

1 The nature of error

Just over 60 years ago, Spearman (1928) grumbled that “crammed as psychological writings are, and must needs be, with allusions to errors in an incidental manner, they hardly ever arrive at considering these profoundly, or even systematically.” Even at the time, Spearman’s lament was not altogether justified (see Chapter 2); but if he were around today, he would find still less cause for complaint. The past decade has seen a rapid increase in what might loosely be called ‘studies of errors for their own sake’.

The most obvious impetus for this renewed interest has been a growing public concern over the terrible cost of human error: the Tenerife runway collision in 1977, Three Mile Island two years later, the Bhopal methyl isocyanate tragedy in 1984, the *Challenger* and Chernobyl disasters of 1986, the capsizing of the *Herald of Free Enterprise*, the King’s Cross tube station fire in 1987 and the Piper Alpha oil platform explosion in 1988. There is nothing new about tragic accidents caused by human error; but in the past, the injurious consequences were usually confined to the immediate vicinity of the disaster. Now, the nature and the scale of certain potentially hazardous technologies, especially nuclear power plants, means that human errors can have adverse effects upon whole continents over several generations.

Aside from these world events, from the mid-1970s onwards theoretical and methodological developments within cognitive psychology have also acted to make errors a proper study in their own right. Not only must more effective methods of predicting and reducing dangerous errors emerge from a better understanding of mental processes, it has also become increasingly apparent that such theorising, if it is to provide an adequate picture of cognitive control processes, must explain not only correct performance, but also the more predictable varieties of human fallibility. Far from being rooted in irrational or maladaptive tendencies, these recurrent error forms have their origins in fundamentally useful psychological processes. Ernst Mach (1905) put it well: “Knowledge and error flow from the same mental sources, only success can tell the one from the other.”

A central thesis of this book is that the relatively limited number of ways in which errors actually manifest themselves is inextricably bound up with the ‘computational primitives’ by which stored knowledge structures are selected and retrieved in response to current situational demands. And it is just these

processes that confer upon human cognition its most conspicuous advantage over other computational devices: the remarkable ability to simplify complex informational tasks.

1. The cognitive ‘balance sheet’

Correct performance and systematic errors are two sides of the same coin. Or, perhaps more aptly, they are two sides of the same cognitive ‘balance sheet’. Each entry on the asset side carries a corresponding debit. Thus, automaticity (the delegation of control to low-level specialists) makes slips, or actions-not-as-planned, inevitable. The resource limitations of the conscious ‘workspace’, while essential for focusing computationally-powerful operators upon particular aspects of the world, contribute to informational overload and data loss. A knowledge base that contains specialised ‘theories’ rather than isolated facts preserves meaningfulness, but renders us liable to confirmation bias. An extraordinarily rapid retrieval system, capable of locating relevant items within a virtually unlimited knowledge base, leads our interpretations of the present and anticipations of the future to be shaped too much by the matching regularities of the past. Considerations such as these make it clear that a broadly-based analysis of recurrent error forms is essential to achieving a proper understanding of the largely hidden processes that govern human thought and action.

2. Errors take a limited number of forms

On the face of it, the odds against error-free performance seem overwhelmingly high. There is usually only one way of performing a task correctly, or, at best, very few; but each step in a planned sequence of actions or thoughts provides an opportunity to stray along a multitude of unintended or inappropriate pathways. Think of boiling an egg. At what stages and in how many ways can even this relatively simple operation be bungled? The list of possibilities is very long. Thoughts such as these make it appear highly unlikely that we could ever adequately chart the varieties of human error.

Fortunately, the reality is different. Human error is neither as abundant nor as varied as its vast potential might suggest. Not only are errors much rarer than correct actions, they also tend to take a surprisingly limited number of forms, surprising, that is, when set against their possible variety. Moreover, errors appear in very similar guises across a wide range of mental activities. Thus, it is possible to identify comparable error forms in action, speech, perception, recall, recognition, judgement, problem solving, decision making, concept formation and the like. The ubiquity of these recurrent error forms demands the formulation of more global theories of cognitive control than are

3. Variable and constant errors

3

usually derived from laboratory experiments. Of necessity, these focus upon very restricted aspects of mental function in rather artificial settings.

Far from leading down countless unconnected or divergent pathways, the quest for the more predictable varieties of human error is one that continually draws the searcher inwards to the common theoretical heartland of consciousness, attention, working memory and the vast repository of knowledge structures with which they interact. And it is with these theoretical issues that the first half of this book is primarily concerned.

