an upward pointing gesture, RTs were slower when participants heard a voice saying "down" compared with "up". These findings clearly indicate that participants are simply unable to ignore deictic gestures when interpreting speech. More specifically, Langton et al. (1996) suggested that these interference effects arise because information from deictic gestures and speech is integrated at some point in the comprehension process.

However, the visual gesture stimuli used by Langton et al. did not simply consist of extended arms and fingers signalling the direction of the deictic gesture. Strong directional cues were also provided by the orientation of the gesturer's head and the direction of his gaze. Thus it is possible that the interference effects exerted by deictic gestures on spoken words are the result of the automatic processing of any or all of these directional signals.

The experiments reported here once again use the Stroop-type interference paradigm in order to examine whether or not head/gaze cues are processed automatically along with the directional signals provided by the hands, and to investigate how all of these cues might contribute to the cross-modal interference effects exerted by deictic gestures on responses to spoken words. Experiment 1 examined whether the interference effect exerted by to-be-ignored hand gestures on spoken words could be moderated by the orientation of the gesturer's head. Experiment 2 employed an intra-modal version of the task to explore whether head orientation could moderate the processing of hand gestures and vice-versa. Finally Experiments 3 and 4 tested the hypothesis that the various directional cues, including gestures, head/gaze signals and arrows,

exert their effects on a stage or on several stages of processing where a directional decision is computed.

Experiment 1

In Experiment 1 we ask whether head/gaze cues are able to moderate the interference exerted by to-be-ignored pointing gestures in a cross-modal interference task similar to that used by Langton et al. (1996). On each trial, participants were presented with either the word "up" or "down" spoken by a male voice. Each word was presented together with a digitised photograph of a male individual making a pointing gesture which could either be congruent or incongruent with the meaning of the spoken word. In addition, the gesturer's head could either be oriented in the same direction as his gesture, in the opposite direction as his gesture, or straight ahead. Participants were asked to make speeded keypress responses contingent on the meaning of the spoken word and to completely ignore the visually presented information. Following Langton et al. (1996), agreeing head and gesture cues were expected to exert large interference effects on responses to spoken words. However, if the orientation of the head contributes to the interfering effect of the gesture on responses to spoken words, then this congruity effect might be expected to interact with the agreement of the head and gesture. More specifically, the effect of gestures on spoken words should be reduced when the head is angled straight ahead or in the opposite direction to that of the gesture.

Method

Participants. These were 18 undergraduates recruited through advertisements, all with normal hearing and normal, or corrected-to-normal vision.

Materials and Apparatus: The visual stimuli consisted of four digitised images of an individual who was instructed to point either upwards or downwards whilst orienting their head in the same direction as their pointing gesture, in the opposite direction to their pointing gesture, or straight ahead. Examples of these images are presented in Figure 1. These stimuli subtended

approximately 11° of vertical visual angle, and 14° of horizontal visual angle and were viewed by participants seated approximately 0.7 m from a 14 inch colour monitor.

Figure 1. Reproductions of the digitised stimuli used in Experiment 1. These images show agreeing and disagreeing head and gesture cues. Neutral head stimuli are not shown.

In addition to the gesture stimuli, the visual materials also contained a large question mark which subtended approximately 3° of vertical visual angle. The verbal stimuli were recorded using Hypercard audio software and edited using "SoundEdit" software on a Macintosh IIci. Two spoken words ("up" and "down") were recorded and edited to be approximately the same length (0.8 secs). These auditory stimuli were combined with each of the visual images to yield the twelve test stimuli which were presented using the "SuperLab" software on a Macintosh Performa 630.

Design. The audio-visual stimuli were presented in a within-subjects design with two factors: Gesture Congruity (congruent or incongruent) was determined by the relationship between the pointing gesture and the spoken word, and Head Agreement (agree, disagree or neutral), determined by the relationship between the orientation of the head and the direction of the gesture.

Procedure. Each trial began with the simultaneous presentation of a visual gesture stimulus and a spoken word. Participants were asked to respond to the verbal information as quickly and

accurately as possible by depressing one of two vertically arranged keys on the keypad area of the keyboard (8 for up, and 2 for down responses respectively). Participants were asked to operate each key with a separate hand and were free to decide which key to operate with which hand. The onset of the auditory stimulus activated the timer which was stopped when the participant pressed either of the response keys. This response also terminated the presentation of the visual stimulus. The intertrial interval was set at 1000 ms following the execution of a response. Participants were also instructed to depress the space bar in response to a large question mark which occasionally appeared on the screen. Seven of these question marks appeared in each response block. These trials were included in order to ensure that the subjects actually watched the screen.

Participants completed a set of 14 practice trials on the task. These comprised two question mark trials and one of each of the 12 test stimuli. Following the practice block, participants were given two blocks of 79 trials (72 test trials plus seven question mark trials). Each of the twelve test stimuli were presented six times in each block, giving a total of 24 trials per condition. Four practice trials were also presented immediately prior to the second experimental block. Trials were randomised within all blocks.

Results

In this and all other experiments reported, outliers were removed from individual participant's scores by eliminating those reaction times greater than two standard deviations from each cell mean. The resulting mean correct reaction time scores and percentage of errors recorded in each condition are reported in Table 1. Overall, RTs were slower when the gesture and voice were incongruent. However, the difference between incongruent and congruent RTs (the congruity effect) was largest when the head direction was in agreement with the gesture compared with conditions with neutral head orientation, or where the head disagreed with the direction of the gesture.

Table 1
Mean Reaction Times (RTs; in milliseconds) and Percentage of Errors in Each Condition of Experiment 1

| Head Agreement | Congruent gesture | | Incongruent gesture | | Congruity Effect |
	RT	% of errors	RT	% of errors	RT
Agree	466	1.11	527	2.00	+61
Disagree	482	0.44	510	1.33	+28
Neutral	482	1.11	516	2.22	+34
M	477	0.89	518	1.85	+41

A 2 (gesture congruity) x 3 (head agreement) analysis of variance (ANOVA) performed on the RT data was consistent with these observations. Performance was affected by the congruity of the gesture with the voice stimuli ($F(1, 17) = 17.88$, $p < 0.01$) but not by the agreement of the head with the gesture ($p = 0.79$). The main effect of gesture congruity was qualified by an interaction with head agreement ($F(2, 34) = 4.21$, $p < 0.05$). The effect of congruity was strong when the head direction agreed with the gesture (61 ms) but weaker when the head direction was either neutral (34 ms), or disagreed with the gesture (28 ms). However, a simple main effects analysis revealed that the effect of gesture congruity was significant at all levels of head agreement (p's < 0.05).

