A REVISED METHOD FOR ANALYSING NEGLECT USING THE LANDMARK TASK

Alessio Toraldo[1], **Robert D. McIntosh**[2], **H. Chris Dijkerman**[3] **and A. David Milner**[2]

(University of St Andrews, Scotland, UK; [1]Now at: Cognitive Neuroscience Sector, SISSA-ISAS, Trieste, Italy; [2]Now at: Department of Psychology, University of Durham, UK; [3]Now at: Psychological Laboratory, University of Utrecht, Netherlands)

Abstract

In order to better disentangle 'perceptual' and 'response' biases in neglect patients, Bisiach and his co-workers developed a new version of the 'landmark task'. In their version, subjects are required to choose which is the longer (first condition) or the shorter (second condition) of the two portions of a pre-bisected horizontal line. Two indices were proposed, for the purpose of measuring perceptual and response bias respectively. The perceptual bias index (PB) is the constant error across conditions, while the response bias index (RB) is the degree of response consistency between conditions. Although valuable in a clinical context, these indices are not mathematically independent of one another. Furthermore, they do not exploit all of the information available in a given set of landmark data, since the responses made at the different landmark locations are all averaged together. To overcome these problems, we propose two new indices that can be derived from the revised landmark task. Our perceptual bias index is the Point of Subjective Equality (PSE) – the mean landmark location that appears to be halfway along the line. The response bias index, M, is the mean probability of making a response that opposes the patient's subjective midpoint. PSE and M are mathematically independent of each other and use most of the landmark information. The method and its theoretical foundation are summarized, and illustrative data obtained from brain damaged patients and control subjects are presented. Finally, computational procedures are provided for both PSE and M.

Key words: landmark task, unilateral neglect, spatial cognition, perceptual bias, response bias

Introduction

Bisiach and his colleagues (for example, Bisiach, 1993) have argued persuasively that many tasks traditionally used to diagnose and measure visuospatial neglect are 'impure', in that different patients perform abnormally on them for different reasons. It was first proposed by Heilman and Valenstein (1979) that rightward errors made by patients with left-sided neglect in the standard line bisection task might not necessarily reflect a *perceptual* neglect of the leftward portion of the line, but rather a disorder that they called '*directional hypokinesia*' (Heilman et al., 1985). According to this idea, the patient might make line bisection errors due to a response bias favouring rightward over leftward movements. Bisiach et al. (1990) made the first attempt to separate what they called 'perceptual' and 'premotor' effects in the line bisection task, by pitting the two in opposition to each other. They devised a task involving pulleys, in which the subject had to make a leftwards movement in order to move a bisection pointer rightwards, and vice versa. They found that some neglect patients were dominated by the visual position of the pointer, and would make large leftward movements in order to make their usual rightward line bisections. Others were reluctant to make such leftward movements, so that their response bias partly or completely overcame their usual tendency to make rightward bisection errors.

Some authors have, however, criticized methodologies such as this for creating unnatural conflicts which patients might resolve in ways that reflect little about the causation of their neglect symptoms (Mattingley et al., 1998). A simple task that avoids such problems of interpretation was devised by Milner and his colleagues (Milner et al., 1992, 1993). In what they called the 'landmark task' (in recognition of its formal resemblance to a task used in non-human primate studies: Pohl, 1973), the subject was asked to make a simple decision on each trial, either to respond to the left or to the right. The subject was presented with a series of horizontal lines each of which had a mark located part-way along it, and was asked to point to the end of the line that the mark was closer to. The critical trials were those on which the landmark was located exactly midway along the line. Milner and colleagues reasoned that if a left neglect patient had a perceptual bias, then responses should be made mainly leftwards – the patient should indicate that the left half of the line looked shorter. If, however, the patient had no such perceptual bias, but instead a response bias, then if anything pointing responses should be made rightwards, as the patient should be biased against making leftward responses. Harvey et al. (1995) subsequently reported that seven out of eight left neglect patients did indeed point left on nearly all of the critical landmark trials, although one patient performed in the converse fashion. This result might, however, have provided an underestimation of the extent to which response biases cause line bisection errors. If patients first made a cognitive decision as to the location of the landmark before

responding, they might then be less biased toward making a rightward response than when making a bisection mark on an empty line.

By adding a simple extension to the landmark task procedure, Bisiach et al. (1998a) overcame this objection, and provided a means for making commensurate estimates of perceptual bias and response bias. As well as asking patients in one set of trials to point to the shorter section of a pre-bisected line, they asked them to point to the longer section in a separate set of trials. Patients with a response bias would presumably still tend to point rightwards in *both* sets of trials, i.e., irrespective of instructions; while those with a perceptual bias should point leftwards in the 'shorter' trials, but rightwards in the 'longer' trials. Thus, averaging the two rightward scores should give an improved measure of response bias (RB), while averaging the first leftward score with the second rightward score should yield an equally fair measure of perceptual bias (PB). Therefore,

$$RB = (RS + RL)/2$$
$$PB = (LS + RL)/2$$

where LS (Left Shorter) and RS (Right Shorter) are the overall percentages of leftward and rightward responses made in the 'shorter' judgement condition, while LL (Left Longer) and RL (Right Longer) are the percentages of such responses made in the 'longer' condition. (Here and henceforth, indices computed on the overall set of data, i.e. averaging across all of the landmark locations, will be named with upper-case letters, e.g., LS, RS, LL, RL, while indices computed on the data from a single landmark location will be named with lower-case letters, e.g., *ls*, *rs*, *ll*, *rl*.)

These measures allowed the investigators to estimate response bias and perceptual bias separately, whereas Milner et al.'s measure (Milner et al., 1992, 1993) could only provide an estimate of the *relative* influence of the two factors. Thus for example, a non-significant behavioural asymmetry in their original version of the test could not distinguish between (a) the presence of equally large opposing biases or (b) an absence of any bias at all.

Bisiach et al.'s (1998a) new version of the landmark task, and the two measures of bias that they extracted from it, constitute a valuable methodological advance, which has already yielded several new insights (Bisiach et al., 1998a, 1998b, 1999). There are, however, two limitations associated with their procedures. First, unlike the original method of Milner and colleagues, the new version sums all of the responses made by the subject over all of the trials, regardless of the location of the landmark. This averaging procedure wastes an opportunity that Bisiach et al.'s method provides. Since equal numbers of trials are given at landmark positions across a wide range, there is much more information available for estimating the perceptual bias than is used by the averaging method.

The second limitation concerns the lack of *mathematical independence* between the two measures of bias (see Appendix C for a more detailed discussion of this concept). This can best be seen by plotting the range of possible co-variation of the indices of response bias (RB) and perceptual bias (PB) that emerge from Bisiach et al.'s method of analysis. The indices of rightward bias (RB and PB) derived as indicated above have built-in limits to their mutual variability. As shown in Figure 1, the more extreme a subject's PB, the less the possible range of variation of RB: that is, the value of RB is artificially truncated at high or low values of PB (and vice versa). In other words, the two indices of bias are *in principle* not independent of each other.

This means that the use of these particular measures will inevitably produce diagnostic errors. For instance, a patient with severe output-related neglect (ORN) will always be classified as having little or no input-related neglect (IRN). In reality, however, we would know almost nothing about the patient's IRN. As an extreme example, consider the case of a patient responding rightwards in all trials (in both test conditions): he would obtain PB = 50, i.e. no IRN at all, whereas the correct conclusion should be 'IRN unknown'. Necessarily, a response that never changes from trial to trial can convey no information about the patient's perception of the different lines used. But even in less extreme cases of ORN, the artificial truncation prevents the subject from achieving extreme PB scores; for instance, if RB = 75, it is impossible to obtain a PB score higher than 75 or lower than 25. In general, therefore, the absolute degree of IRN will be underestimated. Only subjects with no ORN at all would be free from this bias in the estimation of IRN. Another diagnostic error can occur for different reasons. Using Bisiach et al.'s PB score, a patient who guesses in some of the trials (e.g. as a consequence of a comprehension disorder, or perhaps of a 'frontal' syndrome) would obtain an underestimate of his actual degree of IRN deficit. In the extreme case of a patient who *always* guesses, the mean PB value would be, in the long run, 50, i.e. an apparent absence of bias. Yet in reality of course we know nothing about his perceptual bias.

The indices RB and PB have a practical utility in the clinical context, in that they provide a quick and easy differential diagnosis between IRN and ORN. Nevertheless, for the reasons listed above, they are less useful for deriving scientific conclusions as to the nature of the two deficits, and more generally for the classification of patients in experimental studies with respect to both IRN and ORN.

The intention of the present paper is to propose a different mathematical model for analysing

