Behavioural Interpretation of Cognition

David C. Palmer
Smith College
Northampton, Massachusetts

Ullin Place impressed his colleagues by his philosophical work on the mind-body problem, but, by his own account, he bewildered them by his unswerving advocacy of behaviourism. In this essay, dedicated to his memory, I suggest that this bewilderment arises from a widespread misunderstanding of the modern behavioural approach to cognition. It is not commonly appreciated that the behaviourist distinguishes the interpretation of behaviour from the experimental analysis of behaviour. Only the latter enterprise requires that variables be public, measurable, and reliable; the interpretation of behaviour serves a different purpose and is differently constrained.

Introduction

Behavioural phenomena embraced by the term "cognition" are distinctive in that commonly some portion of the performance is unobserved, and that portion that is observed is often insufficient to permit prediction or control. Appropriate behaviour appears as if by magic. If this sense of magic is to be dispelled, there is no alternative to speculating about unobserved phenomena. The behaviourist does not object to such speculations, but he does insist that the interpretive tools used in such exercises be restricted to those that have been validated with observable phenomena. Thus he sees complex behaviour as an orderly web of units of behaviour governed by established behavioural principles and nothing more. This psychological uniformitarianism sets behaviourism apart from more permissive paradigms that invoke constructs that do not have an independent empirical status. The behaviourist argues that the psychological functions assigned to these constructs can be accomplished by covert behaviour; additional terms are unnecessary. My purpose here is not to debate the merits of different approaches to cognition but simply to discuss the conceptual foundation of the behavioural approach alone. To do so requires that we identify the place of that approach within the wider enterprise of science.

Further contents:

The Two Purposes of Science

Science serves two purposes, and these purposes are in some ways quite distinct. The first purpose, the mastery of nature, has led to remarkable improvements, or at least remarkable changes, in human welfare, and the pace of such changes appears to be continually accelerating. But perhaps more importantly, science offers beautiful, satisfying, and elegant accounts for how the world works. As much as we love technological advances, we tend to hold in highest regard those scientists who have contributed most to this second purpose, resolving mysteries about nature.

The two purposes of science are served by intertwined pursuits that may have entirely different standards of control. The mastery of nature requires precise control of every relevant variable; a stray enzyme in an assay can doom an experiment on genetically engineered tomatoes; the misplacement of a transistor by a few microns on a sliver of silicon could render useless a new particle accelerator. Consequently, advances in science often rest on advances in technology that permit ever greater control of experimental variability.

The understanding of nature, however, is a more liberal enterprise. We resolve a mystery about a puzzling phenomenon by identifying at least one pathway that nature might have taken to produce that phenomenon, invoking only familiar principles and tools; we need not demonstrate control. Newton explained ocean tides, not by achieving experimental control over them, but by showing that they were consistent with his inverse-square law. Newton might be wrong; perhaps the tides are caused by the respiration of a gigantic sea serpent, unknown to science. We dismiss such a claim as nonsense, not because we can prove otherwise, but because it appeals to unknown and untested variables; presumably there are an unlimited number of such empty hypotheses. Newton's account is satisfying because his principles of motion have been abundantly confirmed, at least to an approximation suitable for such interpretive exercises, in domains where experimental control is possible. We get the account of ocean tides for free, so to speak.

Much of what we consider to be the scientific understanding of the world is of this tentative but plausible sort. The disciplines of cosmology, meteorology, geology, and ethology make liberal use of such interpretation, and they supply gratifying advances in our understanding of the world. Controlled observations reveal principles and conceptual tools that are then invoked to interpret phenomena too complex or too vast to be experimentally manipulated. Our understanding of the genesis of planets, the origin of hurricanes, the formation of mountain chains, and the migration of birds depends on this assumption of uniformitarianism, that is, that the world beyond our narrow laboratory is like that within it.

