British Journal of Psychology (1988), **79**, 13–45 *Printed in Great Britain*

© 1988 The British Psychological Society

Knowing more about what we can tell: 'Introspective access' and causal report accuracy 10 years later

Peter A. White

Department of Psychology, University of Auckland, Private Bag, Auckland, New Zealand

Nisbett & Wilson (1977*a*) proposed that people lack 'introspective access' to their mental processes, and that retrospective causal reports about those processes are in general inaccurate. This paper reviews the literature subsequent to that proposal. This literature has established that: (i) the term 'process' lacks clear, adequate and valid definition; (ii) the proposal depends upon an untested assumption that verbal reports are valid indicators of 'introspective access'; (iii) knowledge of the process is not sufficient for an accurate report about causal relations between stimuli and responses. Hypotheses about causal report accuracy can be tested, but there are severe methodological difficulties associated with measures of actual effects, the conditions under which retrospective causal reports are given and the use of 'observer' groups to rule out specific sources of information. Of the recent studies, none has achieved unambiguous support for, or falsification of, even a weak form of the proposal. There is no justification at present for using the proposal to interpret null results on verbal report measures. Some issues of broader significance are discussed, including the usefulness of verbal reports in research, the place of cultural beliefs in theoretical statements, and the social and practical context of verbal reports.

Ten years ago, Nisbett & Wilson (1977*a*) published a paper on the accuracy of retrospective verbal reports about causal influences upon responses in social judgement tasks. Their main contention was that causal reports were so inaccurate as to suggest little or no insight into the mental processes underlying social judgements. There has been a considerable variety of critical responses to the paper, and several studies have been carried out which yielded relevant evidence. The purpose of this paper is to review the literature following the original proposal, covering in turn critiques of the content of the proposal, and of pertinent methodological issues, and concluding with an assessment of the relevant experimental literature. The main conclusions are that the proposal is for several reasons untestable in its original form; that eliminating much of the theoretical content is necessary for a testable form of the proposal to be constructed; and that the evidence of research on this testable form remains inconclusive because of pervasive problems with the internal validity of the experiments. A final discussion sets this literature in a broader context, suggesting a possible line for future research to take.

The Nisbett–Wilson proposal

The Nisbett–Wilson (N–W) proposal contains distinct but related propositions about the extent of 'access' by actors to mental events, and about the accuracy of causal reports.

(1) The 'access' proposition: 'There may be little or no direct access to higher order cognitive processes' (Nisbett & Wilson, 1977*a*, p. 231).

Nisbett & Wilson also present three subsidiary propositions: 'Subjects are sometimes (*a*) unaware of the existence of a stimulus that importantly influenced a response, (*b*) unaware of the existence of the response, and (*c*) unaware that the stimulus has affected the response' (p. 231).

Nisbett & Wilson link these to the access proposition by arguing that, if subjects are unaware of these things, it follows that they cannot be aware of the intervening mental processes. This argument is used as justification for taking evidence for these types of unawareness as support for the access proposition.

(2) The 'verbal report' proposition: 'The accuracy of subjective verbal reports is generally so poor [as to suggest that any introspective access that may exist is not sufficient to produce generally correct or reliable reports]' (p. 233, parentheses added).

This proposition is also related to some more specific statements designed to enable prediction of exceptions to this generalization: (*a*) '[People] may base their reports on implicit, *a priori* theories about the causal connection between stimulus and response' (p. 233); (*b*) '...instances of correct report...are due to the incidentally correct employment of *a priori* causal theories' (p. 233).

The methodological significance of (2) (*b*) is that, to the extent that actors and observers share the same *a priori* causal theories, they are predicted to produce reports of equivalent accuracy. Thus, under the N–W proposal, an accurate report by an actor is not sufficient to demonstrate 'access' to the process. For this, the report must be *more* accurate than an observer's guess. The comparison between actors and observers is therefore an important ingredient of tests of the proposal. The picture is complicated by the fact that Nisbett & Wilson also allow actors access to several sorts of information *other* than process information, which may also help to make their reports more accurate than observers' guesses (e.g. knowledge of current focus of attention). Identifying the contribution of process information to report accuracy would therefore not be easy, even if the proposal were false.

Nisbett & Wilson review findings which they interpret as support for their proposal from several research areas, including attribution, cognitive dissonance, subliminal perception, problem solving and bystander apathy. Characteristically, the findings they report are of manipulations which led to significant differences between conditions on behavioural measures but not on verbal self-report measures. They then describe a series of studies carried out by themselves (also described in more detail in Nisbett & Bellows, 1977; Nisbett & Wilson, 1976, 1977*b*; Wilson & Nisbett, 1978), all of which show either behavioural effects that did not appear in causal reports, or reported effects that were not evident in behavioural measures, or causal reports that were no more accurate than causal judgements made by observer subjects.

Some of the subsequent literature has been concerned with the main ingredients of the proposal. In the order covered below, these include the distinction between process and content; concepts of 'introspective access' and the relationship between 'introspective access' and causal reports; and the possible contribution of internal events to causal report accuracy. Out of these emerges an alternative form of the proposal, in which no distinction between process and content is made.

The distinction between process and content

Although the access proposition carries the qualifying phrase 'may be little', the force of the proposition is that 'introspective access' is restricted to products, contents and 'interim results' of processes, and processes themselves are beyond the reach of 'introspective access'. Nisbett & Wilson do not define the term 'process' (Ericsson & Simon, 1980; Rich, 1979; White, 1980). They do list a number of things that they regard as content or product (and therefore 'introspectively accessible'). These include personal historical facts, present focus of attention, current sensations, emotions, evaluations, plans, intermediate results in a series of observations, knowledge of prior idiosyncratic reactions to a stimulus category, subcultural theories not shared by observers, and knowledge of intentions.

Some authors have questioned whether the exclusion of all these things from the concept of a process is correct. Smith & Miller (1978) question the distinction between process and intermediate result. The reason they give for this is that there can be so many intermediate results in a process such as mental rotation that the distinction between process and result becomes artificial. They also argue that, at higher cognitive levels, a rule, such as a long division rule, *is* a process. McClure (1983) seems to suggest that an intention is a process: he argues that it is not information or content. White (1980) notes that there is a danger of circularity in the N–W proposal if we use what is conscious or 'accessible' as an *ad hoc* criterion for deciding what is product or content.

In an information-processing model the term 'process' is conventionally applied to an operation which transforms information: the product is then the transformed information. This distinction is similar to the suggestion made by Rich (1979) that content answers 'What?' and process answers 'How?' (but see Sabini & Silver, 1981, described below). Adopting this convention might resolve some of the disagreements noted above. The distinction between process and intermediate result, for example, would then be a function not of the density of intermediate results but of the distinction between operator and operated-upon. A rule in itself is not a process by this convention, but the *application* of a rule on a particular occasion might be. Ability to report the rule itself would not, however, imply awareness of the process of using it. An intention may or may not be content, but it is not process unless it has the function of transforming information in some way.

Unfortunately what may hold for a model or simulation cannot be assumed to hold for mental activity (White, 1980). This convention regarding the use of the term 'process' is at present a theoretical construct of uncertain validity. Unusual use is made of this fact by Nisbett & Ross (1980). Borrowing an argument from Mandler (1975), they claim that, since 'process' is a theoretical construct, there cannot be

conscious awareness of processes, because one cannot have conscious awareness of a theoretical construct. Therefore, they say, the access proposition is true by definition. No one could disagree with this: if they choose to define 'process' as a theoretical construct, then the access proposition must be true. But it would also be devoid of meaning. There would be no need to test the proposal, or to refer to empirical evidence for it, because even evidence of perfect accuracy could not falsify it. But arguably the issue really concerns the limitations on 'introspective access' to *actual* goings-on in the head, irrespective of the theoretical constructs one might employ to try to model them: statements about theoretical constructs are not so much meaningless as beside the point.

Both Nisbett & Ross (1980) and Rakover (1983) advance a definition of 'process' as causal relations between mental events or entities. This potentially offers a reasonable prospect for a definition of the term, although the nature of a causal relation itself is not easy to define (Shaver, 1985). Nisbett & Ross use this definition as the basis for another argument that the N–W proposal is true by definition, this time on the grounds that no causal relation can be observed. But this sceptical point of view, associated in philosophy with Hume, is not universally accepted by philosophers. Several recent authors have argued for a 'causal realist' position, according to which causes actually produce effects and can be known and observed to do so (Shultz, 1982). If this is the case, then a causal relation definition of 'process' is not sufficient to make the N–W proposal true by definition. In addition, Nisbett & Ross assume in their argument that 'introspective access' is the same as, or a type of, observation – an assumption that may not be valid (Helminiak, 1984; Natsoulas, 1981).

The process/content distinction is reminiscent of the philosophical distinction between 'knowing how' and 'knowing that'. Taking up this parallel, one could suggest that a good test of 'knowing how' is ability to perform, whereas a good test of 'knowing that' is ability to report. If so, then causal reports are not appropriate as a test of 'knowing how'. One could 'know how', for example, without being able to translate that knowledge into the 'knowing thats' that a report could express. In this perspective, the question of whether we 'have access to' our processes also seems inappropriate. If we do, and *know how* to do, the process, then that is all the access one could expect anyone to have. They are our processes, after all. Philosophers have debated over whether statements about knowing what one is thinking, and suchlike, are even intelligible. Ayer (1973), for example, paraphrases Wittgenstein's view thus – 'People *have* thoughts and feelings, but this is not a matter of their knowing anything' (p. 59, author's italics) – though Ayer himself goes on to disagree.

It seems, therefore, that there are not less than three candidates for a definition of 'process': an operation in an information-processing system, a causal relation and a type of 'knowing how'. Apart from the need to ascertain which of these may be correct, choice of definition has significant consequences for the N–W proposal. In the extreme case, a 'knowing how' definition may render inappropriate not only the use of verbal reports as a measure of knowledge, but any proposition about access to processes.

In practice, more progress can be made if a distinction between product and process is not required. In his more recent publications (Wilson, 1985; Wilson &

Stone, 1985), Wilson has adopted this tactic. Wilson (1985) considers 'mental states', in particular a type of mental state which he defines as 'psychological dispositions that exert a causal influence on behaviour. Examples include 'attitudes, desires, moods, traits, affects, and evaluations' (p. 10). Wilson & Stone (1985) discuss 'knowledge of the "workings of one's own mind", obtained via introspection. This may consist of memories of prior thoughts and feelings or concurrent observations of one's reactions to a stimulus' (p. 171). These approaches simplify some of the problems of terminology and methodology, since one is no longer required to distinguish empirically between the influence on reports of different sorts of internal events and states. But there is some danger of throwing the baby out with the bath water. There may be good reasons for trying to ascertain valid distinctions between different types of internal event, and questions about which of these different types may be 'conscious' or open to 'introspective access' may still be important. Abandoning the distinction between 'process' and content is a pragmatic step, but not necessarily a correct one.

'Introspective access' and verbal reports

The comment at the end of the previous section begs the question of what exactly 'introspective access' is, and what sort of concept of 'access' is implied in the N–W proposal. In this respect the literature on the N–W proposal is a terminological quagmire. A variety of terms is used by different, and sometimes by the same, authors, who apparently regard them as interchangeable. Nisbett & Wilson (1977*a*) themselves use 'introspective access', 'aware/unaware', 'direct access', 'observe [the workings of their own minds]', 'conscious awareness', 'interrogate [a memory of]', 'introspective awareness', 'conscious', '[hidden from] conscious view', 'genuinely insightful introspection', 'knowledge of' and 'consulting [a memory of]'.