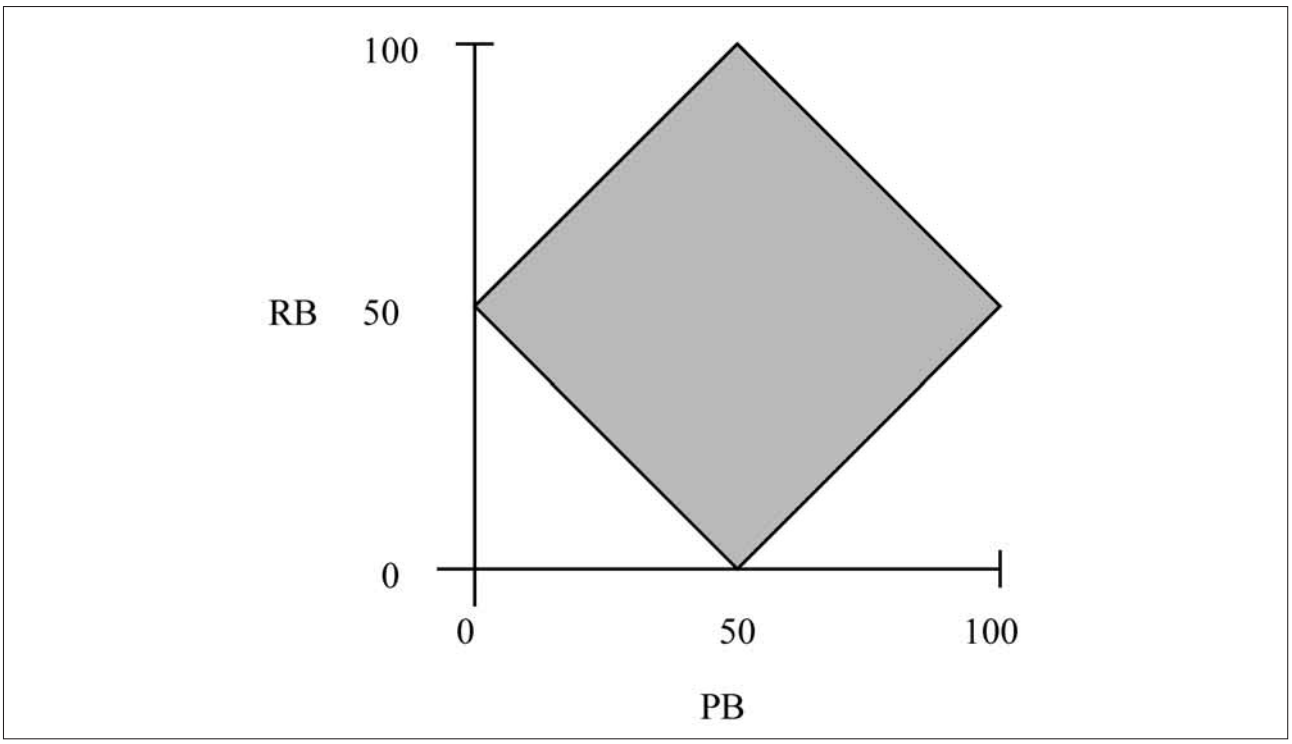

Fig. 1 – *Mutually constraining ranges of variation of Bisiach et al.'s indices PB (horizontal axis) and RB (vertical axis). When the performance of a single patient is represented as a point (PB,RB) in the plot, only points within the shaded diamond are possible.* (Adapted from Bisiach et al., 1998a)

individual data derived from using Bisiach et al.'s improved version of the landmark task. The basic problem of Bisiach et al.'s model is that the relationship between the deficits to be measured (IRN, ORN) and the indices used to measure them (PB, RB) is not transparent. To remedy this lack of transparency, our intention is to specify more explicitly how IRN and ORN might operate to determine behaviour in the landmark task. This line of thought leads to the formulation of two new, mathematically independent, indices, which avoid the problems of dependence between the ranges of variation of the two measures.

There is also a more general problem that has often been 'neglected' in neuropsychological research. This is the sample-to-sample instability of a patient's test score, i.e. the level of uncertainty about its 'real' value. Consider, for instance, a patient who when faced with two 180-mm lines, bisects one 30 mm to the left and the other 50 mm to the right of the objective midpoint. The mean bisection error would be + 10 mm, well beyond the normal range. One might be tempted to conclude that this patient has a rightward lateral bias on the basis of such a 'local estimate'. Yet given the huge variability of the patient's bisection performance, a second testing session could easily produce a mean bisection point of –10 mm! The problem of uncertainty about the real value of a parameter is particularly important in the context of Bisiach et al.'s revised landmark task. As we saw, when a patient has a very severe ORN, the task provides very little information for determining his real level of IRN. A new method for analysing data from the revised landmark task should provide a means of taking this uncertainty into account, in order to minimize diagnostic errors.[1]

IRN, ORN, AND THE LANDMARK TASK

Input-Related Neglect (IRN)

The landmark task was initially conceived in classical psychophysical terms as a method of determining the point of subjective equality (PSE) between the two sections of a bisected line (Milner et al., 1992, 1993). Asking an observer which of two collinear segments is the longer, and varying the position of the landmark defining the two segments, should allow an estimation of the point where the two segments are subjectively identical. By its very nature, the experiment implies the existence of such a point, and provides all of the data needed for estimating it. IRN, in this context, could only cause a shift, either to the left or to the right, of the PSE. Of course IRN may have a different significance with respect to other tasks sensitive to neglect, for example search tasks. However in the present context of the line-bisection task and the landmark task, we propose to identify IRN operationally with a pathological displacement of the PSE.

[1]Nevertheless, local estimates that are too unstable to be used for diagnostic purposes, may still be useful in group studies.

As for the specific input-related mechanisms inducing a shift of the PSE, there are two main possibilities. The horizontal dimension of space might be misperceived by, e.g., *left* IRN patients, in such a way that *physically* equal distances look progressively shorter the more on the left they are located (see, e.g., Milner, 1987; Bisiach et al., 1996). Therefore, the PSE will lie some distance to the right of the objective midpoint. Alternatively, a rightward shift of the PSE might result from an abolished representation of the leftmost regions of the line (see Bisiach and Vallar, 1988; Bisiach et al., 1998b).

Whichever is the case, the PSE is assumed to reflect the outcome of the visual processing of the stimulus, i.e., its internal spatial representation. IRN is defined as an alteration of this representation.

Output-Related Neglect (ORN)

Whether or not there is a shift of the PSE, there could be a bias on the 'output' side, i.e., a bias in favour of 'Right' (or 'Left') responses. This bias might be manifested as a reduction in (explicit or implicit) eye movements, limb movements, or both, towards the opposite side. Consider a subject who has a tendency to favour 'Right' responses (rightward ORN). When the *right* segment looks longer, he will have no difficulty in selecting it as his response in the 'longer' condition. The same holds for trials in which the *right* segment looks shorter in the 'shorter' condition. The deficit will only arise when the *left* segment looks shorter in the 'shorter' condition, or longer in the 'longer' condition. Here a 'leftward' outcome of visual processing will need to be strong enough to overcome the patient's baseline tendency to respond rightwards. Sometimes it will be strong enough, sometimes it will not. The proportion *m* of times in which the tendency to respond rightwards is *not* overcome is our proposed measure of rightward ORN.

Symmetrically, a measure of leftward ORN would be the proportion of times in which a 'rightward' outcome of visual processing is not strong enough to overcome a baseline tendency to respond leftwards. Here the leftward direction of *m* will be indicated by negative sign (– *m*). Like previous investigators, we assume that ORN will have the same influence whether the subject is being asked which is 'the longer' or 'the shorter' of the two line segments. While perhaps questionable, this assumed constancy of ORN across conditions is implicit in all attempts to disentangle IRN and ORN. Thus the difference between the percentage of rightward responses in the two conditions ('longer' vs. 'shorter') was taken to measure IRN by Bisiach et al. (1998a).[2] Similarly Bisiach et al. (1990) made an analogous assumption in their 'pulley' experiment. It was assumed that a patient with a purely 'premotor' neglect would produce equal and opposite bisection errors in the 'congruent' and 'incongruent' conditions (since in their task the perceptual result of identical motor acts would be equal and opposite). The constancy assumption was thus built into the logic of these experimental designs – without it, analysis would have become impossible.

[2]E.g., RS = 70%, RL = 100%. PB = (30% + 100%)/2 = 65%, which leads to the diagnosis of a rightward IRN.

Derivation of the Two Indices

The subjective midpoint of the line (SM) is the location that, *on a given trial*, would generate two subjectively identical segments. In accordance with standard psychophysical models, the SM is assumed to vary randomly across trials, and be distributed along the line following a normal probability density function with a mean equal to the PSE (our measure of IRN), and a certain standard deviation SD (see Figure 2A). Consider a *particular* landmark location T. When the subject's SM lies to the left of T the left segment will look longer (and the right will look shorter). This will happen in a proportion *p* of cases, corresponding to the area under the curve lying to the left of T. Conversely, in the remaining proportion 1 – *p* of cases, when SM falls to the right of T, the right segment will look longer (and the left will look shorter). Now, if we plot *p* as a function of landmark location, we obtain the cumulative curve of the normal distribution, the inflection point of which lies at the PSE (see Figure 2B). To estimate the PSE it is thus necessary to plot the *p* values that have been obtained experimentally on a similar graph, and to construct a cumulative normal curve that best fits that cloud of points. In Bisiach et al.'s (1998a) version of the task, we have nine landmark locations, so we need to compute the *p* value for each landmark location, partialling out the effects of any local variation in ORN. How can we estimate *p* values avoiding any response bias effects? There is a direct analytic solution, which also has an intuitive geometrical analogue.

The lower part of Figure 2 shows proportions of responses as vertical bars. Shaded areas indicate 'Right' responses, open areas indicate 'Left' responses. The whole vertical bar represents the overall set of responses in a specific condition. Figure 2C illustrates the proportions of responses given to landmark location T by a subject whose SMs vary as shown in Figure 2A. This subject has no ORN: thus, he will choose the Left as the shorter in a proportion *p* of cases, and, consistently, the Right as the longer in a same proportion *p* of cases (the shaded section in the 'shorter' condition is of the same size as the open area in the 'longer' condition). If, however, this same subject had some

