

# **Actions Speak No Louder Than Words: Symmetrical Cross-Modal Interference Effects in the Processing of Verbal and Gestural Information**

Stephen R.H. Langton and Claire O'Malley

ESRC Centre for Research in Development, Instruction and Training

Department of Psychology, University of Nottingham

and

Vicki Bruce

Department of Psychology, University of Stirling

Running Head: SYMMETRICAL CROSS-MODAL INTERFERENCE EFFECTS

This article may not exactly replicate the final version published in the APA journal. It is not the copy of record.

## **Abstract**

Five experiments are reported which investigate the distribution of selective attention to verbal and non-verbal components of an utterance when conflicting information exists in these channels. A Stroop-type interference paradigm is adopted in which attributes from the verbal and non-verbal dimensions are placed into conflict. Static directional (deictic) gestures and corresponding spoken and written words show symmetrical interference (Experiments 1, 2 and 3) as do directional arrows and spoken words (Experiment 4). This symmetry is maintained when the task is switched from a manual key press to a verbal naming response (Experiment 5) suggesting the mutual influence of the two dimensions is independent of spatial stimulus response compatibility. It is concluded that the results are consistent with a model of interference where information from pointing gestures and speech are integrated prior to the response selection stage of processing.

In the mid 1980's a number of authors published work which criticised the widely held view of gestures and other non-verbal behaviours as "body-language". For instance Rimé (1983) and McNeill (1985) challenged the notion, largely established by Argyle (e.g. Argyle, 1975), that gestures form part of a system of body movements which might offer a privileged means of knowing and perceiving one another, a system thought to follow its own laws and transmit affective, cognitive and regulating mechanisms distinct from those carried by any accompanying speech. McNeill's (1985) article suggested that gestures and speech, far from being psychologically distinct, "share a computational stage; they are, accordingly, parts of the same psychological structure" (p. 350). This prompted rebuttals from Feyereisen (1987) and Butterworth & Hadar (1989) with accompanying replies from McNeill (McNeill, 1987b, 1989). Most seem to agree that gesture production depends, to some extent, on the mechanisms responsible for speech production (see also Rimé, 1983; Kendon, 1983). The arguments centred around specifying the locus of the interaction, elaborating McNeill's conception of *inner speech* as the shared computational stage. This work represented a shift in emphasis from the social impact of non-verbal behavior to an approach which sought to examine the processes underlying the performance of body movements and, in particular, the relationships between these processes and the structures mediating vocal behavior. However, despite a relatively large amount of research on gesture and speech *production*, the field of gesture *comprehension* remains a "neglected field in cognitive psychology" (Feyereisen 1991, p.57). The main aim of this study was to begin to redress this imbalance by studying the comprehension of gestures within an information processing framework. More specifically we ask whether gestures performed concurrently with spoken and written words influence the processing of that verbal signal and reciprocally whether verbal processing modifies the processing of the gestural component of the utterance.

### **Gesture and Speech**

For researchers such as McNeill and Kendon (e.g. McNeill, 1985, 1987b, 1992; Kendon, 1983, 1986), gestural and verbal behaviors serve to represent different aspects of the underlying meaning that a speaker is striving to express. Gesture and speech might both refer to the same event but

each offers a somewhat different view of it. One form of expression is thought to complement the other. Certain gestures might emphasise the shape of the referent, spatial arrangements or actions whilst others might stress intonation, pauses or the logical structure of an utterance. In this way gestures depict a more holistic or “global” aspect of the speaker’s meaning and might serve to structure or “chunk” sections of the discourse. Thus in gesticulation “we observe components of the utterance content that are *not* represented in words” (Kendon, 1986, p.12). An important point for the listener then, is that neither the gestural nor the verbal outcome alone is necessarily sufficient to specify the speaker’s underlying cognitive representation. The complete picture can only be fully appreciated by considering jointly both gestural and verbal behaviours. Under this view a listener might be expected to process both the gestural and verbal components of an utterance, combining this information at some point in processing to provide an integrated representation of the speakers intended meaning. Thus gesture and speech may well share a common computational stage in comprehension as well as in production.

A number of empirical studies have yielded results consistent with the suggestion that listeners process both gestural and verbal information in comprehension.

Graham & Argyle (1975) had speakers describe a set of drawings under conditions where a folded arm posture was maintained and a second where gestures were allowed. The task for the audience was to reproduce these drawings based on the verbal description of the speakers. Drawings resulting from the gesture condition were rated as more similar to the original than those from the no-gesture condition. Moreover an analysis of the discourses produced in the two conditions showed very few differences, suggesting that the advantage to the listeners in the gesture condition resulted from the speaker’s use of hand movements and not because of a reduction in the fluidity in speech that has been noted when movements are restricted (e.g. Graham & Heywood, 1975; Rimé et al., 1984).

Other studies have demonstrated an influence of gestural context on the recall of verbal material. Riseborough (1981) noted that free recall of a list of verbs and cued recall of words from a short story were improved when the presentation of the verbal material was accompanied by

illustrative gestures, often referred to as iconics (e.g. McNeill, 1985). These are the kinds of gesture that present some figural aspect of the object evoked in the accompanying speech (Efron, 1941). An example might be a speaker making a spiralling motion of the finger whilst simultaneously referring to a spiral staircase. Similarly Woodall & Folger (1981, 1985) found superior recall of the verbal content of a conversation when representational gestures (again, iconics) were included.

Finally Thompson & Massaro (1986) demonstrated that both adults and children relied heavily on referential gestures when making decisions concerning the identity of an object simultaneously referred to by, sometimes ambiguous, synthesised speech syllables.

Taken together, these studies suggest that performance in comprehension is influenced by both gestural and verbal information. In this paper the intention was to further explore this hypothesis by adopting a well established procedure familiar to experimental cognitive psychology, but to our knowledge novel to the study of gesture and speech comprehension.

### **Interference Effects**

One way of investigating selective attention or the extent to which certain dimensions are processed separately is to put them into conflict as in the McGurk effect (McGurk & McDonald, 1976) or using a Stroop type interference paradigm (Stroop, 1935; for a recent review see MacLeod, 1991). In the original version of the Stroop colour word interference task subjects 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 colour words in black ink. However, reading the colour word was largely unaffected by the ink colour in which it was printed. This asymmetry has also been demonstrated in picture-word interference tasks where printed words are combined with line drawings of familiar objects. Picture naming is slowed by the presence of an incongruent or unrelated word but word reading is unaffected by the line drawing (e.g. Glaser & Dünghoff, 1984; Smith & Magee, 1980). There is also some evidence that this type of interference exists when opposing dimensions are presented cross-modally (Cowan & Barron, 1987; Shimada, 1990

but see Miles, Madden and Jones, 1989) thus naming visually presented ink colours is slowed when hearing incongruent colour words. Schriefers & Meyer (1990) have demonstrated cross-modal, visual-auditory picture-word interference- longer picture naming latencies were observed when the picture was paired with an associated or unrelated word than in a silence condition.

In a further variation of the Stroop task, Seymour (1973,1974) presented the words ABOVE, BELOW, LEFT and RIGHT to his subjects and demonstrated an interference effect when subjects were required to name the position of the word relative to a dot. Indeed the use of conflicting spatial attributes of various stimuli has been popular in the Stroop literature, for instance Shor (1970, 1971) embedded the words LEFT, RIGHT, UP and DOWN in arrows pointing in directions incongruent with the embedded word and obtained interference in naming the direction of the arrow. Dyer (1972) obtained similar interference effects from the same four directional words which he moved in either the congruent or incongruent direction.

Very few of these studies using stimuli with spatial attributes have investigated whether the interference effect operates symmetrically or not i.e. whether interference is obtained when the task is changed to identifying the verbal dimension of the stimulus as opposed to the pictorial element. Virzi & Egeth (1985), however, managed to reverse the usual asymmetry, obtaining an interference effect from the spatial location (to the left or right of fixation) of a directional word (LEFT or RIGHT) when the task was to make a *manual* response to the meaning of the word. On the other hand the response to the position of the word was unaffected by its meaning.

This particular observation has much in common with the so-called “Simon effect” (after Hedge & Marsh, 1975) first noted by Simon & Rudell (1967). In this paradigm subjects might typically 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. The results indicate that the location of the stimulus provides an irrelevant cue that interferes with the processing of the target stimulus. Thus in Simon & Rudell’s original paper subjects were slower to respond with a right keypress to the word “right” presented in the *left* ear than to an identical stimulus presented to the *right* ear. Thus the reverse Stroop-type effect

observed by Virzi & Egeth (1985) might also be treated as an example of the Simon effect, the irrelevant spatial location of the directional words produced an interference effect when making a manual response to the meaning of the word.

To summarise, Stroop-type interference has been observed with a large variety of stimuli, the usual asymmetrical effect has been noted with colour-word/colour and picture/word stimuli both within and across modalities. The use of the technique has been extended to manual judgements based on the spatial attributes of various stimuli where examples of reverse Stroop-type or Simon interference have been obtained. In view of the diversity of stimuli which have been used in the study of the various interference phenomena and, in particular, the spatial component evident in the Simon task and Stroop variants, it was felt that the application of a cross modal Stroop-type method would be appropriate in the study of the comprehension of deictic gestures and verbal information.

In the various gestural taxonomies, pointing (deictic) gestures have proved rather difficult to pin down. Efron (1941/1971) suggested that they belonged to the class of gestures which could convey meaning independently of verbally expressed utterances whilst others have suggested that they belong in a class of speech-related iconic gestures which carry no such independent meaning (e.g. Ekman & Freisen, 1972; McNeill, 1985; for a review of gesture classification see Rimé & Schiaratura, 1991). So it appears that pointing gestures may fall somewhere on this taxonomic boundary, they may very well carry some autonomous semantic content whilst frequently appearing in a verbal context. For these reasons pointing gestures form a class of non-verbal behaviour which might be expected to influence the processing of information in the verbal channel. Furthermore deictic gestures, unlike many other forms of body movement, can be represented as static images. These images can then be paired with single word utterances forming stimulus pairings similar to those used in the picture-word and spatial variants of the Stroop task reviewed above.

The existence of an interference effect is usually interpreted as evidence that the to-be-ignored attribute of the stimulus has received some degree of processing. Of particular interest in the study

of the interaction of gesture and speech is whether an irrelevant gesture will interfere with the processing of a spoken word. If movements of the hands and/or body provide only redundant information we might not expect them to be attended and as such they should not interfere with the processing of verbal information. On the other hand, McNeill's ideas on the complementarity of gesture and speech suggest that both dimensions should be attended. In this case one might expect a mutual influence of the two sources of information. Gestures should interfere with the processing of information in the verbal channel and reciprocally, verbal information should interfere with the processing of gestural information.

Thus in what follows a Stroop-type paradigm is adopted to investigate any mutual influence of verbal information and deictic gestures which together form two of the components of an utterance. Verbal attributes were paired with either congruent or incongruent non-verbal attributes with a manual response required to both dimensions presented in separate blocks. Experiments 1, 2 and 3 made use of static directional (deictic) gestures with their auditory and visual (written) verbal equivalents respectively. Experiment 4 investigated the "linguistic" status of the deictic gestures used in Experiments 1-3 by replacing them with arrows. Finally in Experiment 5 the response mode was changed from manual to verbal in order to assess the contribution of stimulus-response mechanisms to the effects.