In order to pursue further the interaction between head agreement and gesture congruity, an additional one-way ANOVA was conducted to compare congruity effects in the three head agreement conditions. Congruity scores were calculated by subtracting each participant's mean RT for congruent stimuli from their mean RT for incongruent trials in each of the head agreement conditions. The one-way ANOVA revealed a main effect of head agreement ($F(2, 34) = 4.28$, $p < 0.05$). Post hoc Newman-Keuls tests ($\alpha = 0.05$) indicated that the gesture congruity effect in the head-agree condition was significantly larger than those effects in both the disagree and neutral conditions which did not differ from one another.

In Table 1 the error scores generally mirror the RT data; the overall mean error score was only 1.37%. The correlation between overall mean RTs and mean error rates was 0.77 suggesting

no evidence of a trade-off between speed and accuracy which might compromise interpretation of the RT data. Because of the low rate of errors, no further analysis was conducted on these data.

Discussion

The results of this experiment replicate those of Langton et al. (1996) who obtained interference effects from irrelevant pointing gestures when participants were asked to make speeded keypress responses to spoken directional words. Pointing gestures produced reliable congruity effects in all conditions of the experiment. However, the congruity effect observed when the head and gesture were oriented in the same direction (as in Langton et al., 1996) was significantly reduced when the head was in a neutral position, or oriented in the opposite direction to the gesture. Clearly head direction does indeed modulate the interfering effect of irrelevant gestures. Both pointing gestures and head orientation therefore receive some kind of obligatory processing by the cognitive system, even when attention is directed to stimuli appearing in a completely separate modality.

Although the findings of Experiment 1 indicate that both head orientation and gesture are able to influence the processing of spoken words, it is not clear how these stimuli exert their effects. Logically there seem to be two possibilities. Information encoded from gesture and head cues may be combined prior to some integrated representation exerting an influence on the processing of spoken words. Alternatively, gesture and head cues may be processed in parallel, exerting *independent* effects on the processing of these directional words.

In order to tease apart these alternatives, the data from Experiment 1 were re-analysed to examine whether the effect of the congruity of the gesture with the direction words interacted with the congruity of the head orientation with these words, or whether these effects were additive with one another. To achieve this, the Head Agreement factor was replaced with a Head Congruity factor which referred to the congruent/incongruent relationship between the head orientation and the directional word, rather than the relationship between the head and the

gesture. The RTs for each subject were re-scored accordingly, and the resultant means entered into a 2(Gesture Congruity) x 3(Head Congruity) ANOVA. This analysis yielded main effects of Gesture Congruity ($F(1, 17) = 17.88$, $p < 0.01$) and Head Congruity ($F(2, 34) = 4.65$, $p < 0.05$), and critically, no interaction between these factors ($F(2, 34) = 0.55$, $p = 0.58$). Gesture direction and head orientation therefore exert independent effects on responses to spoken directional words. So, rather than reflecting the influence of a combined gesture/head signal on spoken words, the results are more consistent with a model where gesture and head orientation are evaluated in parallel, with the outputs of each system influencing the processing of spoken directional words at some stage of processing.

Some evidence from developmental psychology has indicated that the comprehension of gaze may indeed be independent of the comprehension of pointing. Whilst infants seem to respond to gaze cues from as young as three months (Hood, Willen & Driver, 1998), and to head/gaze cues from six months (e.g., Butterworth & Jarrett, 1991; but see Corkum & Moore, 1995.), comprehension of manual pointing does not seem to emerge until about 12 months of age (Butterworth, 1991; cited in Butterworth, 1995). Butterworth has argued that, unlike the comprehension of gaze, understanding pointing gestures requires additional cognitive developmental changes - notably the comprehension of signs - which are not available to the developing child until around 10-12 months. The comprehension of manual pointing may therefore emerge much later than the understanding of the more "social" cues, and furthermore may be underpinned by a more general ability to appreciate symbolic meaning.

Langton et al. (1996) obtained symmetrical interference effects between pointing gestures and spoken words. That is, responses to spoken words were influenced by to-be-ignored gestures and reciprocally, responses to pointing gestures were influenced by irrelevant directional words. They suggested that this pattern of results was best accounted for by a model where gesture and words were identified in parallel with the outputs of each system somehow combined before a directional response could be determined, programmed and executed. The findings of the present experiment suggest that the processing of these gestures should be further broken down into the

parallel processing of head/gaze orientation and the actual pointing gesture itself. On this basis one would expect to observe symmetrical interference effects between head and gesture signals when placed into conflict in an intra-modal version of the interference task. This hypothesis was examined in Experiment 2.

Experiment 2

In Experiment 2, participants were presented with the agreeing and disagreeing head/gesture stimuli used in Experiment 1. They were asked to make manual keypress responses contingent on the direction of the gesture in one block of trials, and the orientation of the head in a separate block of trials. If head and gesture cues are processed automatically and in parallel then a symmetrical pattern of effects might be expected. Responses to pointing gestures should be influenced by to-be-ignored head cues, and conversely responses to head cues should be similarly influenced by irrelevant pointing gestures.

Method

Participants. These were twenty undergraduates who were recruited through advertisements. All had normal, or corrected to normal vision.

Apparatus and Materials. The visual stimuli consisted of four digitised images of an individual who was instructed to point either upwards or downwards whilst orienting their head in the same direction as their pointing gesture or in the opposite direction to their pointing gesture. Examples of these images are presented in Figure 2. These stimuli subtended approximately 11˚ of vertical visual angle, and 14˚ of horizontal visual angle and were viewed by participants seated approximately 0.7 m from a 14 inch colour monitor.

Design. The materials were tested in a within subjects design with three factors: Response dimension (head or gesture), Congruity (congruent or incongruent) and Target direction (up or down).

Procedure. On each trial, participants were asked to respond as quickly and accurately as possible to the either the gesture or head direction of the visual stimulus which appeared in the centre of the screen. Participants executed their responses by depressing one of two vertically arranged keys on the keypad area of the keyboard (8 for up, and 2 for down responses respectively). They were asked to operate each key with a separate hand and were free to decide which key to operate with which hand. Depression of either key terminated the presentation of the visual stimulus and stopped the timer. Following the response, the screen remained blank for 1000 ms before the start of the next trial.