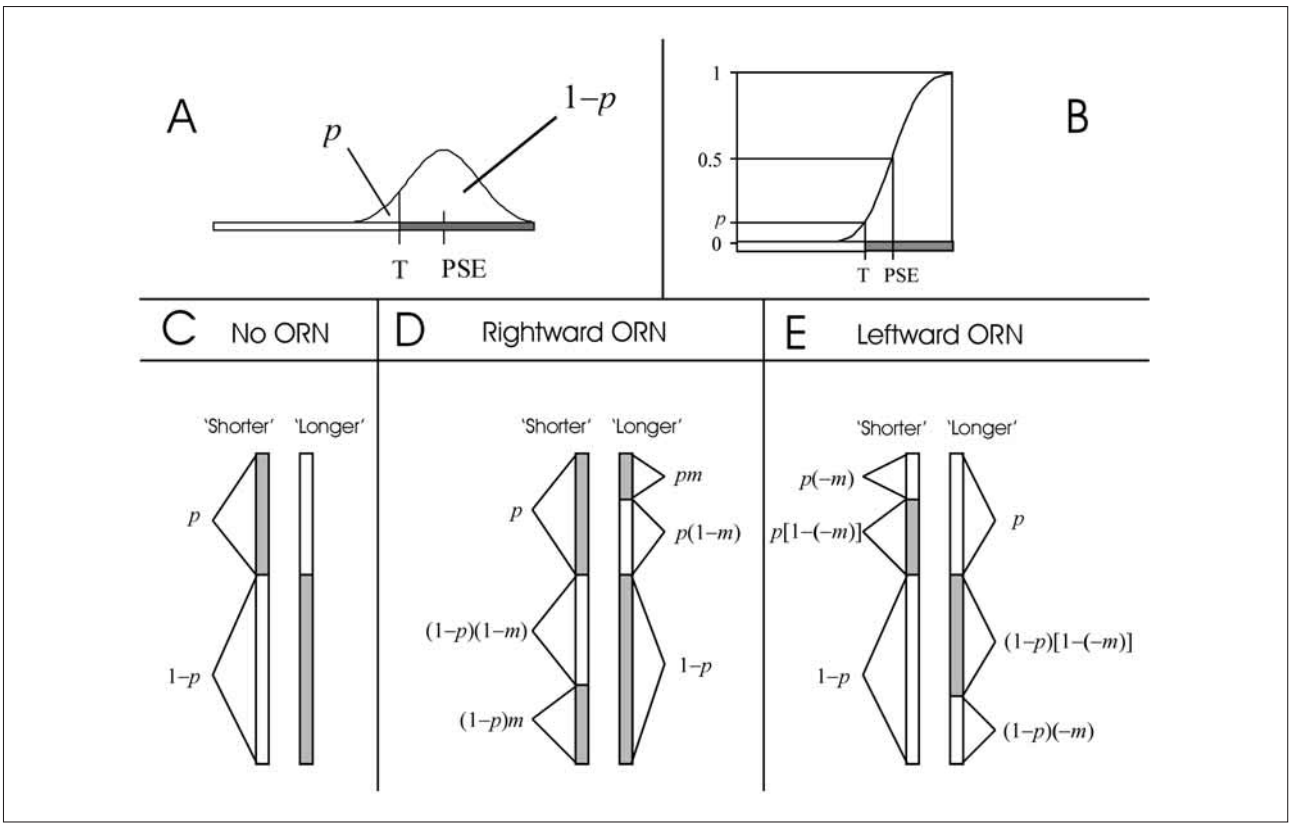

Fig. 2 – *A: The two collinear stimulus segments are illustrated in white (left) and grey (right), T being the landmark location. The subjective midpoint (SM) varies across trials according to a normal distribution, whose mean is the Point of Subjective Equality (PSE). B: The PSE can be estimated by fitting the experimental points that relate landmark locations to p proportions by means of a cumulative curve. The PSE will be the abscissa of the inflection point. C, D, E: The determination of the probability of 'left shorter' and 'right longer' responses according to the model is shown by representing probabilities as vertical extensions, for a patient without ORN (C), with rightward ORN (D) and with leftward ORN (E).*

rightward ORN, he would be reluctant to give 'Left' responses. On a proportion *m* of trials in which his 'ORN-free' twin (Figure 2C) said 'Left', he would instead say 'Right'. Thus, a proportion *m* of the open bars in Figure 2C would become shaded. The result is shown in Figure 2D. Conversely, if our patient had *leftward* ORN, he would be reluctant to give 'Right' responses. Thus, a proportion *m* of the *shaded* bars in 2C would become *open*, obtaining the pattern in Figure 2E (we conventionally put the negative sign in front of *m* here, to indicate the leftward direction of ORN and distinguish it from rightward ORN).

We do not directly know *p* and *m* from the experiment, but we can deduce them. The quantities we know from the experiment are the proportions of 'left shorter' (*ls*) and of 'right longer' (*rl*) responses for that landmark location. (As noted earlier, quantities computed on the data from a single landmark location are expressed with lower-case letters). As illustrated in Figure 2D, when ORN is rightward,

$$ls = (1 - p)\,(1 - m)$$
$$rl = (1 - p) + pm$$

These two equations constitute a simple first-degree system, with only two unknowns (*m* and *p*), *ls* and *rl* being known from the experimental data. It is thus easy to calculate in an unequivocal way the values of *m* and *p*. They will be given by:

$$p = (1 - rl) / (1 + ls - rl) \quad (1)$$
$$m = rl - ls \quad (2)$$

Similar reasoning can be applied in the case of a leftward ORN (Figure 2E). This time,

$$ls = (1 - p) + p\,(- m)$$
$$rl = (1 - p)\,(1 + m)$$

The solution of these equations is:

$$p = (1 - ls) / (1 + rl - ls) \quad (3)$$
$$m = rl - ls \quad (2')$$

It will be noticed that the estimation of *m* is *rl* – *ls* for both rightward and leftward ORN [formulae (2) and (2')] are identical]. For *p*, however, the two formulae are different. Formula (1) must be applied when ORN is rightward, i.e. when *m* is positive, and formula (3) must be applied when ORN is leftward, i.e. when *m* is negative. In mathematical terms:

$$m = rl - ls$$

and

$$\text{if } m > 0,\ p = (1 - rl) / (1 + ls - rl)$$
$$\text{if } m < 0,\ p = (1 - ls) / (1 + rl - ls)$$

These equations can be combined to give a compact formula for *p*:

$$p = [1 - \max(ls,rl)] / [1 + \min(ls,rl) - \max(ls,rl)] \quad (4)$$

We have thus obtained the sample estimates of the two parameters, *m* and *p*, for a single landmark location T.

We are now able to compute our final indices of IRN and ORN. After having obtained the *p* values for each landmark location (as shown above), it will be possible to derive the PSE by interpolating them with a normal cumulative curve (Figure 2B). And for an overall score of ORN, we can average out the *m* values from each landmark location, and obtain M.

How do our Indices Relate to those of Bisiach et al. (1998a)?

The ORN index M is a linear transformation of Bisiach et al.'s RB score [M = (RB/50) – 1]. It thus contains exactly the same information. According to our arguments, in other words, RB was already an unbiased estimator of ORN. M (ranging from 0 to 1 in its absolute value) is the *probability that a response will be made in the direction that 'contradicts' the patient's subjective midpoint*. For example, if on a given trial the SM falls to the right of T (i.e. the left segment looks shorter than the right) and yet the subject responds 'right shorter', then the response directly 'contradicts' the SM. The sign of M is an indicator of the direction of such 'contradictory' responses (+ = rightward, – = leftward).

On the other hand, our parameter *p* has an important difference in meaning from Bisiach et al.'s PB. PB is a measure of a subject's constant bias throughout the test, irrespective of landmark location. In contrast, *p* is defined as the probability of SM falling to the left of a *given* landmark location. Thus *p* values can be plotted against landmark locations to obtain the PSE.

The theoretical definition of our two parameters M and PSE guarantees their mathematical independence. M and PSE do not have the problem of the artificial truncation of the ranges (as PB and RB, see Figure 1). Every possible value of PSE can combine with every possible value of M, and vice versa. Thus, patients with ORN will never obtain underestimates of their degree of IRN (as can happen with PB), but will always be assigned an unbiased IRN score. Another advantage of the elimination of the artificial truncation is that any statistically significant correlation between the two indices M and PSE in a group of patients can be interpreted in neuropsychological terms and not as the result of a mathematical artefact. Furthermore, our parameters allow the researcher to understand whether a patient guessed in many or all of the trials. Such a patient will yield an especially high SD in the cumulative normal function (which would thus appear abnormally shallow). This should prevent misleading underestimates of the degree of IRN, and false negative decisions of no IRN. Appendix B reports more details as to how guessing behaviour can be detected.

Confidence Intervals

As we pointed out in the Introduction, inferences about the visual experience of a subject who has a very severe ORN or guesses on most of the trials are uncertain, and become impossible when ORN is maximal, or when there is guessing behaviour across-the-board. Therefore, it is necessary to provide the researcher with a measure of the uncertainty (i.e. sample-to-sample variability) of the obtained estimates, in order to minimize the likelihood of incorrect diagnoses. We will propose a diagnostic criterion based on confidence intervals for this purpose.

Confidence Interval for the PSE

The 95% confidence interval for the PSE is the region on the stimulus line with a 95% probability of including the true PSE position. This CI gives a clear idea of the level of uncertainty about the PSE location: if it extends, say, from –30 to +70 mm with respect to the line's centre, then we would have little useful information about the location of PSE!

Making use of the CI, we studied the level of uncertainty of PSE estimates as a function of ORN. It is clear that the presence of ORN makes less information available about the PSE. We already discussed the extreme case where a patient has the maximum possible rightward ORN (M = 1): he would choose the right segment on all of the trials. The information available about PSE is zero here: its CI is infinite. By means of data simulation, we found that in cases of 'very severe' ORN, i.e., when the absolute value of M is higher than 0.95, the CI could be as wide as 100 mm. As |M| decreases, however, the width of the CI decreases, falling to between 2 and 22 mm when M = 0. The use of the confidence interval in the diagnostic procedure thus allows the investigator to take the level of uncertainty into account. Instead of diagnosing an IRN when the simple PSE value is outside of the normal range (as in the 'local estimate' diagnostic criterion), we can require that the entire CI for the PSE lie outside of the normal range ('confidence interval' diagnostic criterion). In other words, to diagnose a deficit, the normal range and the CI must *not* overlap (see Armitage and Berry, 1994, pp. 98-99, for a similar logic, and Appendix A of this paper for further details; see also Toraldo, 2003).

More generally, this criterion solves the problem of unstable scores. In the Introduction, we offered the example of a patient bisecting two lines at – 30 and + 50 mm from the true centre. The mean bisection point (+ 10 mm), which is outside of the normal range, might suggest a diagnosis of a lateral bias. The CI of the mean bisection point would instead overlap the normal range, and therefore, lead to an opposite conclusion. In this way, a possible false positive is avoided.