Evolutionary biology provides some of the most elegant examples of scientific interpretation and reminds us of some of its limitations. We account for the countless wonderful adaptations of organisms by appealing to differential contingencies of survival iterated over many generations. We note, however, that any one evolutionary story usually rests on flimsy evidence: Most of the details of an organism's phylogenetic history have been irretrievably lost. We find such accounts satisfying, not because they are complete, but because they appeal to nothing but familiar processes that can be readily verified. An extraordinary puzzle has been unlocked with an ordinary key. However, evolutionary accounts of biological adaptations share a curious property with other scientific interpretation: Even though they serve to allay curiosity about the phenomenon in question, any particular account might be entirely wrong. The function of such interpretations is to bring the incredible into the domain of the believable, not to offer certainty. They serve this function when they identify just one mechanism that might account for a phenomenon without resorting to occult or unknown processes, one mechanism, possibly among many. We are utterly baffled to see a magician's assistant pop out of a box that a moment ago seemed empty; our curiosity is satisfied if someone points out that the box might be fitted with a diagonal mirror that provides the illusion of emptiness while hiding a compartment. It may be that this explanation is wrong, that some other device is used in the trick, but we no longer care: The impossible has been reduced to the merely clever.

In the domain of cognition, everyone must resort to interpretation, regardless of paradigm. No one can exert the kind of experimental control over human subjects necessary for an experimental analysis. Like the phylogenetic history of a species, most of the relevant history of an individual subject is beyond recovery. Therefore, our accounts will be tentative, like those of the evolutionary biologist. Their relative merits will depend upon the stature of the principles that are adduced. Behaviourism is well-placed to offer interpretations in this domain because its basic science has revealed a set of atomic principles of great scope and reliability.

The Three-term Contingency as an Interpretive Tool

The workhorse of behavioural interpretation is the three-term contingency, an analytical unit comprising a setting, a behaviour, and a consequence. As an analytical tool it is analogous to the environmental niche, variation, and differential reproduction of evolutionary biology. Behaviour followed by a reinforcing consequence is more likely to occur again in that setting and in similar settings. This principle has been abundantly confirmed in a wide variety of species, and it serves as the nucleus of operant theory. (Of course, the science of behaviour embraces other principles as well, such as Pavlovian conditioning, habituation, laws of the reflex, higher-order discriminations, extinction, avoidance, and so on, but for purposes of illustration, we will discuss only the three-term contingency here, since it is a particularly versatile and powerful tool in the interpretation of complex behaviour.) Much human behaviour can be understood by appeal to such contingencies. When a child replies, “Fifty-six,” to the question, “What is 8 times 7?” we assume that the child has encountered the question before, has responded correctly, and has been given appropriate feedback; such contingencies are nearly inevitable in the life of an educated child. Recalling a name, reciting a poem, using a tool, closing a door, catching a ball, and innumerable other examples can all be plausibly interpreted in terms of a history of such contingencies. In the laboratory, programs of gradually changing contingencies can shape highly unusual and finely differentiated behaviour; it is likely that comparable programs are implicit in the experience of the skilled craftsman, artist, sportsman, and scientist, among others.

All of our actions are embedded in contingencies of reinforcement, and such contingencies are atomic units that form the building blocks of our experience. Moreover, the analysis of contingencies of reinforcement makes no appeal to hypothetical processes. We might have to guess about one's history of exposure to various contingencies of reinforcement, but the constructs about which we guess are real enough. Thus, the three-term contingency is a powerful interpretive tool because such contingencies are ubiquitous, elementary, and objective.

However, not all behaviour conforms to this formula. Consider the following cases:

1. We often respond appropriately to novel questions. If asked, “What is 80 times 70?” most of us would respond correctly after a bit of thought, but it is unlikely that we have encountered just that question before. We cannot explain the answer by invoking a history

of reinforcement of that answer to that question; rather, we are likely to appeal to “mental arithmetic.”

2. When asked, “Where did you park?” our response cannot be considered to be directly evoked by the question as a discriminative stimulus, since our answer will vary from occasion to occasion, although the question, and perhaps the entire context, remains unchanged. Rather, we appear to consult a stored “memory” of our actions.

3. Even though one, and only one, response is reinforced in a particular setting, we do not always emit that response when the setting is presented again. For example, suppose we were to lavishly reward a friend for telling us, correctly, that today's date is June, 30th. Even if we were to hold constant every detail of the setting as well as all motivational variables, we would be astonished if he did not respond, “July 1st,” when asked the same question tomorrow. The response appears to be evoked, not by the current setting, but by a “mental representation of the calendar.”

4. Our behaviour often appears to be governed, not by past contingencies, but by future ones. We plan menus a week in advance; we make travel arrangements to attend future meetings; we decline rich desserts lest we be disappointed at the scales later on; we exercise, hoping to avoid an untimely demise. All “intentional” behaviour appears to require reference to future consequences, real or hypothetical, not past consequences.