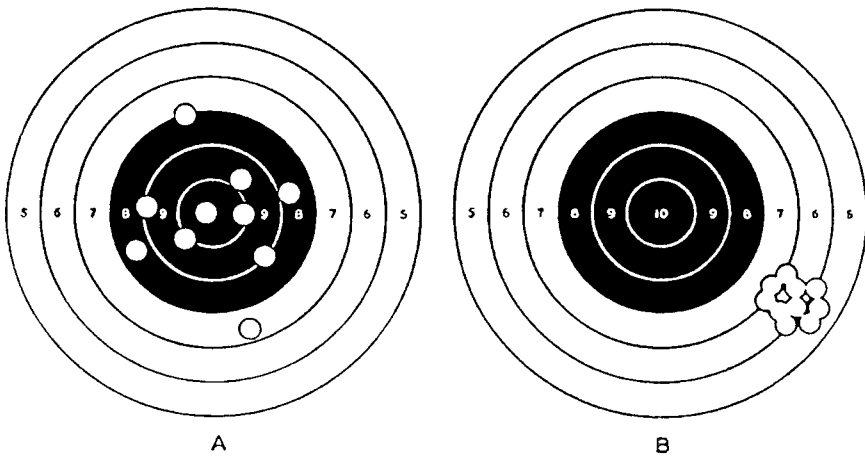


Figure 1.1. Target patterns of ten shots fired by two riflemen. *A*'s pattern exhibits no constant error, but rather large variable errors. *B*'s pattern shows a large constant error, but small variable errors (from Chapanis, 1951).

3. Variable and constant errors

Although it may be possible to accept that errors are neither as numerous nor as varied as they might first appear, the idea of a *predictable* error is a much harder one to swallow. If errors were indeed predictable, we would surely take steps to avoid them. Yet they still occur. So what is a predictable error?

Consider the two targets shown in Figure 1.1 (taken from Chapanis, 1951). Each shows a pattern of ten shots, one fired by rifleman *A*, the other by rifleman *B*. *A* placed his shots around the bull's eye, but the grouping is poor. *B*'s shots fell into a tight cluster, but at some distance from the bull's eye.

These patterns allow us to distinguish between two types of error: *variable* and *constant* errors. *A*'s pattern exhibits no constant error, only a rather large

amount of variable error. *B* shows the reverse: a large constant error, but small variable error. In this example, the variability is revealed by the spread of the individual shots, and provides an indication of the rifleman's consistency of shooting. The constant error, on the other hand, is given by the distance between the group average and the centre of the target.

What do these patterns tell us about the relative merits of these two individuals? If we should rely only on their respective scores, then *A* would appear the better shot, achieving a total of 88 to *B*'s 61. But it is obvious from the groupings that this is not the case. A more acceptable view would be that *A* is a rather unsteady shot with accurately aligned sights, while *B* is an expert marksman whose sights are out of true.

It is also evident that the errors of these two marksmen differ considerably in their degree of predictability. Given another ten shots each, with *B* still aiming at the target's centre and his sights still unadjusted, we could say with a high degree of confidence whereabouts his shots would fall; but the variability of *A*'s shooting makes such a confident forecast impossible. The difference is very clear: in *B*'s case, we have a theory that will account for the precise nature of his constant error, namely, that he is an excellent shot with biased sights. But our theory in *A*'s case, that he has accurate sights but a shaky hand, is not one that would permit a precise prediction of where his shots will fall. We can anticipate the poor grouping and have some idea of its spread, but that is all.

The lesson of this simple example is that the accuracy of error prediction depends very largely on the extent to which the factors giving rise to the errors are understood. This requires a theory which relates the three major elements in the production of an error: the nature of the task and its environmental circumstances, the mechanisms governing performance and the nature of the individual. An adequate theory, therefore, is one that enables us to forecast both the *conditions* under which an error will occur, and the particular *form* that it will take.

For most errors, our understanding of the complex interaction between these various causal factors is, and is always likely to be, imperfect and incomplete. Consequently, most error predictions will be probabilistic rather than precise. Thus, they are liable to take the form: "Given this task to perform under these circumstances, this type of person will probably make errors at around this point, and they are likely to be of this variety," rather than be of the kind: "Person *X* will make this particular error at such-and-such a time in such-and-such a place." Nevertheless, predictions of this latter sort can be made in regard to certain types of error when they are deliberately elicited within a controlled laboratory environment. This is especially true of many perceptual illusions. Not only can we predict them with near certainty (given an intact sensory system), we can also forecast with considerable accuracy how

4. Intentions, actions and consequences

5

their experience will vary with different experimental manipulations. But these are exceptions.

