EUROPEAN JOURNAL OF COGNITIVE PSYCHOLOGY, 1994, *6* (2) 149–169

# Another Look at Eliminative and Enumerative Behaviour in a Conceptual Task

Mike Oaksford
*Department of Psychology, University of Wales, Bangor, UK*

Nick Chater
*Department of Psychology, University of Edinburgh, Edinburgh, UK*

There are currently two competing interpretations of hypothesis testing in Wason's (1960) 2-4-6 task: the *positivity heuristic* (Klayman & Ha, 1987; 1989) and the *counterfactual strategy* (Farris & Revlin, 1989a; 1989b). We argue that an extension of the counterfactual strategy—the *iterative* counterfactual strategy—should be preferred over the positive test heuristic because it may resolve the paradox of why subjects succeed on this task while apparently adopting an irrational strategy. We argue that an account of hypothesis *generation* is required to explain these data and that only the counterfactual strategy is of help here. We discuss the strategy and the 2-4-6 task in the light of contemporary history and philosophy of science, highlighting the rational basis of the strategy and some unrealities of the task.

## INTRODUCTION

There are currently two competing interpretations of the hypothesis testing behaviour observed in Wason's (1960) 2-4-6 task. One is due to Klayman and Ha (1987; 1989) and the other is due to Farris and Revlin (1989a; 1989b). Distinguishing experimentally between these accounts is difficult because both theories make similar empirical predictions. In this paper, we argue that Farris and Revlin's (1989a; 1989b) *counterfactual strategy*

---

Requests for reprints should be addressed to Mike Oaksford, Department of Psychology, University of Wales, Bangor, Gwynedd LL57 2DG, UK. E-mail: PSS027@bangor.ac.uk *or* mike@cogsci.ed.ac.uk.

We would like to thank Mike Gorman, Joshua Klayman, Russell Revlin, Ryan Tweney, Norman Wetherick and an anonymous reviewer for their helpful comments on an earlier draft of the manuscript.

© 1994 Lawrence Erlbaum Associates Limited

should be preferred over Klayman and Ha's (1987; 1989) *positive test heuristic*. This is because of the counterfactual strategy's potential to resolve the paradox of why subjects are so successful at the task. We argue that both strategies, at least prima facie, only address the *context of justification* (i.e. how hypotheses are tested) rather than the *context of discovery* (i.e. how hypotheses are generated). However, success at the 2-4-6 task could only be explained by an account of the latter. We argue that the positive test heuristic provides little information about how to revise a hypothesis, and where it does it is likely to prove a hindrance rather than a help.[1] We also argue that while there is a logical inconsistency in Farris and Revlin's account of the counterfactual strategy, repairing this problem leads to a test strategy which may facilitate the discovery process. On this basis, we suggest that an extension of the counterfactual strategy is to be preferred over the positive test heuristic as an explanation of the hypothesis testing behaviour observed in the 2-4-6 task.

We begin by outlining the 2-4-6 task. We observe that although prima facie an irrational *enumerative* strategy is adopted in this task, success rates are nonetheless very high.[2] Klayman and Ha (1987; 1989) and Farris and Revlin (1989a; 1989b) both attempt to resolve this paradox by claiming that despite superficial appearances, an *eliminative* procedure is being used. We outline and criticise both the positive test heuristic and the counterfactual strategy. We then present an extension of Farris and Revlin—which we call the *iterative counterfactual strategy*—in the form of a flow diagram. The diagram reveals that only the inputs to hypothesis generation processes are identified by this strategy. The nature of these processes, however, remains to be elucidated. We discuss the strategy in terms of some contemporary history and philosophy of science and some unrealities of the 2-4-6 task are highlighted. We conclude that while the extension of the counterfactual strategy we propose may more adequately explain the 2-4-6 task, it may not reflect general scientific practice.

## THE 2-4-6 TASK

In the 2-4-6 task, a number theoretic rule that applies to various number triples must be discovered. Subjects are told that the experimenter has a simple rule in mind for generating triples of which an initial "seed" triple (usually 2-4-6) is an instance. The task is to discover the experimenter's "target" rule by proposing number triples and receiving feedback as to whether they are instances of the experimenter's rule or not. Once the

---

[1]These points are also made by Klayman and Ha (1989).