Natsoulas (1981) distinguishes 11 'problems' of consciousness, and his discussion helps to clarify the jungle of terms in which the N–W proposal is embedded. The following summarizes the way in which Natsoulas uses various terms:

'Experience': the qualitative dimension of awareness (particularly as by Sperry, 1969, 1970, 1976), e.g. 'qualia', 'raw feels'; not propositional in nature, and not necessarily conscious;

'Awareness': must have propositional content, according to Natsoulas; a necessary ingredient of one's being conscious of something, but an awareness is not necessarily conscious; the problem of awareness has two components – the content of awareness, and the intrinsic character of awareness (what is left when a description of content is exhausted);

'Conscious experience' or 'conscious awareness': specifically experiences or awarenesses which are introspected upon;

'Introspection': one's awareness of being aware, or of having awarenesses (Natsoulas also uses the terms 'second-level awareness' and 'metaconsciousness' here); it is a faculty possessed by a 'self' or 'subject' of awareness; Natsoulas distinguishes the substantive problem of introspection – how this type of awareness is possible – from the methodological problem of introspective verbal report.

It is possible to disagree with both Natsoulas' preferred uses of these words and the validity of the theoretical notions he puts forward. For instance Strange (1978) describes different concepts of consciousness, some from other cultures. Helminiak (1984) proposes a concept of consciousness involving 'awareness of self as subject', a direct awareness that is not relational in the sense that Natsoulas' 'conscious awareness' is relational. White (1986) proposes a non-relational concept of consciousness in which there is no 'self' or 'subject'.

Despite these legitimate differences of opinion on conceptual matters, the distinctions drawn by Natsoulas are of some use in the present context. They reveal that several proposals relating these terms to verbal reports are possible. On one side, one could draw up a set of propositions concerning what, among all internal events, can be 'contents of awareness', or 'experiences', or 'introspected upon', and so on. On the other side, one could draw up a set of propositions about the possible relationships between 'contents of awareness' or 'experiences' or 'introspections' and so on and the accuracy of verbal reports.

It is possible to discern in the writings of Nisbett & Wilson (and of some of their critics) a consistent point of view on both sides of the question. The 'access' proposition specifically concerns introspection (as defined by Natsoulas) rather than awareness. It states that introspective access to mental processes is not possible. But the relationship between the access proposition and the verbal report proposition depends upon an assumption that introspective access to mental processes would be sufficient for an accurate verbal report about those processes, subject to certain methodological caveats; and, by implication, that an inaccurate verbal report would be evidence for lack of introspective access, subject to the same caveats. The use of causal report accuracy as an indicator of introspective access, as in their own studies, requires this assumption. Yet it is neither stated nor justified by Nisbett & Wilson.

Many other assumptions and hypotheses are possible. For example, one could hypothesize that, on the one side, all types of mental events can be contents of awareness, and on the other side contents of awareness can be reported if and only if they are stored in short-term memory (i.e. that report accuracy is limited by the limited capacity of short-term memory, cf. Ericsson & Simon, 1980). One could also hypothesize that actual links between any sort of internal events and verbal report content do not depend upon any type of introspective access or awareness, conscious or not, but upon other sorts of factors, such as information-processing characteristics of the system.

Given this range of possibilities, why do Nisbett & Wilson implicitly accept one very particular assumption about the relationship between introspective access and verbal report accuracy? White (1980, 1982, 1986) argues that it is because this assumption reflects a generally accepted cultural theory. According to White, this theory is founded upon a distinction between a hypothetical entity often called the 'self' or 'subject' and the remainder of mental activity. Consciousness, in this theory, is a faculty possessed by the self, which mediates between the self and (some of) the remainder of mental activity. 'Introspective access' refers to the gaining of knowledge by the self about mental activity through the perceptual aspect of

consciousness – 'conscious awareness'. The analogy with perception is detectable in the writings of several authors in the present debate. For example, Nisbett & Wilson (1977*a*) use the phrase 'observe the workings of their own minds', and the phrase 'perception of cognitive processes' appears in the title of the paper by Smith & Miller (1978). Natsoulas (1981) provides a critical analysis of this analogy. Through the aspect of consciousness called 'conscious control', the self controls or directs whatever of mental activity is accessible to it. In this way the self communicates what it observes of mental activity through verbal reports. The 'self' makes occasional appearances in statements by these authors: e.g. '. . . *we* may have no direct access to higher order mental processes' (Nisbett & Wilson, 1977*a*, p. 232, italics added).

Despite its plausibility to common sense, this theory is a cultural artifact. For this reason, the ease and automaticity with which it slips into our statements and postulates about consciousness and related matters should not be taken as evidence of its validity. There are many possible relationships between each of experiences, awarenesses, consciousness and introspection and, on the one hand, internal events of any sort, and, on the other hand, verbal reports, which are not demonstrably less valid than this cultural theory (Baars, 1983; Ericsson & Simon, 1980; Rakover, 1983). White (1986) argues that issues of communication of information are not issues of consciousness at all, and indeed information-processing models (e.g. models of verbal report production: Ericsson & Simon, 1980) do not require reference to consciousness or to introspection. Under this view, propositions about consciousness (and related concepts) carry no implications for verbal report accuracy, and propositions about verbal report accuracy carry no implications for consciousness. Any attempt to test a hypothesis about 'introspective access' which uses verbal report content or accuracy as a dependent measure requires some kind of assumption about the relationship between 'introspective access' and the verbal report parameter being used. If an assumption about this relationship cannot be validated, then there is little prospect for using verbal reports to test hypotheses about 'introspective access'.

Rakover (1983) provides an analysis of the assumptions made by Nisbett & Wilson that somewhat resembles that of White. In his view, Nisbett & Wilson claim that the traditional theory of verbal reports has two assumptions: (i) mental entities are the causes for the subject's behaviour (Rakover defines mental processes as causal relations among mental entities); (ii) the subject has introspective access to his or her mental entities. Nisbett & Wilson then argue that these two assumptions imply that the subject should be able to provide a 'true explanation' for his or her behaviour. Nisbett & Wilson reject this 'true explanation' conclusion on empirical grounds and therefore, by the *modus tollens* rule, infer that assumption (ii) is false.

Rakover criticizes this argument in two ways. First, he points out that the falsity of the 'true explanation' conclusion would imply the falsity of *one* of the assumptions, but is not sufficient for the rejection of (ii) in particular. Second, he argues that the inference of the 'true explanation' conclusion from (i) and (ii) actually requires one of two further assumptions: either that the subject can infallibly induce the true explanation from available information, or that the subject directly knows the causes of his or her own behaviour. Rakover argues that neither of these is tenable: the

former because induction is never infallible, and the latter because causes are unknowable. For this reason, even if the 'true explanation' conclusion is false, this does not imply the rejection of either of the original two assumptions.

Rakover's own rejection of the two further assumptions may, however, not be justified. The 'true explanation' inference requires only the assumption that the subject's inductions are reliably correct, which is possible: subjects are not required to be infallible. And it is not certain that causes are unknowable (Shultz, 1982). Nonetheless Rakover is correct in arguing that one cannot infer necessary lack of awareness from a contingent failure to give a 'true explanation': reports can fail to be accurate for many other reasons, as he goes on to show.

In Rakover's own model, he adds to the original two assumptions the following: (iii) introspection is strongly related to the focusing of attention; (iv) the subject's explanation for his/her behaviour is no more than a hypothesis which stems from his or her introspections. Rakover argues that, because of the involvement of attention, one can be aware of mental entities despite not being aware of the stimuli that gave rise to them; or of the stimulus but not the mental entities; or of the mental entities but not the response; or of both the stimuli and the mental entities but not the relationship between them. It all depends on where attention happens to be focused. So, if Rakover's assumptions are valid, then inaccurate verbal report shows only that subjects developed a poor hypothesis from their introspections: it does not prove necessary lack of awareness.

It is very hard to imagine how one could attempt to establish the validity of any proposition about the relationship between 'experience' or 'awareness' or 'conscious awareness' or 'introspection' or 'consciousness' and causal report accuracy. The issue is not unimportant, but again for practical purposes the only immediate way forward involves withdrawing references to 'introspective access' and suchlike from the N–W proposal. This would leave us with a proposal about the possible enhancement of causal report accuracy by internal events and states of any kind. This is somewhat similar to the step taken by Wilson (1985). Although it deals with the issue of consciousness by avoiding it, it has the advantage of perhaps being testable now. The question remains, however, what sort of information is the right sort to enhance causal report accuracy. The next section deals with this question.

The relationship between internal events and causal reports

If references to consciousness and related terms are eliminated, the N–W proposal concerns the availability of internal events to causal reports. In practice, because of their preferred methodology, the proposal concerns the extent to which the accuracy of verbal reports about causal influences on responses can be enhanced by the use of internally available information about events between the stimuli and response in question. This is an important difference, because it raises the problem of how causal influences are to be identified. First, do stimulus–response contingencies enable the *experimenter* to identify accurately the causes of the subject's responses (an issue to be considered in more depth in the methodology section); and, second, would access to the process (however process may be defined) enable the *subject* to identify the causes of his or her responses?

Smith & Miller (1978) opened this debate with a critique of the position effect study (Nisbett & Wilson, 1977*a*; Wilson & Nisbett, 1978). In this study passers-by appraised four actually identical pairs of stockings in a linear display and chose the pair they judged of best quality. The results showed a marked preference for the pair furthest to the right. Subjects did not report that position had influenced their choice but referred instead to aspects of the quality of the garments. Smith & Miller argue that the position of the stockings was not a causal influence on subjects' responses: instead the 'position effect' was a by-product of the process used to appraise the stockings, and it was the appraisal process that caused the response. Subjects would therefore be right to deny that position had influenced their choice. Smith & Miller claim, in effect, that the experimenters have drawn the wrong causal inference from correlational data.

Sabini & Silver (1981) show that things are more complex than this argument makes them appear to be. They begin by pointing out that there is a difference between having access to a process and understanding how it works. According to them, the latter is what is necessary for an accurate report of the process. Without this understanding, which is not itself part of the particular process in question, we might start our account at the wrong place or include things that are actually irrelevant. Everything in the account might be accurate in itself, but not accurate as an account of the process. So having access to a process is not sufficient for giving an accurate account of the process.

Nisbett & Wilson did not ask their subjects for an account of the process, but rather for causal influences on the outcome of the process. Sabini & Silver then argue that access to the process is not sufficient to answer causal questions of this sort either. They distinguish between a programme, which specifies conditionals, and a trace, which is a record of particular events determined by the programme in response to specific input. They argue that if the trace is the process then introspecting upon it will not be sufficient to answer causal questions: for this, knowledge of the programme would be required. That is, one must know the conditionals specified in the programme in order to know *how* the process worked. But if the programme is the process then one could answer causal questions accurately *without* knowledge of or 'introspection upon' the trace, merely by knowing the conditionals in the programme. That is, the way to achieve causal report accuracy does not involve knowledge of the trace, or events occurring between stimulus and response, at all.

Now let us apply this reasoning to the position effect study. Suppose we have a programme of this form: if the current stockings are no worse than the previous pair on dimensions x, y and z, then the current pair is the choice so far. For a trace dealing with stockings that were identical, operation of this programme would result in choice of the pair considered last. This would be the right-hand-most pair if subjects read from left to right. Subjects using this programme who reported that their choice was determined by the attributes of the stocking on dimensions x, y and z would be *correct*; they would also be correct in denying that the position of the stockings had influenced their choice, even though a position effect emerged out of the operation of the programme. But their correct report would derive not from 'access to the process' or to the trace, but from knowledge of the programme.