## **Experiment 1**

In the first experiment we paired static directional gestures (up, down, left and right) with their congruent or incongruent verbal equivalents and measured subjects' reaction times in a *manual* response to both verbal and non-verbal dimensions in separate blocks. The "complementarity" hypothesis suggests that listeners combine information in the gestural and verbal channels. In this case we predict a mutual interaction of the two dimensions. Irrelevant gestures should influence the processing of the target verbal stimulus whilst to-be-ignored verbal information should also influence the processing of gestural information. Thus a symmetrical pattern of interference effects is predicted.

There are rather few examples of symmetrical interference to be found in the Stroop literature. Indeed the reverse Stroop effect (interference in word reading caused by an incongruent colour) is notoriously difficult to obtain (e.g. Glaser & Glaser, 1982; Glaser & Dünghoff, 1984; Dunbar & MacLeod, 1984; MacLeod & Dunbar, 1988) and we know of no examples of a reverse effect of picture-word interference (i.e. the interference in word reading caused by an irrelevant picture). The few cross modal studies that exist did not investigate the influence of the colour or picture on the processing of the auditorily presented word (Cowan & Barron, 1987; Shimada, 1990; Schriefers & Meyer, 1990).

When the task is switched from a verbal to a manual naming response the original Stroop effect persists although perhaps slightly attenuated (e.g. Logan et al., 1984; Keele, 1972; Roe et al., 1980). The examples of the Simon effect, which involves manual responses, demonstrate a reversal of the usual asymmetry although in these examples it is the irrelevant location of the target stimulus and not an incongruent picture which contributes to the interference (e.g. Simon & Rudell, 1967; Virzi & Egeth, 1985)

Symmetrical interference patterns appear to be more often reported in the spatial variants of the Stroop task. Shor (1971) reported such effects using directional words embedded in arrows employing a naming task whilst Melara & Marks (1990) observed symmetrical congruity effects using the words HI and LO presented above or below the midline of a computer screen although using the “Garner Interference” paradigm (see Pomerantz, 1983, 1986). More recently O’Leary & Barber (1993) managed to obtain symmetrical interference using the words LEFT and RIGHT to the left and right of the screen with verbal naming responses to either the location or the meaning of the word. Symmetrical effects were also obtained using a left/right keypress response again to the meaning or location of the stimulus word (the Simon effect and reverse Simon effect respectively) but only when the salience of the irrelevant word in the location judgement task was marked.

Thus the predictions arising from the literature are a little confusing. If the gestural stimuli used here are considered as pictures, the available evidence from studies investigating the various

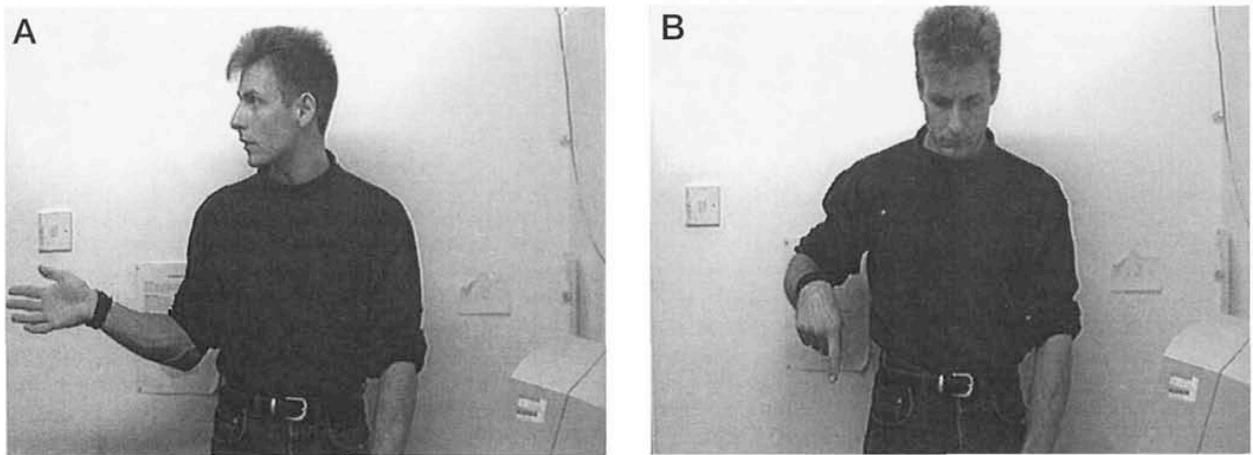
forms of the Stroop effect would predict an asymmetric pattern of interference with little hindrance from the gesture in responding to the auditorily presented word. If, on the other hand, the gestural stimuli indicate a spatial location, the literature on the Simon effect would suggest that the asymmetry should be in the reverse direction, whilst McNeill's theoretical position on the production of gesture and speech predicts symmetrical interference effects. To recapitulate, the complementarity hypothesis predicts a symmetrical pattern of interference effects, congruity effects of equal sizes should be observed when responding to gesture and verbal stimuli respectively manifesting as a main effect of congruity with no interactions with response.

## **Method**

*Subjects:* 14 undergraduate and postgraduate volunteers participated in this experiment, all were between 18 and 30 years of age with normal or corrected to normal vision.

*Materials and Apparatus:* Digitised images of a person gesturing to the left, right, up and down were obtained. Examples of these images are presented in Figure 1. These stimuli subtended approximately 13 degrees of vertical visual angle. The verbal stimuli were recorded using audio software on hypercard and edited using "SoundEdit" software on a Macintosh IIci. Four spoken words (left, right, up and down) were recorded and edited to be approximately the same length (0.8 secs). The visual and auditory stimuli were presented together using the "SuperLab" software on the Mac IIci, this enables the two types of stimuli to be presented simultaneously and claims to record reaction times to millisecond accuracy. The onset of the presentation of the speech stimuli coincided with the presentation of the visual gesture stimulus.

Two types of test stimuli were prepared from the frame grabbed images and the auditory speech stimuli. Congruent trials consisted of the spoken word "up", "down", "left" and "right" paired with their respective gestures whilst incongruent trials consisted of the spoken words "up" and "down" paired with *down* and *up* gestures respectively along with "left" and "right" paired with *right* and *left* gestures respectively.



**Figure 1.** Examples of “gesture” stimuli from Experiment 1.

*Design:* The experiment took the form of a 2 x 2 x 2 repeated measures design. The three factors were: Response (either to a gesture or to a voice), Congruity (congruent or incongruent pairs of stimuli) and Decision (either left/right or up/down decisions). The stimuli were blocked by response, either to gesture (4 cells made up of all combinations of decision and congruity) or to voice (again 4 cells). Half of the subjects responded to the voice first and half to the gesture. With 10 trials per cell, subjects responded to 40 trials in each response block.

*Procedure:* The stimuli were presented in a random order using the “SuperLab” software with the visual and auditory stimuli being presented simultaneously. The subjects were required to make a response by pressing one of four buttons on the keyboard. These keys were selected to give a correspondence between stimulus attributes and response. Thus the keys used were 4, 5, 6 and 8 on the keypad area of the keyboard. These keys possess the same spatial relationships as the responses required in the experiment i.e. “left”, “right”, “up” and “down” and were assigned these values accordingly.

The subjects were instructed to use their preferred hand to depress the keys, using their first and third fingers for the left/right decisions (keys 4 and 6 respectively) and their second finger for the up/down decisions (keys 8 and 5 respectively). The visual stimuli persisted until the subject had made a response which was then followed by a 500 ms interval and then the next trial. The subjects were also instructed to depress the space bar in response to a large question mark which

occasionally appeared on the screen. Eight of these question marks appeared in each response block. These trials were included in order to ensure that the subjects actually watched the screen as the task of responding to the voice could, it was realised, be made trivially easy by shutting ones eyes or staring away from the screen. A set of 10 practice trials were presented before each block, these included two question mark trials and a sample of the other trials, ensuring that all the experimental conditions were represented at least once. Both reaction times and percentage of errors were recorded as dependent variables in the experiment.

## Results

In this, and all other experiments reported, outliers were removed from individual subject's scores by removing those reaction times greater or less than two standard deviations from the mean. The mean correct reaction time scores and percentage of errors recorded in each condition are reported in Table 1.

Table 1.  
Mean RT's (in milliseconds) and Percentage of Errors for Up/Down and Left/Right Decisions to Voices and Gestures in Congruent and Incongruent Conditions of Experiment 1.

Congruity	Up/Down Decisions		Left/Right Decisions		Overall Mean	
	RT	Percentage of Errors	RT	Percentage of Errors	RT	Percentage of Errors
	Voice Responses					
Congruent	687	2.14	739	2.86	713	2.50
Incongruent	773	1.43	798	7.14	786	8.57
<i>M</i>	730	1.79	769	5.00	750	5.54
	Gesture Responses					
Congruent	652	1.43	633	0	643	0.72
Incongruent	707	7.14	658	2.86	682	5.00
<i>M</i>	680	4.29	646	1.43	663	2.86

A 2 (response) x 2 (congruity) x 2 (decision) ANOVA conducted on the reaction time scores revealed a main effect of congruity ( $F(1,13)=21.44, p<0.001$ ) of some 57 ms in the predicted

direction. The absence of an interaction between the response and congruity factors confirmed the symmetrical nature of the interference or congruity effects ( $p=0.24$ ). Subjects were generally faster to respond to the gestural information, an observation confirmed by the main effect of response ( $F(1,13)=6.11, p<0.05$ ). However the interaction of response and decision also reached significance ( $F(1,13)=5.49, p<0.05$ ). Left/right decisions were slowed by some 123 ms when the relevant response dimension was switched from gesture to voice, a difference which proved to be significant ( $p<0.01$ ). Up/down decisions were also slowed but the difference of 50 ms did not reach significance ( $p=0.2$ ).

From Table 1 the error scores generally appear to mirror the reaction time data, the overall mean percentage error was 3.13%. The correlation between RT's and errors was -0.04 suggesting no evidence of a trade-off between speed and accuracy. Because of the relatively low rate of errors in the cells of the design (including scores of 0% in some cases) made by subjects in this and other experiments no further analysis was conducted on the error data.

## **Discussion**

The main result of this experiment was the demonstration of symmetrical Stroop-like interference effects consistent with McNeill's ideas on the complementarity of gesture and speech. This pattern of congruity effects is consistent with the notion that deictic gestures and verbal information are mutually influential in the comprehension process i.e. gestures appear to influence the processing of verbal information and reciprocally verbal information influences the processing of gestural information. This was perhaps a rather striking result in view of the overwhelming number of studies involving Stroop-type methodologies which tend to produce asymmetric effects. The reverse Stroop effect is only obtained by either manipulating the SOA between the two dimensions, dramatically slowing down the reading process, or by practising colour naming (e.g. Stroop, 1935). Here we have demonstrated a reverse, as well as the normal Stroop-type effect merely by manipulating the task instructions. The fact that we have obtained an interference effect at all is interesting given the exchanges between Cowan (e.g. Cowan & Barron, 1987; Cowan, 1989a, 1989b) and Miles (Miles, Maddon & Jones, 1989; Miles & Jones, 1989) concerning the

existence of cross modal Stroop-type interference. Clearly the results of this experiment suggest that a visual-auditory interference effect is at least possible, supporting the conclusions of Cowan and his colleagues (see also Schriefers & Meyer, 1990; Shimada, 1990 for evidence of cross modal effects).