Confidence Interval for M

There are two types of 'judgemental inconsistency' between the two response conditions of the Bisiach landmark task. The first type, 'left inconsistency', is when a subject says that the left is the shorter (in the 'shorter' condition), and that the left is the longer (in the 'longer' condition), for the same stimulus. The second type, 'right inconsistency', is when a subject says that the *right* is the shorter in one condition, and that the right is the longer in the other, again for the same stimulus. Our index M is the difference between the frequencies of these two types of inconsistency. When M = 0, the two types of inconsistency are equally frequent. When M > 0, there are more 'right' than 'left' inconsistencies; and when M < 0, 'left inconsistencies' are more frequent.

Both of the factors, ORN and perceptual uncertainty, will influence the parameter M. ORN induces a preference for one side, and thus will increase the frequency of only one type of inconsistency (e.g. 'right inconsistencies'). This will move the M score one way (e.g. towards positive values). By contrast, if a subject is uncertain about the stimulus, he will produce inconsistencies of *both* types ('left' and 'right'). This will move the M score (from sample to sample) both ways, thus increasing its variance.

So, perceptual uncertainty renders our measure of ORN unstable. The higher the perceptual uncertainty (measurable as SD, i.e. the shallowness of the cumulative curve, Figure 2B), the less stable the M estimate. Since brain damaged patients are likely to be perceptually uncertain (see e.g. their increased variance in bisection responses, Marshall and Halligan, 1990, table 1), their M estimate will be unstable, and vary widely from sample to sample. This means that a patient could easily obtain an M value out of the normal range, not because of an ORN, but because he is perceptually uncertain. A repetition of the same experiment might well produce an M value within the normal range, or even outside it on the opposite side.

This perceptual uncertainty argument is a possible explanation of the surprisingly high rate at which patients were diagnosed as suffering from ORN in Bisiach et al.'s (1998a) sample, and especially, of the relatively high proportion of patients showing a paradoxical leftward ORN after *right* hemisphere damage.

The 'confidence interval' diagnostic criterion can thus be used in this context as well. Since it takes into account the sample-to-sample variability of the score, it will prevent a diagnosis of ORN on grounds of an unstable M local estimate. In summary, an ORN will be judged to be present only when the CI for M, and not just M itself, is entirely outside of the normal range.

Accumulating Data from more Testing Sessions

If the CIs for the PSE or M are wide, the researcher can consider giving the patient further testing sessions. Under optimal conditions, i.e. when the real PSE and M do not vary across sessions, the width of the CI will reduce by approximately 35% with a second session, and by another 20% with a third session. For example, a CI for the PSE as wide as 30 mm on session 1 will reduce to 19 mm on session 2 and to 15 mm on session 3.

Empirical Examples of the Use of the New Indices

The Landmark-V task (Bisiach et al., 1998a) was administered to a set of 18 brain-damaged patients. Seven of them showed neglect on the Albert line cancellation test. They had all suffered a right hemisphere stroke, with the exception of D.L., who showed persistent neglect following a stroke in the left hemisphere (see: Pritchard et al., 2001). We also tested 12 controls, from whom we obtained ranges of normal scores (Table II). Table I reports the raw Landmark data from the patients; Table II reports the PSE and M estimates obtained from each subject by applying the techniques specified in Appendix A (see the website: www.masson.it/cortex/database/PC_Toraldo_40_3.htm or the website: www.toraldo.it/landmark/index.htm for automatic computation of the indices). The interested reader can apply those techniques to the individual data sets of Table I and compare his results with those listed in Table II.

Output-Related Neglect (M values)

Table II reports both the local estimate and the confidence interval for M, to be compared against the normal range at the bottom of the column [see 'M; CI(M)' in Table II]. A diagnosis relying on the M local estimate (completely equivalent to Bisiach et al.'s criterion using their RB index) defines five patients as suffering from ORN. However, only one of them suffers from ORN according to our confidence interval criterion. If one considers the high instability of M, in other words, the local estimate diagnostic criterion produces four false

TABLE I

Landmark-V Data (length of the line = 180 mm) for 17 Right Hemisphere Patients and one Left Hemisphere Patient (D.L.)

		Transection position (mm from the centre)								
		– 60	– 30	– 15	– 5	0	5	15	30	60
S.L.	LS	6	6	6	6	4	0	0	0	0
	RL	6	6	6	6	5	1	0	0	0
D.B.	LS	6	5	5	0	0	0	0	0	0
	RL	6	6	6	0	0	0	0	0	0
J.B.	LS	6	6	6	6	6	0	1	0	0
	RL	6	6	6	6	5	1	0	0	0
J.C.	LS	6	6	6	6	6	0	0	0	0
	RL	6	6	6	6	6	0	0	0	0
J.L.	LS	6	6	6	6	6	5	0	0	0
	RL	6	6	6	6	6	6	0	0	0
A.L.	LS	6	6	6	6	5	0	1	0	0
	RL	6	6	6	5	4	3	0	0	0
M.M.	LS	6	6	6	6	4	0	0	0	0
	RL	6	5	6	6	3	0	0	0	0
C.S.	LS	6	6	6	5	6	5	3	2	0
	RL	6	6	6	3	3	2	0	1	0
R.D.	LS	6	6	6	6	5	3	1	0	0
	RL	6	6	6	5	3	1	0	0	0
N.McL.	LS	6	6	4	4	3	3	1	0	0
	RL	6	6	6	4	4	4	2	1	1
L.McN.	LS	6	6	6	4	3	2	0	0	0
	RL	6	6	6	6	4	2	0	0	0
W.McI.	LS	6	6	5	6	4	5	1	0	0
	RL	6	6	5	6	6	6	1	0	0
M.McP.	LS	6	6	6	6	6	5	2	5	0
	RL	6	6	5	6	6	6	3	5	1
W.P.	LS	6	6	6	6	6	6	6	6	0
	RL	6	6	6	6	6	6	6	3	0
J.H.	LS	6	6	6	5	4	4	3	0	0
	RL	6	6	6	5	6	4	3	2	0
R.M.	LS	6	6	6	6	5	2	0	1	0
	RL	5	5	5	5	2	2	0	2	1
L.A.	LS	6	6	6	5	4	4	4	0	0
	RL	6	4	6	5	6	4	2	2	0
D.L.	LS	6	6	1	0	0	0	0	0	0
	RL	6	3	0	0	0	0	0	0	0

Legend: LS = Left Shorter responses (out of 6); RL = Right Longer responses (out of 6).

TABLE II

Results of the Analyses

		Bisiach et al.'s method		New model's parameters				
Patient	Can	PB;	CI(PB)	PSE	CI (PSE)	SD	M;	CI (M)
S.L.		53.7;	[47.4, 60.0]	0.94	[0.01, 1.87]	1.07	0.04;	[– 0.03, 0.10]
D.B.		*31.5*°;	[*26.8*, *36.2*]°	*– 10.00*°	[*– 12.25*, *– 7.75*]°	*1.71*	0.04;	[– 0.01, 0.08]
J.B.		*56.5**;	[50.7, 62.2]	*2.50**	[1.48, 3.52]	0.85	– 0.02;	[– 0.08, 0.04]
J.C.		55.6;	[55.6, 55.6]	*2.50**	[1.48, 3.52]	0.85	0.00;	[0.00, 0.00]
J.L.		*65.7**;	[*62.4*, *69.1*]*	*10.00**	[*7.75*, *12.25*]*	*1.71*	0.02;	[– 0.01, 0.05]
A.L.		55.6;	[47.2, 63.9]	0.94	[0.02, 1.86]	1.07	0.00;	[– 0.08, 0.08]
M.M.		50.0;	[43.0, 57.0]	0.42	[– 0.59, 1.43]	*1.42*	– 0.04;	[– 0.11, 0.03]
C.S.	+	*61.1**;	[49.8, 72.4]	*10.13**	[0.94, 19.31]	*17.13*	*– 0.22*°;	[*– 0.34*, *– 0.11*]°
R.D.		55.6;	[46.4, 64.7]	*3.13**	[0.46, 5.79]	*4.46*	*– 0.11*°;	[– 0.20, – 0.02]
N.McL.	+	*56.5**;	[43.1, 69.8]	*5.00**	[– 1.04, 11.04]	*12.25*	*0.13**;	[0.00, 0.26]
L.McN.		52.8;	[43.3, 62.3]	*3.00**	[0.12, 5.88]	*5.34*	0.06;	[– 0.04, 0.15]
W.McI.		*63.9**;	[55.4, 72.4]	*10.00**	[*3.80*, *16.20*]*	*11.64*	0.06;	[– 0.03, 0.14]
M.McP.		*79.6**;	[*70.0*, *89.2*]*	*33.75**	[*21.55*, *45.95*]*	*19.72*	0.04;	[– 0.06, 0.13]
W.P.	+	*86.1**;	[*81.7*, *90.6*]*	*45.00**	[*37.25*, *52.75*]*	*5.12*	– 0.06;	[– 0.10, – 0.01]
J.H.	+	*66.7**;	[55.2, 78.1]	*12.50**	[*6.05*, *18.95*]*	*11.46*	*0.07**;	[– 0.04, 0.19]
R.M.	+	54.6;	[42.5, 66.8]	*7.83**	[– 0.20, 15.86]	*16.73*	*– 0.09**;	[– 0.21, 0.03]
L.A.	+	*64.8**;	[52.8, 76.9]	*12.50**	[*6.10*, *18.90*]*	*11.46*	0.00;	[– 0.12, 0.12]
D.L. (*lbd*)	–	*20.4*°;	[*14.8*, *25.9*]°	*– 22.50*°	[*– 26.61*, *– 18.39*]°	*2.56*	– 0.07;	[– 0.13, – 0.02]
Normal range (n = 12)		[44.0, 55.8]		[– 2.23, 2.11]		≤ 1.41	[– 0.076, 0.061]	

Legend: Values in *italic* indicate scores (or confidence intervals) outside the normal range reported at the bottom. Normal ranges were derived from the mean ± 1.96 standard deviations of our 12 controls' scores, with the exception of SD where a unidirectional 5% cut-off was set at the mean + 1.645 standard deviations. *Left Neglect, °Right Neglect. Can: Albert line cancellation performance, + Left neglect, – Right neglect. PB, CI(PB): Bisiach et al.'s (1998a) parameter and its 95% confidence interval. PSE: estimated point of subjective equality, in mm of distance from the true centre, – leftwards, + rightwards. CI(PSE): 95% confidence intervals for PSE. SD: estimated standard deviation (a measure inversely proportional to the slope of the normal cumulative). M: measure of Output Related Neglect, a linear transformation of Bisiach et al.'s RB; CI(M): 95% confidence interval for M.

positives out of 18 patients. Considering only patients with neglect on cancellation, three out of seven cases are false positives. This strengthens our suspicion that some proportion of the ORN cases in Bisiach et al.'s (1998a) sample (all of whom had neglect on classical clinical tests) were false positives. As in Bisiach et al.'s group, we also had a non-negligible proportion of paradoxical leftward ORN (three out of five patients showing ORN), as predicted by the hypothesis that these originate mainly from the instability of M, rather than from a true 'contraversive' ORN. Also in favour of the instability hypothesis is the fact that four out of five patients 'with ORN' (according to the local estimate criterion) showed considerable perceptual uncertainty (SD values higher than 11 mm). That is, their subjective midpoints could fluctuate from trial to trial across a range of more than 44 mm on a 180-mm line). As discussed in the previous section, a patient with strong perceptual uncertainty is quite likely to obtain a M value artificially out of the normal range, even in the absence of any reliable ORN.