There are countless other examples of behaviour that seem to defy interpretation by elementary behavioural processes, but these suffice to make the point. At the level of casual observation, we find no orderly relationship between a set of objective conditions and the behaviour that occurs under those conditions. Appropriate behaviour seems to simply emerge from the individual as an autonomous agent. This puzzle is commonly resolved by appealing to constructs forbidden to the behaviourist: schemas, intentions, representations, storage registers, retrieval processes, cognitive maps, and the like. These constructs are presumed to mediate the relationship between the context and appropriate behaviour. Unfortunately, each such construct introduces a qualitatively new element that must itself be understood if our explanation of behaviour is to advance. Unless such terms can be understood better than behaviour itself, the overall explanatory burden has been increased, not reduced.

In contrast, the behaviourist insists that the relationship between behaviour and the context in which it occurs is in fact an orderly one. The lack of order is an illusion arising from the circumstance that when we observe behaviour we inevitably see only a portion of the performance. As an analogy, the motion of a child on a swing conforms to the principles of dynamics and occasions no surprise. However, if we could not see, or otherwise detect, the swing itself, or its ropes, the performance would appear wonderful indeed. The observer is in much this position with respect to behaviour. If all of the behaviour of the organism could be observed, the behaviourist argues, the illusion of autonomy would vanish, and one would need not appeal to additional hypothetical constructs.

It is beyond the scope of this paper to address such examples, except to point to an approach. Let us consider instead a simpler problem for illustrative purposes, chosen because we are all likely to agree on the origin of the appropriate response. Suppose we ask, “What is the 10th letter after F?” Although the question is surely novel, most of us are able to respond, after a bit, with a common answer, “P.” If we restrict our data to only those behaviours that we have observed, now or in the past, the response appears magical. We have uttered the response “P” many times, but never in response to such a question. However, it is evident that we solve the problem by engaging in mediating behaviour that, together with the question, is sufficient to evoke the answer. The posing of any question for which no answer has previously been learned sets the occasion for an individual to engage in collateral behaviour. In this case, most people covertly recite the alphabet past F while ticking off numbers on their fingers. Each collateral response provides supplementary stimuli that accumulate to eventually potentiate the target response. “P” is not evoked by the question, but it is strongly evoked by the intraverbal chain, “L-M-N-O . . .” Since it is the letter that coincides with the ticking off of the tenth digit, it fits the requirements of the question.

None of this is remarkable. When the sequence of behaviour is overt, as it is in some subjects who are asked the question, it is clear that the target response is evoked by mediating behaviour. When it is covert, we infer the existence of the mediating behaviour. The inference is plausible: We note the latency of the response, and the lack of competing behaviour; we remark that the intraverbal chains of reciting the alphabet and counting to ten are surely in the subject's repertoire. The behaviourist regards such inferences as legitimate, since the terms of the account are restricted to familiar processes. The only controversy concerns the status of covert behaviour.

What is Behaviour?

Laboratory studies reveal that under controlled conditions observable behaviour enters into orderly relationships with antecedent and consequent events. We define "behaviour" as any activity of the organism that can enter into such orderly relationships. The scope of this definition is therefore an empirical matter and cannot be determined in advance. This definition is not circular; rather, it is analogous to terms such as "marine organism," "allergen," or "carnivore;" we cannot tell if something is a candidate without investigation. Our definition is empirical, not prescriptive. It does not specify that behaviour be peripheral, or muscular, or motor. Nor does it specify that it be observable. We must assume that whether something is sensitive to behavioural principles is independent of whether we are observing it (except insofar as our techniques of observation intrude as disturbance variables). Organisms evolved a sensitivity to contingencies of reinforcement presumably because of contingencies of survival; experimental observation played no role in such contingencies.

Whether a response is observable is not an intrinsic property of the response; rather, it depends upon the sensitivity of the techniques of the observer and the advantages of one's point of view. A whisper can be heard a few feet away, but not across a room, whereas a bellow can be heard across a room but perhaps not across a noisy fairground. Instrumental amplification permits the observation of otherwise undetectable muscle twitches. It may be that there are units of behaviour, by our definition, that cannot be detected with our current tools but could be measured by techniques to be developed in the future. Observability, then, is a criterion for experimental control, but it is not an intrinsic property of a response; every response becomes unobservable if we step back far enough. For any behaviour, we can conceive of a threshold of observability above which it can be detected and below which it cannot, but in every case, this threshold depends upon the circumstances and equipment of the observer at a particular time, not on the response itself.