In general the switch from gesture to voice as the relevant response dimension resulted in a significant increase in RT. This may well be due to the relative temporal parameters of the verbal and gestural stimulus presentations. Because visual stimuli are presented almost instantaneously whereas, by their nature, verbal stimuli are temporally extended, the complete identity of a relevant gesture will become available to a subject before the identity of a corresponding verbal stimulus in the voice response condition. Regardless of the relative processing speeds of visual and verbal information, the head start given to the gestural stimuli is likely to produce faster overall RT's to this dimension. However the effect of response was found to interact with that of decision. Left/right decisions were slowed by a larger magnitude than up/down RT's when the relevant response dimension was switched from gesture to voice. This differential effect on left/right decisions was probably due to a left/right confusion effect. A number of studies have shown that it takes longer to make locational discriminations when the relevant spatial dimensions are given the *verbal* labels "right" and "left" than when they are described by terms such as "above" and "below" (e.g. Farrell, 1979; Maki, Grandy & Hauge, 1979; Sholl & Egeth, 1981). This effect is often referred to as the right/left confusion, and is thought to result from the lack of a natural referent in the horizontal dimension caused by the symmetrical right-to-left axis of the human body (Corballis & Beale, 1976). When the response is switched from the gestural to the verbal dimension any relevant visual cue which might be used to resolve the left/right ambiguity is removed. The consequent left/right confusion ensures that decisions to left/right voices are particularly slowed.

To summarise, in this experiment we have demonstrated a normal Stroop-type effect, i.e. interference caused by an irrelevant, or to-be-ignored verbal stimulus in responding to a gestural stimulus. Notably, in this case, the dimensions in question were presented cross modally and

required a manual response. Significantly we have also demonstrated a reverse effect, i.e. interference caused by an irrelevant gesture when responding to a verbal stimulus. This suggests that subjects were able to monitor the visual stimulus and extract spatial information from the gesture even though this was not required by the task demands. This symmetrical pattern of effects is consistent with the complementarity hypothesis derived from McNeill's ideas on the production of gesture and speech. Furthermore, the results of this experiment are suggestive of the existence of cross-modal, visual-auditory interference effects in a picture-word interference task (e.g. Schriefers & Meyer, 1990).

## Experiment 2

It might be the case that the symmetrical effects obtained in Experiment 1 were due to the *cross-modal* nature of these stimuli rather than specifically to the combination of gestural and auditory materials. For instance it is the cross-modal presentation of the stimuli in Experiment 1 which largely contributes to the "head start" in the processing of the gestural dimension discussed above. This processing advantage may well be the cause of the interference effect exerted by the irrelevant gestures in Experiment 1. Relative speed of processing accounts of the Stroop effect rely on just such processing mismatches (e.g. Morton, 1969; Morton & Chambers, 1973; Posner & Snyder, 1975). These "horse race" models are discussed more fully in the general discussion. Briefly, they assume that the faster dimension will interfere with a slower dimension but not vice-versa.

To investigate whether or not the cross modal nature of the stimuli were responsible for the pattern of effects obtained, Experiment 1 was repeated intra-modally. By presenting both dimensions within the visual modality we can ensure that both gestural and verbal information is available to the subject at the same time. Any processing advantage enjoyed by the gestural stimuli due to the temporal extent of the auditorily presented verbal material is therefore eliminated. Thus in Experiment 2 the words "up", "down", "left" and "right" were printed across the chest of the gesturer rather than being presented auditorily. This arrangement is more like the classic picture-word interference paradigm with less confusing predictions. Rosinski et al. (1975) obtained classic

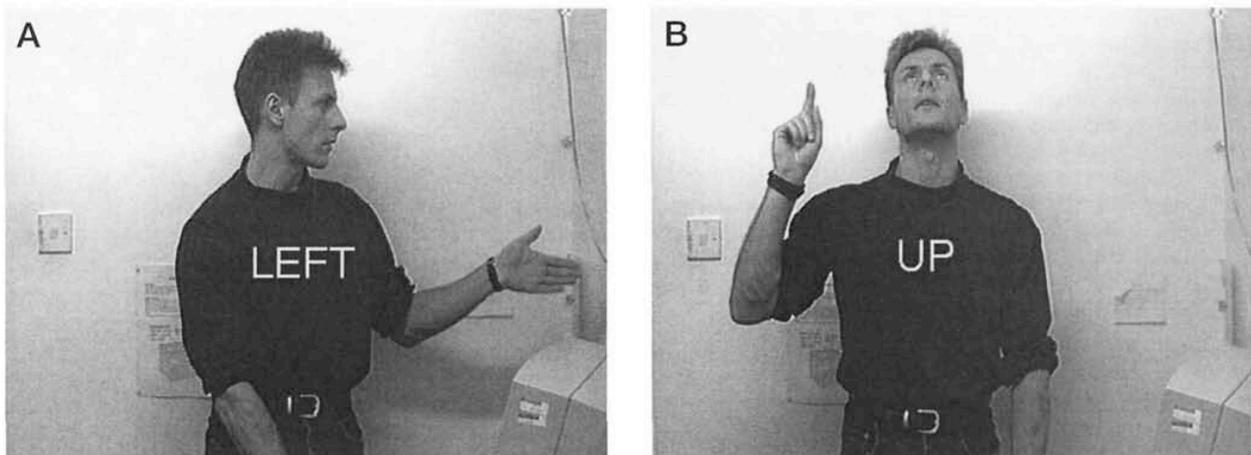
Stroop interference with combinations of picture-word stimuli, that is, incongruent words caused a strong inhibition in the picture naming task whereas pictures had only weak effects on word reading. These effects were also demonstrated by Glaser & Dünghoff (1984) who also largely failed to induce reverse effects comparable in magnitude to the normal interference pattern by pre-exposing the distracting picture. Indeed when both stimuli were presented together incongruent pictures had no effect on word reading.

Thus in this experiment, assuming the stimuli to be analogous to a picture-word paradigm, a normal asymmetric pattern of interference effects was expected, incongruent words should slow responses to the gestures (pictures) but gestures might be expected to have no effect on responding to the words.

## **Method**

*Subjects:* 14 undergraduate and postgraduate volunteers participated in this experiment none of whom had taken part in Experiment 1.

*Materials, Design and Procedure:* These were identical to Experiment 1 with two exceptions. The auditory stimuli were replaced by the equivalent words which were printed across the chest of the gesturer (see Figure 2). The height of each letter subtended approximately one degree of visual angle (the horizontal angle depended on the length of each word). This process was completed using image processing software on the Macintosh. Secondly, as completion of the task required subjects to attend to the visual stimuli presented on the screen the question mark task was not included in the design of this experiment.



**Figure 2.** Examples of congruent and incongruent stimuli from Experiment 2

As before the experiment took the form of a 2 x 2 x 2 within subjects design. The three independent variables were: response (to gesture or word), congruity (congruent or incongruent stimulus pairs) and decision (up/down or left/right decisions). Again both reaction time and percentage of errors were recorded as dependent variables.

## Results

Table 2.

Mean RT's (in milliseconds) and Percentage of Errors for Up/Down and Left/Right Decisions to Words and Gestures in Congruent and Incongruent Conditions of Experiment 2.

Congruity	Up/Down Decisions		Left/Right Decisions		Overall Mean	
	RT	Percentage of Errors	RT	Percentage of Errors	<i>M</i> RT	Percentage of Errors
	Word Responses					
Congruent	675	2.86	633	0	654	1.43
Incongruent	755	2.14	722	5.71	739	3.93
<i>M</i>	715	2.50	678	2.86	697	2.68
	Gesture Responses					
Congruent	626	3.57	561	2.14	593	2.86
Incongruent	686	2.14	628	0.71	657	1.43
<i>M</i>	656	2.86	595	1.43	625	2.15

The average reaction time and error scores obtained under all the experimental conditions are presented in Table 2. There are clear effects of congruity on the RT data for all combinations of response and decision factors. Overall congruent trials received 75 ms faster RT's than incongruent trials. Left/right decisions also appear to be consistently faster than up/down decisions. These observations were largely confirmed by a 2 (response) x 2 (congruity) x 2 (decision) ANOVA . This revealed a main effect of congruity ( $F(1,13)=41.09$ ,  $p<0.001$ ) and an effect of decision which approached significance ( $F(1,13)=4.19$ ,  $p=0.06$ ). No other effects were reliable (all  $p$ 's  $> 0.1$ ). Notably the interaction between response and congruity did not reach significance ( $p=0.33$ ). Thus the results show a symmetrical pattern of interference effects.

The overall mean error score was only 2.42%. The correlation between RT's and errors was 0.19 again suggesting no evidence of a trade-off between speed and accuracy. Because subjects made relatively few errors including an overall mean error rate of 0% in one of the cells in the design, no further analysis was conducted on these data.

## **Discussion**

Here again symmetrical Stroop-type interference has been demonstrated despite the predictions arising from the picture-word interference literature. It seems that the pattern of effects found in Experiment 1 cannot be attributed solely to the cross modal relationship between the stimuli but rather more to the nature of the stimuli per se. It appears that deictic gestures at least, do indeed influence the processing of verbal information even when the "head start" they enjoyed due to the cross-modal presentation in Experiment 1 is eliminated. At the same time, processing of verbal information influences that of non-verbal gestural information. These findings support the idea that listeners attend to both verbal and gestural information in the comprehension process, combining the two sources of information at some point in processing.

The switch from a cross-modal to a within modality presentation of the stimuli also eliminated the response by decision interaction found in Experiment 1. Recall that the main cause of this

interaction was the slowing of responses to the verbal stimulus dimension, particularly to left/right decisions, caused by the temporal extent of the voice stimuli and by a left/right confusion effect.

In Experiment 2, however, there was no significant effect of response and no evidence of a left/right confusion effect. The absence of a response effect might be expected as this was originally thought to be due in part to the temporal extent of the voice stimuli in Experiment 1. In the present experiment, all the stimulus information required by a subject to identify a printed word is available at the same time as that required to identify the gesture. Thus, in contrast to the cross-modal case, we might expect written words to show no RT advantage over gestures. Nevertheless, although non-significant, gesture responses were performed 72 ms faster than average word responses in Experiment 2. In general there appears to be an advantage for keypress responses to gestural over verbal stimuli regardless of the modality of presentation. It is suggested that this might be due to a spatial stimulus-response compatibility (SRC) effect (e.g. Fitts & Seeger, 1953; Simon et al., 1981). We return to the issue of SRC later.

In Experiment 1 left/right decisions were adversely slowed when the response was changed from the gestural to the verbal dimension. However no such left/right confusion was evident in Experiment 2. Instead, left/right decisions showed a consistent trend towards an advantage over up/down decisions for both gestural and verbal response dimensions. This may be due to a “fingering” advantage for left/right over up/down decisions caused by the arrangement of the response keys. This required subjects to respond to up/down stimuli by moving a single finger to one of two keys. In contrast two fingers were used for the left/right decisions so that no initial movement to the location of the correct key was necessary. This presumably results in a natural advantage for left/right decisions.