Therefore Nisbett & Wilson are wrong in taking report accuracy as an indicator of 'access to the process', or indeed of access to any sort of events between stimulus and response. In fact, it would reflect a special type of understanding. Conversely, inaccurate report does not indicate a lack of 'access to the process'.

But, even if subjects have knowledge of programme, process and anything else that may be relevant, they may still give a report that appears inaccurate, as the following argument shows. An important feature of the original proposal was the notion that actors may often rely upon *a priori* cultural theories, rather than upon process information, when making their reports. It is possible, however, that *a priori* theories may act as supplements, rather than alternatives, to internally available information. One of the many things that can happen to information on its way to a verbal report is that it may be transformed or operated upon in some way, particularly if it is not in verbal form to begin with (cf. Ericsson & Simon, 1980). Let us consider an example. While walking along the street I overheard this snippet of conversation: 'I could just *feel* the adrenalin flowing through me'. Taken at face value, this report is obviously wrong: one cannot directly (consciously?) feel minute amounts of chemical in the bloodstream. But this does not mean that it would be correct to infer that no kind of internally available information was being used as a basis for the verbal report. The report equally obviously did not come out of thin air: the woman must have felt something, even if no more than a sense of excitement.

The point is that the verbal report necessarily involved some kind of *translation* of available information. First, the translation was from the form of the experience or available information into words and, second, the translation was into a form that would be readily understood by the listener. To facilitate this understanding, it would be appropriate for the speaker to make use of contemporary cultural beliefs. Although in this instance the use of the belief led to an incorrect statement about what the woman could feel, nevertheless she succeeded in communicating her feeling to her listener, because of their shared cultural beliefs. To do this is not inaccurate: it is practical, effective use of metaphor. Such interpretive, translative or symbolic uses of cultural beliefs by experimental subjects may mislead experimenters into inferring lack of access, or use of cultural theories instead of access, when the proper inference is that the relationship between internal information and reports is merely indirect.

Comment on substantive issues

The published literature since Nisbett & Wilson (1977*a*) has established that the 'access' proposition cannot be maintained. It depends upon a distinction between 'process' and 'content' that has not so far proved susceptible to valid definition; and, both in itself and in its proposed relationship with the verbal report proposition, it implies and depends upon a set of beliefs about 'introspective access' that have not been substantiated. In addition, the proposed relationship between the 'access' proposition and the verbal report proposition has been fractured by the argument that process information is not suitable for enhancing the accuracy of causal reports.

The verbal report proposition, which is more purely empirical, has not been undermined. If we define 'poor' to mean 'no better than that of judgements made

by suitable "observer" subjects', then the verbal report proposition is testable. In that form, however, it has lost all of its theoretical foundation. It implies that, of all possible differences between actors and observers, none can be used by actors to enhance the accuracy of their causal reports.

There are, of course, many differences between actors and observers. Some were noted by Jones & Nisbett (1972): differential salience of actor and situation, visual orientation differences, differences in the need to predict and control behaviour (and perhaps other differences of a broadly motivational nature), and overall differences in amount of information about the actor possessed by each. Wilson & Stone (1985), concentrating on the last of these, identify three types of information available to actors but not to observers: (*a*) idiosyncratic theories, i.e. the actor's personal beliefs about his or her own peculiarities; (*b*) knowledge of the 'workings of one's own mind' obtained via introspection; and (*c*) observation of covariation between the actor's responses and antecedent stimulus conditions.

There are additional possibilities. For example, and in fact in some of the studies discussed below, the observer may sometimes have an inferior or no knowledge of the actor's behaviour on the occasion in question. Also, the observer will often know less about the history of the actor's behaviour, irrespective of the question of covariation assessment.

It would of course be desirable to distinguish empirically between the possible effects of these different sources on report accuracy. This can be achieved to some extent with the use of different sorts of 'observer' groups, subject to other methodological limitations. In particular, in deference to the spirit of the original proposal, it would be desirable to assess the possible effects on report accuracy of information about internal events between stimulus and response, irrespective of whether this information concerns process, content, product or programme, and irrespective of whether its utilization requires any form of consciousness or 'introspective access'. A modified form of the N–W proposal would therefore be that report accuracy cannot be significantly improved by use of this type of information. The review of literature that follows is aimed mainly at assessing this form of the proposal. First, however, general aspects of methodology will be discussed.

Aspects of methodology

The verbal report proposition is unusual in that it leads to predictions of null results: reports by actors and observers should not differ in respect of accuracy. This feature has the advantage that it makes the verbal report proposition comparatively easy to falsify. On the other hand, it leads to the risk that all sorts of dubious experiments that produce null results may be passed off as evidence in favour of the proposal. As Cotton (1980) observes, it may be all too tempting for researchers to use the N–W proposal to explain away uncomfortable data, or a null result on a verbal measure. For this reason aspects of methodology assume an unusual degree of importance in tests of the verbal report proposition, and even more so in attempts to differentiate between different possible sources of actors' superiority over observers.

Tests of the proposition tend to have a similar pattern. Initially the subject is exposed to certain stimuli and makes some kind of response (e.g. a judgement of

liking). The experimenters assess the actual effects of features of the stimulus on the response by looking at stimulus–response contingencies in various possible ways. After the response has occurred, the experimenters elicit a verbal report from the subject by asking questions about the influence of features of the stimuli on the subject's response. The accuracy of the reports thus elicited is assessed by comparison with the 'actual' effects as measured by the experimenters. Since the identification of effects of individual features of the stimuli requires some kind of manipulation of features, samples of stimulus presentations, responses and causal reports are required, fitting some statistical design. Finally, isolation of sources of report accuracy is achieved by comparing report accuracy of actors with the accuracy of judgements made by 'observers' under various possible conditions. Each of these ingredients of the study may influence the validity of the results obtained, and each will be considered in turn below. At least one author, however, has claimed that this entire approach is inappropriate, and his argument is summarized first.

The 'positivistic' bias

McClure (1983) argues that Nisbett & Wilson based their approach on a positivistic conception of the causes of behaviour, under which the subject or 'actor' is merely a passive mediator between the experimenters' machinations and their own responses. Internal events make no contribution to stimulus–response linkages. It is only under this assumption that it can be regarded as appropriate to ask the actor about causal influences of stimuli on responses. If instead one conceptualizes the actor as an active determiner of his or her behaviour, then one does not view stimuli as 'causing' responses. A different causal model becomes appropriate, in which the actor's knowledge of his or her intentions is a critical element.

McClure argues that actors are far superior to observers in terms of knowledge of their intentions. Nisbett & Wilson (1977*a*) also say this, but McClure argues, contrary to Nisbett & Wilson, that intentions are not information but a form of process. He argues that intentions and information are not the same because one can pursue an intention one already has, but one cannot pursue information unless one does not have it. This argument is the victim of an ellipsis, however. What one pursues is not an intention as such but its realization or accomplishment: and, arguably, the pursuit of realization is more likely to be the process than is the intention itself.

The general thrust of McClure's argument, however, is familiar from the 'crisis' literature (e.g. Harré & Secord, 1972): traditional laboratory methods take no account of the active nature of the subject, for example, of the 'causal powers' that reside in the subject. The subject's own account of his or her behaviour may validly refer to these causal powers, in the form of intentions or reasons (cf. Locke & Pennington, 1982). The experimenters' request to report causal influences of extrinsic events not only excludes an entire class, perhaps the major and certainly the final one, of determinants of behaviour, but also pushes the subject into a way of talking about themselves that is essentially foreign to them.

This is not to deny that there is a relationship of some sort between stimuli and responses. But researchers have the wrong causal model of the relationship, in that

it fails to take account of the contribution of events, states, processes and powers internal to the subject. The obverse failure by subjects to abandon *their* causal model, in which these internal things do play the most important part, and to conform to that of the experimenters, does not justify the conclusion that subjects know nothing about what goes on between stimuli and response. It could equally well mean that they know more about that than the experimenters do.

This general critique of methodology is echoed by some of the criticisms of more specific components of the method used by Nisbett & Wilson.

The 'actual effects' measure

The basic requirement for a test of any proposition about causal report accuracy is some means of assessing accuracy. This means that, ideally, the experimenter should know what factors influenced the subject's response, how, and to what extent. The experimenter, not having direct access to events inside the head of the subject, is confined to the use of observables to work out causal influences, and to set a standard for the assessment of report accuracy. The standard used by Nisbett & Wilson is provided by the sophisticated assessment of contingencies between independent and dependent variables.

Obviously this standard has limitations of its own. First, there may be causal influences that are not detected or detectable in stimulus–response contingencies. Second, the procedure depends upon the validity of causal inferences drawn from covariation data. Covariation data are not adequate to justify any particular causal inference, and so making the simplest inference, that manipulating levels of the independent variable caused (or failed to cause) a difference on the dependent variable, is always risky. For example, the inference may take inadequate account of the peculiar way in which the intervening process operates upon the stimuli (McClure, 1983; Sabini & Silver, 1981; both discussed above).

To borrow an example from White (1984), at a certain point in a church service the minister may say 'Let us pray', following which everyone in the congregation goes down on their knees. We may also observe that kneeling behaviour occurred only after this command. The ANOVA-type inference here would be that the behaviour was caused by the command. But one could say with equal validity that the cause of the behaviour was the knowledge, shared by all participants, that going down on one's knees is a conventional part of the act of praying in a church service. A worshipper who was asked why he/she kneeled and who replied 'Because I know that's what one is supposed to do when one prays' would be correct, if incomplete, but would be judged incorrect by the equally and oppositely incomplete ANOVA standard.

In addition, even if it is accurate, the experimenter's causal inference may identify only a very small factor among many that contribute to the response. It is hardly reasonable to expect subjects to refer to a stimulus factor that accounts for only 4 per cent of the variance in responses. This led Kraut & Lewis (1982) to recommend that the actual effects of the manipulated variables should be major (in terms of percentage of variance accounted for).

But let us suppose that subjects do have the ability to report accurately on

stimulus–response contingencies. What does this imply about the influence of internal events on reports? Given that the experimenter is using the stimulus–response contingency standard, the best way for a subject to achieve an accurate report is to use the same method that the experimenters use. To achieve this the subject does not need internal information at all, as indeed the experimenters do not. Naturally the experiment is usually designed in such a way that the subject cannot use this strategy. But the point is that it is not certain what sorts of internal events, if any, might best contribute to the subject's attempts to meet the experimenter's standard. Whatever level of accuracy subjects can attain by that standard, it would not be legitimate to infer that that is the highest level of accuracy they can attain by *any* standard. Much internal information that subjects may have and could use for other sorts of report may well be useless for the sort of report they are requested to give.

Circumstances of reporting

A null report cannot be interpreted as support for the verbal report proposition if subjects might not have been *trying* to use internal information, or were unable to do so because of the conditions under which the report was made. Given that there are several possible sources of information, subjects do not try to utilize internal events if it is less effortful to utilize other sources (Adair & Spinner, 1981; White, 1980). Ericsson & Simon (1980) point out that Nisbett & Wilson provided considerable background information in their experiments, and subjects could have chosen to base their reports on the background information rather than on internal information. Reliance upon *a priori* causal theories is likely to increase when those theories are readily available and when the subject is not highly motivated to take time and trouble at the report stage.