[2]Especially in comparison to some other hypothesis testing paradigms, e.g. Wason's (1966) selection task where only 4% of subjects reach the correct solution (although, we should note that the task demands are rather different).

initial hypothesis is formed (e.g. ascending by twos), two kinds of test can be performed. Either triples that conform to a hypothesis (e.g. 20-22-24) can be proposed, or triples that do not conform to a hypothesis can be proposed (e.g. 20-22-23). Subjects typically propose triples that conform to the rule, that is, they *enumerate* potential instances of their hypothesis rather than propose instances which could lead to its *elimination*. After a series of such tests, subjects make a "rule announcement" indicating what they believe the target rule to be. They are usually highly confident that they have identified the target rule.

In Wason's (1960) original paper on the 2-4-6 task, he observed that early success was associated with more eliminative behaviour. This interpretation was questioned by Wetherick (1962),[3] who showed that subjects were proposing instances which supported their hypotheses.[4] These findings created a paradox that Klayman and Ha (1987; 1989) and Farris and Revlin (1989a; 1989b) could potentially resolve. Enumerators propose instances that confirm their hypotheses. However, while logically it is possible to unequivocally eliminate a hypothesis, it is impossible to unequivocally confirm a hypothesis (Popper, 1959). Thus enumerative behaviour would appear irrational. Yet paradoxically up to 75% of subjects solve the 2-4-6 problem (Wason, 1960).[5] If they were using such a maladaptive strategy, how is it that they are so successful? Both Klayman and Ha (1987; 1989) and Farris and Revlin (1989a; 1989b) argue that despite the apparently *enumerative* behaviour observed on this task, subjects are actually adopting an *eliminative* strategy (Popper, 1959).

## THE POSITIVE TEST HEURISTIC

Other than a perfect correspondence, there are four possible relationships between the set of number triples defined by the target rule (T) and the set of number triples defined by a subject's hypothesis (H) about the target

[3]The title of the present paper is borrowed from Wetherick's (1962) "Eliminative and enumerative behaviour in a conceptual task".

[4]Wetherick pointed out that Wason's procedure could not determine whether subjects were enumerators or eliminators, as they were not asked whether the instances they proposed conformed to their hypothesis or not. In Wetherick's modification, subjects had to file a proposed instance under the heading "Conforms" or "Does not conform".

[5]Although we should be careful to distinguish between success at the first announcement of the rule and subsequent success. In studies where more than one rule announcement is allowed, success rates typically reach around 75% (Wason, 1960; 73.9%: Farris & Revlin, 1989a). However, when only one rule announcement is made, success rates are still between 21.4% (Wason, 1960) and 40.9% (Farris & Revlin, 1989a). Thus even when only one rule announcement is made, performance on the 2-4-6 task compares favourably with performance on, for example, Wason's other classical hypothesis testing task, the selection task (Wason, 1966), where performance is typically as low as 4% correct.