Input-Related Neglect (PSE)

Table II reports the results of the computational method proposed in Appendix A.

The 'confidence interval' diagnostic criterion was applied: CIs were compared to the normal range (reported at the bottom of Table II), and an IRN was diagnosed when a CI did not overlap the normative interval. According to the local estimate criterion, Bisiach et al.'s PB diagnosed IRN in 11 out of 18 patients, while the PSE did so in 15 out of 18 patients (see relevant columns). Therefore, the PSE appears a more sensitive measure of IRN already as a local estimate. Nevertheless, as can be seen in the column CI(PSE), the CIs for the PSE can be quite large, thus justifying the use of the prudent CI criterion. Applying this method, we found that while the CI for PB did not overlap the normal range in five patients, the CI for PSE did so in eight patients (see relevant columns).

The comparison between patients J.L. and C.S. illustrates the advantage of the CI criterion. Both patients showed a displacement of the PSE around 10 mm to the right of the objective midpoint. Nevertheless, they differed widely in the stability of their judgements. Patient J.L. obtained $p = 0$ for the landmark located 5 mm to the right of centre, and $p = 1$ for the neighbouring location, 15 mm to the right of centre. In other words, he judged, with only one exception, the + 5 mm landmark as to the left of centre, and the + 15 landmark as to the right of centre. The high stability of his judgements is reflected in his CI, which is very narrow (4.5 mm) and does not overlap the normal range. A rightward IRN is thus reasonably diagnosed in this case. On the other hand, patient C.S. was very uncertain (high SD): her judgements were inconsistent with the landmark located 5 mm to the *left* of centre ($p = 0.25$), and even with the landmark located 30 mm to the *right* of centre ($p = 0.8$). At 18 mm, the CI for her PSE reflects this high uncertainty, and it slightly overlaps the normal range. This casts doubts on the true presence of a rightward IRN: on a repetition of the same test in similar conditions C.S. might well have produced, for instance, a displacement of 1 (and not 10) mm to the right, i.e. a PSE location within the normal range. Therefore, the stricter CI-based diagnostic criterion avoids the misinterpretation of an eccentric PSE due to perceptual uncertainty as the genuine effect of an IRN.

To compare our CI diagnostic criterion to that of Bisiach et al. *ceteris paribus*, confidence intervals for PB were also computed (see Table II)[3] and the same 'non-overlap' logic was applied. Our PSE index proved to be a more sensitive measure of perceptual bias than PB (three more patients were classified as having a perceptual bias, 8/18 vs. 5/18). Of course we expected that our measure would be more sensitive to IRN than PB, because PSE does not have the constraint of the 'diamond' (see Figure 1). Nonetheless, this advantage would be even greater if we had patients with more severe ORN (the higher ORN, the stronger the constraint on PB scores).

A Fortuitous Advantage

Case L.A., on the other hand, illustrates another advantage of our PSE method. L.A. had a zero ORN score (M = 0, i.e. RB = 50). When as in this case no ORN is present at all, one would expect to find an identical sensitivity of our method and that of Bisiach et al. in diagnosing IRN. Nonetheless, L.A. was classified as normal using the CI for PB and as suffering from IRN using the CI for PSE (see Table II). L.A. behaved in a bizarre way when presented with landmarks at – 30 and + 30 mm (see Table I). These were perceptually unambiguous to him: according to his PSE and SD estimates, – 30 mm had probability 0% to be perceived as to the right of centre, and + 30 mm had probability 6% to be perceived as to the left of centre. On + 30 mm, he chose the right as longer twice out of six trials: this might in principle be explained in terms of rightward ORN. Except that on – 30 mm, he twice chose the *left* as longer! Overall, his behaviour is more consistent with the idea that he simply got 'distracted' (perhaps confusing 'longer' with 'shorter' at some conceptual or linguistic level). Although rare, these errors can be observed also in normal subjects: one

[3]Confidence intervals for PB were computed on the basis that PB is, in essence, the average of nine average proportions, each average [(*ls* + *rl*)/2] being relative to one landmark location. Therefore, the logic for obtaining CI(PB) is the same as that used for obtaining CI (M) (see Appendix A for details). In fact, CI(PB) is always 100 times as wide as CI(M).

of our control subjects made four of them, one on each of the four more extreme landmark locations (– 60, – 30, + 30, + 60). In any case, while Bisiach et al.'s method gives weight to such bizarre responses in the computation of PB, the mathematics of our method fortuitously eliminates their effect on the estimation of *p* (and thus of PSE). Specifically, this mathematical cancellation happens when distractions (i) occur on landmark locations unambiguous for the subject, and (ii) affect the trials of a specific landmark location in only *one* response condition.

In conclusion, the PSE measure is more stable than PB, in that its variance is less increased by 'distractions'. This might explain why it discriminates better between normal subjects and patients also when ORN is zero.

General Discussion

We are proposing an alternative method for analysing the data from Bisiach et al.'s revised version of the landmark task. Our parameter M and Bisiach et al.'s RB both express the degree to which the subject makes logically inconsistent judgements across conditions, favouring *one* direction of response. The sole advantage of M is that it has a clearly defined theoretical meaning: it says how likely it is that a response inconsistent with the spatial representation of the line will be made.

On the other hand, our new parameter *p* is different from Bisiach et al.'s parameter PB in a number of ways. As previously noted, *p* eliminates the logical problems with PB. When a patient does not provide any information about his visual processing (e.g., choosing the right on all trials), parameter *p* is consistently 'unknown', while Bisiach et al.'s PB classifies such a patient as having no perceptual bias. When some information about visual experience *is* available in the data, parameter *p* extracts it by partialling out the effects of ORN. Thus, it provides an unbiased estimate of IRN. Parameter PB instead underestimates the degree of IRN as a function of the absolute degree of ORN, as shown in Figure 1. Therefore, *p* provides, in general, a more sensitive measure of IRN.

A second difference regards the information used to arrive at a measure of IRN. This deficit may reflect a pervasive perceptual distortion of the patient's visual experience, which is exemplified in other perceptual tasks as well, such as matching (e.g. Milner and Harvey, 1995) and line extension (e.g. Bisiach et al., 1998b). By its very nature, the landmark task is a method for deriving the point on a line that produces two subjectively equal sections (the PSE). Thus, the landmark task measures the perceptual distortion associated with IRN purely in terms of a shift of the PSE. The derivation of the PSE requires the analysis of differential response frequencies across the various landmark positions: it therefore exploits all of the information provided by the experiment. By contrast, Bisiach et al.'s PB is computed as an average asymmetry score across the whole range of stimulus lines, thus losing much of the available information.

In group studies, the use of PSE guarantees an important advantage. Since the PSE is *mathematically* independent of M, it will thus unambiguously allow the identification of any *empirical* relations between the phenomena of IRN and ORN, which can thus be fruitfully investigated without problems of interpretation.

For diagnosing a deficit in single-case studies, on the other hand, we would recommend the use of confidence intervals, because PSE and M can be quite unstable. A diagnosis relying on the PSE and M 'local' estimates may in fact be seriously risky. Bisiach et al. (1998a) reported that a relatively large number of patients showed paradoxical response biases, in the context of a surprisingly large number of subjects with a response bias in one direction or the other. This even occurred in their 'Landmark-V' task, where a verbal rather than a motor response was required. By using the 'confidence interval' criterion, which partials out the effects of behavioural instability, the frequency of ORN cases in our sample diminished dramatically with respect to that obtained by means of the 'local estimate' criterion. We therefore surmise that many of Bisiach et al.'s 'response bias' cases were false positives, obtained because a 'local' estimate of RB – a linear transformation of M – was used instead of an 'interval' estimate in the diagnostic procedure.

Acknowledgements. The authors are grateful to the Wellcome Trust (grant 048060), the Leverhulme Trust (grant F/268/T) and the Ministero dell'Università e della Ricerca Scientifica (MURST), for their financial support of this research. Preliminary versions of the work were presented at the Euroconference on *Cognitive and Neural Basis of Spatial Neglect*, Como (Italy), September 2000; at the *19th European Workshop of Cognitive Neuropsychology*, Bressanone (Italy), January 2001; and at the *12th Annual Meeting of Theoretical and Experimental Neuropsychology* (TENNET), Montreal, June 2001.

References

Armitage P and Berry G. *Statistical Methods in Medical Research*. Oxford: Blackwell, 1994.

Bisiach E. Mental representation in unilateral neglect and related disorders: The Twentieth Bartlett Memorial Lecture. *Quarterly Journal of Experimental Psychology, 46A:* 435-461, 1993.

Bisiach E, Geminiani G, Berti A and Rusconi ML. Perceptual and premotor factors in unilateral neglect. *Neurology, 40:* 1278-1281, 1990.

Bisiach E, Pizzamiglio L, Nico D and Antonucci G. Beyond unilateral neglect. *Brain, 119:* 851-857, 1996.