Smith & Miller (1978) argue that reports and behaviour differ in the extent to which they are prone to the influence of response artifacts such as evaluation apprehension (Rosenberg, 1965) and demand characteristics (Orne, 1962). Wilson & Nisbett (1978) argue against a demand characteristics interpretation of their experiments on the grounds that the subjects did not tend to show the actual effects predicted and expected by the experimenters. But demand characteristics are no longer thought of only as cues which reveal the experimenter's hypothesis. Adair & Spinner (1981) contend that the experiments by Nisbett & Wilson contain cues, and through heeding these the subject would acquire a conceptualization of the experiment *different* from that possessed by the experimenter. The subjects then give reports that conform to their conceptualization of the experiment. Thus, their reports appear to be inaccurate because report content is determined by demand characteristics which *mislead* subjects as to the purpose of the experiment. Adair & Spinner show that each of the Nisbett–Wilson experiments can be reinterpreted to support this demand characteristics view.

Type of probe

The causal report is given in response to a question, or probe, from the experimenter. Subjects' responses can be significantly influenced by the question asked by the

experimenter (Salancik, 1974; Salancik & Conway, 1975). Therefore the phrasing and format of the probe may influence the accuracy of the report. The subject is, after all, responding to the question asked, with its implicit demands: the question must therefore be phrased in such a way that a person answering it properly would stand a chance of being accurate by comparison with the actual effects measure. Adair & Spinner (1981) share this concern and recommend research into the effects of probes of different types.

Ericsson & Simon (1980) claim that probes used by Nisbett & Wilson are not closely related to stored information about specific instances of processing. Given a probe with this drawback, a subject presumably has little choice but to use other sources of information. Indirect probes of this sort are more likely to lead to inaccurate reports than are direct probes.

White (1980) comments that a probe may mean different things to experimenter and subject. For an experimenter, a probe acquires meaning from its relation to the experimenter's general background of psychological knowledge and terminology. Subjects share none of these things and are therefore sure to have a different conceptualization which may influence their interpretation of the meaning of the probe. Obviously subjects respond in accordance with their interpretation, not with that of the experimenter.

Type of report

There are many possible types of verbal report content. Those discussed in connection with the present topic include reports of causal influence, programmes (Sabini & Silver, 1981), process (subject to problems of definition), strategies (e.g. in problem solving or memory tasks – Morris, 1981*a*), attentional focus (Kellog, 1982), rehearsal or other current activity (Lieberman, 1979), intention (McClure, 1983; Morris, 1981*b*); and hypotheses or rationalizations about any of the above (Morris, 1981*a*).

In principle the N–W proposal was originally concerned with reports about processes, but a concern this specific leads to two distinct methodological problems. First, in tests of the proposal it has been customary to ask for reports about the causal influence of stimuli on responses, and the role of process information in relation to the accuracy of these causal reports is not at all clear (Sabini & Silver, 1981). For this reason the N–W proposal has effectively become a proposal about *causal* report accuracy, so that it does not cover other types of report. Second, it is very difficult in practice to identify what sort of report a subject is giving on any particular occasion. For example, Morris (1981*a*) argues that it is hard to tell the difference between a rationalization and a direct report of a strategy; but if a subject elects to give a rationalization, it may not be correct to infer that they could not have done otherwise.

Retrospective reports

In the Nisbett & Wilson experiments the report is always retrospective. Ericsson & Simon (1980) argue that, in order to appear in a retrospective report, information

must enter short-term memory. When the limit to short-term memory capacity is reached, information can be transferred to long-term memory, but with the attendant danger that it may not be retrieved when the report is made. Ericsson & Simon point out that verbalization tends to be incomplete when subjects are under a high cognitive load: yet in several experiments the load placed on subjects has been excessive. In Nisbett & Bellows (1977), for example, subjects were required to read three pages of information and to make several judgements before any reports were taken. Under these circumstances much internal information must be unavailable at the time of report.

But the operating characteristics of memory may have more severe and complex effects. White (1985) reanalysed data from a problem-solving experiment by Maier (1931). In this experiment Maier presented subjects with a hint and then measured the amount of time that passed between the hint and the discovery of the solution. White's analysis showed that subjects who reported that the hint had *not* been useful solved the problem significantly more quickly than subjects who reported that the hint *had* been useful. Maier suggested that the sudden appearance of the solution 'dominated consciousness'. Developing this suggestion, the counter-intuitive trend observed in Maier's experiment can be likened to a suffix effect (Baddeley, 1976; Crowder, 1971); that is, the appearance of the solution has a role analogous to that of a suffix in a serial recall task. The effect of a suffix is to impair memory for the most recent events, but to leave memory for earlier events intact. Under this idea, subjects in Maier's experiment who reported that the hint was not useful were suffering failure of memory for the effects of the hint. This was because they had solved the problem more quickly, so that the hint was in the comparatively recent past (about 35 seconds). Other subjects reported the hint as useful because they had taken longer to solve the problem, so that memory for the effects of the hint was less impaired by the appearance of the solution.

This interpretation can be generalized to form a hypothesis about causal report inaccuracy. The possibility is that any product (defined for present purposes as the subject's overt response) of a process has the effect of a suffix. If so, the emergence of the product destroys memory for internal events immediately antecedent to it. Therefore no report taken after the product has emerged can be informed by memory for those events, because no such memory remains. Since this hypothesis predicts inaccurate retrospective verbal reports, none of the studies that have been claimed as support for the N–W proposal can be accepted as such unless and until this suffix effect hypothesis can be ruled out. Future studies need to take careful account of the operation of memory.

Ericsson & Simon (1980) argue that concurrent verbalization, as in the 'think aloud' procedure, may be a more suitable means of revealing information about the process, and the history of problem-solving research shows that the think aloud procedure provides enough information to generate accurate computer simulations of performance on various problems.

However, the think aloud procedure has drawbacks. It has been claimed that it may alter the process being investigated (Kellog, 1982). Ericsson & Simon (1980) argue against this, but their arguments may not hold in the realm of social judgement (Taylor & Fiske, 1981). Taylor & Fiske doubt the usefulness of the think aloud

procedure in social judgement studies. They say that the procedure works best when thought is already in verbal code, a point that applies also to the verbal measures used by Nisbett & Wilson. They also say that the procedure is susceptible to the effects of social desirability, a drawback that does not arise in the field of problem solving. Most pertinent in the present context, think aloud protocols could be accurate by a kind of self-fulfilling effect, whereby the report itself determines the process put into operation. Finally, accurate reports in think aloud studies are not necessarily directly informed by the process itself. Subjects could deduce the strategy being used from available information about the order in which subproblems are attempted (Evans, 1981). They could also use knowledge of the programme to interpret available information into a report about the particular process being used (Sabini & Silver, 1981).

Design

Several authors (Kraut & Lewis, 1982; Smith & Miller, 1978; White, 1980; Wright & Rip, 1981) have objected to the use of between-subject designs in the studies by Nisbett & Wilson. Smith & Miller argue that subjects are required to be aware of the one factor that differentiates their situation from that of other subjects, and this is the very thing that the experimenter is trying to conceal from them. Morris (1981*b*) counters this by arguing that subjects should not need to know the design of the study in order to be aware of the effects of a stimulus factor on their response. Smith & Miller also argue that, since awareness of a mental process is something that happens for one individual, hypotheses about awareness are better tested by within-subject designs, because these permit the accuracy of each individual's reports to be assessed. Several of the more recent studies (Kraut & Lewis, 1982; Sprangers, van den Brink, van Heerden & Hoogstraten, 1987; White, 1980; Wright & Rip, 1981), have used within-subject designs.

Nisbett & Ross (1980) counter these arguments with the claim that within-subject designs make clues to the actual effects public and thereby destroy the utility of 'introspection'. That is, in a within-subject design, subjects could work out the actual effects by assessing stimulus–response contingencies, as the experimenters do. If observer subjects are given the same information about stimuli and responses that real subjects have, then the potential effects of this can be taken into account. The danger, however, is that a null result on a subject–observer comparison would be ambiguous: it could reflect either lack of 'access' or use by both groups of stimulus–response contingencies.

Nisbett & Ross (1980) also comment that within-subject designs are unrepresentative of social reality: 'Life itself has a between design' (p. 220). It is perhaps doubtful whether life has any kind of design.* But the real issue here is the role that

* Variations in conditions can approximate both between and within designs. For example, I could go to a party, holding the situation constant, and observe differences in my responses to different individuals I meet there. It is surely easier for me to do that (quasi-within design) than for me to assess the responses of 10 of my friends to person *X* compared with the responses of the same 10 to person *Y* (quasi-between design). Kelley's multiple observation model of attribution (1967, 1972*a*) has two within-subject dimensions (consistency and distinctiveness) and one between-subject dimension (consensus); and most research has shown that the latter is used by judges less than the other two (Harvey & Weary, 1984; Major, 1980).

'ecological' validity should play in these studies. The verbal report proposition concerns possibility, not common practice, so designing experiments that are representative of social reality is less important than using a design that can best reveal what subjects are and are not capable of.

Comparisons with observer subjects

To summarize, there are many reasons why causal reports may be inaccurate, or judged inaccurate (correctly or not), by the experimenters. The more important of these are listed in Table 1. To infer support for any one of these, the contributions of the others must be ruled out or accounted for in some way. As usual, it is virtually impossible to make any justified interpretation of a null result.

Table 1. Reasons why a retrospective causal report might be judged inaccurate by the experimenter

1. Lack of introspective access to relevant information (Nisbett & Wilson)
2. Lack of availability of internal events between stimulus and response (modified N–W position)
3. Lack of knowledge of programme (as opposed to process) (Sabini & Silver)
4. Translation of information by means of cultural beliefs
5. Inaccurate/imperfect measure of actual effects
6. Incorrect causal inference from stimulus–response contingencies by experimenter (Smith & Miller)
7. Failure of memory for relevant information (Ericsson & Simon, White)
8. Misleading cues to purposes of experiment (Adair & Spinner)
9. Inappropriate probe question (Adair & Spinner, Ericsson & Simon)
10. With observer comparison, simulation by observer of actor's processes (Rakover)
11. Practical concerns lead subjects deliberately to be inaccurate, or not to try (White)

On the other hand, it is not easy to interpret a positive result as evidence against the N–W proposal in its original or modified form. There is no doubt that retrospective causal reports can be very accurate. For example, Nisbett & Bellows (1977) found a correlation of +0.94 between reported and actual effects of information about intelligence on judgements about a hypothetical target. Using a similar task, White (1980) found a correlation of +0.98 between reported and actual effects of information about selfishness. But high accuracy alone is not sufficient to infer that internally available information about events between stimuli and response was the reason for accuracy. This is why Nisbett & Wilson (1977*a*) added the two supplements to the verbal report proposition: to justify such an inference, reports by actors should be *more* accurate than guesses made by outside observers.

But in fact, even this is not enough. Nisbett & Wilson also stated that actors' reports could be more accurate than observers' guesses because of actors' knowledge of content (as opposed to process), past actions or idiosyncratic personal theories. None of these is available to observers, but none is included in the original 'access' proposition. Even the modified form of the proposal does not include the latter two categories.