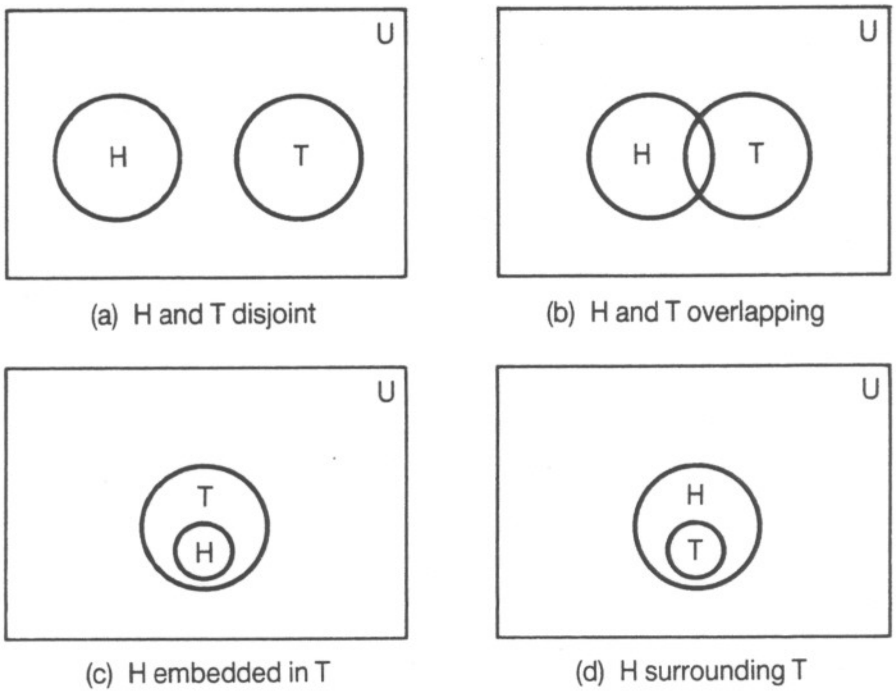


FIG. 1 Four possible relationships between hypothesis and target rule. U, universe of possible number triples; H, the set of number triples identified by the hypothesis; T, the set of number triples identified by the target rule. After Klayman and Ha (1989).

rule (see Fig. 1): H and T may *overlap*. H may be included in T (*embedded*); T may be included in H (*surrounding*); or T and H may be *disjoint*.

Klayman and Ha (1987) observed that these four situations lead to different outcomes depending on whether instances are being proposed which conform to a hypothesis (what Klayman and Ha refer to as +H-tests), or whether instances are being proposed which do not conform to a hypothesis (what Klayman and Ha refer to as −H-tests). Their analysis revealed that falsification is as likely with positive feedback as with negative feedback. Crucially, it also revealed that falsification is as likely with +H-tests (and "No" feedback) as with −H-tests (and "Yes" feedback). Moreover, on making two further assumptions, it was possible for Klayman and Ha (1987) to show that +H-tests actually lead to a greater probability of falsificatory feedback. These assumptions were an equiprobabilty and a minority phenomenon assumption.

The *equiprobability* assumption involves two probabilities: the probability that a triple is in the target set (in the 2-4-6 task this is usually given by the target rule "any ascending numbers"), and the probability that a triple is in the hypothesised set (e.g. even numbers). Little faith will be put in a hypothesis that severely under- or over-estimates the target set. Hence it is reasonable to assume that these two probabilities are roughly equal. If it is also assumed that the probability of being in the target set is less than *not* being in the target set—that is, a *minority phenomenon* assumption is made—then "you are more likely to receive falsification using +H-tests than −H-tests" (Klayman & Ha, 1987, p. 217). Thus despite superficial appearances, the behaviour observed on the 2-4-6 task is rational *because* falsification is more likely. Klayman and Ha (1987), therefore, resolve the paradox of why subjects are so successful while employing a prima facie irrational strategy.

In empirically evaluating some predictions of their *positive test strategy*, Klayman and Ha (1989) observed an apparently maladaptive response. For example (for others, see Klayman & Ha, 1989), let the target rule be *even numbers* and assume that the hypothesis *ascending numbers* is generated. This hypothesis overlaps the target rule. The confirmatory triple 3-5-7 is then generated, which receives "No" feedback. Rather than wholly abandon the initial hypothesis, Klayman and Ha (1989) observed that subjects would *qualify* it, e.g. they would generate the hypothesis *ascending*, ***even*** *numbers*. Positive instances of this hypothesis will always receive "Yes" feedback because the set of triples satisfying it is now *embedded* in the set of triples satisfying the target rule. Klayman and Ha (1989) refer to this situation as "embedded hypotheses".

If the qualification response was invariably adopted on receiving negative feedback to a +H-test, then subjects would rarely succeed at this task. The standard target rule is quite general, i.e. "any ascending numbers". Hypotheses, therefore, most frequently need to be expanded. However, the qualification strategy can only lead to *narrowing* a hypothesis. Klayman and Ha suggest that the qualification strategy emerges in response to receiving "No" feedback to a +H-test. The positive test heuristic may, therefore, often prove counter-productive. Klayman and Ha (1989, p. 601) themselves observe that, "even when the positive test strategy is effective in showing that a hypothesis should be revised, it can produce one-sided information about *how* it should be revised".