The more usual type of prediction is illustrated by the following example. It can be forecast with near certainty that during next January the banks will return a large number of cheques with this year's date on them. We cannot necessarily predict the exact number of the misdated cheques (although such information probably exists for previous years, so the approximate number could be estimated), nor can we say precisely who will make this error, or on which day. But we do know that such *strong habit intrusions* are among the most common of all error forms; that dating a cheque, being a largely routinised activity (at least with respect to the year), is particularly susceptible to absent-minded deviations of this kind; and that the early part of the year is the period in which these slips are most likely to happen. Such *qualitative* predictions may seem merely banal, but they are nonetheless powerful. Moreover, the regular recurrence of this error form is extremely revealing of the covert processes controlling practised activities.

4. Intentions, actions and consequences

The notions of intention and error are inseparable. Any attempt at defining human error or classifying its forms must begin with a consideration of the varieties of intentional behaviour.

One psychologically useful way of distinguishing between the different kinds of intentional behaviour is on the basis of yes-no answers to three questions regarding a given sequence of actions (Figure 1.2):

Were the actions directed by some prior intention?

Did the actions proceed as planned?

Did they achieve their desired end?

Notice that all of these questions are capable of being answered. In contrast to issues like basic motivation or detailed execution, the nature of the prior intentions, knowledge of whether or not the subsequent actions deviated from them and an appreciation of their success or failure are potentially available to consciousness. Indeed, one of the primary functions of consciousness is to alert us to departures of action from intention (Mandler, 1975; 1985) and, though less immediately, to the likelihood that the planned actions currently underway will not achieve their desired goal.

The notion of intention comprises two elements: (a) an expression of the end-state to be attained, and (b) an indication of the means by which it is to be achieved. Both elements may vary widely in their degree of specificity. For most everyday actions, prior intentions or plans consist of little more than a series of verbal tags and mental images. With the repetition of an action sequence, fewer and fewer 'intentional tags' come to stand for increasingly larger

6

The nature of error

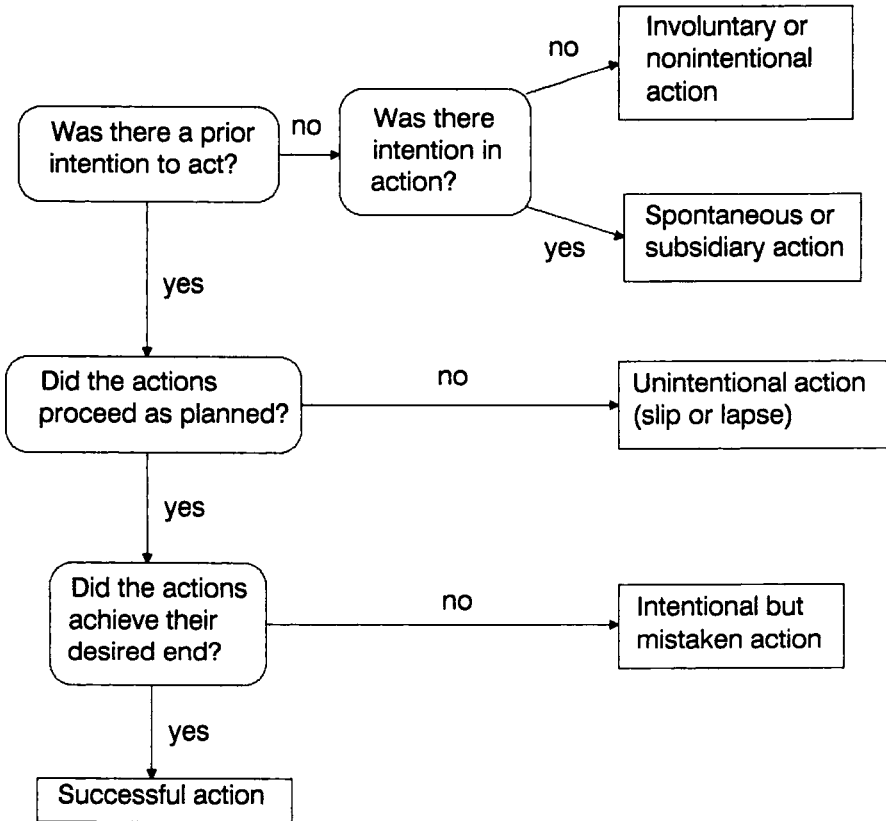


Figure 1.2. Algorithm for distinguishing the varieties of intentional behaviour. The three main categories are non-intentional behaviour, unintentional behaviour (slips and lapses) and intentional but mistaken behaviour.

amounts of detailed movement. The more routine the activity, the fewer the number of low-level control statements required to specify it. In novel activities, however, we are aware of the need to ‘talk ourselves through’ the actions. Under these circumstances, our activities are guided by the effortful yet computationally-powerful investment of conscious attention.