Adair & Spinner (1981) argue that observers' judgements could be less accurate because the experiment had been inadequately described to them. In the experimental conditions employed by Nisbett & Wilson, observers go through a procedure similar to that for actors, except that the critical manipulation is omitted. They are essentially a control group, and are used as such to assess the effect of the manipulation, in the between-subject designs favoured by Nisbett & Wilson. They are then presented with the manipulation and asked how they would have been influenced by it if it had been included in their procedure. This appears adequate to counter the main force of Adair & Spinner's point, though it is true that the observers' knowledge of the manipulation is out of context and hypothetical. But Adair & Spinner add that the observer condition used by Nisbett & Wilson is equivalent to a 'non-experiment', and that actor–observer similarity under such circumstances is usually interpreted as suggestive of demand characteristics. This makes it very difficult to interpret actor–observer similarity of report accuracy.

Wilson & Stone (1985) argue that actors can achieve superior report accuracy by assessing covariation between stimulus conditions and their own responses, something which observers are not generally given the information to be able to do. This would be a reasonable strategy for an actor who was trying to achieve an accurate report, because it is closer to the method that the experimenters use to assess actual effects. Wilson & Stone (1985) report a study in which they find evidence for actors' use of this strategy, and it may account for some of the other evidence of actors' superiority (see Table 3).

Possible reasons why actors' reports may be superior to those of observers are summarized in Table 2. Before superiority of actors' reports over observers' judgements can be taken as evidence against the N–W proposal, all of these possible reasons except the first must be accounted for.

Table 2. Reasons why actors' causal reports may be more accurate than observers' guesses

1. Availability to actor of relevant internal events between stimulus and response
2. Actors' knowledge of programme (as opposed to process) (Sabini & Silver)
3. Actors' knowledge of past history of own reactions (Nisbett & Wilson)
4. Actors' idiosyncratic personal theories (Nisbett & Wilson)
5. Actors' superior knowledge of own behaviour in experiment
6. Actors' superior knowledge of covariation between stimuli and own responses (Wilson & Stone)
7. Actors' superior knowledge of conditions of experiment (Adair & Spinner)

Two variations on standard observer conditions have been used in attempts to account for the effects of these various reasons. Wright & Rip (1981) attempted to exclude actors' knowledge of their own past reactions and personal idiosyncratic theories by using observers who knew individual subjects well. Although an improvement, this method would only provide a complete control for those things if the observers knew as much about the actors as the actors did. Also it fails to rule out some of the other factors listed in Table 2. The observers did not participate in

the experiment directly, so their knowledge of the manipulated variables might have been inadequate.

White (1980) and Sprangers *et al.* (1987) used a procedure called the 'actors' control condition' devised by White (1980). In this, a group of subjects goes through the same procedure and manipulations as the actor subjects, except that they do not go through the report phase until some time later (one week in White's study). By this time memory for internal events at the time of judgement is presumed to have declined significantly. The actors' control condition subjects are given back all of the original stimuli and their responses when they make their reports. Kraut & Lewis (1982) used a similar procedure with a delay of half an hour and a distractor task, but subjects did not see their responses again. In the actors' control condition, knowledge of manipulated variables is not hypothetical, and the subjects know both the stimuli and their own responses, as the actors do.

All forms of observer comparison run a certain risk, however. Observers can achieve a level of accuracy equivalent to that of actors by a form of active role playing in which they simulate the processes through which the actor has gone. In such a case the observer is just as privileged as the actor, and may do just as well in consequence (Rakover, 1983). Wilson & Stone (1985) make a similar argument that observers may be helped by memories of their own past reactions to similar stimuli. The actors' control condition is perhaps especially susceptible to these weaknesses, which again make null results almost impossible to interpret. On the other hand, evidence of actors' superiority would be sufficient to rule out these possibilities.

Experimental results

We can now proceed to a survey of studies of causal report accuracy. One such survey has already been published (Wilson & Stone, 1985). Wilson & Stone state that they summarized 'all known research addressing the accuracy of actors' versus observers' causal reports, conducted since the Nisbett & Wilson studies' (p. 169). In fact their survey included one study that did not ask for a causal report (Wilson, Laser & Stone, 1982, in which the dependent measure was a predictive judgement of covariation) and excluded two that did (Kraut & Lewis, 1982, actors' control comparison; White, 1980). Two more studies have appeared since that review, one by Wilson & Stone themselves.

Wilson & Stone opted for a quantitative assessment, converting (where necessary) accuracy data to a common format and averaging across studies. This sort of summary is only appropriate when each of the studies included in it has high internal validity. In general, in the present literature this is not the case. (For instance, Wilson *et al.*, 1982, p. 540, comment that the Weiss & Brown, 1977, study, which is included in the quantitative summary by Wilson & Stone, 1985, has 'a number of methodological problems'.) For this reason, a quantitative summary may give a misleading impression of the state of the empirical case for or against a proposal. The main objectives in the present review are to assess the internal validity of the studies and to identify the factors that may have contributed to the results obtained.

Considering first what may be called null results, several studies have found actors' reports either of zero accuracy (four studies in Nisbett & Wilson, 1977*a*, *b*) or no

more accurate than causal judgements made by 'observers' (two studies in Kraut & Lewis, 1982; Nisbett & Bellows, 1977; Nisbett & Wilson, 1977*a*; Sprangers *et al.*, 1987; Weiss & Brown, 1977, actors' control comparison; Wilson & Stone, 1985, interaction judgements). In most of these studies it is impossible to rule out more than one or two of the 11 factors listed in Table 1. The exception is the study by Wilson & Stone (1985), which involved a comparison with another condition in which actors were more accurate than observers. This comparison will be discussed shortly.

Turning to what may be called positive results, some studies have found actors' causal reports either of greater than zero accuracy (Smith & Miller, 1978) or more accurate than causal judgements by 'observer' subjects (Kraut & Lewis, 1982, observer comparison; White 1980; Wilson & Stone, 1985, main effect judgement; Wright & Rip, 1981, two studies). These studies are listed in Table 3, together with the possible sources of the positive result in each case (see Table 2). Again, in most cases there are several possible explanations for the results obtained. It is true that the only factor common to all cases is availability to the actor of relevant information about events between stimuli and response. Moreover, one of the positive results (White, 1980) did exclude all factors other than this one. But there are reasons for doubting whether any weight can be attached to either of these observations, as consideration of individual studies will show.

Smith & Miller (1978) reanalysed data from Nisbett & Bellows (1977) on a within-subject basis, and found positive correlations between ratings and reports. Despite the use of an observer condition in the original study, however, Smith & Miller did not incorporate an observer comparison in their analysis, so the source of the observed accuracy is unknown.

Table 3. Studies in which actors' reports were either significantly accurate or more accurate than observers' guesses

	Possible sources of accuracy						
	1	2	3	4	5	6	7
Smith & Miller (1978)	A	A	A	A	A	A	A
White (1980)	A	AO	AO	AO	AO	AO	AO
Wright & Rip (1981)							
Study 1	A	A	AO	AO	A	A	AO
Study 2	A	A	A	A	AO	AO	AO
Kraut & Lewis (1982)							
Observer comparison	A	A	A	A	A	A	A
Wilson & Stone (1985)							
Main effect judgement	A	A	A	A	A	A	AO
Gavanski & Hoffman (1987)	A	A	A	A	AO	AO	AO

Note. The seven possible sources of accuracy are as listed in Table 2, and in the same order. A indicates that the source in question was available only to actors in that study; AO indicates that the source in question was available to both actors and observers: therefore, actors' superiority in a given study is attributable to any cell marked A in the row for that study.

White (1980) used an actors' control condition to rule out other sources of accurate report. The statistical analysis was, however, flawed (N. Anderson, personal communication). A reanalysis of White's data by Sprangers *et al.* (1987) found no significant differences between actors and actors' control subjects, although in all four comparisons the actors were the more accurate group.

In their first study, Wright & Rip (1981) used an observer condition containing people who knew individual subjects well. The observers did not see the actors' responses, however. The task in this study was to rank order 32 stimuli, and no report was made until the completion of the entire task. This task must overload short-term memory, thereby reducing the likelihood that the positive result was due to information about relevant internal events (Ericsson & Simon, 1980). In their second study, Wright & Rip ruled out contingently accurate impression-management as a source of report accuracy, but this time used observers who were not familiar with individual subjects.

The observers used by Kraut & Lewis (1982) were not familiar with individual subjects, nor did they see actors' responses, or participate in a proper control condition. They merely guessed how they would have weighted the manipulated factors.

The remaining positive result was obtained by Wilson & Stone (1985). The authors had the idea that actors could be superior to observers by means of assessing covariations between stimulus features and their responses. They tested this by manipulating seven variables in stimulus presentations and asking actors to report effects of (*a*) each factor alone (main effects), and (*b*) each factor in combination with each other factor (two-way interactions). Each of these was compared with judgements made by observers who did not see subjects' responses. The reasoning was that covariation assessment should be easier for main effect judgements than for interaction judgements: therefore actors should do better than observers on the main effect judgement but not on the interaction judgement, if they use covariation assessment to inform their report. The results were consistent with this reasoning.

It is not easy to assess the merit of a study reported so briefly, and it is to be hoped that a fuller account will be published. For example, it would be valuable to know what steps were taken to ensure that subjects understood the probe question used at the report stage: it is often difficult to explain to someone what an interaction is, even when the explainer knows well enough. Also, subjects made a large number of reports only at the completion of a long series of judgements. Again, the amount of information they would be required to store far exceeds the capacity of short-term memory, so it would not be surprising if subjects opted for an easier method of making a causal report. This point has been made of some of the other studies reviewed above. Wilson & Stone make a good case for covariation assessment as one of the sources of actors' superiority in these studies (although this cannot be the case for Wright & Rip, 1981, Study 2, because the observers here were able to assess covariation between stimulus features and actors' judgements). But this does not count against the possibility of using other sources and means of judgement, under circumstances less unfavourable to them.*

* Since writing the above, I have read a paper by Gavanski & Hoffman (1987) that provides relevant evidence. Gavanski & Hoffman compared actors with a group of observers who were able to assess covariation between, for

The actors' control condition method provides the best prospects for a test to isolate the contribution of relevant internal information to causal reports. In addition to the study by White (1980), the method has been used by Kraut & Lewis (1982) and by Sprangers *et al.* (1987), in both cases with null results. In the former case, however, no report was made until all of 50 stimuli had been presented and responded to, leading to the familiar problem of overloading short-term memory. In the latter case, Sprangers *et al.* actually made things too easy. They used stimulus presentations in which three factors were held constant while a fourth was manipulated (four levels). With this sort of presentation it would be easy for both actors and observers to work out the effects of the manipulated factor by looking at the covariation between their responses and the levels of the factor, a total of only four pairs of observations. This was, in fact, the basis for the authors' assessment of actual effects. Consistent with this reasoning, both actors and observers made highly accurate judgements, equivalent to a reported vs. actual effects correlation of 0.80.

How can one find the happy medium between stimuli that are too complex and stimuli that are too simple? One way is to request a report of influence after each stimulus presentation, instead of waiting until the end of a series. Also, a slightly different means of assessing actual effects can be employed. For instance, suppose one presents a stimulus with pieces of information A, B, C, D, E and F. In the report, the subject is asked *only* about the influence of one of these, let us say F. Then, much later (or earlier) in the stimulus series, one has a stimulus consisting of A, B, C, D and E. The difference between these two on the dependent measure gives a rough indication of the contribution of F to the first judgement. It is rough, because it is bound to be contaminated by a certain amount of noise, and because it does not distinguish main from interactive effects. But it is not clear that subjects make this distinction in their reports, and the roughness is equal for actors and observers. If anyone thinks the issue is worth pursuing empirically, this procedure might assist the chase.