The problem Klayman and Ha (1989) identify is ubiquitous: it militates against any strategy which is concerned only with the *context of justification*. In the philosophy of science, it has been generally accepted that there is a distinction between the *context of discovery* (i.e. how hypotheses are generated) and the context of justification (i.e. how hypotheses are tested) (Popper, 1959). The quotation from Klayman and Ha (1989) implicitly suggests a closer relationship, where the method of testing a hypothesis may suggest how to discover a new hypothesis. We believe that this is a constructive move.[6] It does, however, militate against the view that subjects succeed at the 2-4-6 task *because* they adopt a strategy which leads to a greater probability of falsification. Falsifying a hypothesis provides no

[6]It also concurs with much contemporary philosophy of science (Brown, 1989).

useful information concerning plausible alternative hypotheses. Moreover, as Klayman and Ha (1989) concede, the positive test strategy may actually lead to a *maladaptive* selection of alternative hypotheses to test.

In summary, despite the excellence and ingenuity of Klayman and Ha's (1987; 1989) analysis, we still seem no closer to resolving the paradox of why subjects are so successful on this task. Although success is associated with eliminative behaviour, this could not be the *cause* of the success. As Wetherick (1962) observed (see also Klayman & Ha, 1989), those subjects who adopt an eliminative strategy also consider more alternative hypotheses (roughly three times as many in Klayman & Ha, 1989). Prima facie, the more hypotheses one tries, the higher the probability of success. So perhaps eliminators succeed *because* this leads them to generate more hypotheses. However, the space of possible hypotheses is extremely large if not infinite. Therefore, there is only a vanishingly small difference in the probability of being correct given you test, say, 10 rather than 2 hypotheses. We suggest that more emphasis will have to be placed on the *context of discovery* if this paradox is to be resolved. What is required is an account of how one strategy or another enables the discovery of ever more *relevant* hypotheses. We now look at Farris and Revlin's *counterfactual strategy*. We argue that although there is a logical error in the analysis of the counterfactual strategy, it nonetheless provides a fruitful source of likely hypotheses.

## A COUNTERFACTUAL STRATEGY

Farris and Revlin (1989a) argue that subjects may be adopting a *counterfactual* strategy in the 2-4-6 task. On this strategy, a hypothesis is first generated based on some property of the seed triple, e.g. even numbers. The *complement* of this rule is then generated, i.e. odd numbers. This rule is assumed to be true and subjected to a confirmatory test.[7] If "Yes" feedback is received, then the original hypothesis, even numbers, is incorrect, and its complement, odd numbers, may be correct.[8] Conversely, if "No" feedback is received, then the original hypothesis may be correct

[7]Note that "complement" is being used imprecisely in this context. The set union of the set of even natural numbers and the set of odd natural numbers yields the set of all natural numbers (i.e. $A \cup A' = U$). However, the set union of the set of even numbered triples and the set of odd numbered triples does *not* yield the set of all number triples (i.e. $A \cup A' \neq U$). Hence, "complement" is not meant as logical or set theoretic complement in the universe of number triples. In this context, "number theoretic *opposite*" may have been a more precise term.

[8]Thus the logic of the counterfactual strategy is given by *reductio ad absurdum* rather than *modus tollens*, as in the falsification strategy.

and the process is repeated. As Farris and Revlin (1989a) observe, this strategy would generate a series of positive instances of any local hypotheses a triple is intended to test with the consequence that subjects would appear to be confirming.

Farris and Revlin (1989a) contend that a "Yes" response from the experimenter will lead to the complement rule (e.g. odd numbers) being regarded as plausible. However, this is logically inconsistent. Since properties of the seed triple are chosen for the initial hypothesis, the complement rule is invariably falsified "at birth". By definition, the complement rule will concern a property *not shared* with the seed triple, in which case the complement rule *could not be* the target rule. For example, finding out that the triple 3-5-7 is an instance of the target rule, indicates that the even numbers rule is false, but since 2-4-6 is not an instance of the odd numbers rule, the odd numbers rule must also be false.

More recently, Farris and Revlin (1989b; see also Gorman, 1991) have argued for a revised model of the hypothesis testing process where the role of the counterfactual strategy is delimited. Rather than being the single strategy adopted by subjects, the counterfactual strategy is only invoked when subjects are attempting to find falsifying instances. The assumption is that because any hypothesis will be rare by definition, subjects attempt to determine its boundaries, that is, they want to know what kind of instances fall within and without T. One way to generate an instance that may fall outside T is to generate an instance you currently think might falsify your current hypothesis H about T, i.e. an instance of H's complement H′. In response to "Yes" feedback, Farris and Revlin (1989b) now propose that rather than regard H′ as plausible, it should be rejected "because the expanded set was not large enough". Note that while allowing that H′ is rejected, this is not for the logically correct reason. Moreover, within their modified strategy, H is apparently not rejected at this stage, rather only "a new *counterfactual* hypothesis is generated" (our italics). Again this is logically inconsistent.