Participants completed 160 experimental trials, comprising 20 trials in each of the eight cells of the experimental design. These trials were blocked by the response factor, and were presented in a random order within each block. Thus, in one block of trials participants responded to the orientation of the head and were asked to ignore the gesture, and in a second block of trials they responded to the gesture direction and were asked to ignore the orientation of the head. The order in which these blocks were presented was alternated between successive participants. A set of 16 practice trials were presented before each block, comprising two repetitions of each of the eight experimental stimuli. Both reaction times (RTs) and the proportion of errors were recorded as dependent variables in this experiment.

Results

The interparticipant mean correct RTs and percentage of errors recorded in each condition of the experiment are summarised in Table 2. An inspection of this table reveals that, overall, participants responses were slower and less accurate in incongruent compared with congruent conditions of the experiment. However, this difference was more marked when head direction was the relevant response dimension.

Table 2
Mean Reaction Times (RTs; in milliseconds) and Percentage of Errors for Responses to Head and Gesture Direction in Each Condition of Experiment 2.

Congruity	Down Targets		Up Targets		Overall Mean	
	RT	% of Errors	RT	% of Errors	RT	% of Errors
Responses to Head						
Congruent	486	0.50	496	1.25	491	0.88
Incongruent	521	3.00	520	2.50	521	2.75
M	504	1.75	508	1.88	506	1.82
Responses to Gesture						
Congruent	423	0.50	435	1.25	429	0.88
Incongruent	441	1.00	445	1.25	443	1.13
M	432	0.75	440	1.25	436	1.00

A 2 (response dimension) x 2 (congruity) x 2 (target direction) ANOVA conducted on participants' mean RT's supported the above observations. Performance was equivalent for up (474 ms) versus down (467 ms) targets ($p = 0.24$). However, there were main effects of both response dimension ($F(1, 19) = 34.28$, $p < 0.001$) and congruity ($F(1, 19) = 34.41$, $p < 0.001$) which were qualified by a significant interaction between these factors ($F(1, 19) = 5.23$, $p < 0.05$). Simple main effects analysis revealed that the effects of congruity were significant for both response dimensions (p's < 0.05). The cause of this interaction therefore appears to be the relatively strong effect of congruity for responses to head direction (30 ms) compared to the weaker effect of congruity on responses to gesture direction (14 ms)[1].

Error scores closely mirror the RT data. Indeed, the correlation between overall mean RTs and mean error rates in each condition was 0.73 offering no evidence of a trade-off between speed and accuracy. Because of the low rate of errors (the overall mean error rate was 1.41%), no further analyses were performed on these data.

Discussion

Experiment 1 demonstrated that participants were unable to ignore gesture and head cues to direction even when responding to verbal stimuli. The findings of Experiment 2 also suggest that

head and gesture cues are processed automatically by observers. There were significant effects of congruity for responses to both gesture and head/gaze direction stimuli. Thus, to-be-ignored head/gaze direction was able to influence participants' responses to the direction of a pointing gesture and reciprocally, irrelevant pointing gestures influenced speeded judgements of head/gaze direction.

However, despite being bi-directional, the interference effects were not exactly symmetrical; responses to head orientation were affected more by irrelevant gestures than vice-versa. One possibility is that this asymmetry may have been the result of mismatching discriminabilities of the two dimensions of the test stimuli. Melara and Mounts (1993) have shown that the direction of the interference effects between interacting dimensions is a function of their relative baseline discriminabilities. In particular the more discriminable dimension (i.e., the dimension with greater trial-to-trial variation) will cause greater disruption on classification of the less discriminable dimension than vice versa. It could be argued that in the present experiments the perceptual variation between up and down gestures is greater than that between up and down head angles. Indeed, the fact that participants' responses to gesture stimuli were significantly faster than those to head stimuli (436 ms vs. 506 ms) provides some evidence for this position. Thus the more discriminable gesture dimension will exert a greater influence on the less discriminable head orientation dimension than vice-versa, resulting in the asymmetrical pattern of interference observed in Experiment 2. However, it is noteworthy that despite being disadvantaged by relatively poor discriminability, irrelevant head direction cues still produced significant interference effects on classification of the more discriminable pointing gestures.

The findings of this experiment add to those of other studies conducted in our laboratory in which a variety of directional signals produce symmetrical or bi-directional interference effects when placed into conflict in a Stroop-type task. As well as pointing gestures and spoken words, symmetrical interference effects have been obtained between gestures and written directional words, between directional arrows and spoken words, and between head orientation and gaze direction (Langton et al., 1996; Langton, submitted). This pattern of results is perhaps best

accounted for by a model in which signals such as head orientation, gaze direction, pointing gestures, and spoken words etc. are processed in parallel by a number of separate systems. These systems feed their outputs into a processing stage, or processing stages where the directional information automatically encoded from the cues is somehow pooled before a response can be selected and executed. Thus, interference effects will emerge whenever a directional decision has to be made to a directional signal when conflicting information is available from any other co-present cue.

According to this model, it is the nature of the target stimulus and the meaning extracted from this signal which is important in producing the interference effects. If the target requires a non-directional decision, the processing of this stimulus and the parallel automatic processing of the task-irrelevant directional cue will result in the encoding of non-conflicting information and, as a consequence, no interference effects will be observed. Therefore, interference should occur whenever participants are asked to make decisions based on direction, but non-directional decisions should remain uninfluenced by to-be-ignored attentional cues. This hypothesis was examined in Experiment 3.

Experiment 3

In this experiment the pointing gestures used in Experiment 2 were replaced by "thumbs-up" and "thumbs-down" gestures. These are so-called emblematic or symbolic gestures which have a precise meaning understood by a particular group or culture, and convey this meaning in the absence of any accompanying speech. Conventionally, throughout Europe the thumbs-up gesture means "good" or "okay" whilst the thumbs-down gesture means "bad" or "no-good" (Morris, Collett, Marsh & O'Shaughnessy, 1979). Thus, although these gestures are visually similar to the pointing gestures used in Experiments 1 and 2 (see Figure 3), they carry no intrinsic directional meaning. Once again participants were asked to respond to the gestures and head/gaze orientation in separate blocks of trials. However, instead of deciding whether the gesture was directed upwards or downwards, participants were asked to make one response if the gesture was

"thumbs-up" and another if the gesture was "thumbs-down". Because participants were not asked to make a decision based on the direction indicated by the gesture, responses were expected to be uninfluenced by the accompanying head/gaze orientation.