To summarize, in its broadest, atheoretical form, the verbal report proposition has been falsified by the results of four studies. Under some circumstances, actors can make causal reports that are both intrinsically highly accurate and more accurate than the causal judgements of observers. The studies conducted so far, however, do not entitle us to draw any conclusions about the possible contribution to causal report accuracy of information about internal events between stimulus and response. If the results of studies do not support the N–W proposal because of methodological imperfections, neither do they go against it because of the failure to rule out other possible sources of accurate report. On the other hand, there is some evidence for utilization by actors of other sources – *a priori* cultural beliefs (Kraut & Lewis, 1982), and assessment of covariation between stimuli and responses (Wilson & Stone,

each observer, the target stimuli and one actor's response. Gavanski & Hoffman found that covariation assessment contributed to report accuracy, but that even when both covariation assessment and shared cultural theories were taken into account, actors' reports were still significantly more accurate than those of observers. Gavanski & Hoffman interpret this as evidence that people have some sort of privileged knowledge concerning influences on their own judgements. However, because the observers did not know the individual actors whose responses they had been given, any of the first four sources listed in Table 2 could be the source of actors' superiority in that study. Again, therefore, the results are inconclusive.

1985). The questions remain when and why people use these sources and when they might choose others instead.

Discussion

Current status of the N–W proposal

The subsequent literature has established that the N–W proposal cannot be maintained. The access proposition suffers from a failure to define the key term 'process', from dependence upon untested assumptions about the nature of and limitations upon 'access', and from uncertainty about the *type* of information or 'access' that would be beneficial to the accuracy of a causal report. Evidence against the verbal report proposition, considered on its own, has been found in four studies. Although several other studies have found null results consistent with the verbal report proposition, the difficulty in designing a proper test of the proposition and the many reasons for which null results may occur (see Table 1) mean that no inferences can be drawn from these null results.

Some authors in other areas of research have made use of the N–W proposal, for example, to interpret a null result on a verbal report measure, or to support a claim that people are not aware of processes occurring in the researcher's area of interest (e.g. Groff, Baron & Moore, 1983; Lewicki, 1982). These authors typically do not refer to any of the critical sequelae to the N–W proposal. A survey of recent volumes of three leading journals in social psychology revealed 24 papers on other topics that cited Nisbett & Wilson (1977*a*): of these, only four cited any of the critical sequelae (White, 1987). From the present review it can be concluded that there is no justification for such practices. On the contrary, such citation tendencies may help to spread and perpetuate myths about the 'unreliability' of verbal reports, and conjectures about the limitations upon 'introspective access' – effects which would not be in the best interests of science.

Verbal report accuracy

Verbal reports *other* than causal reports can be extremely accurate and informative, in ways that outside observers could not hope to emulate. Several authors have argued that we should be asking when, and not if, reports are reliably accurate.

For example, Lieberman (1979) claims that the best measure of amount of rehearsal in memory tasks is the subject's report of the occurrence of rehearsal. This argument is based on data from Kroll & Kellicutt (1972), who found correlations ranging from 0.93 to 0.97 between subjects' reports of how much they were rehearsing target material and a subsequent recall measure. In this case the 'report' was a button press, made whenever rehearsal was occurring. A more indirect measure, based on observed performance on a distractor task, had a much lower correlation with subsequent recall.

Warshaw & Davis (1984) give self-prediction as an example of verbal reports that tend to be accurate. Self-predictions could not be informed by specific internal events because, by definition, they concern things that have not yet happened. But Warshaw &

Davis show that people who report themselves as having high self-understanding are better at self-prediction than people who report themselves as low in self-understanding. This suggests that consistent individual differences in report accuracy may be worthy of further study.

Kellog (1982) refers to the distinction made by Shiffrin & Schneider (1977) between 'controlled' and 'automatic' processing, and argues that reports can be informative and useful when they concern 'controlled' processing, but not when they concern 'automatic' processing. Kellog describes an experiment in which he took a verbal report measure of the amount of attention given to a stimulus. In conditions where processing was ascertained to be 'controlled', the report was significantly positively correlated with degree of conceptual memory for the stimuli. In conditions where processing was ascertained to be 'automatic', there was no such correlation. Kellog's argument raises the possibility that the null results in the causal report literature may be explained by the fact that subjects were asked for reports on processes that were automatic. Langer (1978) argued that much of social behaviour is automatic (her term was 'mindless'), but she did not look specifically at social inference. A possibly related argument by Morris (1981*b*) is that Nisbett & Wilson studied only non-intentional processes, by which he means processes for which words like choice and responsibility are not usually considered appropriate. Explanations are usually expected only for intentional components of actions, and Morris argues that we can expect reports to reveal insight into these.

Overall, it seems likely that people have more difficulty with causal reports than with reports of other kinds: causal reports could turn out to be a small island of inaccuracy in a sea of insight.

Recent work by Wilson

Wilson (1985) has continued to pursue the issue of 'access to mental states'. No longer distinguishing between process and content or product, Wilson defines the states with which he is concerned as 'psychological dispositions that exert a causal influence on behavior' (p. 10). Wilson suggests a model in which there are two mental systems: one which is non-conscious and which mediates behaviour; and one which is conscious and which attempts to verbalize and communicate mental states. The latter, verbal system may often have limited access to mental states, and then makes inferences about what these might be.

Wilson (1985) interprets the results of studies by Wilson, Dunn, Bybee, Ryman & Rotondo (1984); Wilson, Hull & Johnson (1981) and Wilson & Linville (1982) in the light of this idea. With the emphasis shifting to 'mental states', these studies have not employed causal verbal reports but instead self-report measures of the states under study. These have included liking for a puzzle (Wilson *et al.*, 1981) and liking for photographic slides of scenery (Wilson *et al.*, 1984). The studies have mainly been concerned with demonstrating the relative independence of the behavioural and self-report measures from each other (to support the suggestion of two separate mental systems). But accuracy of self-report is inferred by comparison with the behavioural measure, on the assumption that the behavioural measure is diagnostic of the mental states in question. In these studies, subjects are asked to 'reflect consciously upon the

reasons for their behaviour' (Wilson, 1985, p. 19). This instruction leads sometimes to greater accuracy (Wilson *et al.*, 1981) and sometimes to lesser (Wilson *et al.*, 1984).

The issues raised by this work are exceedingly complex. It has long been understood that verbal and behavioural measures may be affected differently and by different types of factor. A great deal of research effort in the area of verbal and behavioural measures of attitude has been devoted to identifying these factors and the roles they play (Ajzen, 1982; Ajzen & Fishbein, 1977; Cialdini, Petty & Cacioppo, 1981; Cooper & Croyle, 1984; Fazio, Powell & Herr, 1983; Fishbein & Ajzen, 1975, 1980; Wicker, 1969). Methodological factors have been responsible for many of the low correlations observed between verbal and behavioural measures of attitude. For example, the more closely the two measures correspond in their specification of action, target, context and time, the higher the correlation between them (Ajzen & Fishbein, 1977; Cialdini *et al.*, 1981; Jaccard, King & Pomazal, 1977). In addition, behaviour may be determined by many factors as well as attitudes, such as intentions (Fishbein & Ajzen, 1975) and habits (Bentler & Speckart, 1981) – factors which may not influence responses on the verbal measure. What holds for attitudes may hold for the other mental states in Wilson's list. Wilson (1985) also acknowledges that factors such as self-presentation may influence verbal but not behavioural responses. The question is, when the contributions of all of these factors have been taken into account, is the dislocation between behavioural and verbal responses sufficiently great to justify the postulation of two separate mental systems?

Wilson's exposition is peppered with references to 'access' and such terms as 'conscious scrutiny', 'conscious control', 'consciously to search', 'to reflect consciously', and so on. If we ask who or what it is that has 'access', his answer is always 'people'. From this it appears that Wilson is persisting in using the common-sense theory discussed by White (1980, 1982, 1986, see above). Essentially the two 'systems' proposed by Wilson are the 'self', with its tool or faculty of consciousness, and the remainder of mental activity. Why is the remainder of mental activity set aside from the person, rather than characterized as *part of* the person? What exactly is this 'access' of which the person is supposed to have a limited amount? The meaning of the term, the actual operation of access, is never discussed by Wilson. The point is that, when we start to take these terms out from the cultural belief system in which they are embedded, and to ponder what they could actually mean, it becomes apparent (*a*) that Wilson's suggestion of two systems is not novel at all, but merely a reworking of common sense, and (*b*) it contains almost nothing: it only appears to contain something because our cultural beliefs fill the gaps for us. The latter comment echoes the assessment by Ericsson & Simon (1980) of the original N–W proposal:

> ...[Nisbett & Wilson] did not propose a definite model of the cognitive processes as a framework for interpreting the findings they surveyed. Their theoretical interpretations of these findings are entirely informal, resting on an undefined distinction... (p. 245).

Ericsson & Simon were advocating an information-processing approach to models of verbal report production. While such an approach is not without

drawbacks, it has the advantage that operations involved in verbal report production, and the relations between them, are specified with some exactness. One does not suggest that there are two 'systems', and talk about one having 'limited access' to the other: one describes the operating characteristics of the systems and the functional, communicative links between them. In fact, models of, or incorporating, 'consciousness' are increasingly met with in cognitive/information-processing psychology (Baars, 1983; Hilgard, 1976; Mandler, 1984; Marcel, 1983; Posner & Warren, 1972; Shallice, 1972, 1978; Shiffrin & Schneider, 1977), although whether they have been or even can be successful as models of consciousness is another question (White, 1986).

Cultural belief in psychology

The issue of tacit cultural beliefs, discussed above in relation to the N–W proposal, surfaces again in Wilson's recent work. Both proponents and critics of the N–W proposal have been influenced in their writings by tacit, untested, cultural beliefs, particularly about consciousness but also about the structure and *modus operandi* of the mind in general. According to White (1986) these beliefs involve an entity (called the self) which through its faculty of consciousness is acquainted with other parts of mental activity ('conscious awareness') and controls or directs them ('conscious control'). The relationship between the self and whatever of mental activity is linked with it through consciousness is intimate and mutually influential. By contrast, whatever of mental activity is *not* linked to the self through consciousness is both unknown to and uncontrolled by the self. It is so divorced from the self as to be considered not part of it. The self is therefore identified or closely associated with a subset of mental activity, and the argument in the literature has tended to concern the size of that subset, not the validity of the cultural theory of which it is part.

This distinction between self and conscious mental activity and not-self and non-conscious mental activity in cultural belief can be related to other dichotomies of a cultural origin, for example, that between reason and emotion. In Averill's (1974) analysis, reason and emotion are culturally identified by the attribution to them of opposed characteristics. For example, we are supposed to be the 'passive victims' of emotions, meaning that they control us, rather than we them, whereas we are active users of the faculty of reason. Averill shows that the opposition of emotion and rationality can be traced back at least as far as Anaxagoras (*c.* 500 BC), and has continued to influence researchers on emotion down to the present day. He argues that the distinction is of symbolic origin, meaning that it involves symbolic relationships which are extrinsic from a scientific point of view. He provides strong arguments against the validity of ascribing to emotion the properties it has in common-sense belief (see also Averill, 1983).