It is incontrovertible that, even under controlled conditions, some behaviour of the organism is unobserved. Since the threshold of observability is an arbitrary and variable point, we must assume that the principles of behaviour do not respect it; they apply to all behaviour, whether or not it is being observed. However, we can demonstrate the control of these principles only when behaviour is observed.

It is possible, of course, that some classes of hitherto unobserved behaviour have special properties. It appears, for example, that some patterns of neural activity might meet our definition of behaviour. Ordinary responses such as lever-presses are mediated by populations of neurons, and orderly environment-behaviour relationships are surely mediated by the nervous system. Whether silent speech, visualization, recall, and so on, are entirely central is unclear, but even so, it may be that such activities are qualitatively different from other behaviour. However, this possibility cannot be evaluated with our present technology. It is defensible, even desirable, to assume that such examples fall under the domain of established behavioural processes and nothing more. If nature can accomplish an end with the tools at hand, there is no pressure to develop new ones. We may speculate about other processes, but we should recognize that in the absence of controlled observation one can generate many other hypotheses of equal merit.

A Strategy for Understanding Cognition

Inferences about covert events play no role in the experimental analysis of behaviour.

We would be unable to discover order in nature if we allowed relevant variables to wander freely and unmeasured. Our general principles must be forged in the tight constraints of the laboratory, and all of our measures must be observable and reliable. Rather, inferences about unobserved behaviour serve science's second function: They help us make sense of the world. To the extent that inferences about covert behaviour can offer a web of behavioural relationships capable of producing some phenomenon, we have offered a tentative explanation of that phenomenon. The account will stand as one route that nature might have taken to produce the result. It is not a demonstration; it is a proposal. However, it will be a proposal that rests upon a set of empirical principles and invokes only variables that have an independent status. When a steam engine explodes, an engineer might suggest that the volume of water in the boiler was too low. This is more than just a guess; it is an interpretation from thermodynamic principles that resorts to no terms that themselves are mysterious or poorly understood. Behavioural interpretations that invoke covert events have a similar status. They help us understand the world by showing how our interpretive tools, arising from experimental analysis, are adequate to the task of accounting for puzzling phenomena.

The behaviourist then attempts to interpret the full spectrum of human experience in terms of behavioural processes, appealing to unobserved events as necessary. Covert speech, conditioned perception, operant visualization, and self-prompting, all interpreted as units of behaviour, carry the load commonly assigned to cognitive processes. Examples that defy interpretation as simple discriminative operants arising from single three-term contingencies, such as mental arithmetic, recall, problem solving, planning, etc., are interpreted as mosaics of elementary units of behaviour. Such interpretive exercises offer a shadowy outline of the entire panorama of human behaviour, rather like the dots in a child's connect-the-dots puzzle. Some problems are tractable to experimental analysis, and fill in the picture in detail wherever possible, connecting the dots, as it were. These exiguous lines, nearly meaningless by themselves, take their place in the interpretive framework and become understandable only in that context. Empirical work may reveal additional principles of behaviour or refine existing ones; if so, they will take their place in the corpus of behavioural laws on which our interpretations rest. These discoveries will require that the panorama be touched up, or perhaps entirely redrawn, but such revisions are normal progress in the scientific understanding of nature.

This program might seem to be no more than an awkward and ideologically hidebound endorsement of the practices of traditional cognitive science. However, covert responses are not simply representations, schemas, intentions, scripts, or retrieval processes. Such hypothetical constructs may be helpful in guiding research, but they do not serve as an adequate foundation for interpretive exercises. A kind of translation between the two explanatory systems is often possible, but differences are crucial. Units of behaviour are defined by the orderliness of their relationships with other events. Covert responses are therefore constrained and are not free to carry whatever explanatory burden is required by the example in question. They must obey principles of behaviour and must be plausible with respect to both controlling variables and the individual's history. Since all of the interpretive tools of the behaviourist must arise from an independent experimental analysis, a behavioural interpretation of cognition is analogous to scientific interpretation in other sciences. We understand nature by offering tentative stories about it, invoking only terms that have empirical validity.

By endorsing appeals to covert behaviour, not in the experimental analysis of behaviour, but in its interpretation, the behaviourists' program is more powerful than commonly understood. The boundary between behaviourism and other paradigms is still clear, but it is not drawn at the threshold of observability. In his conviction that the behavioural program is sound, I believe that Ullin Place was ahead of his time, not behind it.