In summary, Experiment 2 has again yielded a symmetrical pattern of interference effects when both stimulus dimensions were presented in the visual modality and consequently both available for identification at the same time. Again this is suggestive of a mutual influence of gestural and verbal information in comprehension.

### Experiment 3

A possible problem with the interpretation of the deictic gestures used in Experiments 1 and 2 was that subjects may have been somewhat confused as to whose left or right the gestures referred to. The responses “left” and “right” referred to the side of the *image* rather than to the side of the *gesturer* on the image. It is possible that a tendency existed to interpret the left/right gestures from the point of view of the gesturer and not with respect to the image. In order to investigate this possibility Experiment 1 was repeated but with the gesturer turned around to face away from the subject. The ambiguity as to whose left or right the gesture referred to was therefore no longer present.

#### Method

*Subjects:* 14 undergraduate and post-graduate volunteers participated in this experiment, all had normal or corrected to normal vision and hearing.

*Materials, Design and Procedure:* The auditory stimuli were the same as those used in Experiment 1 however the gesture stimuli were replaced by similar images but with the gesturer facing away from the subject. Both the design and procedure were identical to Experiment 1.

#### Results

The mean correct reaction time scores and percentage of errors recorded in each condition are reported in Table 3. Again there are clear effects of congruity across all response and decision conditions. Indeed congruity effects of 85 ms were obtained for each response condition suggesting an exactly symmetrical pattern of interference. Responses to the gesture dimension appear to be made faster than to the voice stimuli. However, as in Experiment 1, this effect of response appears to be greater for left/right as opposed to up/down decisions (109 versus 59 ms).

Table 3.

Mean RT's (in milliseconds) and Percentage of Errors for Up/Down and Left/Right Decisions to Voices and Gestures in Congruent and Incongruent Conditions of Experiment 3.

Congruity	Up/Down Decisions		Left/Right Decisions		Overall Mean	
	RT	Percentage of Errors	RT	Percentage of Errors	RT	Percentage of Errors
			Voice Responses			
Congruent	651	1.43	679	2.86	665	2.15
Incongruent	748	5.00	752	7.14	750	6.07
<i>M</i>	700	3.22	716	5.00	708	4.11
			Gesture Responses			
Congruent	600	1.43	562	1.43	581	1.43
Incongruent	681	5.00	651	1.43	666	3.22
<i>M</i>	641	3.22	607	1.43	624	2.33

These observations were supported by a 2 (response) x 2 (congruity) x 2 (decision) ANOVA conducted on the RT data. This analysis revealed a main effect of congruity ( $F(1,13)=18.13$ ,  $p<0.01$ ) and a main effect of response ( $F(1,13)=7.70$ ,  $p<0.05$ ). The interaction between response and congruity did not reach significance supporting the suggestion of a symmetrical pattern of interference. Finally the interaction between response and decision was also significant ( $F(1,13)=10.82$ ,  $p<0.01$ ). Further analysis of this interaction revealed a significant effect of response for left/right decisions ( $p<0.001$ ) but no effect of response for up/down decisions ( $p=0.07$ ).

As can be seen from Table 3 the error rates in this experiment were fairly low, indeed the overall percentage of errors was 3.22%. The correlation between RT's and percentage of errors was 0.005 offering no evidence of a speed/accuracy trade-off.

## Discussion

The results of this experiment were essentially identical to those of Experiment 1, both in terms of the symmetry of the interference and with regard to the interactive effects of response and

decision factors. Irrelevant gestures influenced RT's to voice stimuli to the same extent that to-be-ignored voices influenced responses to the gestural dimension. Left/right decisions were also differentially slowed when the relevant response was switched from gesture to voice, presumably because of the left/right confusion effect discussed above.

These findings suggest that subjects did not have a problem interpreting the various spatial attributes of the gestures in Experiments 1 and 2 whilst also providing more evidence for the mutual interaction of deictic gestures and verbal information in comprehension.

### **Experiment 4**

The results of Experiments 1 to 3 indicate that subjects do not ignore information contained in the deictic gestures used as irrelevant stimuli in these studies. A relevant question is whether or not this information enjoys a peculiarly linguistic status. For instance it is unclear whether the interference effects caused by the presence of incongruent gestures reflects the operation of a system processing *gestures* per se or one concerned with the manipulation of *visuo-spatial* images. Thus in Experiment 4 we ask whether gestures provide anything different from a non-linguistic, nonverbal spatial cue to direction such as an arrow.

Shor (1971) obtained symmetrical interference effects with directional words embedded in arrows. That is, the direction of the arrow interfered with reading of the word, and word reading disrupted responses to the direction of the arrow. These effects were replicated in a study using a manual response. In view of this result a similar pattern of symmetric interference effects were predicted in Experiment 4.

#### **Method**

*Subjects:* 12 subjects participated in this experiment. All had normal or corrected to normal vision and normal hearing.

*Materials:* The pointing gestures used in the previous experiments were replaced by arrows pointing left, right, up and down. The arrows were black presented on a white background. The

non-pointed end of the arrow was positioned in the centre of the screen, thus a “left” arrow extended from the centre to the left of the screen, subtending a visual angle of some 2.9 degrees. The spoken auditory stimuli were as in the previous experiments.

*Design and Procedure* : These were identical to Experiment 1 with the exception that subjects were now asked to respond to the direction of the arrow in one block and the meaning of the voiced word in a second block. Half of the subjects responded to the voice first and half to the arrow. The order of presentation within the blocks was completely randomised.

## Results

The mean reaction time scores and percentage of errors recorded in each condition of the experiment are presented in Table 4.

Table 4.  
Mean RT's (in milliseconds) and Percentage of Errors for Up/Down and Left/Right Decisions to Voices and Arrows in Congruent and Incongruent Conditions of Experiment 4.

Congruity	Up/Down Decisions		Left/Right Decisions		Overall Mean	
	RT	Percentage of Errors	RT	Percentage of Errors	RT	Percentage of Errors
	Voice Responses					
Congruent	618	4.17	644	0.83	631	2.50
Incongruent	635	2.50	661	3.33	648	2.92
<i>M</i>	627	3.34	653	2.08	640	2.71
	Arrow Responses					
Congruent	513	1.67	476	1.67	495	1.67
Incongruent	571	3.33	496	3.33	533	3.33
<i>M</i>	542	2.50	486	2.50	514	2.50

A 2 (response) x 2 (congruity) x 2 (decision) ANOVA conducted on the reaction time data yielded a main effect of congruity ( $F(1,11)=10.47, p<0.01$ ). Responses to congruent stimuli were made 28 ms faster than to incongruent stimuli. The analysis also yielded a main effect of response

( $F(1,11)=23.03$ ,  $p<0.01$ ) and a significant interaction between response and decision ( $F(1,11)=19.29$ ,  $p<0.01$ ). Again the cause of this interaction appeared to be that left/right decisions were slowed by 167 ms ( $p<0.001$ ) whereas up/down decisions were slowed by 85 ms ( $p<0.01$ ) when the relevant response was changed from arrow to voice. Left/right decisions were also made 56 ms faster than up/down decisions to arrow stimuli ( $p<0.01$ ).

The overall error rate was reasonably low (2.61%). The correlation between RT's and error scores was 0.04 which was not suggestive of a trade-off between speed and accuracy.

## **Discussion**

The results of this experiment again indicate a symmetrical pattern of interference effects. Responses to voice stimuli were slowed in the presence of an incongruent arrow relative to a congruent arrow and vice-versa.

The pattern of reaction times to the various decision and response dimensions also closely paralleled the earlier experiments. The left/right fingering advantage resulted in faster RT's to left/right, compared to up/down decisions when the arrow was the relevant dimension. RT's were slowed when the relevant response was switched from arrow to voice because of the temporal extent of the verbal stimuli. This effect was more marked for left/right as opposed to up/down decisions because of the ambiguity of the horizontal dimension when represented verbally.

On the face of it these observations attest to the similarity of the processing of the gestural and arrow stimuli. It would appear that the gestures used in Experiments 1-3 were processed by a system concerned with the manipulation of visuo-spatial material, rather than one processing gestures per se. However, the overall 28 ms interference effect observed in this experiment is rather smaller than the congruity effects obtained in Experiments 1-3 (56, 75 and 85 ms respectively). Indeed, a between-experiments ANOVA comparing the overall congruity effects of Experiments 1 and 4 largely confirmed this observation, yielding a near-significant interaction between experiment and congruity ( $F(1,24)=3.46$ ,  $p=0.075$ ). In particular the arrows used in this experiment caused only a 17 ms interference effect on responses to voice stimuli compared with

the 73 ms effect of irrelevant gestures in Experiment 1. It seems that arrows cause much less interference than gestures. There may be several reasons for this. Firstly the overall reaction times recorded here (the overall mean RT was 577 ms) were somewhat faster than in the previous experiments. The reduction in the magnitude of the interference effect might simply be a result of these faster responses. Secondly, the relative sizes of the arrows and gestures differ. The arrows of Experiment 4 subtended a visual angle of  $2.9^\circ$  compared with the  $13^\circ$  subtended by the gestures. The greater eccentricity enjoyed by the gestures may contribute to their relatively larger influence on voice responses.

The possibility remains, however, that the quantitative differences in the effects of gestures and arrows reflect a qualitative difference in their processing. Thus it may be premature to conclude, on the basis of the present results, that gestures receive the same kind of visuo-spatial analysis as arrows. Instead it is conceivable that deictic gestures are processed by a specialised system concerned with the identification of gestural material per se. Clearly more direct experimental comparisons are needed to fully explore this question. For now we should simply be mindful of the *possibility* that the effects of pointing gestures are mediated by some form of spatial encoding.

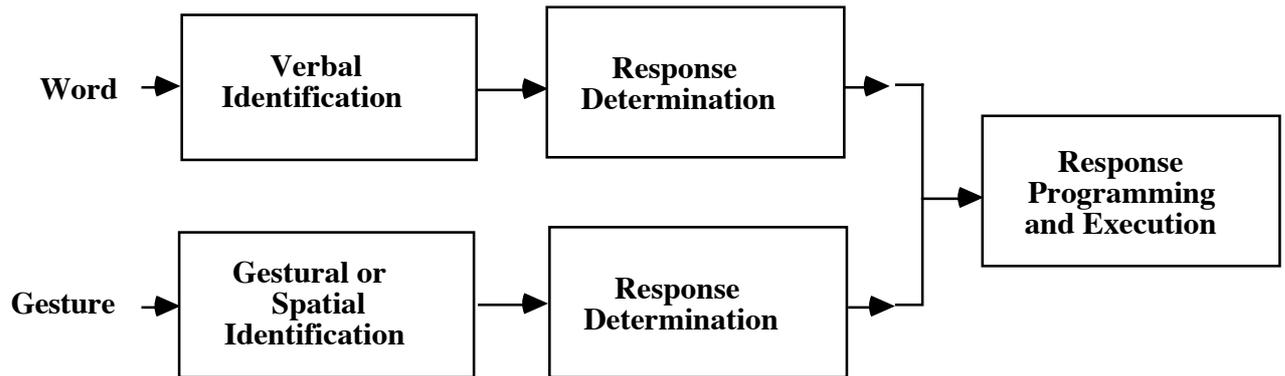
## Experiment 5

Having demonstrated the mutual influence of this particular combination of gestural and verbal information, possible sources of the interaction, in terms of the information processing of the two dimensions, need to be explored. There appear to be several possibilities: gestural and verbal information may interact at a perceptual level of analysis, at a level where information from the early perceptual analysis of the two dimensions contacts semantic memory or at the response selection stage of processing.