In fact it can be argued that the distinction between emotion and reason can be seen as part of the larger cultural theory of the mind. In calling emotions uncontrolled, we mean to say that they fall outside the subset of mental activity that is subject to 'conscious control', i.e. they are outside the self. Rationality, on the other hand, is inside that subset. A major decision for any adherent of this larger theory who engaged in research on an attribute or function of the mind would be in

which of those two subsets that attribute or function should be placed. The N–W proposal in its original form was essentially the postulate that all mental processes belonged in the not-self subset. Wilson's (1985) more recent two-system idea is just a different way of describing the cultural theory as a whole. Other areas of psychology may have incorporated similar decisions: I have suggested that Averill's analysis of the treatment of emotion in science constitutes one.

It is important that what is tacit should be made fully explicit. What is untested should be tested, otherwise it should not be used as an assumption on which to erect scientific theory. This applies to theories of 'introspective access' and verbal reports, just as it does to theories of emotion. I would venture to suggest that further real progress in this area depends largely upon the elimination of cultural beliefs from our theoretical statements.

The importance of accuracy to the lay-person

The giving of a verbal self-report is a social act. To understand the determinants of report content, and therefore of its accuracy, we have to understand the actor's involvement in the social context in which it is given. Although some authors have pointed out the possibility of self-presentational effects (Warshaw & Davis, 1984) and self-deception (Wilson, 1985), it is assumed in most studies that actors are *trying* to give accurate reports, as if accuracy were the only, or at least the most important, consideration. The assumption of a concern with accuracy is a characteristic of the 'lay scientist' approach to human inference and judgement. In this approach, the lay-person is depicted as a naïve version of a professional scientist, aspiring to scientific and statistical standards of accuracy and rigour, but using more informal methods and possessing an imperfect grasp of 'normative' scientific and statistical principles, as well as unavoidable cognitive limitations. Originating in the work of Kelly (1955) and Heider (1958), this approach was developed in attribution theory (Kelley, 1967, 1972*a*, *b*, 1973) and in human inference research (Nisbett & Ross, 1980; Ross, 1977). In the 'lay scientist' approach, judges are guilty of errors or biases when their judgements fall short of the normative standard. From this perspective, in the causal report accuracy literature subjects are attempting to provide the most scientifically accurate reports possible, but fall short because of inescapable cognitive shortcomings.

This approach has been criticized by some authors. In the view of White (1984), for example, people are concerned not with making judgements that are accurate but with making judgements that best suit the practical concerns that pertain for them at the time. Accurate judgement is sometimes, and to some extent, in the interests of the judge's practical concerns. But people will aspire to accurate judgement *only* to the extent that this is appropriate to their practical concerns. Often, other considerations may carry more practical force than accuracy. The actor may, for example, wish to excuse or justify an action, or explain why it is not as mysterious or odd as it seemed. They may wish simply to communicate clearly, dramatically, or with style, by making use of a contemporary cultural belief (as in the adrenalin example cited earlier, on p. 22). They may wish to pre-empt or prevent some undesirable response from their listener. They may wish to gain approval or power.

They may wish to avoid making fools of themselves (Evans, 1981; Morris, 1981*a*). There are many other possibilities.

To take an example, several of the studies have involved presenting people with information about real or fictitious target persons, and asking for a liking judgement (Nisbett & Bellows, 1977; Sprangers *et al.*, 1987; White, 1980; Wilson & Stone, 1985). Here, subjects may be suffering from evaluation apprehension, or concerned with self-presentation (Baumeister, 1982), e.g. should I admit to liking someone a lot more simply because they are described as highly intelligent, or physically attractive? What will the experimenter think of me if I do? Perhaps I should tone that down in my report and play up something else instead?

Inaccuracy in a causal report is therefore not necessarily an error or bias. The appropriate standard for assessing error or bias in causal reports is set by the practical concerns that pertain for the subject at the time, not by the experimenter's measure of scientific accuracy. Practical concerns may have two effects. They may lead subjects deliberately to give reports that are inaccurate by the scientific standard of the experimenters; and, to the extent that the subject is thinking about pertinent practical concerns during the report task, these thoughts may interfere with memory for, or withdraw attention from, the relevant events between stimulus and response, thus reducing the extent to which those events may contribute to report accuracy. In addition, the subjects may be concerned not to waste too much time and effort on a task that has little personal meaning: careful consideration of internal information may just not be worth the bother. Observers do not have the same practical concerns as actors, because the judgements they are making do not relate to their own behaviour. For them, therefore, there is less likelihood that accuracy will be low on the list of priorities.

A methodological solution to the effects of practical concerns on scientific accuracy would be to ensure that it is in the practical interests of the subjects to make scientific accuracy their highest priority. This can be achieved with a standard manipulation of evaluation apprehension (Rosenberg, 1969). For example, one could include in the task instructions information to the effect that previous research has shown that psychologically mature individuals use their knowledge of the workings of their own minds to make very accurate reports about the effects of pieces of information on their judgements. The chosen observer group would have to be given a similar kind of push, to make a fair comparison.

But it is the theoretical implications that are more important. The N–W proposal has reached a dead end largely because the need to make it testable resulted in the elimination of most of the theoretical content. Further consideration of the social and practical aspects of the giving of causal reports would provide a means of revitalizing the proposal by injecting a different kind of theoretical content. An emphasis on the comparison between scientific standards of report accuracy and practical standards of report effectiveness could lead to improved understanding of the determinants of causal report content, and thereby to more accurate assessment of the degree and type of insight people have into external causal influences on their own actions.

Conclusions

This topic has focused the attention of researchers on a confluence of several issues of deep and general significance: the nature of consciousness and 'introspective access', the relationship between internal and external determinants of behavioural and verbal responses, the cognitive and social factors involved in the production of verbal reports, the circumstances under which verbal reports are and are not likely to be accurate, the methodology of research with active human subjects, and others. Issues relating to consciousness/introspection and to verbal reports are of relevance to most if not all of psychology. Consciousness, for example, has been investigated from a bewildering variety of points of view in the years since it emerged from the shadow of behavourism (Baars, 1983; Davidson & Davidson, 1980; Helminiak, 1984; Hilgard, 1980; Natsoulas, 1981; Pope & Singer, 1978; Sperry, 1969, 1970; Underwood, 1982), though some of this variety may be due to the fact that the authors are addressing different problems. Much research in psychology uses some kind of verbal response as a dependent measure (Ericsson & Simon, 1980), and a strong case has been made for the usefulness of 'introspective' verbal reports as measures (Kellog, 1982; Lieberman, 1979). It therefore seems unprofitable for different areas to proceed in isolation from each other. The literature on the present topic is closely related to that on the attitude–behaviour relationship, attribution and social inference, and to cognitive/information-processing psychology, and scarcely less closely to other areas. Although the original N–W proposal may be defunct, perhaps its most valuable contribution would be to ensure that the swirls, eddies and currents of this confluence continue to be explored, in all their fascinating complexity.

References

Adair, J. G. & Spinner, B. (1981). Subjects' access to cognitive processes: Demand characteristics and verbal report. *Journal for the Theory of Social Behaviour*, **11**, 31–52.

Ajzen, I. (1982). On behaving in accordance with one's attitudes. In M. P. Zanna, E. T. Higgins & C. P. Herman (Eds), *Consistency in Social Behavior: The Ontario Symposium*, vol. 2. Hillsdale, NJ: Erlbaum.

Ajzen, I. & Fishbein, M. (1977). Attitude–behaviour relations: A theoretical analysis and review of empirical research. *Psychological Bulletin*, **84**, 888–918.

Averill, J. R. (1974). An analysis of psychophysiological symbolism and its influence on theories of emotion. *Journal for the Theory of Social Behaviour*, **4**, 147–190.

Averill, J. R. (1983). Studies on anger and aggression: Implications for theories of emotion. *American Psychologist*, **38**, 1145–1160.

Ayer, A. J. (1973). Privacy. In A. J. Ayer, *The Concept of a Person and Other Essays*. London: Macmillan.

Baars, B. J. (1983). Conscious contents provide the nervous system with coherent, global information. In R. J. Davidson, G. E. Schwartz & D. Shapiro (Eds), *Consciousness and Self-Regulation*, vol. 3. New York: Plenum.

Baddeley, A. D. (1976). *The Psychology of Memory*. New York: Basic Books.

Baumeister, R. F. (1982). A self-presentational view of social phenomena. *Psychological Bulletin*, **91**, 3–26.

Bentler, P. M. & Speckart, G. (1981). Attitudes 'cause' behaviours: A structural equation analysis. *Journal of Personality and Social Psychology*, **40**, 226–238.

Cialdini, R. B., Petty, R. E. & Cacioppo, J. T. (1981). Attitude and attitude change. *Annual Review of Psychology*, **32**, 357–404.

Cooper, J. & Croyle, R. T. (1984). Attitudes and attitude change. *Annual Review of Psychology*, **35**, 395–426.

Cotton, J. L. (1980). Verbal reports on mental processes: Ignoring data for the sake of the theory? *Personality and Social Psychology Bulletin*, **6**, 278–281.

Crowder, R. G. (1971). Waiting for the stimulus suffix: Decay, delay, rhythm and readout in immediate memory. *Quarterly Journal of Experimental Psychology*, **23**, 324–340.

Davidson, J. M. & Davidson, R. J. (Eds) (1980). *The Psychobiology of Consciousness*. New York: Plenum.

Ericsson, K. A. & Simon, H. A. (1980). Verbal reports as data. *Psychological Review*, **87**, 215–251.

Evans, J. St B. T. (1981). A reply to Morris. *British Journal of Psychology*, **72**, 469–470.

Fazio, R. H., Powell, M. C. & Herr, P. M. (1983). Toward a process model of the attitude–behaviour relation: accessing one's attitude upon mere observation of the attitude object. *Journal of Personality and Social Psychology*, **44**, 724–735.

Fishbein, M. & Ajzen, I. (1975). *Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research*. Reading, MA: Addison-Wesley.

Fishbein, M. & Ajzen, I. (1980). Acceptance, yielding, and impact: Cognitive processes in persuasion. In R. E. Petty, T. M. Ostrom & T. C. Brock (Eds), *Cognitive Responses in Persuasion*. Hillsdale, NJ: Erlbaum.

Gavanski, I. & Hoffman, C. (1987). Awareness of influences on one's own judgments: The roles of covariation detection and attention to the judgment process. *Journal of Personality and Social Psychology*, **52**, 453–463.

Groff, B. D., Baron, R. S. & Moore, D. L. (1983). Distraction, attentional conflict, and drivelike behaviour. *Journal of Experimental Social Psychology*, **19**, 359–380.

Harré, R. & Secord, P. F. (1972). *The Explanation of Social Behaviour*. Oxford: Blackwell.

Harvey, J. H. & Weary, G. (1984). Current issues in attribution theory and research. *Annual Review of Psychology*, **35**, 427–459.

Heider, F. (1958). *The Psychology of Interpersonal Relations*. New York: Wiley.

Helminiak, D. A. (1984). Consciousness as a subject matter. *Journal for the Theory of Social Behaviour*, **14**, 211–230.

Hilgard, E. R. (1976). Neodissociation theory of multiple cognitive control systems. In G. E. Schwartz & D. Shapiro (Eds), *Consciousness and Self-Regulation*, vol. 2. New York: Plenum.

Hilgard, E. R. (1980). Consciousness in contemporary psychology. *Annual Review of Psychology*, **31**, 1–26.

Jaccard, J., King, G. W. & Pomazal, R. (1977). Attitudes and behaviour: An analysis of specificity of attitudinal predictors. *Human Relations*, **30**, 817–824.