4.1. Distinguishing prior intention and intentional action

Searle (1980, p. 52) made an important distinction between ‘prior intentions’ and ‘intentions in action’: “All intentional actions have intentions in action

4. Intentions, actions and consequences

7

but not all intentional actions have prior intentions.” Actions without prior intentions fall into two broad classes: *intentional* and *nonintentional* actions.

4.1.1. Intentional actions without prior intention

Searle (1980) gives two instances of intentional actions without prior intention: spontaneous and subsidiary actions. Someone might hit another on the spur of the moment without forming any prior intention. In this case, the intention resides only in the action itself, as Searle (1980, p. 52) states “the action and the intention are inseparable.” Similarly, in executing well-practised action sequences, only the ‘major headings’ are likely to be specified in the prior intention (e.g., “I will drive to the office”). We do not, indeed cannot, consciously fill in the ‘small print’ of each component operation in advance (i.e., opening the car door, sitting down, putting on the seat belt, inserting the ignition key, starting the engine, etc.). For such subsidiary actions, writes Searle (1980, p. 52), “I have an intention, but no prior intentions” (see also Reason & Mycielska, 1982, p. 9).

4.1.2. Non-intentional or involuntary actions

Textbooks of jurisprudence and criminal law are replete with accounts of intentionless behaviour (Hart, 1968; Smith & Hogan, 1973). As Hart (1968, p. 114) put it: “All civilized penal systems make liability to punishment for at any rate serious crime dependent not merely that the person to be punished has done the outward act of the crime, but on his having done it in a certain frame of mind or will.” A crime has thus two elements: the *actus rea* and the *mens rea*. To prove a criminal liability, it must be shown not only that the consequences of the criminal act were intended, but also that the act itself was committed voluntarily.

The defence of ‘automatism’ rests on demonstrating the absence of “a vital link between mind and body” (Smith & Hogan, 1973, p. 35). In such instances, “the movements of the human body seem more like the movements of an inanimate thing than the actions of a person. Someone unconscious in a fit of epilepsy, hits out in a spasm and hurts another; or someone, suddenly stung by a bee, in his agony drops and breaks a plate he is holding” (Hart, 1968, pp. 91-92).

Not only is the issue of volition fundamental to notions of criminal responsibility, it is also crucial to the psychological definition of human error. Thus the term error can *only* be applied to intentional actions. It has no meaning in relation to nonintentional behaviour because error types depend critically upon two kinds of failure: the failure of actions to go as intended (slips and lapses) and the failure of intended actions to achieve their desired consequences (mistakes). To clarify these basic error forms, we need to consider the distinction between intended and unintended actions.

4.2. Distinguishing between intended and unintended actions

Assuming a prior intention and/or an ‘intention in action’, it is possible to build a psychologically meaningful taxonomy of error and correct performance on the basis of the answers to the second and third questions listed above: Did the actions go as planned? And did the actions achieve their desired consequences? Since our primary concern is with error, we will look first at those cases where actions deviate from intention: unintended actions or actions-not-as-planned.

4.2.1. Unintended actions

Actions that deviate from intention fall into two classes: those that nevertheless achieve their intended goal and those that do not. While just conceivable, the former are highly unlikely. Searle (1980) provides an example of one such ‘successful yet unintended action’: A man intends to murder someone by shooting at him. He misses, but the shot stampedes a herd of wild pigs, which tramples the intended victim to death. I will not consider these curiosities further. Psychologists (as distinct from lawyers, philosophers or even theologians) are more interested in the ‘acts of man’ than in the so-called ‘acts of God’.

Common among the acts of humans are moments of absent-mindedness when we become aware that our actions have strayed from their intended path. Two conditions appear to be necessary for the occurrence of these slips of action: the performance of some largely automatic task in familiar surroundings and a marked degree of attentional ‘capture’ by something other than the job in hand (see Reason, 1979; Norman, 1981).

4.2.2. Intended actions and mistakes

Even when the intended actions proceed as planned, they can still be judged as erroneous if they fail to achieve their intended outcome. In this case, the problem resides in the adequacy of the plan rather than in the conformity of its constituent actions to some prior intention. Errors of this kind are termed *mistakes* (Norman, 1981, 1983; Reason & Mycielska, 1982).

Norman (1983) summarised the distinction between mistakes and slips (or lapses) very succinctly: “If the intention is not appropriate, this is a mistake. If the action is not what was intended, this is a slip.” Mistakes involve a mismatch between the prior intention and the intended consequences. For slips and lapses, however, the discrepancy is between the intended actions and those that were actually executed.