Traditional accounts of both the Simon and Stroop effects have suggested that the stage where the response is selected is the most likely source of the interference effect. Models of this type assume that the two dimensions (e.g. colour and word) are processed in parallel culminating in the determination of two separate response codes, one of which must be selected in order for the

correct response to be programmed and executed. Explanations of the Simon effect, for example, have typically centred around the notion of *stimulus-response (S-R) compatibility* which refers to the degree to which responses map onto particular dimensions of the stimulus, possibly as a result of highly over-learned associations between a stimulus and a response. In the Simon effect, it is assumed that both the relevant and irrelevant aspects of the stimulus undergo parallel S-R transformations, thus subjects perform a word-to-position transformation in order to make the correct response whilst simultaneously making a position-to-position transformation in processing the irrelevant location of the stimulus. This irrelevant S-R transformation will proceed automatically if there exists a mapping, or S-R compatibility (SRC), between the possible irrelevant locations of the stimuli and the position of the response keys. For example the directional words LEFT and RIGHT might appear on either the left or right of the screen whilst the response to the directional word must be made by making a left or right manual response. Having encoded both stimuli and their associated responses some arbitration process operates to select the correct response. The Simon effect arises because of the longer arbitration process which occurs if the responses do not correspond. (e.g. Craft & Simon, 1970; Simon, 1982; Mewaldt et al., 1980).

Returning to the present experiments, the S-R model is capable of explaining the interference caused by irrelevant gestures. It has been suggested that subjects might extract spatial information from the pointing gestures since a similar pattern of effects was obtained using arrows as visual stimuli (Experiment 4). This type of spatial code would be suited to the type of manual response employed in Experiments 1 to 4. If this is the case, it is possible that irrelevant deictic gestures exert their effects on directional words because of a spatial SRC which exists between them and the location of the response keys. Under this model, the response code consistent with the irrelevant gesture is automatically activated by virtue of the SRC. Thus two response codes are present at the decision stage, one of which must be selected, programmed and executed. Interference occurs when a conflict between non-corresponding (incongruent) codes must be resolved (see Figure 3).



**Figure 3.** A possible 3-Stage representation of the information processing of gestural and verbal dimensions locating the source of the interference at the response selection stage of processing.

In Experiment 5 the nature of the subject's task was changed from a manual keypress to a verbal naming response. This will have the effect of removing any spatial compatibility which existed between the directional gestures and the response. If gestures exert their effects on verbal information solely by virtue of their spatial compatibility with the response, removing this spatial SRC should have the effect of eliminating any interference effect caused by the irrelevant gestures. However, with the introduction of the verbal naming task, a compatibility now exists between the verbal stimuli and the verbal response. In terms of a SRC account one would therefore expect asymmetry, the irrelevant gestures will no longer interfere with responses to the voice stimuli as the removal of any spatial SRC will ensure that no response will be automatically encoded. On the other hand interference should persist from the to-be-ignored voice stimuli either because of the SRC or because the effects of verbal stimuli are mediated prior to the response selection stage of processing. Thus under the SRC account a normal Stroop-type asymmetry should be obtained i.e. interference from irrelevant *verbal* stimuli but none from irrelevant *visual* stimuli.

A further addition in Experiment 5 was the inclusion of a neutral condition to examine any contribution of facilitation and/or inhibition to the interference effects. If, as suggested, a complete representation of the intended "message" is only obtained from both gestural and verbal information, a congruent gesture, for example, should provide a performance gain compared with a neutral gesture. An incongruent gesture, on the other hand, might be expected to result in a

detriment in performance over a neutral gesture as the listener integrates inconsistent information. Therefore if listeners routinely integrate information from both gesture and speech in comprehension, we might expect facilitation from congruent stimuli as well as inhibition from incongruent stimuli.

## **Method**

*Subjects* : 14 subjects participated in this experiment. All had normal or corrected to normal vision and normal hearing.

*Materials and Apparatus* : In addition to the gesture stimuli used in Experiment 1, a neutral “gesture” was included. This consisted of the same gesturer standing with his arms by his side. The word “blank” was also recorded and edited as before and included as the neutral verbal stimulus.

SuperLab 1.6, which has the capacity for auditory input from voice-key apparatus, was used as the software for this experiment. The voice-key consisted of a normal Macintosh microphone driven by the SuperLab software. Otherwise the specifications of the software were similar to the version used in the previous experiments. The microphone was attached to the subject’s clothing as near as possible to the throat. Output from this microphone was detected by the computer and used to stop the timer.

*Design and Procedure* : The design was similar to Experiment 1 however in this experiment a neutral condition was added to the congruent and incongruent levels of the congruity factor, giving a 2 (response) x 3 (congruity) x 2 (decision) repeated measures design.

Again the stimuli were blocked by response. However, with the addition of the neutral condition, each block now contained 6 cells made up of all combinations of decision and congruity. With 10 trials per cell, subjects responded to two blocks of 60 experimental trials. Each block was preceded by 14 practice trials containing a cross section of the experimental trials as well as two question mark trials. A number of these question mark trials were included in each

response block, upon which subjects were asked to press the space bar. The order of presentation of the blocks was alternated between subjects with the order of presentation of the stimuli within each block randomised.

In this experiment subjects were instructed to either name the direction of the gesture or repeat the auditorily presented direction word depending on the response block in question. The vocal response of the subject stopped the timer which, as before, measured the RT from the onset of the stimulus pair. The response terminated the stimulus display which was replaced by a blank screen. Subjects' responses were recorded on a response sheet by the experimenter. A 500 ms inter trial interval intervened between the subject's response and the presentation of the following trial.

## **Results**

Again both reaction times and error scores were recorded under each condition and these are presented in Table 5. Trials were discarded where subjects' vocal responses failed to reach the threshold of the voice key microphone. The initial response of each subject was recorded, any self-corrections were marked as errors. Again individual RT scores under each condition were rejected if they deviated by more than two standard deviations from the cell mean.

Table 5.

Mean RT's (in milliseconds) and Percentage of Errors for Up/Down and Left/Right Decisions to Voices and Gestures in Congruent, Incongruent and Neutral Conditions of Experiment 5.

Congruity	Up/Down Decisions		Left/Right Decisions		Overall Mean	
	RT	Percentage of Errors	RT	Percentage of Errors	RT	Percentage of Errors
	Voice Responses					
Congruent	684	0	711	0	698	0
Neutral	690	0	703	0	697	0
Incongruent	722	0	731	0.71	727	0.36
<i>M</i>	699	0	715	0.24	707	0.12
	Gesture Responses					
Congruent	650	0.71	680	0.71	665	0.71
Neutral	650	0.71	675	0.71	663	0.71
Incongruent	657	2.14	705	2.14	681	2.14
<i>M</i>	652	1.78	687	1.19	670	1.19

The reaction time data were entered into a 2 (response) x 3 (congruity) x 2 (decision) ANOVA. This analysis yielded a main effect of decision ( $F(1,13)=23.92, p<0.001$ ) with up/down responses being made 25 ms faster than left/right responses (676 versus 701 ms). The main effect of congruity was also significant ( $F(2,26)=16.08, p<0.001$ ). Post hoc Tukey tests indicated reliable differences between congruent and incongruent conditions and between incongruent and neutral conditions (both  $p$ 's<0.01) but no significant difference between congruent and neutral conditions. This translates to an overall congruity effect (i.e. the difference between incongruent and congruent conditions) of 23 ms (681 versus 704 ms) which consists of inhibition (24 ms) but no facilitation.

As in previous experiments, the error rate was very low (the overall mean percentage error was 0.66%). The correlation between RT's and errors was -0.1 suggesting no evidence of a trade-off between speed and accuracy.

## Discussion

Again the most notable finding from this experiment was the symmetrical nature of the interference effect as in the previous experiments. In particular interference from an irrelevant gesture was obtained despite the lack of any spatial SRC with the verbal response. Irrelevant verbal stimuli also interfered with responses to gesture stimuli as predicted. Furthermore the analysis of the congruity effects is consistent with a pattern of inhibition without facilitation.

Recall from the introduction to this experiment that under the S-R model, an interference effect should only be obtained if two conflicting codes are present at the response selection stage of processing. Under this model, removal of the spatial SRC between the gesture responses should prevent the encoding of the “gesture” response, leaving the word-based response code unopposed. However, whilst the manipulation made in this experiment failed to eliminate this interference, it did result in a reduction in the size of the congruity effect compared with Experiments 1-3. Again, a between-experiments ANOVA comparing overall congruity effects in this experiment with those of Experiment 1 yielded a significant interaction between experiment and congruity ( $F(1,26)=6.68$ ,  $p<0.05$ ) lending statistical support to this observation. This is somewhat at odds with the literature on Stroop-type interference which suggests that the size of the effect should be larger when the response modality is switched from manual to oral (MacLeod, 1991). Moreover, since mean RT's in this experiment are similar in magnitude to those of Experiments 1-3, the reduction in the size of the congruity effect cannot be attributed to an overall decrease in RT. Instead it remains possible that spatial SRC *contributes* to the interference effects reported in this paper. Indeed, a number of other experiments conducted in our laboratory have led to similar conclusions (Langton, in prep). In these experiments the S-R mapping between the stimulus and the manual responses was manipulated. For example, in an incompatible mapping condition, a left/right stimulus would require an up/down manual keypress response. Under these conditions the size of the interference effect caused by an irrelevant gesture was reduced, but not completely eliminated compared to a compatible mapping condition. The somewhat inconclusive findings of these experiments prompted the change in the nature of the response adopted in Experiment 5. Here the

spatial mapping between a gesture and a response was completely eliminated rather than simply manipulated. Taken together, these findings suggest that spatial stimulus response compatibility may well play a role in producing interference between the two dimensions in question. However, since the effect persists in its absence, spatial SRC is clearly not the only factor involved.

One other possibility, consistent with an S-R model, is that a response code was indeed activated from an irrelevant gesture, not because of *spatial* SRC, but because of some sort of *symbolic* SRC between the verbal labels associated with the gestures and the verbal responses (e.g. Simon, Sly & Vilapakkam, 1981). That is, some form of compatibility might exist between the directional gestural stimuli and both the keypress *and* verbal responses. This ensures that a response is always encoded from the gestural input regardless of the response modality (keypress or verbal). An interference effect caused by an irrelevant gestural stimulus will, therefore, always result. However this explanation emphasises a potential problem with the SRC account of interference effects. The difficulty is in defining exactly what constitutes compatibility (e.g. faster RT's to compatible versus incompatible stimuli, past experience with the pairings involved etc.). With no *a priori* definition it becomes impossible to generate direct predictions about performance (McClain, 1983) which rather limits the application of SRC to models of interference. In summary we can conclude that the interference effects caused by to-be-ignored gesture stimuli, whilst probably influenced by the spatial relationship between the various stimuli and the responses, are not caused solely by *spatial* SRC. An explanation based on *symbolic* SRC is, however, possible.