Bisiach E, Ricci R, Berruti G, Genero R, Pepi R and Fumelli T. Two-dimensional distortion of space representation in unilateral neglect: Perceptual and response-related factors. *Neuropsychologia, 37:* 1491-1498, 1999.

BISIACH E, RICCI R, LUALDI M and COLOMBO MR. Perceptual and response bias in unilateral neglect: Two modified versions of the Milner landmark task. *Brain and Cognition, 37:* 369-386, 1998a.

BISIACH E, RICCI R and MÒDONA MN. Visual awareness and anisometry of space representation in unilateral neglect: A panoramic investigation by means of a line extension task. *Consciousness and Cognition, 7:* 327-355, 1998b.

BISIACH E and VALLAR G. Hemineglect in humans. In F Boller and J Grafman (Eds), *Handbook of Neuropsychology, Vol. I*. Amsterdam: Elsevier, 1988, pp. 195-222.

FINNEY DJ. *Probit Analysis*. Cambridge: Cambridge University Press, 1971.

HARVEY M, MILNER AD and ROBERTS RC. An investigation of hemispatial neglect using the landmark task. *Brain and Cognition, 27:* 59-78, 1995.

HEILMAN KM, BOWERS D, COSLETT HD, WHELAN H and WATSON RT. Directional hypokinesia: Prolonged reaction time for leftward movements in patients with right hemisphere lesions and neglect. *Neurology, 35:* 855-859, 1985.

HEILMAN KM and VALENSTEIN E. Mechanisms underlying hemispatial neglect. *Annals of Neurology, 5:* 166-170, 1979.

MARSHALL JC and HALLIGAN PW. Line bisection in a case of visual neglect: Psychophysical studies with implications for theory. *Cognitive Neuropsychology, 7:* 107-130, 1990.

MATTINGLEY JB, HUSAIN M, RORDEN C, KENNARD C and DRIVER J. Motor role of human inferior parietal lobe revealed in unilateral neglect patients. *Nature, 392:* 179-82, 1998.

MILNER AD. Animal models for the syndrome of spatial neglect. In M Jeannerod (Ed), *Neurophysiological and Neuropsychological Aspects of Spatial Neglect*. Amsterdam: Elsevier, 1987, pp. 259-288.

MILNER AD, BRECHMANN M and PAGLIARINI L. To halve and to halve not: An analysis of line bisection judgements in normal subjects. *Neuropsychologia, 30:* 515-526, 1992.

MILNER AD and HARVEY M. Distortion of size perception in visuospatial neglect. *Current Biology, 5:* 85-89, 1995.

MILNER AD, HARVEY M, ROBERTS RC and FORSTER SV. Line bisection errors in visual neglect: Misguided action or size distortion? *Neuropsychologia, 31:* 39-49, 1993.

POHL W. Dissociation of spatial discrimination deficits following frontal and parietal lesions in monkeys. *Journal of Comparative and Physiological Psychology, 82:* 227-239, 1973.

PRITCHARD CL, DIJKERMAN HC, MCINTOSH RD and MILNER AD. Visual and tactile size distortion in a patient with right neglect. *Neurocase, 7:* 391-396, 2001.

TORALDO A. A ruler for measuring representational space. *Visual Cognition, 10:* 567-603, 2003.

Alessio Toraldo, Cognitive Neuroscience Sector, SISSA – ISAS,Via Beirut 2-4, 34014 Trieste, Italy. e-mail: alessio.toraldo@unipv.it

(Received 10 September 2001; reviewed 5 December 2001; revised 5 August 2002; accepted 20 October 2002; Action Editor Giuseppe Vallar)

Appendix A

Estimation of IRN and ORN from Landmark Data

An electronic worksheet for the automatic computation of all the indices can be downloaded from the website www.masson.it/cortex/database/PC_Toraldo_40_3.htm or from www.toraldo.it/landmark/index.htm. It executes all the procedures specified in this Appendix.

Computation of the index M of ORN and of its confidence interval

The mean value M across all landmark locations may be computed as follows, where rl_1, …, rl_t, ls_1, …, ls_t are sample proportions from individual landmark locations 1 to t (t is the overall number of landmark locations: t = 9 in Bisiach et al., 1998a).

$$M = \frac{1}{t}(rl_1 + \ldots + rl_t - ls_1 - \ldots - ls_t)$$

Parameter M, ranging from – 1 (maximum leftward ORN) to + 1 (maximum rightward ORN), has a precise meaning: it is the average *probability* of a response going in the direction opposite to that indicated by the input processing. Such a definition in probability terms allows an estimation of the standard error for M, i.e. of its reliability. In fact, since M is the average of all the m values, which in turn are differences between independent proportions (rl and ls), M will have a distribution close to the normal. Therefore, M's standard error can be estimated as:

$$SE(M) =$$

$$= \frac{1}{t}\sqrt{\frac{rl_1(1-rl_1) + \ldots + rl_t(1-rl_t) + ls_1(1-ls_1) + \ldots + ls_t(1-ls_t)}{N}}$$

where N is the number of repetitions of each stimulus line in each response condition (N = 6 in Bisiach et al.'s procedure). The 95% CI for M will then be [M – 1.96 SE (M), M + 1.96 SE (M)].

Graphical Representation for Deriving PSE and SD

In order to derive PSE and SD, data need to be represented on a plot like that illustrated in Figure 3A. The p-values obtained for each landmark location by means of formula (4) should be entered graphically in the plot. A small *dot* (■) should be drawn at the appropriate height over the corresponding landmark location. No dot should be marked if p is unknown. If all the p-values are unknown, the plot will remain empty and the computation procedure will stop with the conclusion 'unknown PSE'. If, instead, some p-values are known, and therefore, there are some dots in the plot, this should be evaluated according to the following rules:

(1) Find the 'anchor points'. Scan the plot from left to right, and mark with a black triangle (▲, e.g. see Figure 3A) the last consecutive dot with p = 0, starting from the leftmost landmark location (– 60). This will be the 'left anchor'. Then scan the plot from right to left, and mark in the same way (▲) the last consecutive dot with p = 1. This will be the 'right anchor'. If the leftmost dot does not have p = 0, then the left anchor (▲) will be placed on the point (– 90, 0), i.e., on the left endpoint of the line at height p = 0. If the rightmost dot does not have p = 1, then the right anchor (▲) will be placed on the point (90, 1), i.e., above the right endpoint of the line at height p = 1.

(2) Mark with a black circle (●) all the dots lying between the two '▲' points.

(3) If there are *at least two* '●' points, special treatment has to be given to any landmark locations yielding unknown p values and which lie between a '▲' and the closest '●'. An open triangle (△, 'left open anchor') should be drawn with height 0 at the rightmost of any such locations lying between the left '▲' and the closest '●'. Symmetrically, if there are any unknown-p landmark locations between the right '▲' and the closest '●', the leftmost of them must be assigned a '△' at height 1 ('right open anchor').

Computation of the PSE

The PSE is the inflection point, and SD reflects the shallowness, of the normal cumulative curve that best fits the points of the plot. To find that curve, a regression procedure should be applied. Any regression procedure can converge to a meaningful solution only in particular situations, viz. when there are at least two '●' points. We propose solutions (a) and (b) for other cases, in which there are one (a) or zero (b) such points. Solution (c) is proposed to substitute for a regression procedure in cases where there are at least two '●' points.

(a) One '●' point.

In this case a fair solution is to regard the normal cumulative function as passing through the '●' point and as having an 'intermediate' gradient. The 'intermediate' gradient is defined as halfway between the *maximum* gradient and the *minimum* gradient still compatible with the experimental data. The *maximum* possible gradient is the function passing through '●' by means of a vertical segment (SD = 0). In this case, the PSE is the abscissa of '●' itself. The *minimum* gradient compatible with the data was defined as the steeper of two lines (approximating the curved gradients), one passing through '●' and the left '▲', and another passing through '●' and the right '▲' (see Figure 3B). After having chosen the steeper line, its PSE is computed and averaged out with the

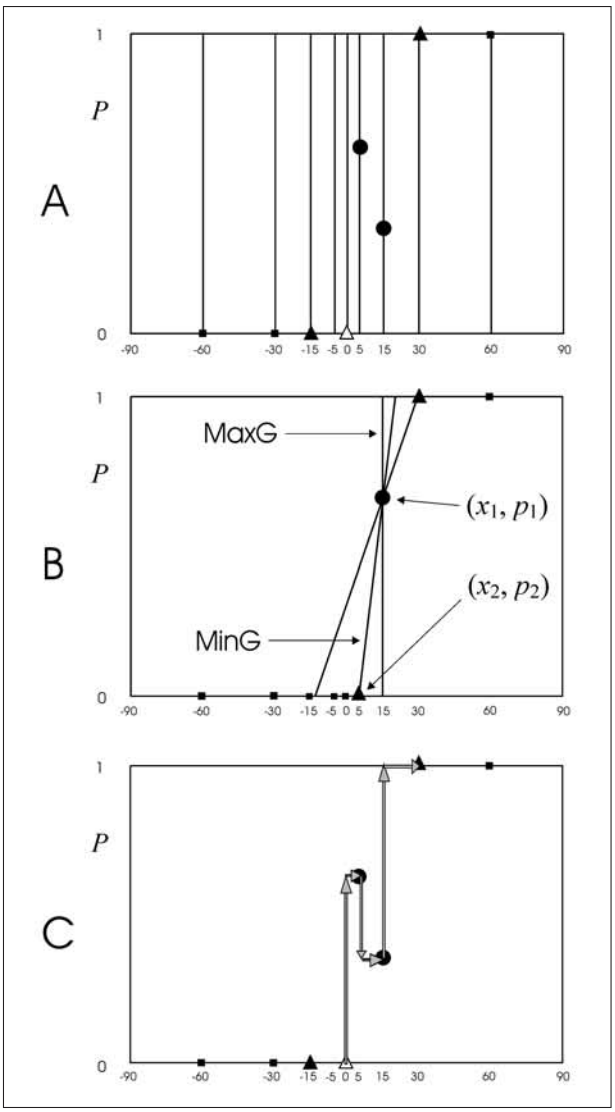

Fig. 3 – *A: Example of plot representing the experimental data, with landmark location (i.e. the physical line) on the horizontal axis and p on the vertical axis. P-values for the landmark locations –5 and 0 are unknown, therefore no dots (■) are placed in the plot for those landmarks. Following the rules specified in the text, a white anchor (△) is placed on landmark 0. B: Maximum (MaxG) and minimum (MinG) gradients compatible with the experimental data. C: Example of path to be traced across the plot, for the same data set as in A. The steps are represented as grey arrows.*

PSE of the *maximum* possible gradient. Therefore, given points '●' (x_1, p_1) and '▲' of the steeper line (x_2, p_2), the PSE local estimate is:

$$PSE = x_1 + \frac{0.5\,(0.5 - p_1)\,(x_1 - x_2)}{(p_1 - p_2)}$$

The approximate value of SD (of the 'intermediate' gradient) will be:

$$SD = 0.171\,\frac{(x_1 - x_2)}{(p_1 - p_2)}$$

(b) No '●' Points

The PSE is estimated as midway between the two landmark locations where the step-change of p from 0 to 1 (from left to right '▲') occurs. Thus, if x_L and x_R are the abscissae of the two anchors '▲', the PSE will equal $(x_L + x_R)/2$. As for the SD, the same logic of the 'intermediate' gradient is applied [see point (a) above]. Therefore, SD = $0.171\,|\,x_L - x_R\,|$.