Another way of distinguishing these two basic error forms is as *planning failures* (mistakes) and *execution failures* (slips and lapses). This categorization emerges not only from the logic of the three-question algorithm, but also from a consideration of cognitive levels. Planning failures are likely to arise from higher-level processes than either slips or lapses.

5. Some working definitions

9

5. Some working definitions

We are now in a position to sketch out some working definitions of error and its principal types. A working definition is serviceable rather than ideal. It tries to embody the essential psychological characteristics of the phenomena without struggling too hard to mark out their exact semantic boundaries. The study of error, being largely an inductive mode of enquiry, does not demand precise axioms and definitions at the outset, as do the deductive sciences. The following working definitions have proved their utility in a psychological rather than a philosophical sense.

Error will be taken as a generic term to encompass all those occasions in which a planned sequence of mental or physical activities fails to achieve its intended outcome, and when these failures cannot be attributed to the intervention of some chance agency.

As we have seen, a series of planned actions may fail to achieve their desired outcome because the actions did not go as planned or because the plan itself was inadequate. It is also possible that both types of error could occur within the same sequence of planning, storage and execution. This distinction gives rise to two further working definitions.

Slips and lapses are errors which result from some failure in the execution and/or storage stage of an action sequence, regardless of whether or not the plan which guided them was adequate to achieve its objective.

Whereas *slips* are potentially observable as externalised actions-not-as-planned (slips of the tongue, slips of the pen, slips of action), the term *lapse* is generally reserved for more covert error forms, largely involving failures of memory, that do not necessarily manifest themselves in actual behaviour and may only be apparent to the person who experiences them.

Mistakes may be defined as deficiencies or failures in the judgemental and/or inferential processes involved in the selection of an objective or in the specification of the means to achieve it, irrespective of whether or not the actions directed by this decision-scheme run according to plan.

It is evident from this definition that mistakes are likely to be more subtle, more complex and less well understood than slips. As a result, they generally constitute a far greater danger. By their nature, mistakes are also far harder to detect. Consciousness is specifically tuned to picking up departures of action from intention (Mandler, 1975), but mistakes can pass unnoticed for lengthy periods (Woods, 1984), and even when detected they sometimes remain a matter of debate. Not only is the quality of the plan open to a diversity of opinion, it is also something that can be judged at two distinct stages: before and after it has been implemented. Prior to its execution, it can be assessed according to whether or not it has sufficient contingencies, displays

soundness of judgement, imagination, flexibility, awareness of detail and the like. But once put into action, we judge it primarily according to how well it achieved its stated objectives.

Unfortunately, the results of these two sets of judgements do not necessarily correspond. Plans rated as good prior to their execution can fail to attain their objectives; whereas plans judged as inadequate by any reasonable criteria can turn out successfully, though not necessarily to the credit of their originators. Consider, for example, the relative qualities and eventual outcomes of the German and French plans for waging the First World War. By most a priori standards, the Schlieffen Plan was a masterpiece of military thinking. It was bold, imaginative and worked out to the last logistic detail; but it ultimately failed. The corresponding Plan XVII was, by comparison, crude and ill-considered, but it eventually succeeded. Janis (1972, p. 11) put the matter succinctly in a discussion of high-level policy making: "Defective decisions based on misinformation and poor judgement sometimes lead to successful outcomes . . . we must acknowledge that chance and the stupidity of the enemy can sometimes give a silk-purse ending to a command decision worth less than a sow's ear."

6. The classification of errors

A central problem in error classification is the difficulty of reconciling the often highly specific contextual triggers of a particular error form with the fact that it may also manifest the influence of some very general adaptive process or basic error tendency. A classification that emphasises the former at the expense of the latter is likely to overlook the broad regularities in the more predictable forms of error. Yet one that neglects the local contextual factors (e.g., task and situational considerations) will not only be of limited use to the practitioner, it will also fail to accommodate the theoretically important causal interactions between basic error tendencies and immediate task features.

There is no universally agreed classification of human error, nor is there one in prospect. A taxonomy is usually made for a specific purpose, and no single scheme is likely to satisfy all needs. Nearly everyone who has published in this field has devised some form of error classification. Consequently, the literature abounds with such taxonomies, reflecting a variety of practical concerns and theoretical orientations and ranging from the highly task specific to broad statements of underlying error tendencies.

Notwithstanding the number and diversity of existing error taxonomies, it is possible to penetrate beyond their surface idiosyncracies to distinguish three levels at which classifications are attempted: the behavioural, contextual and conceptual levels. These correspond approximately to the "What?", "Where?" and "How?" questions about human errors.