The second point of note arising from this experiment was the observation of inhibition without facilitation from both irrelevant gesture and verbal stimuli. This finding does not seem consistent with the hypothesis that listeners routinely integrate information from both gesture and speech in comprehension. If this were the case we might expect facilitation from congruent stimuli as well as inhibition from incongruent stimuli. Instead the current findings appear to be more in keeping with the ideas of Bernard Rimé (e.g. Rimé, 1983; Rimé & Schiaratura, 1991) who suggests that hand gestures provide redundant information to speech and as such are largely

ignored in the comprehension process. Under this model a listener does not integrate information from both channels but, in the interests of processing economy, diverts the lion's share of his or her attentional resources to the speech channel, whilst the speaker's gestures remain in the periphery of the listener's attention. However Rimé adds that this normal relationship between the verbal and non-verbal channels may slip and reverse temporarily so that the non-verbal information becomes the figure and the verbal data the ground. This situation is likely to arise when either the intensity of the non-verbal channel rises e.g. the use of unfamiliar, bizarre or discordant gestures with regard to the context or situation, or the intensity of the verbal channel falls, possibly in noisy, confusing or complex situations. Therefore, under "concordant" conditions attention is allocated to the verbal channel whilst under "gesture neutral" conditions the same will be true. The attention shift only occurs when some discordant or bizarre gesture appears in the non-verbal channel. Thus there should be no processing "gain" for concordant gestures but a detriment in performance with incongruent gestures. This translates to a pattern of performance characterised by inhibition without facilitation as found in this experiment.

Alternatively it may simply be the case that verbal identification in this experiment is at ceiling level, so that no benefit of a congruent gesture could occur. On the other hand, conflicting gestural information could still be capable of slowing down voice responses, producing an inhibitory effect. This same argument could well apply to the interfering effect of voice stimuli on responses to the gestural images.

However we should be cautious about making any strong claims on the basis of this pattern of results for at least two reasons. Firstly the absence of an effect (in this case facilitation) is not sufficient to reject any hypothesised model. It may be that the choice of neutral condition was not appropriate or that the procedure was not sufficiently sensitive to reveal any such facilitative effects. Secondly, and relatedly, Lindsay & Jacoby (1994) have suggested that facilitation and inhibition in the Stroop task cannot be measured accurately using "neutral" control items, instead they suggest that the so-called "process dissociation procedure" (Jacoby, 1991) is more

appropriate. In view of these points we should be cautious in making any claims based on the failure to demonstrate a facilitative effect of congruent gestures in Experiment 5.

In summary Experiment 5 has demonstrated the persistence of an interfering effect of irrelevant gestures despite the removal of any spatial stimulus response compatibility between the gestures and their associated manual responses present in the previous experiments. However an account based on a more general symbolic compatibility cannot be ruled out. The observation of inhibition without facilitation is somewhat problematic for the “complementarity hypothesis” but the well established difficulties in teasing out the effects of facilitation from inhibition make any firm conclusions based on this finding impossible.

### **General Discussion**

The primary aim of these experiments was to investigate the possibility that listeners process and integrate information from both the gestural and verbal channels in the comprehension of an utterance based on David McNeill’s ideas on the complementary nature of gesture and speech (the complementarity hypothesis). Experiment 1 revealed a symmetrical pattern of interference effects in a Stroop-type paradigm, a finding consistent with the hypothesis that listeners attend to both sources of information. Auditorily presented verbal information interfered with the processing of concurrently displayed static deictic gestures and vice-versa. Furthermore the effects were found not to be due to the specific cross-modal nature of the stimuli (Experiment 2) nor to any confusion arising from the interpretation of the directional gestures (Experiment 3). The results of Experiment 4 suggested that the symmetrical effects were not peculiar to deictic gestures but are also obtained using directional arrows and words, raising the possibility that subjects were extracting spatial information from the gestures used in Experiments 1 to 3. Finally symmetrical interference effects persisted in Experiment 5 despite a switch from a manual to a verbal mode of response, suggesting that the effects of to-be-ignored deictic gestures persist independently of any spatial stimulus response compatibility (SRC).

To the best of our knowledge the application of a Stroop-type interference procedure to the study of gesture and speech is a novel one whilst the use of cross-modal stimuli within such a procedure is also rare. The observation of Stroop-type interference with such cross-modal stimuli is also interesting given that the existence of such an effect has been challenged (e.g. Miles, Maddon & Jones, 1989). Finally, the results of the present experiments make an addition to a number of recent studies which have yielded symmetrical interference effects within both the Stroop and Garner interference paradigms and with a variety of dimensional combinations (e.g. Melara & O'Brien, 1987; Melara, 1989; Melara & Marks, 1990; Melara & Mounts, 1993; O'Leary & Barber, 1993). As we shall see, such observations represent a challenge to many models of Stroop and picture-word interference.

In the remainder of this discussion we address other response selection accounts of Stroop interference as well as models locating the interference effect prior to response selection. It is suggested that models of this sort might well be more successful in accounting for the symmetrical nature of the interference. We also discuss the possibility that referential gestural and verbal stimuli are integrated by an algorithm which places more "weight" on the most discriminable, or least ambiguous, source of information. We conclude by discussing the spatial nature of deictic gestures and the processing of other types of gestures with rather different properties.

### ***Response Selection Accounts***

An account based on SRC is not the only explanation of Stroop-type interference. At least two other models place the locus of the effect at a late, response selection stage of processing. These are the relative speed-of-processing model and the translational model.

#### ***Relative Speed of Processing Models***

Speed-of-processing models (e.g. Morton, 1969; Morton & Chambers, 1973; Posner & Snyder, 1975) assume that the two codes in question (e.g. colour and word, or perhaps "gesture" and word) "race" to gain entry to a single channel response buffer. Priority for entry into this buffer is determined by time of arrival. Interference occurs when an irrelevant response code wins the race

to the response buffer and has to be overcome in order for the correct response to be selected. For instance, in the original Stroop task it is assumed that words are processed more quickly than colours, thus the word-based code must be overcome in favour of the colour-based response code. This mismatch in the processing speeds of colours and words is what produces the typical asymmetry of the Stroop effect. The fact that gesture stimuli were processed faster than verbal stimuli in Experiments 1 and 3 might explain the effects of irrelevant gestures, however the slower dimension (verbal) was still able to exert effects on the dimension receiving faster processing (gesture) in both these experiments. Moreover when the relative response times for the critical dimensions were matched (Experiments 2 and 5) symmetrical interference was obtained. Thus any model reliant upon assumptions of unequal speeds of processing will not support the findings of the experiments reported here. Indeed the speed of processing account has been cogently criticised by a number of authors (e.g. Glaser & Glaser, 1982; Glaser & Dünghoff, 1984; Dunbar & MacLeod, 1984; MacLeod & Dunbar, 1988).

### *The Translational Model*

Virzi & Egeth's (1985) translational model (cf. Glaser & Glaser, 1989) also contains multiple processing systems which operate in specific domains and handle specific types of information e.g. colour, pitch, verbal etc. Each system contains an input analyser, decision stage and in some cases a specific output or response stage. Responses are slowed if input to one system requires translation for output in a separate system. In the Stroop effect, for instance, the theory postulates one system devoted to the analysis of verbal input stimuli and verbal (speech) output along with a second system concerned with colour analysis. If the colour of a stimulus has to be named, the encoded colour information must undergo translation into a verbal code for speech output in the verbal system. Competing codes, one translated and one untranslated, will both be present at the same verbal decision stage. The resolution of this conflict is what causes the interference effect. On the other hand when a translation is not needed, for instance when the task is to read the word of a colour-word stimulus, only one code will be present at each of the separate decision stages.

Consequently there is no conflict to resolve and the correct response can be programmed and executed without interference.

In Virzi & Egeth's model of processing, we might assume that the spatial components of the gestural stimuli are extracted by some spatial processing system (as suggested in Experiment 4) producing spatial codes suited to the manual responses. When the required response is to the verbal stimuli, the verbal code must be translated into a spatial code for a manual response to be performed. Thus, in terms of the model, two spatial codes would be present at the decision stage of the spatial processing system, resulting in interference. However, no translation of the verbal code is necessary when subjects are required to make a verbal response to a verbal stimulus as in Experiment 5. Thus the translation model predicts no interference from an irrelevant gesture when the relevant stimulus and response are verbal in nature. The results of Experiment 5, however, suggest that such an interference effect is evident. Moreover the interference effects caused by irrelevant verbal stimuli in Experiments 1 to 4 are not predicted by the translation model.

In short, neither the speed-of-processing nor the translational account can cope with interference effects observed in the present experiments. These models were essentially developed to account for the normal asymmetry of the Stroop effect and, consequently run into difficulty when faced the symmetrical nature of the interference noted here.

### *Alternative Accounts*

#### *Perceptual Encoding Accounts*

It may be that models placing the locus of the interference effect prior to the response selection stage might be more successful in accounting for the current findings. The McGurk effect (McGurk & McDonald, 1976) has demonstrated how visual information can influence the *perception* of speech sounds. In a typical “McGurk” experiment, subjects might see the shape of a mouth uttering the syllable /ga/ whilst being simultaneously presented with the sound /ba/. In this instance subject's often report hearing a blend of the two pieces of information, the syllable /da/. Although it is unlikely that hand gestures could facilitate the phonemic discrimination of speech

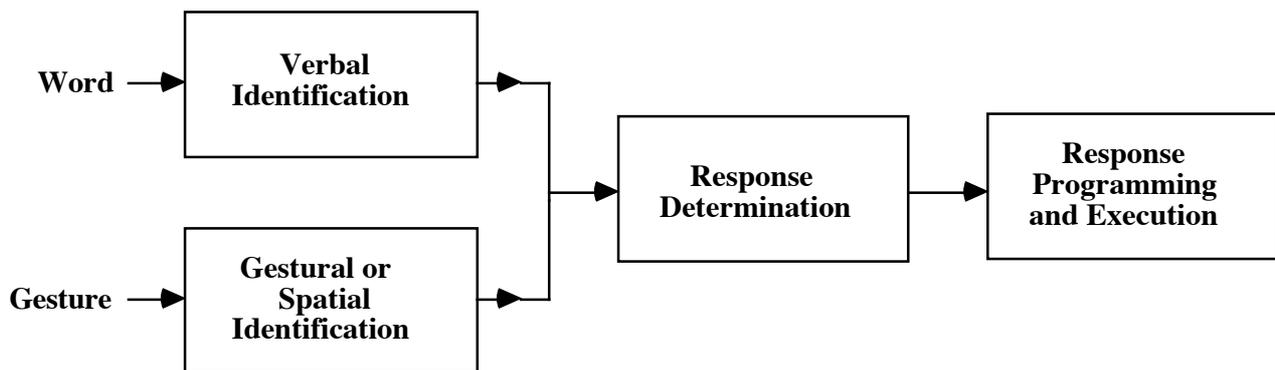
sounds, certain “beat” like gestures (McNeill, 1987a) might act as an aid to the processing of intonation patterns, or prosody, in speech. These types of hand movements frequently occur in close coordination with the rhythmic nature of speech. Moreover the maximal amplitude of a such a gesture will often correspond to a loudness peak in the concurrent speech stream (e.g. Chang & Hammond, 1987). Thus these gestures might provide a visual cue to prosody in the same way that the shape of the mouth provides a cue to the identity of a phoneme.