Method

Participants. These were twenty Open University students attending a residential summer school in Stirling. All had normal or corrected-to-normal vision.

Apparatus and Materials. The visual stimuli consisted of digitised images of an individual making a thumbs-up, or thumbs-down gesture whilst in each case, orienting their head upwards or downwards. Examples of these images are presented in Figure 3. These stimuli subtended approximately 9° of vertical visual angle, and 7° of horizontal visual angle and were viewed by participants seated approximately 0.7 m from a 14 inch colour monitor.

Figure 2. Reproductions of some of the incongruent stimuli used in Experiment 3.

Design and Procedure. These were identical to those of Experiment 2 with one notable exception. In contrast to the gesture response condition of Experiment 2 where participants based their responses on the direction indicated by the model's pointing gesture, participants in the corresponding condition of Experiment 3 were asked to decide whether the model was making a

"thumbs-up", or a "thumbs-down" gesture. The instructions in the head response condition were identical to those of Experiment 2.

Results

The interparticipant means and percentage of errors for the eight conditions of Experiment 3 are shown in Table 3. When responding to the orientation of the head, participants were generally faster and more accurate in the congruent compared with incongruent conditions. However, the RT difference was more marked when participants responded to heads oriented downwards (27 ms) compared with upwards (11 ms). In contrast, there is no consistent effect of congruity for responses to the gestures. Responses to thumbs-up gestures were just 3 ms faster in congruent compared with incongruent conditions, whereas RTs to thumbs-down gestures were actually 7 ms faster when the head was oriented in an incongruent, compared with a congruent direction.

Table 3
Mean Reaction Times (RTs; in milliseconds) and Percentage of Errors for Responses to Head and Gesture Direction in Each Condition of Experiment 3.

Congruity	Down Targets		Up Targets		Overall Mean	
	RT	% of Errors	RT	% of Errors	RT	% of Errors
			Responses to Head			
Congruent	459	1.25	469	1.75	464	1.50
Incongruent	486	2.50	480	4.50	483	3.50
M	473	1.88	475	3.13	474	2.50
			Responses to Gesture			
Congruent	421	2.25	429	2.50	425	2.38
Incongruent	414	2.50	432	2.50	423	2.50
M	418	2.38	431	2.50	424	2.44

These observations were supported by a 2 (response dimension) x 2 (congruity) x 2 (decision) ANOVA conducted on participants' mean RTs. This analysis revealed main effects of congruity ($F(1, 19) = 5.98$, $p < 0.05$), response ($F(1, 19) = 29.86$, $p < 0.001$), and decision ($F(1,$

19) = 4.55, p < 0.05) and also an interaction between response and congruity (F(1, 19) = 14.00, p < 0.01). However, these effects were qualified by a significant higher order interaction between the three factors (F(1, 19) = 5.08, p < 0.05). This interaction seems to arise because of the large effect of congruity for responses to heads oriented downwards compared with upwards (27 ms vs. 11 ms), and the absence of any such congruity effect when participants responded to either type of gesture. Indeed, further analysis of this interaction indicated that the effect of congruity was significant for responses to heads oriented downwards (p < 0.001), but only marginally significant for heads oriented upwards (p = 0.07). There were no effects of congruity for responses to the gesture dimension (p = 0.59).

The overall mean error rate was just 2.47%. The correlation between mean RTs and mean error scores was 0.22 suggesting no evidence of a speed-accuracy trade-off. No further analysis was conducted on these data.

Discussion

The results of Experiments 1 and 2 have established that directional information obtained from the orientation of the head and/or the eyes is processed automatically, and is able to influence directional decisions to both spoken words and pointing gestures. In Experiment 3, however, these same head/gaze cues failed to influence non-directional decisions to certain emblematic gestures which nevertheless contained visually similar information to the pointing gestures studied in Experiments 1 and 2. This finding supports the hypothesis that interference effects between signals will emerge only when a directional decision has to be made to any one of these cues.

However, when participants were asked to make a response to the head/gaze cues, the emblematic gestures did produce a reliable interference effect (marginally significant for "down" responses). Why should this be so, particularly when these gestures did not seem to be treated as directional signals when they formed the target stimulus dimension? One possibility is that when responding to the head/gaze dimension, the gestures would have appeared in the periphery of

participants' vision and may therefore have only received an analysis based on low spatial frequency information, an analysis which might not have been sufficient to distinguish between a thumb and index finger. Thus, without the benefit of focal attention, the gestures could easily be processed as pointing signals and exert their effects on decisions to head/gaze direction accordingly.

On the other hand, it is possible that the default is to process these kinds of gestures as directional stimuli rather than assigning them their intended meanings. When these stimuli form the irrelevant dimension, it may be that it is their directional meanings which become active and hence exert an interfering effect on directional responses to head orientation. When these gestures form the target dimension, the demands of the task dictate that they are not processed directionally, and so are uninfluenced by the irrelevant directional head cues. The asymmetrical nature of the effects in this experiment could therefore be attributable to some intrinsic directional meaning contained in these hand gestures, rather than their visual confusability with pointing gestures[2].

Regardless of how the to-be-ignored gestures exerted their effects on responses to head orientation, the main finding of this experiment was that irrelevant head cues produced no interfering effect on non-directional decisions to gestural stimuli. The elimination of this effect is notable given that the same cues produced reliable interference effects in Experiment 1, where participants were responding to targets in an entirely separate modality, as well as in the intra-modal task used in Experiment 2.

Experiment 4

Experiment 3 has established that non-directional decisions to a social gestural signal are not influenced by directional information carried by head and gaze cues. In Experiment 4 we ask whether these same head/gaze cues will influence directional decisions to ostensibly non-social directional signals. It is possible, for example, that cues such as head and gaze orientation will only produce interference effects when participants have to make a decision based on cues which

communicate something socially about direction, e.g., "Which way is he pointing?", or "Which way is he telling me to look?". However, it seems more likely that it is the directional nature of the decision which is important, rather than the social status of the directional cue. If participants are required to extract some directional meaning from any target and make a response contingent on this, then any other available directional cues may produce an interference effect as described earlier.

Thus, in Experiment 4, the pointing gestures used in Experiment 1 were replaced by arrows which were printed across the chest of the model. Once again, participants were asked to make speeded keypress responses contingent on the direction of the arrow, and the orientation of the head in separate blocks of trials. If the effects observed in Experiments 1 and 2 were caused by the social nature of the directional decision then head cues would not be expected to interfere with responses to arrows. However, if it is the directional decision per se that is critical, then an interference effect should be observed.