These inconsistencies in Farris and Revlin's account may be due to a confusion between set theoretic "complement" and the notion of an "opposite" (see footnote 7). As we pointed out in footnote 7, the set union of a set and its complement yields the universal set, e.g. the set union of even and odd natural numbers yields all natural numbers. However, the set union of the set of even natural number *triples* and odd natural number *triples* does not yield the set of all natural number *triples*, e.g. 2-5-7 falls outside the union of these two sets. Thus it would be better to describe "odd number triples" as the *opposite* of "even number triples" rather than as its *complement*.

Confounding these two concepts may be responsible for the suggestion that on "Yes" feedback only a new counterfactual hypothesis need be generated. This is because there will be other "counterfactual" hypotheses that while outside the scope of the original hypothesis are not included in its opposite. So taking even numbers as the original hypothesis and odd numbers as its opposite, on receiving "Yes" feedback for the triple 3-5-7, a subject may generate the new "counterfactual" hypothesis "alternating even–odd–even" and propose the triple 2-5-7. This seems consistent with Farris and Revlin's contention that on "Yes" feedback subjects generate a new counterfactual hypothesis while making no mention of what happens to the original hypothesis. However, while this is possible, it is illogical. "Yes" feedback indicating that the opposite is in T, the target set, still falsifies H, because the relationship between opposites and complements is one of inclusion; that is, while it is not the case that all complements are opposites, all opposites are complements. So if T includes an instance of the opposite of a hypothesis H, H cannot be T.

The value of an opposite in this context is that they are derived from familiar *antonymic* structures in language that suggest obvious triples to try out as instances of what is not in H. Contrast, for example, generating an instance of the complement of the hypothesis H1 : even numbers, and the hypothesis H2 : that the number triples are derived from the Fibonacci series. Only the former has an opposite defined by its complement in the domain of natural numbers (rather than triples). This makes for ready identification of an appropriate instance of the complement of the hypothesis under test. So if you want an instance of a number that is *not*-even, then an odd number is the best bet. Arriving at an instance of a number that is *not* in the Fibonacci series is altogether more complicated (see Oaksford & Stenning, 1992, for more on the proper interpretation of negations).

## The Proper Function of the Counterfactual Strategy

While there are logical inconsistencies in the counterfactual strategy as described by Farris and Revlin (1989a; 1989b), we argue that these may be avoided, and that when this is done this strategy may have an important role in hypothesis testing. We suggest that by concentrating on the *instances* thrown up by "Yes" feedback in the counterfactual strategy, a means of generating ever more appropriate H's and their complements may be specified [we will continue to use the term "complement" because the logical force of an opposite is the same as a complement (opposite ⊂ complement)].

Take again the example of the hypothesis H′ : odd numbers. The triple 3-5-7 has been proposed and "Yes" feedback received. The complement rule H′ *and* the original rule H are therefore eliminated. However, the instance used to test H′ provides an additional source of constraint on generating a new hypothesis to test. Having received "Yes" feedback, it is now known that both 2-4-6 *and* 3-5-7 are instances of the target rule. This process could be iterated to enumerate a set of partially incompatible instances of the target rule.[9] The process began by selecting a property of the seed triple, so now a property *common to both triples* is selected and tested by the counterfactual strategy. Suppose that the property chosen is *ascending numbers*. Its complement, *descending numbers*, is checked by proposing 7-5-3 as an instance of the target rule. Given "Yes" feedback, it is now known that neither *ascending* nor *descending numbers* is the target rule. *But* it is also known that 2-4-6, 3-5-7 *and* 7-5-3 are instances of the target rule. The next iteration selects a property common to all three instances, e.g. *separated by twos*, and so on.