The perceptual encoding account of the Stroop effect (Hock & Egeth, 1970) has, however, been rejected by Dyer (1973) whilst more recently Simon & Berbaum (1990) have discounted this as an explanation of both the Stroop and Simon effects. Indeed Jacoby (1991) has recently used a “process dissociation procedure” to suggest that colour naming and word reading can operate *independently* to determine responses in the Stroop task. This evidence seems to challenge Hock and Egeth’s notion that the Stroop effect occurs because word meanings affect the perception of colour. In view of this it seems unlikely that the interaction between the two dimensions considered here, dimensions which share commonalities with both Stroop-type and Simon procedures, occurs at this precategorical level of processing.

### ***Conceptual Encoding Accounts***

An alternative explanation places the locus of the effect between perceptual encoding and response selection. For example Melara & Marks (1990) interpret their findings of cross modal congruity effects, using the Garner interference paradigm, as having a semantic basis, whilst Seymour (1977) concluded that the Stroop effect is located during conceptual encoding, a view supported by Simon & Berbaum (1988) and endorsed as a plausible explanation of both Stroop and Simon effects by Hasbroucq & Guiard (1991). In models of this type (see also Logan, 1980; O’Leary & Barber, 1993) the separate response determination or decision stages of the speed-of-processing and translation models are replaced by a single decision mechanism (see Figure 4). Information from multiple sources (e.g. words and colours, or words and gestures) is processed in parallel but is integrated prior to response selection. The appropriate response is determined on the basis of this pooled information. For instance in Logan’s (1980) model, each dimension

contributes “evidence” over time for a particular decision. The relative contribution of each dimension to the decision is determined by two weights: a stable automatic weight; and a flexible, strategic attentional weight. A response can only be made when the threshold of activation for a particular decision is reached. Symmetrical interference could be obtained in this model if the stable automatic weights for both dimensions are similar.



**Figure 4.** A possible 3-Stage representation of the information processing of gestural and verbal dimensions locating the source of the interference prior to the response selection stage of processing.

### *The Integration of Gestural and Verbal Information*

It is suggested that the gestural and verbal information is processed in parallel and integrated at a point prior to response selection. This is consistent with the conclusions of Thompson & Massaro (1986, 1993) in their studies of speech and pointing gestures in referential understanding. They suggest that the evaluation of pointing gestures occurs independently of the evaluation of linguistic material. Furthermore these sources of referential information were found to be integrated with a rule best described as a multiplicative algorithm. That is, the two types of information are combined in such a way that the least ambiguous source is more influential in the decision process. Along these lines, Rogers (1979) noted that the facilitatory effects of visual cues (i.e. from face and body) on the comprehension of audio messages (assessed by forced choice responses to factual questions) were inversely related to the signal-to-noise ratio, i.e. the noisier the speech the more subjects relied on visual cues.

The idea that visual and verbal information is combined multiplicatively is similar to a view of dimensional interaction which relies on the relative discriminability of the various dimensions (e.g. Garner & Felfoldy, 1970; Melara & Mounts, 1993). Briefly Melara & Mounts (1993) have found that when the classic Stroop dimensions of colour and colour word were mismatched in baseline discriminability, the more discriminable (or less ambiguous) dimension disrupted classification of the less discriminable (or more ambiguous) dimension. In Logan's (1980) terms this equates to unequal stable automatic weights governing the contributions of the two dimensions. When baseline discriminabilities were matched, Melara & Mounts obtained small but symmetrical congruity effects in a Stroop task.

In order to examine whether an account based on discriminability could be applied to the present findings, we obtained measures of relative discriminability (calculated by subtracting overall gesture RT's from voice RT's) from each of 10 experiments. These included Experiments 1, 2, 3 and 5 of the present paper as well as six unpublished experiments of similar design conducted in our laboratory (Langton, in prep). Positive discriminability scores are obtained when gestures are more discriminable than words, and on Melara and Mounts' analysis, these scores should correlate positively with the interfering effect of gestures on responses to words, and negatively with the interfering effect of words on gestures. It should be noted, however, that these baseline measures are not exactly equivalent to those used by Melara & Mounts (1993). They obtained their baselines by examining subject's choice RT's to a particular dimension (e.g. colour) whilst the irrelevant dimension remained constant (e.g. the word GREEN). Nevertheless it was felt that overall mean RT's to each dimension would provide a baseline which would be sufficient to explore the hypothesis. In line with Melara & Mounts' analysis, the discriminability scores we obtained were found to correlate positively with the size of the interference effect caused by to-be-ignored gestures ( $r=0.81$ ,  $p<0.01$ ). In other words, as gestural stimuli became more discriminable than verbal stimuli, they caused increasing larger interference effects when subjects were asked to classify the verbal dimension. The correlation between discriminability and the magnitude of the interference effects caused by irrelevant verbal stimuli was only  $-0.3$  ( $p=0.38$ ). Thus, although not as compelling, it appears that the interference effect caused by irrelevant verbal information is

large when voices are *more* discriminable than gestures and decreases slightly as voices become *less* discriminable than gestures.

Thus, it may well be that the relative discrimination of the two dimensions is critical in determining the size of the interference effect. This finding is in accordance with Melara & Mounts (1993) and, we suggest, is consistent with a multiplicative integration of the gestural and verbal dimensions (e.g. Thompson & Massaro, 1986, 1994).

In summary, it is suggested that the symmetrical nature of the interference and the persistence of this effect in the face of the manipulation to the S-R relationship, is best explained within a model of processing where the gestural and verbal stimuli are analysed in parallel with a single response determined by some integration of the two sources of information. Moreover this integration process appears to be sensitive to the relative discriminability, or ambiguity, of the dimensions in question and may well ensure that the least ambiguous source of information contributes more “evidence” to the decision.

### ***Visuo-Spatial Processing of Gestures***

The results of Experiment 4, accompanied by the fact that we have employed static rather than dynamic gestures, raise the possibility that the gestural stimuli are processed by some spatial system rather than one devoted to the analysis of gestural material per se. This putative system might well be the locus of the effects reported here. Indeed Toms, Morris & Foley (1994) have suggested that such a system might exist within the central executive component of working memory (e.g. Baddeley, 1981, 1983). This device would be free to operate on spatial material residing in any of the passive perceptual stores. Thus information from both the phonological and visuo-spatial stores might be integrated at this central component where a spatial response is encoded. Others have suggested the existence of a subsystem of working memory that is specific for movement configuration (Smyth et al., 1988; Smyth & Pendleton, 1989) but distinct from memory of movement to spatial targets. This leaves open the possibility that information from other types of gestures might receive some specialised analysis. Moreover there is some evidence

to support the notion of a gesture “module”. Rothi et al. (1986) have identified a patient with so-called gesture agnosia, impaired discrimination and comprehension of gestures with no associated auditory comprehension deficit or general visual impairment.

### ***Conclusions***

We have presented some evidence for the mutual influence of deictic gestures and verbal information in comprehension. Further work needs to be done to investigate the processing of other forms of gesture and the possibility that they too are integrated with information contained in the speech channel. For instance we might expect so-called iconic gestures (e.g. the type of hand movement a fisherman might make in exaggerating the size of “the one that got away”) to interact with verbal information as these gestures are only performed with speech (e.g. McNeill, 1985), and cannot be understood without reference to the verbal component of the utterance (e.g. the words, “It was this big”). Conversely, emblematic gestures (e.g. the hitchhiking gesture) might not be expected to influence the processing of verbal information since they are often performed in the absence of speech. These questions could well be addressed by extending the scope of the present methodology to include *dynamic* emblematic and iconic gestural stimuli.

## References

- Argyle, M. (1975). *Bodily Communication*. London: Methuen.
- Baddeley, A. D. (1981). The concept of working memory: A view of its current state and probable future development. *Cognition*, *10* , 17-23.
- Baddeley, A. D. (1983). Working memory. *Philosophical Transactions of the Royal Society of London, B* *302* , 311-324.
- Butterworth, B., & Hadar, U. (1989). Gesture, speech and computational stages: A reply to McNeill. *Psychological Review*, *96* , 168-174.
- Chang, P., & Hammond, G. R. (1987). Mutual interactions between speech and finger movements. *Journal of Motor Behavior*, *19* , 265-274.
- Corballis, M. C., & Beale, I. L. (1976). *The psychology of left and right*. Hillsdale, N.J.: Erlbaum.
- Cowan, N., & Barron, A. (1987). Cross-Modal auditory visual Stroop interference and possible implications for speech memory. *Perception and Psychophysics*, *41* , 393-401.
- Cowan, N. (1989a). A reply to Miles, Madden, and Jones (1989): Mistakes and other flaws in the challenge to the cross-modal Stroop effect. *Perception and Psychophysics*, *45* , 82-84.
- Cowan, N. (1989b). The reality of cross-modal Stroop effects. *Perception and Psychophysics*, *45* , 87-88.
- Craft, J. L., & Simon, J. R. (1970). Processing symbolic information from a visual display: Interference from an irrelevant cue. *Journal of Experimental Psychology*, *83* , 415-420.
- Dunbar, K., & MacLeod, C. M. (1984). A horse race of a different colour. Stroop interference patterns with transformed words. *Journal of Experimental Psychology: Human Perception and Performance*, *10* , 622-639.

- Dyer, F. N. (1972). Latencies for movement naming with congruent and incongruent word stimuli. *Perception and Psychophysics*, *11* , 377-380.
- Dyer, F. N. (1973). The Stroop phenomenon and its use in the study of perceptual, cognitive and response processes. *Memory and Cognition*, *1* , 106-120.
- Efron, D. (1941/1972). *Gesture, race and culture*. The Hague: Mouton.
- Ekman, P., & Freisen, W. V. (1972). Hand movements. *Journal of Communication*, *22* , 353-374.
- Farrell, W. S. (1979). Coding left and right. *Journal of Experimental Psychology: Human Perception and Performance*, *5* , 42-51.
- Feyereisen, P. (1987). Gestures and speech, interactions and separations: A reply to McNeill. *Psychological Review*, *94* , 493-498.
- Feyereisen, P. (1991). Brain pathology, lateralization, and nonverbal behavior. In R. S. Feldman. & B. Rimé (Eds.), *Fundamentals of nonverbal behavior* . Cambridge: Cambridge University Press.
- Fitts, P. M., & Seeger, G. M. (1953). S-R compatibility: Spatial characteristics of stimulus and response codes. *Journal of Experimental Psychology*, *46* , 483-491.
- Garner, W. R., & Felfoldy, G. L. (1970). Integrality of stimulus dimensions in various types of information processing. *Cognitive Psychology*, *1* , 225-241.
- Glaser, M. O., & Glaser, W. R. (1982). Time course analysis of the Stroop phenomenon. *Journal of Experimental Psychology: Human Perception and Performance*, *8* , 875-894.
- Glaser, W. R., & Dünghoff, F. J. (1984). The time course of picture-word interference. *Journal of Experimental Psychology: Human Perception and Performance*, *10* , 640-654.
- Glaser, W. R., & Glaser, M. O. (1989). Context effects in Stroop-like word and picture processing. *Journal of Experimental Psychology: General*, *118* , 13-42.