(c) At least Two '●' Points

Regression procedures such as Probit Analysis (Finney, 1971) might be applied here to obtain PSE and SD, with the use of appropriate computer software. We offer instead a method that gives good approximate solutions and, like the previous ones (a) and (b), requires only a pocket calculator. The general idea that gave rise to this method exploits the fact that the PSE and the SD are, respectively, the *mean* and the square root of the *variance* of the distribution of the SMs. We will use two general properties of these parameters, i.e., the fact that the mean is the sum of the products between the single scores and their probabilities, and the fact that the variance is the mean of the square scores minus the square of the mean score.

The method works as follows. A path needs to be traced across the plot representing the data (see Figure 3C for an example). The path starts either from the left '▲' or, if present, from the left '△', and ends either at the right '▲', or, if present, at the right '△'. The first step of the path is a vertical line from the starting point up to the height of the first '●'; the second step is a horizontal line reaching the '●'. Then, the process is repeated: first a vertical line is traced to reach the height of the next '●'; second, a horizontal line reaches the '●' itself. This process must be repeated until the right end of the path (right '▲' or '△') is reached. The path thus consists of pairs of steps (the first of a pair is vertical, the second horizontal).[4] The PSE is estimated as follows. For each pair of steps the height of the vertical step (positive when the step is upward, 0 when it is null, negative when it is downward) must be multiplied by the midpoint value of the horizontal step. The sum of these products will be the estimate of PSE. For deriving SD, it is also necessary to compute the products *height of the vertical step* × *squared position of the horizontal step*. If G is the sum of these products, then

$$SD = \sqrt{G - PSE^2}$$

Figure 5 reports an example of these computations for obtaining PSE and SD.

General Formula for the PSE Confidence Interval

By means of a data simulation study, we could see that a good approximation of a 95% confidence

[4]Note that the vertical step can be also null (this happens when two consecutive '●' are at a same height).

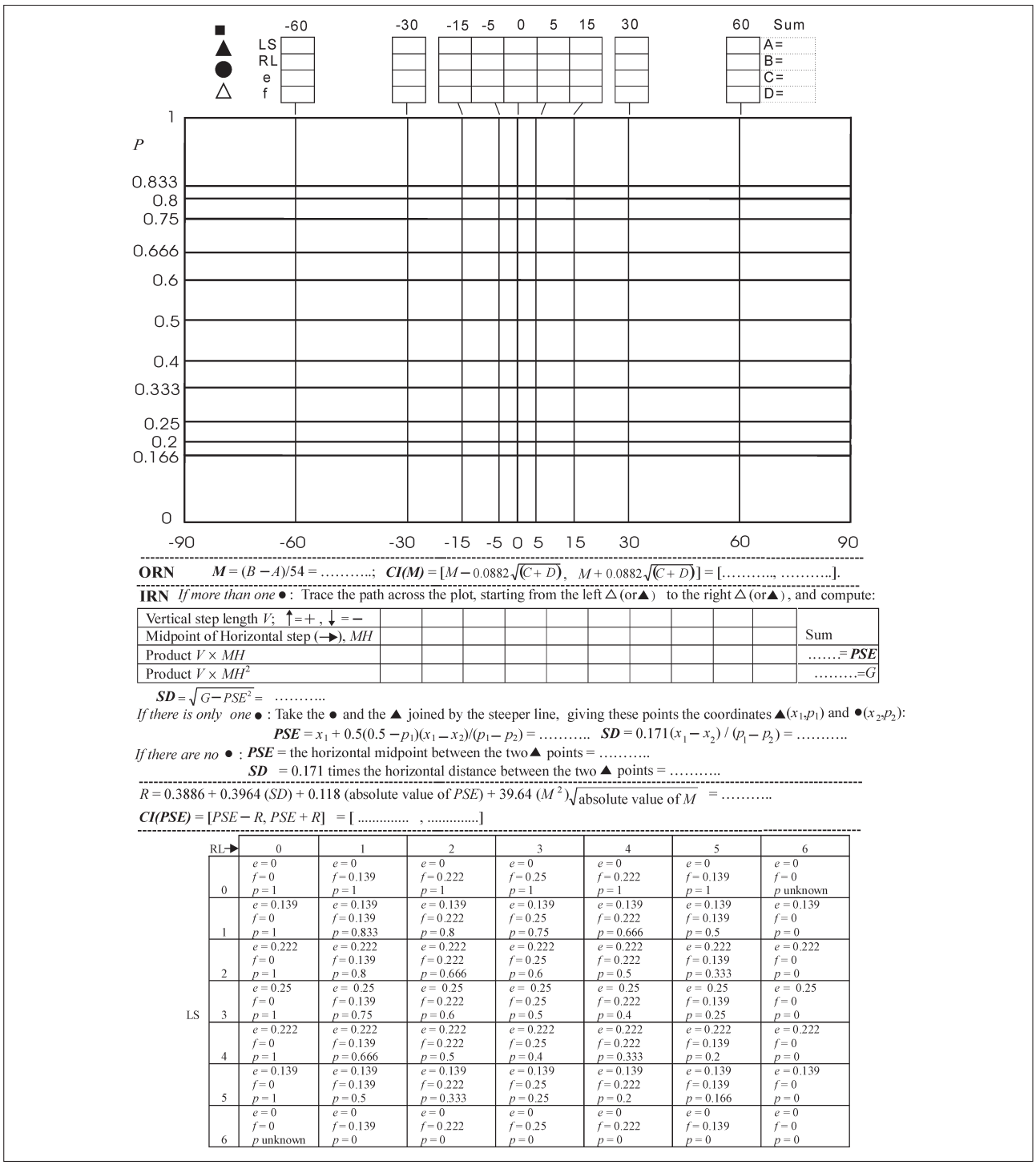

Fig. 4 – *Computation sheet for a single experimental session.*

interval for PSE is obtained by applying the following formula:

$$\mathrm{PSE} \pm \left(1+\frac{0.0153\,(\sqrt{\mathrm{S}}-1)}{1-|\mathrm{M}|}\right)\left(0.5+\frac{0.07}{\mathrm{S}^2}\right)\times$$
$$\times\left(\frac{0.6817+0.6954\,\mathrm{SD}+0.2071\,|\mathrm{PSE}|+69.54\,|\mathrm{M}|^{0.316\mathrm{S}+2.2662}}{\sqrt{\mathrm{S}}}\right)$$

The formula has as entries PSE, SD and M as estimated from the data set; the new parameter S is the number of testing sessions from which data have been accumulated, each session having N = 6 repeats per stimulus per condition, as in Bisiach et al.'s (1998a) procedure.

This formula gives 95% confidence intervals[5] if (i) Bisiach et al.'s (1998a) stimuli have been used and (ii) the assumptions of our mathematical model hold. If more than one session have been administered (S > 1), the formula gives 95%

[5]More precisely, we obtained intervals with an average confidence level around 95.4%, and ranging from 86% to 100%. None of the confidence levels obtained from each simulation was significantly lower than 95%.