Method

Participants. These were a further twenty Open University students attending the residential summer school. All had normal or corrected-to-normal vision and none had participated in the previous experiment.

Apparatus and Materials. The visual stimuli consisted of digitised images of the same individual who served as a model for the stimuli used in Experiment 4. Images were obtained of this individual orienting his head either upwards or downwards as before, but making no accompanying hand gestures. These stimuli subtended approximately 8° of vertical visual angle, and 6° of horizontal visual angle and once again were viewed by participants seated approximately 0.7 m from a 14 inch colour monitor. For each head orientation, congruent and incongruent stimuli were created by pasting arrows, measuring approximately 3° x 2°, across the chest of the model, either pointing in the same, or in the opposite direction to the orientation of his head. These images are recreated in Figure 4.

Design and Procedure. As in Experiment 1, the materials were tested in a 2 x 2 x 2 within-subjects design. The three variables were response (either to the arrow direction, or to the orientation of the head), congruity (congruent or incongruent head/arrow pairings) and target direction (up or down). Participants were asked to respond to the direction of the arrow in one block of trials and the orientation of the head in a second block. Half of the participants responded to the arrow first and half to the head. The order of presentation within the blocks was completely randomised.
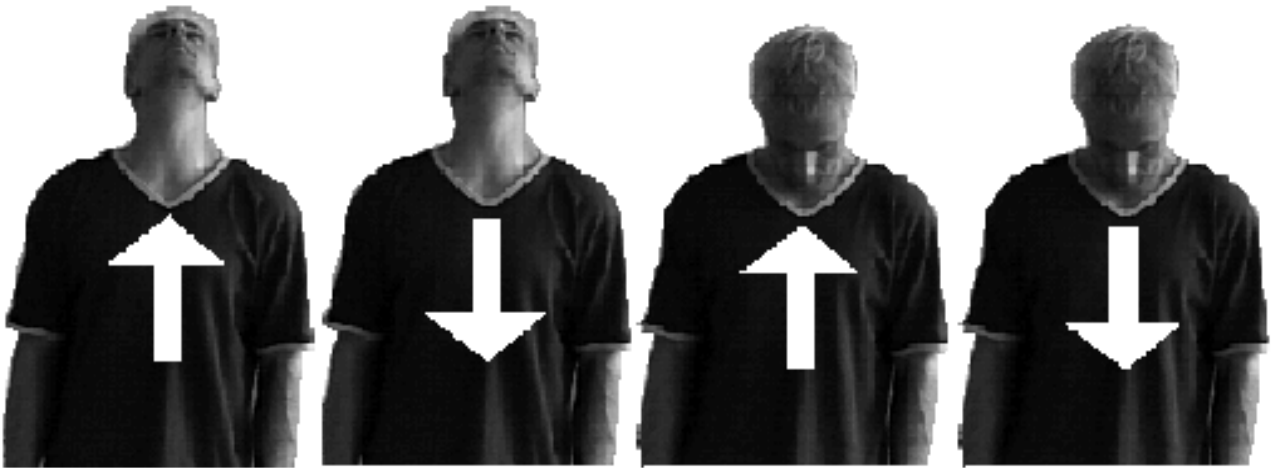

Figure 3. Reproductions of the digitised stimuli used in Experiment 4.

Results

Mean correct RTs and error rates in each condition of the experiment are presented in Table 4. When responding to arrows, participants were generally slower and less accurate in incongruent compared with congruent conditions. However, when responding to head orientation, this RT difference is only apparent for heads oriented downwards.

Table 4
Mean Reaction Times (RTs; in milliseconds) and Percentage of Errors for Responses to Head and Arrow Direction in Each Condition of Experiment 4.

Congruity	Down Targets		Up Targets		Overall Mean	
	RT	% of Errors	RT	% of Errors	RT	% of Errors
			Responses to Head			
Congruent	425	2.25	438	0.25	432	1.25
Incongruent	445	1.50	438	2.00	442	1.75
M	435	1.88	438	1.13	437	1.50
			Responses to Arrow			
Congruent	438	0.75	439	0.75	439	0.75
Incongruent	457	1.00	452	1.00	455	1.00
M	448	0.88	446	0.88	447	0.88

The RT data were entered into a 2 (response) x 2 (congruity) x 2 (target direction) ANOVA which yielded a main effect of congruity ($F(1, 19) = 9.93$, $p < 0.01$) with RTs some 13 ms slower in incongruent compared with congruent conditions. The analysis also yielded a marginally significant interaction between target decision and congruity ($F(1, 19) = 4.15$, $p = 0.06$) which seems to arise because the congruity effect is larger for down targets (19 ms) than for up targets (7 ms). Although Table 4 indicates that performance was equivalent for upward oriented heads paired with congruent and incongruent arrows, the three-way interaction term failed to reach significance ($p = 0.22$). No other main effects or interactions approached significance (p's > 0.1).

The overall mean error score in this experiment was just 1.19%. The correlation between overall mean RTs and mean error rates in each condition was -0.39. Thus, although the correlation fails to reach statistical significance, there is some evidence that participants were trading accuracy for speed. However, because the error rates in all conditions were so low, and generally mirrored the RT congruity effects the small negative correlation was not deemed sufficient to compromise interpretation of the RT data.

Discussion

The most notable finding from this experiment is that to-be-ignored head/gaze cues produced significant interference effects on participants' responses to the arrow stimuli. It seems then that directional signals are able to exert effects on decisions to both social and non-social cues. Clearly it is the directional meaning encoded from the target which is important in producing interference effects rather than the social nature of the decision or the social status of the target stimulus.

Although numerically only apparent for downward decisions, to-be-ignored arrows also exerted an interference effect on responses to head orientation. Langton et al. (1996) showed that irrelevant arrows interfered with spoken words so it is not all that surprising that Experiment 4 has provided some evidence that, like head and gesture cues, the direction of arrows seems to be processed automatically. However, it is unclear why this effect should only occur for downward decisions. One possibility is that with the head-down stimuli, the face and arrow are in closer proximity than in the head-up stimuli (see Figure 3). Directing attention toward the face may also bring the arrow into the attentional focus which may in turn facilitate its processing.