- Graham, J. A., & Argyle, M. (1975). A cross-cultural study of the communication of extra-verbal meaning by gestures. *International Journal of Psychology*, *10*, 57-67.
- Graham, J. A., & Heywood, S. (1975). The effects of the elimination of hand gestures and of verbal codability on speech performance. *European Journal of Social Psychology*, *5*, 189-195.
- Hasbroucq, T., & Guiard, Y. (1991). Stimulus-response compatibility and the Simon effect: Toward a conceptual clarification. *Journal of Experimental Psychology: Human Perception and Performance*, *17*, 246-266.
- Hedge, A., & Marsh, N. W. A. (1975). The effects of irrelevant spatial correspondence on two-choice response time. *Acta Psychologica*, *39*, 427-439.
- Hock, H. S., & Egeth, H. (1970). Verbal interference with encoding in a perceptual classification task. *Journal of Experimental Psychology*, *83*, 299-303.
- Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*, 513-541.
- Keele, S. (1972). Attention demands on memory retrieval. *Journal of Experimental Psychology*, *93*, 549-561.
- Kendon, A. (1983). Gesture and speech: How they interact. In J. M. W. & R. P. Harrison (Eds.), *Nonverbal interaction* Beverley Hills. CA: Sage Publications.
- Kendon, A. (1986). Some reasons for studying gesture. *Semiotica*, *62*, 3-28.
- Langton, S. R. H. (in prep). *Exploring interference effects in the processing of verbal and gestural components of an utterance*. Unpublished doctoral thesis, University of Nottingham.
- Lindsay, S. D., & Jacoby, L. L. (1994). Stroop process dissociations: The relationship between facilitation and interference. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 219-234.

- Logan, G. D. (1980). Attention and automaticity in Stroop and priming tasks: Theory and data. *Cognitive Psychology*, *12* , 523-553.
- Logan, G. D., Zbrodoff, N. J., & Williamson, J. (1984). Strategies in the color-word Stroop task. *Bulletin of the Psychonomic Society*, *22* , 135-138.
- MacLeod, C. M., & Dunbar, K. (1988). Training and Stroop-like interference: Evidence for a continuum of automaticity. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *14* , 126-135.
- MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, *109* , 163-203.
- Maki, R. H., Grandy, C. A., & Hauge, G. (1979). Why is telling right from left more difficult than telling above from below? *Journal of Experimental Psychology: Human Perception and Performance*, *5* (52-67).
- McClain, L. (1983). Stimulus-response compatibility affects auditory Stroop Interference. *Perception and Psychophysics*, *33* , 266-270.
- 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. (1987a). *Psycholinguistics: A new approach*. New York: Harper & Row.
- McNeill, D. (1987b). 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. (1992). *Hand and mind: What gestures reveal about thoughts..* Chicago: University of Chicago Press.

- Melara, R. D. (1989). Dimensional interaction between colour and pitch. *Journal of Experimental Psychology: Human Perception and Performance*, 15, 69-79.
- Melara, R. D., & Marks, L. E. (1990). Processes underlying dimensional interactions: Correspondences between linguistic and nonlinguistic dimensions. *Memory and Cognition*, 18, 477-495.
- 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.
- Melara, R. D., & O'Brien, T. P. (1987). Interaction between synesthetically corresponding dimensions. *Journal of Experimental Psychology: General*, 116, 323-336.
- Mewaldt, S. P., Connelly, C. L., & Simon, J. R. (1980). Response selection in choice reaction time: Test of a buffer model. *Memory and Cognition*, 8, 606-611.
- Miles, C., Madden, C., & Jones, D. M. (1989a). Cross-modal, auditory-visual Stroop interference: A reply to Cowan and Barron (1987). *Perception and Psychophysics*, 45, 77-91.
- Miles, C., & Jones, D. M. (1989b). The fallacy of the cross-modal Stroop effect: A rejoinder to Cowan (1989). *Perception and Psychophysics*, 45, 85-86.
- Morton, J. (1969). Categories of interference: Verbal mediation and conflict in card sorting. *British Journal of Psychology*, 60, 329-346.
- Morton, J., & Chambers, S. M. (1973). Selective attention to words and colours. *Quarterly Journal of Experimental Psychology*, 25, 387-397.
- Nicoletti, R., Anzola, G. P., Luppino, G., Rizzolatti, G., & Umiltà, C. (1982). Spatial compatibility effects on the same side of the body midline. *Journal of Experimental Psychology: Human Perception and Performance*, 8, 664-673.

- Nicoletti, R., & Umiltà, C. (1984). Right-left prevalence in spatial compatibility. *Perception and Psychophysics*, *35*, 333-343.
- Nicoletti, R., & Umiltà, C. (1985). responding with hand and foot: The right-left prevalence in spatial compatibility is still present. *Perception and Psychophysics*, *38*, 211-216.
- Nicoletti, R., Umiltà, C., Tressoldi, E. P., & Marzi, C. A. (1988). Why are left-right spatial codes easier to form than above-below ones? *Perception and Psychophysics*, *43*, 287-292.
- O'Leary, M. J., & Barber, P. J. (1993). Interference effects in the Stroop and Simon paradigms. *Journal of Experimental Psychology: Human Perception and Performance*, *19*, 830-844.
- Pomerantz, J. R. (1983). Global and local precedence: Selective attention in form and motion perception. *Journal of Experimental Psychology: General*, *112*, 515-540.
- Pomerantz, J. R. (1986). Visual form perception: An overview. In E. C. S. & H. C. Nusbaum (Eds.), *Pattern recognition by humans and machines: Visual perception*. (pp. 1-30). New York: Academic Press.
- Rimé, B. (1983). Nonverbal communication or nonverbal behavior? Towards a cognitive-motor theory of nonverbal behavior. In W. D. & S. Moscovici (Eds.), *Current issues in European social psychology* Cambridge: Cambridge University Press.
- Rimé, B., Schiaratura, L., Hupet, M., & Ghysselinckx, A. (1984). Effects of relative immobilization on the speaker's nonverbal behavior and on the dialogue imagery level. *Motivation and Emotion*, *8*, 311-325.
- Rimé, B., & Schiaratura, L. (1991). Gesture and speech. In R. S. Feldman. & B. Rimé (Eds.), *Fundamentals of nonverbal behavior* Cambridge: Cambridge University Press.
- Riseborough, M. G. (1981). Physiographic gestures as decoding facilitators: Three experiments exploring a neglected facet of communication. *Journal of Nonverbal Behavior*, *5*, 172-183.

- Roe, W. T., Wilsoncroft, W. E., & Griffiths, R. S. (1980). Effects of motor and verbal practice on the Stroop task. *Perceptual and Motor Skills*, *50* , 647-650.
- Rogers, W. T. (1979). The relevance of body motion cues to both functional and dysfunctional communicative behavior. *Journal of Communication Disorders*, *12* , 273-282.
- Rosinski, R. R., Golinkoff, R. M., & Kukish, K. S. (1975). Automatic semantic processing in a picture-word interference task. *Child Development*, *46* , 247-253.
- Rothi, L. J. G., Mack, L., & Heilman, K. M. (1986). Pantomime agnosia. *Journal of Neurology, Neurosurgery, and Psychiatry*, *49* , 451-454.
- Schriefers, H., & Meyer, A. S. (1990). Experimental note: Cross-modal, visual-auditory picture-word interference. *Bulletin of the Psychonomic Society*, *28* , 418-420.
- Seymour, P. H. (1973). Stroop interference in naming and verifying spatial locations. *Perception and Psychophysics*, *14* , 95-100.
- Seymour, P. H. (1974). Stroop interference with response, comparison and encoding stages in a sentence-picture comparison task. *Memory and Cognition*, *2* , 19-26.
- Seymour, P. H. (1977). Conceptual encoding and locus of the Stroop effect. *Quarterly Journal of Experimental Psychology*, *29* , 245-265.
- Shimada, H. (1990). Effect of auditory presentation of words on colour naming: the intermodal Stroop effect. *Perceptual and Motor Skills*, *70* , 1155-1161.
- Sholl, M. J., & Egeth, H. E. (1981). Right-left confusion in the adult: A verbal labelling effect. *Memory and Cognition*, *9* , 339-350.
- Shor, R. E. (1970). The processing of conceptual information on spatial directions from pictorial and linguistic symbols. *Acta Psychologica*, *32* , 346-365.

- Shor, R. E. (1971). Symbol processing speed differences and symbol interference effects in a variety of concept domains. *Journal of General Psychology*, 85 , 187-205.
- Simon, J. R., & Rudell, A. P. (1967). The effect of an irrelevant cue on information processing. *Journal of Applied Psychology*, 51 , 300-304.
- Simon, J. R. (1969). Reactions towards the source of stimulation,. *Journal of Experimental Psychology*, 81 , 174-176.
- Simon, J. R., & Sudalaimuthu, P. (1979). Effects of S-R mapping and response modality on performance in a Stroop task. *Journal of Experimental Psychology: Human Perception and Performance*, 5 , 176-187.
- Simon, J. R., Sly, P. E., & Vilapakkam, S. (1981). Effects of compatibility of S-R mapping on reaction toward the stimulus source. *Acta Psychologica*, 47 , 63-81.
- Simon, J. R. (1982). Effects of an auditory stimulus on the processing of a visual stimulus under single and dual task conditions. *Acta Psychologica*, 51 , 61-73.
- Simon, J. R., & Berbaum, K. (1990). Effects of conflicting cues on information processing: The "Stroop effect" vs. the "Simon effect". *Acta Psychologica*, 73 , 159-170.
- Smith, M. C., & Magee, L. E. (1980). Tracing the time course of picture-word processing. *Journal of Experimental Psychology: General*, 109 , 373-392.
- Smyth, M. M., Pearson, N. A., & Pendleton, L. R. (1988). Movement and working memory: Patterns and positions in space. *Quarterly Journal of Experimental Psychology*, 40A , 497-514.
- Smyth, M. M., & Pendleton, L. R. (1989). Working memory for movements. *Quarterly Journal of Experimental Psychology*, 41A , 235-250.
- 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.
- Toms, M., Morris, N., & Foley, P. (1994). Characteristics of visual interference with visuospatial working memory. *British Journal of Psychology*, *85* , 131-144.
- Virzi, R. A., & Egeth, H. E. (1985). Toward a translational model of Stroop interference. *Memory and Cognition*, *13* , 304-319.
- Woodall, W. G., & Folger, J. P. (1981). Encoding specificity and nonverbal cue context: An expansion of episodic memory research. *Communication Monographs*, *48* , 39-53.
- Woodall, W. G., & Folger, J. P. (1985). Nonverbal cue context and episodic memory: On the availability and endurance of nonverbal behaviors as retrieval cues. *Communication Monographs*, *52* , 319-333.

## **Author Notes**

This research was supported by a grant from the Economic and Social Research Council (R000233560) to O'Malley, Bruce and Anderson (Universities of Nottingham, Stirling and Glasgow) and forms part of the research submitted in a PhD thesis by Steve Langton at the Department of Psychology, University of Nottingham, supervised by Bruce and O'Malley. The authors thank Alan Allport, Nelson Cowan, Robert Melara and an anonymous reviewer for their helpful comments on earlier versions of this article. We also thank Simon Garrod for his constructive comments and advice.

Correspondence concerning this article should be addressed to Steve Langton who is now at the Department of Psychology, University of Stirling, Stirling, Scotland, FK9 4LA, U.K. (e-mail: vb1@stirling.ac.uk).