If there were an effective procedure for identifying common number theoretic properties between number triples, this strategy could represent a component of a *discovery* procedure for identifying the target rule. Its termination conditions are two-fold. First, a "No", in which case announce the current positive hypothesis as the target rule (if this is not the target rule, then discount this triple and search for another common property of the remaining instances and iterate). However, announcing the rule immediately may be premature, since the seed triple could be the only positive instance encountered. Some "novel predictions" of the hypothesis may, therefore, be tested before announcing the new hypothesis. Hence, proposing a limited number of positive instances at this stage prior to rule announcement may be wise. Second, no common properties are left, in which case announce "any three numbers" as the target rule. However, if "Yes" feedback has been received, termination may occur before these conditions are met, because among the many instances only one common property can be perceived, i.e. there may be more *in principle*, but this is getting beyond a hypothesis tester's abilities for recognition *in practice*. Thus a rule may confidently be announced, without any falsificatory feedback having been received, because of the diverse range of partially incompatible instances the rule covers. When the conditions for termination are met and whether the first "No" results in the announcement of the correct target rule will depend on what initial property of the seed triple is chosen. Thus individual differences in the number of trials employed would be expected.

In the next section, we propose a flow diagram model of the processes which may be involved in the 2-4-6 task based on the *iterative counterfac-*

---

[9]That is, incompatible along some dimension(s), e.g. ascending *vs* descending. The instances cannot be completely incompatible, since otherwise no target rule could subsume them all (except in the limiting case where the target rule is simply any three numbers).

tual strategy we have suggested. Flow diagram models are not currently in favour. This is largely because of the recent development of lower-level *connectionist* models, which appear more able to model the statistical variation found in the empirical data. However, flow diagram models do have an important role to play. They have the status of conceptual tools for clarifying process issues independent of precise implementations. Moreover, this is the role flow diagram models have in systems analysis (Colin, 1980). They provide a program specification independent of (1) the particular programming language in which a process is to be encoded and (2) the particular machine on which the program is to run. Thus in so far as they are a useful conceptual tool in systems analysis in computer science, they will remain an important functional level tool for clarifying process issues in cognitive psychology. An important role such models can play is identifying those aspects of a process that need further theoretical elaboration. By adhering to the conceptual rigour of specifying such a model, it becomes clear which subprocesses can be specified algorithmically and which are currently vague.

## AN ITERATIVE COUNTERFACTUAL STRATEGY

The *iterative* counterfactual strategy is summarised in the flow diagram in Fig. 2. The strategy begins by adding the seed triple to the list L of instances (1). A common property of the instances in L is then selected to function as the hypothesis H (2). Initially, this will mean just selecting a property of the seed triple. Box (2) is represented as a "cloud" (Colin, 1980), because this is where the crucial and currently unaccounted for processes of hypothesis generation occur. The complement of H, H′, is then generated (3), followed by an instance I′ of H′ (4). I′ is then proposed as an instance of T, the target rule (5). If "Yes" feedback is received, then both H and H′ are rejected and I′ is added to L (6). Conversely, if "No" feedback is received, H′ is rejected (7). Before announcing H as T (12), it may be prudent to generate a few positive instances, $I_i$, of H (8) and test whether they are instances of T (9) in what we have called the "positive sub-loop". If "No" feedback is received, then H is rejected and I *not* added to L (13). Conversely, if "Yes" feedback is received and the number, $i$, of positive instances, $I_i$, is such that $i$ is less than some criterion $x$ (10), then add I to L (11) and iterate and try another positive instance. Conversely, if $i = x$, then announce H as T (12). If "Yes" feedback is received, then the procedure has successfully terminated. Conversely, if "No" feedback is received, then H is rejected and I′ *not* added to L (13). At this point, the lines from (5) converge on a check to ascertain whether any common properties of the instances in L are perceivable (14). If they are, then the procedure iterates. If no common properties are perceivable,