General Discussion

The studies reported in this paper sought to examine further the processing of pointing gestures of the type studied by Langton et al. (1996). Using a Stroop-type interference paradigm, Langton et al. showed that to-be-ignored pointing gestures were able to influence the speed of responses to spoken directional words, suggesting that these directional signals are processed automatically by observers. However, in addition to the outstretched hand, the directional information contained in Langton et al.'s gestures was provided by the orientation of the head and the line of regard of the gesturer's gaze. Any or all of these signals could have received automatic processing and hence influenced responses to spoken words.

Experiment 1 of the present paper replicated Langton et al.'s finding. Once again, to-be-ignored pointing gestures exerted an interference effect on responses to spoken directional words. However, this effect was found to be moderated by the direction in which the gesturer's head and gaze were oriented. Clearly, observers were automatically extracting directional information from the gesturer's head and eyes as well as from his hand. Moreover, the effects of head/gaze cues and gesture on spoken words were found to be additive, implicating a model where pointing gesture and head cues are processed automatically and in parallel by separate systems. In support of this model, Experiment 2 demonstrated that head and gesture cues produced bi-directional interference effects in a intra-modal version of the Stroop-type task. Participants' choice RTs to the direction of pointing gestures were slower when the gesturer's head and eyes were oriented in the opposite direction to the gesture compared with a condition where gesture and head were oriented in the same direction. Reciprocally, responses to head angle were similarly influenced by to-be-ignored gestural information. When participants' made non-directional decisions to emblematic gestures in Experiment 3, interference effects were absent suggesting that the locus of the interference effect between head and gesture cues is at a decision stage of processing following the independent evaluation of the cues. Finally, in Experiment 4 to-be-ignored head cues produced significant interference effects on responses to arrows, suggesting further that it is the nature of the directional decision which is important in producing the effects, rather than the social nature of the target stimulus.

It is clear from these findings and from experiments reported elsewhere (e.g., Driver, Davis, Kidd, Maxwell, Ricciardelli & Baron-Cohen, in press; Friesen & Kingstone, 1998; Langton & Bruce, in press; Langton et al., 1996) that various directional signals are processed automatically by observers. Why should this be so? One possibility is that these signals all carry information regarding the direction of another individual's *social attention*. The remainder of this section is devoted to a discussion of this hypothesis and to speculation about the nature of the interaction or combination of information thought to produce the interference effects noted in Experiments 1-4.

Most organisms will orient their sensory receptors toward objects in the environment which are of immediate interest to them. Human and non-human primates, for example, will direct their eyes and heads toward individuals with whom they might interact, mate or perhaps eat. In order to prepare for any of these encounters, it is critical that the potential interactant, sexual partner, or meal be able to detect when it has become the recipient of another's attention. Alternatively, rather than signalling their own interest in you, another's direction of attention might indicate the presence and the exact location of a third party or object, which again might constitute a threat or potential food source. By computing this individual's angle of gaze and then following this line of regard, the focus of their attention can be perceived, and any necessary action taken. As well as alerting organisms to the presence of a predator or prey, this kind of "attention following" behaviour is critical in human communication such as in understanding otherwise ambiguous deictic expressions such as "here", "there", "that one" or "this one". Gaze following may also facilitate vocabulary acquisition by toddlers as the referent of a new word can be specified by the direction in which the speaker is looking (Baldwin, 1991) or perhaps, pointing. In line with this, several studies have shown that non-human primates (e.g., Emery, Lorincz, Perrett, Oram, & Baker, 1997), infants (e.g., Butterworth & Jarrett, 1991; Hood et al., 1998), and adults (Driver et al., in press; Langton & Bruce, in press), spontaneously redirect their gaze and/or their visual attention in accord with another's gaze or head orientation.

Given their importance to survival and communication, it is perhaps unsurprising that we may have evolved mechanisms which process directional cues such as head and gaze direction automatically. Indeed, neuropsychological, neurophysiological and behavioural evidence is emerging in support of the position that there is a functionally specific mechanism devoted to the task of detecting eyes and computing where in the environment eye-gaze is directed (e.g., Campbell, Heywood, Cowey, Regard & Landis, 1990; Heywood and Cowey, 1992; Perrett, Mistlin, Chitty, Harries, Newcombe & de Haan, 1988; Perrett, Smith, Potter, Mistlin, Head, Milner and Jeeves, 1985). Moreover, Perrett and his colleagues have suggested that something like a gaze detection mechanism forms only part of a system designed to process the direction of

"social attention" (e.g., Perrett & Emery, 1994; Perrett, Hietanen, Oram & Benson, 1992). Their work has indicated that individual cells in the STS region of the macaque brain respond to *conjunctions* of eye, head and body position. Thus, cells which respond preferentially to eyes directed downwards show further preferences when the head is oriented downwards and when the body adopts a quadrupedal posture. Accordingly, Perrett et al. (1992) suggest that one function of the STS region lies in the analysis of social attention direction. This mechanism, termed the "Direction of Attention Detector" (DAD) by Perrett and Emery (1994), is considered to combine information from separate detectors analysing body, head and gaze direction in order to compute the whereabouts of another individual's focus of interest.

Pointing gestures also provide an important additional source of information regarding another's direction of attention. Indeed, they may provide a more accurate cue to the spatial location of a referent than either eye or head orientation (Butterworth, 1991; cited in Butterworth, 1995). Because of the value of these cues, it is possible that structures exist which process pointing gestures, like gaze and head orientation as additional cues to the direction of social attention. Thus it may well be that the automatic processing of head, gaze and gestural information noted in the present experiments reflects the operation of some kind of direction of attention detector. However, further research is needed to establish whether or not participants are actually processing these directional signals as cues to social attention direction.

A possible argument against the suggestion that participants are processing social attention direction is the finding of Experiment 4 that arrows, like pointing gestures and head cues, also receive automatic processing. The argument is that arrows are conventional, non-social cues to direction and so would not be expected to be processed in the same way as social attention cues such as gaze, head orientation or pointing gestures. The fact that arrows do produce an interfering effect on responses to head cues suggests that there is nothing special about these social signals.

However, the evidence that arrows receive obligatory processing in Experiment 4 is equivocal, with numerical effects apparent when participants responded to heads directed downwards but not upwards. The qualitative difference between the effects of these to-be-ignored arrows and to-be-ignored social cues in Experiments 1 and 2 may suggest that arrows are processed rather differently from social signals. Alternatively, one might ask whether arrows are really conventional non-social symbols at all. Rather than acting as arbitrary directional symbols, arrows may have emerged as schematic representations of manual pointing gestures and might also be considered as social in the sense of communicating something about intention or meaning to another individual. In this respect it is perhaps unsurprising that Experiment 4 revealed some evidence that arrows, like head, gaze and gesture signals, are processed automatically.