Jones, E. E. & Nisbett, R. E. (1972). The actor and the observer: Divergent perceptions of the causes of behavior. In E. E. Jones, D. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins & B. Weiner (Eds), *Attribution: Perceiving the Causes of Behavior*. Morristown, NJ: General Learning Press.

Kelley, H. H. (1967). Attribution in social psychology. In D. Levine (Ed.), *Nebraska Symposium on Motivation*, vol. 15. Lincoln, NB: University of Nebraska Press.

Kelley, H. H. (1972*a*). Attribution in social interaction. In E. E. Jones *et al.* (Eds), *Attribution: Perceiving the Causes of Behavior*. Morristown, NJ: General Learning Press.

Kelley, H. H. (1972b). Causal schemata and the attribution process. In E. E. Jones *et al.* (Eds), *Attribution: Perceiving the Causes of Behavior*. Morristown, NJ: General Learning Press.

Kelley, H. H. (1973). The processes of causal attribution. *American Psychologist*, **28**, 107–128.

Kellog, R. T. (1982). When can we introspect accurately about mental processes? *Memory and Cognition*, **10**, 141–144.

Kelly, G. (1955). *The Psychology of Personal Constructs*, vols 1 & 2. New York: Norton.

Kraut, R. E. & Lewis, S. H. (1982). Person perception and self-awareness: Knowledge of influences on one's own judgments. *Journal of Personality and Social Psychology*, **42**, 448–460.

Kroll, N. E. A. & Kellicutt, M. H. (1972). Short-term recall as a function of concept rehearsal and of intervening task. *Journal of Verbal Learning and Verbal Behavior*, **11**, 196–204.

Langer, E. J. (1978). Rethinking the role of thought in social interaction. In J. H. Harvey, W. J. Ickes & R. F. Kidd (Eds), *New Directions in Attribution Research*, vol. 2. Hillsdale, NJ: Erlbaum.

Lewicki, P. (1982). Trait relationships: The nonconscious generalisation of social experience. *Personality and Social Psychology Bulletin*, **8**, 439–445.

Lieberman, D. A. (1979). Behaviorism and the mind: A (limited) call for a return to introspection. *American Psychologist*, **34**, 319–333.

Locke, D. & Pennington, D. (1982). Reasons and other causes: Their role in the attribution process. *Journal of Personality and Social Psychology*, **42**, 212–223.

Maier, N. R. F. (1931). Reasoning in humans. II. The solution of a problem and its appearance in consciousness. *Journal of Comparative and Physiological Psychology*, **12**, 181–194.

Major, B. (1980). Information acquisition and attribution processes. *Journal of Personality and Social Psychology*, **39**, 1010–1023.

Mandler, G. (1975). *Mind and Emotion*. New York: Wiley.

Mandler, G. (1984). *Mind and Body: Psychology of Emotion*. New York: Norton.

Marcel, A. J. (1983). Conscious and unconscious perception: An approach to the relations between phenomenal experience and perceptual processes. *Cognitive Psychology*, **15**, 238–300.

McClure, J. (1983). Telling more than they can know: The positivistic account of verbal reports and mental processes. *Journal for the Theory of Social Behaviour*, **13**, 111–127.

Morris, P. E. (1981*a*). Why Evans is wrong in criticizing introspective reports of subject strategies. *British Journal of Psychology*, **72**, 465–468.

Morris, P. E. (1981*b*). The cognitive psychology of self-reports. In C. Antaki (Ed.), *The Psychology of Ordinary Explanations of Social Behaviour*. London: Academic Press.

Natsoulas, T. (1981). Basic problems of consciousness. *Journal of Personality and Social Psychology*, **41**, 132–178.

Nisbett, R. E. & Bellows, N. (1977). Verbal reports about causal influences on social judgments: Private access versus public theories. *Journal of Personality and Social Psychology*, **35**, 613–624.

Nisbett, R. E. & Ross, L. (1980). *Human Inference: Strategies and Shortcomings of Social Judgment*. Englewood Cliffs, NJ: Prentice-Hall.

Nisbett, R. E. & Wilson, T. D. (1976). *On telling more than we can know: Inaccurate verbal reports on mental processes*. Unpublished manuscript, University of Michigan.

Nisbett, R. E. & Wilson, T. D. (1977*a*). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, **84**, 231–259.

Nisbett, R. E. & Wilson, T. D. (1977*b*). The halo effect: Evidence for unconscious alteration of judgments. *Journal of Personality and Social Psychology*, **35**, 250–256.

Orne, M. T. (1962). On the social psychology of the psychology experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, **17**, 776–783.

Pope, K. S. & Singer, J. L. (Eds) (1978). *The Stream of Consciousness: Scientific Investigations into the Flow of Human Experience*. New York: Wiley.

Posner, M. & Warren, R. E. (1972). Traces, concepts, and conscious constructions. In A. W. Melton & E. Martin (Eds), *Coding Processes in Human Memory*. Washington, DC: Winston.

Rakover, S. (1983). Hypothesizing from introspections: A model for the role of mental entities in psychological explanation. *Journal for the Theory of Social Behaviour*, **13**, 211–230.

Rich, M. C. (1979). Verbal reports on mental processes: Issues of accuracy and awareness. *Journal for the Theory of Social Behaviour*, **9**, 29–37.

Rosenberg, M. J. (1965). When dissonance fails: On eliminating evaluation apprehension from attitude measurement. *Journal of Personality and Social Psychology*, **1**, 28–42.

Rosenberg, M. J. (1969). The conditions and consequences of evaluation apprehension. In R. Rosenthal & R. L. Rosnow (Eds), *Artifact in Behavioural Research*. London: Academic Press.

Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology*, vol. 10. London: Academic Press.

Sabini, J. & Silver, M. (1981). Introspection and causal accounts. *Journal of Personality and Social Psychology*, **40**, 171–179.

Salancik, G. R. (1974). Inference of one's attitude from behaviour recalled under linguistically manipulated cognitive sets. *Journal of Experimental Social Psychology*, **10**, 415–427.

Salancik, G. R. & Conway, C. (1975). Attitude inferences from salient and relevant cognitive content about behaviour. *Journal of Personality and Social Psychology*, **32**, 829–840.

Shallice, T. (1972). Dual functions of consciousness. *Psychological Review*, **78**, 383–393.

Shallice, T. (1978). The dominant action system: An information processing approach to consciousness. In K. S. Pope & J. L. Singer (Eds), *The Stream of Consciousness: Scientific Investigations into the Flow of Human Experience*. New York: Wiley.

Shaver, K. G. (1985). *The Attribution of Blame: Causality, Responsibility, and Blameworthiness*. New York: Springer-Verlag.

Shiffrin, R. M. & Schneider, W. (1977). Controlled and automatic human information processing. II. Perceptual learning, automatic attending, and a general theory. *Psychological Review*, **84**, 127–190.

Shultz, T. R. (1982). Rules of causal attribution. *Monographs of the Society for Research in Child Development*, **47**, (1, serial no. 194), 1–51.

Smith, E. R. & Miller, F. D. (1978). Limits on perception of cognitive processes: A reply to Nisbett and Wilson. *Psychological Review*, **85**, 355–362.

Sperry, R. W. (1969). A modified concept of consciousness. *Psychological Review*, **76**, 532–536.

Sperry, R. W. (1970). An objective approach to subjective experience: Further explanation of a hypothesis. *Psychological Review*, **77**, 585–590.

Sperry, R. W. (1976). Mental phenomena as causal determinants in brain function. In G. G. Globus, G. Maxwell & I. Savodnik (Eds), *Consciousness and the Brain*. New York: Plenum.

Sprangers, M., van den Brink, W., van Heerden, J. & Hoogstraten, J. (1987). A constructive replication of White's alleged refutation of Nisbett and Wilson and of Bem: Limitations on verbal reports of internal events. *Journal of Experimental Social Psychology*, **23**.

Strange, J. R. (1978). A search for the sources of the stream of consciousness. In K. S. Pope & J. L. Singer (Eds), *The Stream of Consciousness: Scientific Investigations into the Flow of Human Experience*. New York: Wiley.

Taylor, S. E. & Fiske, S. T. (1981). Getting inside the head: Methodologies for process analysis in attribution and social cognition. In J. H. Harvey, W. Ickes & R. F. Kidd (Eds), *New Directions in Attribution Research*, vol. 3. Hillsdale, NJ: Erlbaum.

Underwood, G. (Ed.) (1982). *Aspects of Consciousness*, vol. 3. London: Academic Press.

Warshaw, P. R. & Davis, F. D. (1984). Self-understanding and the accuracy of behavioural expectations. *Personality and Social Psychology Bulletin*, **10**, 111–118.

Weiss, J. & Brown, P. (1977). Self-insight error in the explanation of mood. Unpublished manuscript, Harvard University.

White, P. A. (1980). Limitations on verbal reports of internal events: A refutation of Nisbett and Wilson and of Bem. *Psychological Review*, **87**, 105–112.

White, P. A. (1982). Beliefs about conscious experience. In G. Underwood (Ed.), *Aspects of Consciousness*, vol. 3. London: Academic Press.

White, P. A. (1984). A model of the layperson as pragmatist. *Personality and Social Psychology Bulletin*, **10**, 333–348.

White, P. A. (1985). The awareness issue and memorial influences upon accuracy of verbal report: A re-examination of some data. *Psychological Reports*, **57**, 312–314.

White, P. A. (1986). On consciousness and beliefs about consciousness: Consequences of the physicalist assumption for models of consciousness. *Journal of Social Behaviour and Personality*, **1**, 505–524.

White, P. A. (1987). Selective citation. Unpublished paper. University of Auckland.

Wicker, A. W. (1969). Attitudes v. actions: The relationship of verbal and overt responses to attitude objects. *Journal of Social Issues*, **25**, 41–78.

Wilson, T. D. (1985). Strangers to ourselves: The origins and accuracy of beliefs about one's own mental states. In J. H. Harvey & G. Weary (Eds), *Attribution: Basic Issues and Applications*. New York: Academic Press.

Wilson, T. D., Dunn, D. S., Bybee, J. A., Hyman, D. B. & Rotondo, J. A. (1984). Effects of analyzing reasons on attitude–behaviour consistency. *Journal of Personality and Social Psychology*, **47**, 5–16.

Wilson, T. D., Hull, J. G. & Johnson, J. (1981). Awareness and self-perception: Verbal reports on internal states. *Journal of Personality and Social Psychology*, **40**, 53–71.

Wilson, T. D., Laser, P. & Stone, J. (1982). Judging the predictors of one's own mood: Accuracy and the use of shared theories. *Journal of Experimental Social Psychology*, **18**, 537–556.

Wilson, T. D. & Linville, P. W. (1982). Improving the academic performance of college freshmen: Attribution therapy revisited. *Journal of Personality and Social Psychology*, **42**, 367–376.

Wilson, T. D. & Nisbett, R. E. (1978). The accuracy of verbal reports about the effects of stimuli on evaluations and behaviour. *Social Psychology*, **41**, 118–131.

Wilson, T. D. & Stone, J. I. (1985). Limitations of self-knowledge: More on telling more than we can know. In P. Shaver (Ed.), *Self, Situations, and Social Behaviour: Review of Personality and Social Psychology*, vol. 6. London: Sage.

Wright, P. & Rip, P. D. (1981). Retrospective reports on the causes of decisions. *Journal of Personality and Social Psychology*, **40**, 601–614.

Received 5 February 1987; revised version received 13 April 1987