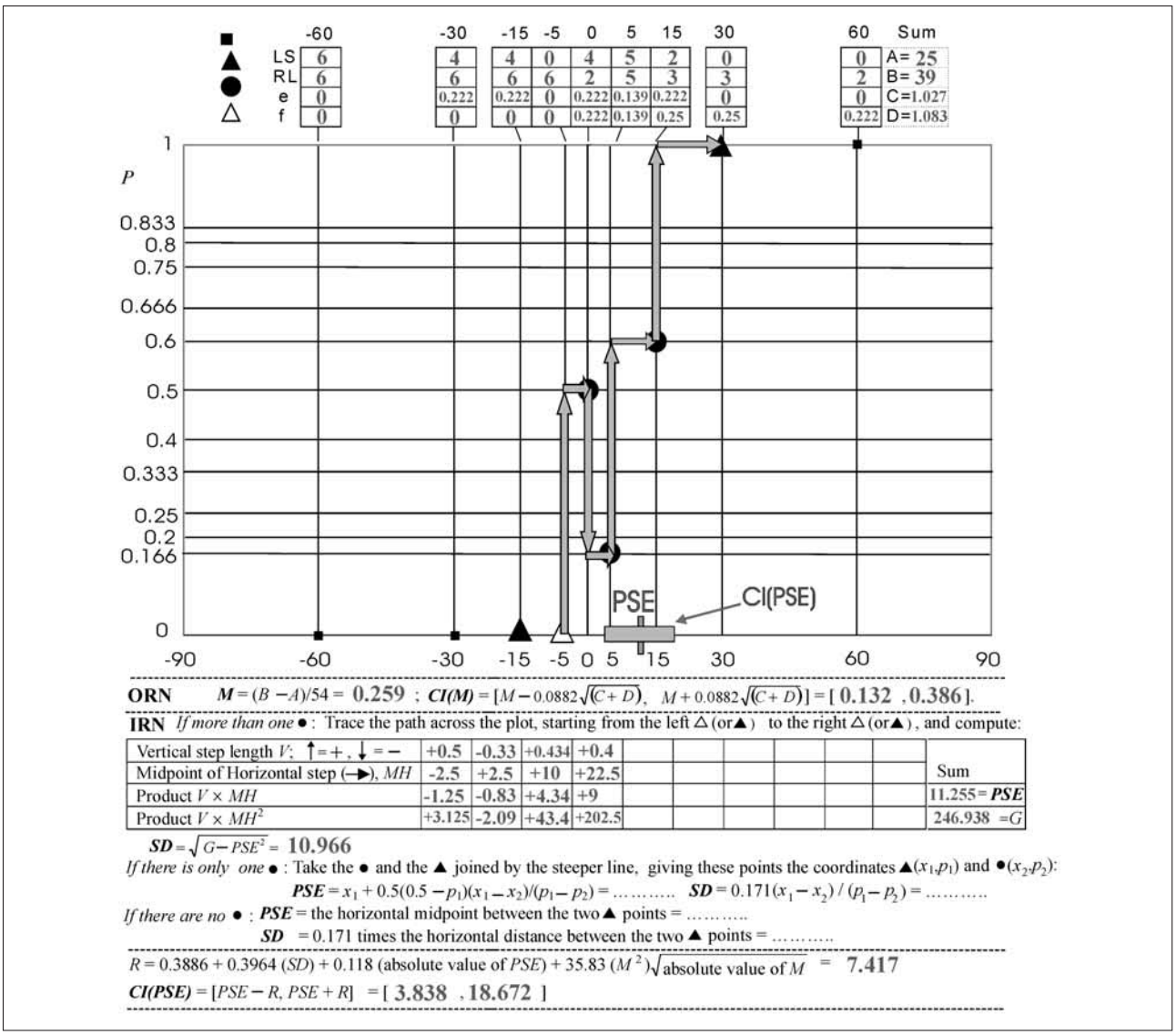

Fig. 5 – *Example of PSE and M estimation. Grey numbers have been added to carry out the calculation. The PSE and its CI are also represented.*

confidence intervals when the patient's 'true' PSE, SD and M have not varied across sessions. Thus, it is advisable to administer further sessions in consecutive days, or on the same day, shuffling the order of the stimuli within each experimental block, as otherwise the formula might underestimate the width of the confidence intervals.

The formula has been obtained by simulating data in the range of S from 1 to 3, of SD from 1 to 20 mm, of | M | from 0 to 0.99 and of | PSE | from 0 to 60 mm. Therefore, if data are obtained from more than three sessions, or from patients with SDs far higher than 20 mm, or with PSEs more eccentric than 60 mm from the true centre, or with M absolute values between 0.99 and 1 (all very rare conditions), our formula might give confidence intervals with an average level of confidence different from 95%. The formula should still work well if S, SD or PSE are beyond the above limits, but we are less sure that it would work if | M | is between 0.99 and 1. However, none of the 121 left neglect patients tested by Bisiach et al. (1998a) obtained an | M | higher than 0.99.

Simplified Versions of the Formula for S = 1, S = 2, S = 3

Much simpler (approximate) versions of the formula for CI(PSE) are available when S = 1, S = 2, S = 3.

These can be worked out with a pocket calculator and give intervals of the same average confidence level as those obtained from the general formula (95.4%).

When S = 1,

$$\mathrm{PSE} \pm (0.3886 + 0.3964\,\mathrm{SD} + 0.118\,|\,\mathrm{PSE}\,| + 39.64\,\mathrm{M}^2\sqrt{|\,\mathrm{M}\,|})$$

When S = 2,

$$\mathrm{PSE} \pm \left(1 + \frac{0.0063}{1 - |\,\mathrm{M}\,|}\right) \times$$

$$\times\,(0.2495 + 0.2545\,\mathrm{SD} + 0.0758\,|\,\mathrm{PSE}\,| + 25.58\,|\,\mathrm{M}\,|^3)$$

When S = 3,

$$\mathrm{PSE} \pm \left(1 + \frac{0.0112}{1 - |\mathrm{M}|}\right) \times$$

$$\times\ (0.2 + 0.2039\,\mathrm{SD} + 0.0607\,|\mathrm{PSE}| + 19.88\,|\mathrm{M}|^3)$$

Simplified Computation Sheet for the Single Session (S = 1)

If data have been obtained from a single session as specified in Bisiach et al. (1998a), i.e. administering N = 6 repeats per landmark location per condition, the worksheet shown in Figure 4 can be used to simplify the computation procedures.

Raw data should be transcribed on to the table at the top of the worksheet. Stimulus landmark locations are listed, from – 60 to + 60 mm. The numbers of Left Shorter (LS, 0-6) and Right Longer (RL, 0-6) responses should be inserted in the first two rows, in the appropriate cells. The third and fourth rows should then be completed according to the values *e* and *f* specified in the table at the bottom of the worksheet, using the specific LS and RL values as entries.

The sums of the four rows should then be computed, obtaining the values A, B, C and D respectively. In the bottom table, the *p* values corresponding to every LS-RL combination are provided. These should be entered graphically in the plot, applying the classification rules listed above to obtain the two lateral anchors (of '▲' or '△' type) and, if present, the '●' points. The formulae and procedures for computing M and its confidence interval CI(M), the PSE and its confidence interval CI(PSE), and SD, are given below the plot. As for PSE and SD, the different formulae are reported for the three possible circumstances (zero, one or more than one '●'). An example of application is shown in Figure 5.

Appendix B

Guessing Strategy

Consequences of Guessing Behaviour

The impact of guessing behaviour on our parameter PSE depends on whether the guessing behaviour is complete, i.e. occurring on each and every trial of the experiment, or partial, i.e. affecting only a subset of the trials. When there is complete guessing, the average estimate of PSE will be 0, i.e., no bias at all, irrespective of the real PSE. Therefore, similarly to what PB would suggest, the conclusion would be 'no IRN' although no information is really available about the subject's stimulus processing. On the other hand, when the guessing behaviour is partial, the PSE estimate will be on average correct (unbiased), although more instable with respect to the case of a subject who never guesses. By contrast, PB underestimates the severity of an actual IRN in this case of partial guessing.

As far as ORN is concerned, guessing behaviour introduces a bias in the estimation of M that depends on the nature of the guessing. While ignoring the metric properties of the stimulus, the subject can either guess the *colour* of the correct response [i.e. give a 'red' or 'black' response by chance, with probability independent of the stimulus, see Bisiach et al.'s (1998a) 'Landmark-V'], or guess the *side* of the correct response (i.e. name the colour of, or point to, the right or the left collinear segment by chance, with probability independent of the stimulus). If the subject guesses the colour, the M estimate will be either zero (complete guessing) or biased towards 0, i.e. underestimating the degree of an actual ORN (partial guessing). If the subject guesses the side, the absolute value of the M score will be biased towards 1: we will therefore have overestimates of the degree of an actual ORN.

In summary, complete guessing behaviour introduces a bias in the estimation of PSE, and partial or complete guessing introduces biases in the estimation of M. Fortunately, a criterion for the detection of partial or complete guessing behaviour is provided by our method.

Diagnostic Criterion for Guessing Strategy

A good means for detecting partial or complete guessing is to consider the shallowness of the cumulative normal curve (SD). Consider the extreme case of a subject who guesses on each and every trial. All of his *p* values would be around 0.5, so that the resulting cumulative curve would be very shallow (SD virtually infinite). Therefore, the higher SD, the more likely it is that the subject guessed at least on some trials. One possible criterion for establishing a cut-off SD score for guessing behaviour is the range of distribution of the subjective midpoints (SMs) along the line. A subject could be defined as guessing if fewer than 99% of the SMs (99% of the area under the gaussian curve, see Figure 2B) lie within the confines of the line. The highest SD value corresponding to this situation is 34.94. Therefore, if a subject shows an SD value higher than 34.94, he should be considered as having guessed on at least some trials, and the consequences of this behaviour in terms of estimation biases (see above) should be considered in the interpretation of the results.

Appendix C

Some Theoretical Considerations

It is important in the present context to bear in mind the difference between *mathematical* and

empirical relationships when any two measures are compared across a neurological population. The interesting question is always whether there is an *empirical* relationship, so that inferences can potentially be made about common underlying mechanisms subserving performance on the two tasks. If, however, there is a relationship for the trivial reason of overlap between the two tasks used to measure performance (e.g. some test items are common to both tasks), then clearly one can make no inferences about the underlying deficits. In other words, if there is *mathematical dependence* between the two scores, it will be impossible to decide whether an experimental correlation between them in a population of patients is due to their mathematical link, or to real communalities between the neuropsychological deficits they measure. If, on the other hand, the two scores are mathematically *in*dependent, any experimental correlation will be interpretable in neuropsychological terms.

For instance, the relationship between comprehension and working memory disorders can be investigated administering the Token test and the Digit Span test. The scores on the two tasks are separate: no information from Token test performance is used to compute the Digit Span score, nor vice versa. Therefore in the present sense, the scores are *mathematically* independent.[6] Such mathematical independence allows the attribution of any experimental correlation between the two scores to a real relationship between the two deficits, e.g., a positive correlation might be found because successful comprehension needs the input sentence to be stored in working memory.

The general point is that mathematical independence between scores is a necessary condition to study whether there is empirical dependence or independence between any two deficits. Such mathematical independence does not (and cannot) require any assumption about the empirical relation between the deficits – on the contrary it is a prerequisite to study that relation.

Just as the Token test and Digit Span provide mathematically independent scores that do not assume (in)dependence between comprehension and working memory disorders, our PSE and M are mathematically independent scores that do not assume (in)dependence between IRN and ORN.

[6]If two scores are derived from different data sets, as are Token and Digit Span, this implies that they are mathematically independent; but not vice versa: different scores from a same data set can be mathematically dependent (as PB and RB) or independent (as PSE and M).