Regardless of whether or not participants in Experiments 1-4 were processing social attention direction, the results of these experiments do indicate that the gestures used by Langton et al. actually comprised signals carried by the gesturer's head, gaze and hand, and that these cues were all processed automatically. The additional suggestion that these cues are initially processed in parallel and interact prior to the programming and execution of a response is similar to a model of information integration developed by Massaro and his colleagues (e.g., Massaro & Friedman, 1990). Their Fuzzy Logical Model of Perception (FLMP) comprises three operations; evaluation, integration and decision. Information from the two sources is evaluated independently and then integrated according to a multiplicative algorithm which ensures that the least ambiguous source of information carries most weight in the decision process. Following integration, a decision is made on the basis of the relative goodness of match of the integrated stimulus information with the relevant prototype descriptions in memory. Massaro's group has applied this model to a variety of domains including interactions between gesture and speech (Thompson & Massaro, 1986, 1994), and between emotional expressions in the face and voice (Massaro & Egan, 1996).

However, it may be that the interaction between attention signals observed in the present experiments is best accounted for by a *non-integrative* interaction. In these experiments participants were asked to attend selectively to one relevant cue and to ignore information carried by the second irrelevant signal. On the whole, participants were able to do this successfully, suffering no loss in accuracy, only a reduction in speed, when the cues carried conflicting information. This suggests that participants' decisions were based on intact rather than integrated representations of information encoded from the target cue. Clearly, the precise mechanisms of the interaction require further investigation.

Another hypothesis, currently under investigation, is that the type of cross-modal interference effects noted in Experiment 2 here, and by Langton et al. (1996), are mediated by the effect which certain social signals can exert on an observer's visuo-spatial attention. A number of research groups, including our own, have recently established that non-predictive head/eye gaze cues (Langton & Bruce, in press) and gaze cues from images of real faces (Driver et al., in press) and schematic faces (Friesen & Kingstone, 1998) can trigger a reflexive, exogenous visual orienting response on behalf of an observer (see Spence & Driver, 1994 for a review of visual orienting). According to one particular theory of spatial attention, the "premotor theory" (e.g., Rizzolatti, Riggio & Sheliga, 1994), in order to shift attention to a particular location entails the programming of an eye movement to that location, whether or not the eye movement is ever actually executed. This motor program contains directional features which becomes a spatial code when a manual response has to be selected. In this way a stimulus containing a social attention cue would automatically generate a spatial code prior to initiating an orienting response. This spatial code could then somehow interfere with a spatial representation encoded from the spoken directional word at some stage in processing[3]. Whether this kind of mechanism could account for interference effects between social attention cues is not clear. However, the fact that head and gaze cues are able to trigger a reflexive orienting response on behalf of an observer, despite this observer's intentions and incentives to ignore these signals,

constitutes further evidence that at least some social attention cues are analysed rapidly and automatically by the information processing system.

Finally, it should be stressed that the visual stimuli used in the present experiments were all static images; poses of manual pointing, head and eye gaze cues which would ordinarily be strongly dynamic events. In addition, gestures naturally occur in close synchronisation with semantically corresponding spoken words (e.g., Condon, 1970; Levelt, Richardson & La Heij, 1985). Clearly, we can only speculate that the effects we have observed will transfer to naturally occurring moving stimuli, but given the attention grabbing properties of such stimuli it is even possible that such effects will be augmented. Design constrains precluded the use of dynamic stimuli in the present studies, but a future challenge is to develop methodologies which could explore similar questions with more naturally occurring materials.

References

Argyle, M. (1988). *Bodily Communication* (2nd ed.). London: Methuen.

Baldwin, D. (1991). Infants' contribution to the achievement of joint reference. *Child Development*, *62*, 875-890.

Butterworth, G. E. (1991). Evidence for the geometric comprehension of manual pointing. Paper presented at the meeting of the Society for Research in Child Development, Seattle. Cited in Butterworth, G. (1995). Origins of mind in perception and action. In C. Moore & P. J. Dunham (Eds.), *Joint attention: Its origins and role in development* Hillsdale, NJ: Lawrence Erlbaum Associates.

Butterworth, G., & Jarrett, N. (1991). What minds have in common is space: Spatial mechansims serving joint visual attention in infancy. *British Journal of Developmental Psychology*, *9*, 55-72.

Campbell, R., Heywood, C. A., Cowey, A., Regard, M., & Landis, T. (1990). Sensitivity to eye gaze in prosopagnosic patients and monkeys with superior temporal sulcus ablation. *Neuropsychologia*, *28*, 1123-1142.

Condon, W. S. (1970). Method of micro-analysis of sound films of behavior. *Behavioral Research Methods and Instrumentation*, *2*, 51-54.

Corkum, V., & Moore, C. (1995). Development of joint visual attention in infants. In C. Moore & P. J. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 61-83). Hillsdale, NJ: Lawrence Erlbaum Associates.

Driver, J., Davis, G., Kidd, P., Maxwell, E., Ricciardelli, P., & Baron-Cohen, S. (in press). Shared attention and the social brain: Gaze perception triggers automatic visuospatial orienting in adults. *Visual Cognition*.

Emery, N. J., Lorincz, E. N., Perrett, D. I., Oram, M. W., Baker, C. I. (1997). Gaze following and joint attention in rhesus monkeys (macaca mulatta). *Journal of Comparative Psychology, 111,* 286-293

Feyereisen, P. (1991). Brain pathology, lateralization, and nonverbal behavior. In R. S. Feldman & B. Rimé (Eds.), *Fundamentals of nonverbal behavior* Cambridge: Cambridge University Press.

Friesen, C. K. & Kingstone, A. (submitted). The eyes have it!: reflexive orienting is triggered by nonpredictive gaze. *Psychonomic Bulletin and Review*.

Heywood, C. A., & Cowey, A. (1992). The role of the 'face cell' area in the discrimination and recognition of faces by monkeys. *Philosophical Transactions of the Royal Society of London, Series B., 335,* 31-38.

Hood, B. M., Willen, J. D., & Driver, J. (1998). An eye direction detector triggers shifts of visual attention in human infants. *Psychological Science,* 9, 131-134.

Kendon, A. (1986). Some reasons for studying gesture. *Semiotica, 62,* 3-28.

Kendon, A. (1994). Do gestures communicate?: A review. *Research on Language and Social Interaction, 27,* 175-200.

Kleinke, C. L. (1986). Gaze and eye contact: A research review. *Psychological Review, 100,* 78-100.

Langton, S. R. H. (submitted). The mutual influence of gaze and head orientation in the analysis of social attention direction. *Quarterly Journal of Experimental Psychology*

Langton, S. R. H., & Bruce, V. (in press). Reflexive visual orienting in response to the social attention of others. *Visual Cognition*.

Langton, S. R. H., O'Malley, C., & Bruce, V. (1996). Actions speak louder than words: Symmetrical cross-modal interference effects in the processing of verbal and gestural information. *Journal of Experimental Psychology: Human Perception and Performance*, *22*. 1357-1375.

Levelt, W. J. M., Richardson, G., & La Heij, W. (1985). Pointing and voicing in deictic expressions. *Journal of Memory and Language*, *24*, 133-164.

Massaro, D. W., & Egan, P. B. (1996). Perceiving affect from the face and voice. *Psychonomic Bulletin and Review*, *3*, 215-221.

Massaro, D. W., & Friedman, D. (1990). Models of integration given multiple sources of information. *Psychological Review*, *97*, 225-252.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, *109* , 163-203.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*, 746-748.

McNeill, D. (1985). So you think gestures are non-verbal? *Psychological Review*, *92*, 350-371.

McNeill, D. (1987). So you *do* think gestures are non-verbal! Reply to Feyereisen (1987). *Psychological Review*, *94*, 499-504.

McNeill, D. (1989). A straight path to where? Reply to Butterworth and Hadar. *Psychological Review*, *96*, 175-179.

McNeill, D., Cassell, J., & McCullough, K (1994). Communicative effects of speech-mismatched gestures. *Research on Language and Social Interaction*, *27*, 223-237.

Melara, R. D., & Mounts, J. R. W. (1993). Selective attention to Stroop dimensions: Effects of baseline discriminability, response mode and practice. *Memory and Cognition*, *21*, 627-645.

Morrel-Samuels, P., & Krauss, R. M. (1994). Word familiarity predicts temporal asynchrony of hand gestures and speech. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 615-622.

Morris, D., Collett, P., Marsh, P., & O'Shaughnessey, M. (1979). *Gestures, their origins and distribution*. London: J. Cape.

Nicoletti, R., & Umilà, C. (1994). Attention shifts produce spatial stimulus codes. *Psychological Research, 56*, 144-150.

Perrett, D. I., & Emery, N. J. (1994). Understanding the intentions of others from visual signals: Neurophysiological evidence. *Cahiers de Psychologie Cognitive*, *13*, 683-694.

Perrett, D. I., Hietanen, J. K., Oram, M. W., & Benson, P. J. (1992). Organisation and functions of cells responsive to faces in the temporal cortex. *Philosophical Transactions of the Royal Society of London, Series B.*, *335*, 23-30.

Perrett, D. I., Mistlin, A. J., Chitty, A. J., Harries, M., Newcombe, F., & de Haan, E. (1988). Neuronal mechanisms of face perception and their pathology. In C. Kennard & F. Clifford Rose (Eds.), *Physiological aspects of clinical neuro-ophthalmology* (pp. 137-154). London: Chapman & Hall.

Perrett, D. I., Smith, P. A. J., Potter, D. D., Mistlin, A. J., Head, A. S., Milner, A. D., & Jeeves, M. A. (1985). Visual cells in the temporal cortex sensitive to face view and gaze direction. *Proceedings of the Royal Society of London*, *B223*, 293-317.

Rimé, B., & Schiaratura, L. (1991). Gesture and speech. In R. S. Feldman & B. Rimé (Eds.), *Fundamentals of nonverbal behavior* Cambridge: Cambridge University Press.

Rizzolatti, G., Riggio, L., & Sheliga, B. M. (1994). Space and selective attention. In C. Umiltà & M. Moscovitch (Eds.), *Attention and Performance Vol. XV* Hillsdale, N.J.: Erlbaum.

Rubichi, S., Nicoletti, R., Iani, C., & Umiltà, C. (1997). The Simon effect occurs relative to the direction of an attention shift. *Journal of Experimental Psychology: Human Perception and Performance, 23,* 1353-1364.

Simon, J. R., & Rudell, A. P. (1967). The effect of an irrelevant cue on information processing. *Journal of Applied Psychology, 51,* 300-304.

Spence, C. J., & Driver, J. (1994). Covert spatial orienting in audition: Exogenous and endogenous mechanisms. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 555-574.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*, 643-662.

Thompson, L. A., & Massaro, D. W. (1986). Evaluation and integration of speech and pointing gestures during referential understanding. *Journal of Experimental Child Psychology*, *42*, 144-168.

Thompson, L. A., & Massaro, D. W. (1994). Children's Integration of Speech and Gestures in Comprehension. *Journal of Experimental Child Psychology*, *57*, 327-354.

Author Note

Footnotes

[1]For ease and clarity of presentation we have chosen to present the analyses of the RT data from Experiments 2-4 collapsed over block order. Neither the pattern of effects, nor the conclusions drawn from these studies are at all influenced when block order is taken out as a between-subjects factor, or when participants' data from each of the two block orders are analysed separately.

[2]The tendency to process the gesture cues directionally seems to be quite strong. Those participants who first received the gesture response task knew that the gestures represented thumbs-up and down, rather than pointing, and yet these stimuli still interfered with responses to head orientation in the second block of trials. This observation was confirmed by an additional ANOVA conducted on the RT data from Experiment 3 with block order as a between-subjects factor, and response, congruity and target direction as within-subjects factors. There was neither a main effect of block order ($p = 0.65$), nor any interactions involving this factor (p's > 0.1). In particular, the three-way interaction between block order, response and congruity failed to reach significance ($p = 0.70$).

[3]This argument is similar to, and is motivated by an account of the Simon effect (e.g., Simon & Rudell, 1967) given by Nicoletti, Umiltà and their colleagues (e.g., Nicoletti & Umiltà, 1994, Rubichi, Nicoletti, Iani & Umiltà, 1997). Briefly, the Simon effect occurs when the location of a target stimulus interferes with the response to that stimulus. Typically, participants are asked to make a left-right keypress response contingent on the identity of a stimulus (e.g., colour of light, direction of arrow, the words *left* or *right*, etc.) presented randomly to the left or right of some central point. Responses are faster when the location of the stimulus matches the location of the response than when it does not.