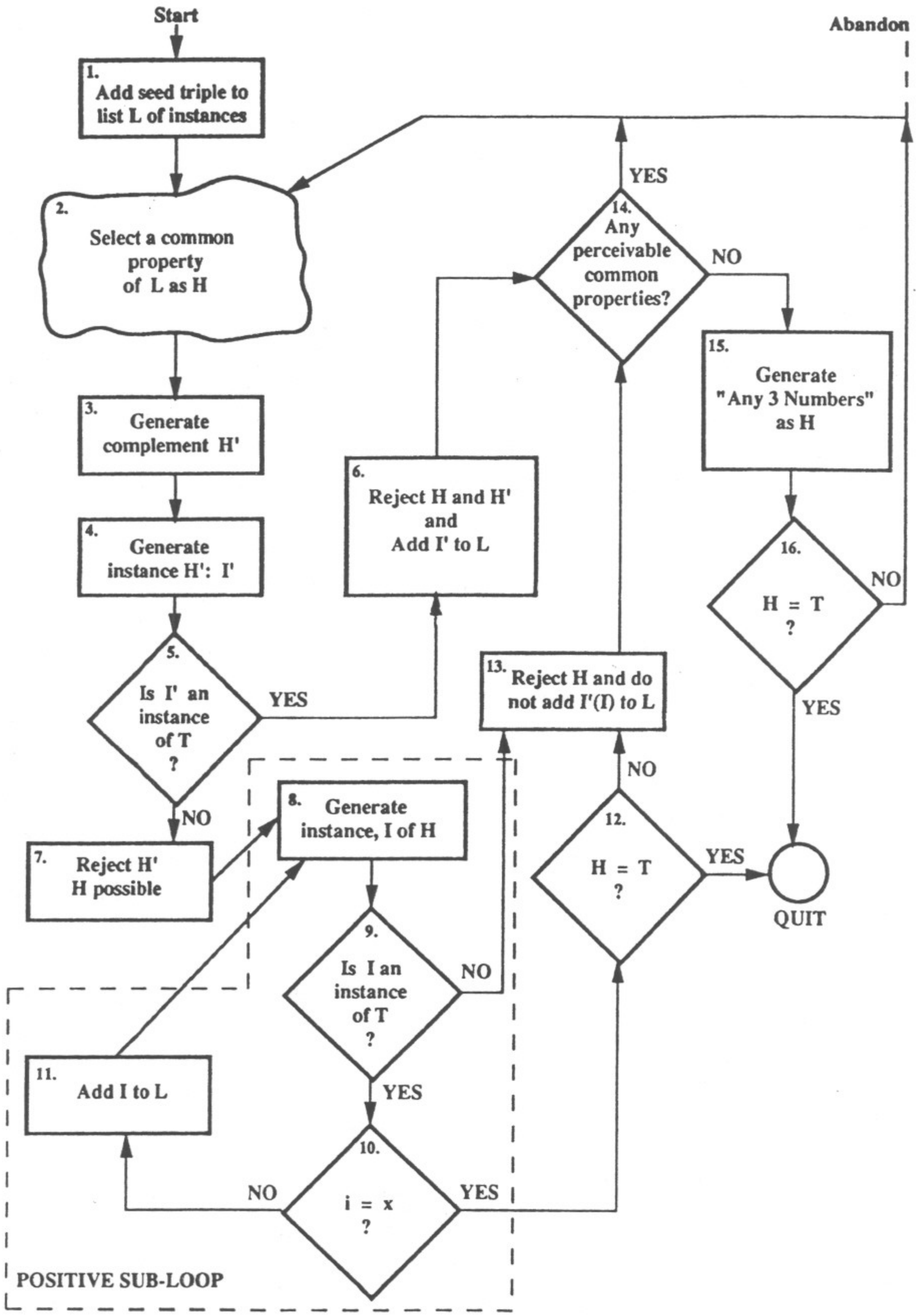


FIG. 2 Flow chart representation of the iterative counterfactual strategy (see text for explanation).

then *any three numbers* is generated as H (15) and announced as T (16). If "Yes" feedback is received, then the procedure has again successfully terminated. If "No" feedback is received, then it must be assumed that although not initially perceivable, there must be some further common properties available. Alternatively, subjects may abandon the whole exercise in exasperation.

This flow diagram reveals the central role played by the subprocesses involved in hypothesis generation (2). In particular, in contrast to Farris and Revlin (1989b), the iterative counterfactual strategy captures the important role played by the actual instances that are thrown up by the counterfactual strategy. These instances are available to constrain the choice of a new H and hence H′. Whether the process iterates depends on the ability to select a new hypothesis to test. Thus some explication of the processes of hypothesis generation is crucial. The counterfactual strategy does provide some clues about the *inputs* to the hypothesis generation process. A diverse set of instances may constrain the choice of likely hypotheses, delimiting the logically possible to the probable. However, although the inputs are clearly delineated, the innards of box (2) remain obscure. There is also an outstanding problem with the strategy as represented in Fig. 2 for which the qualification response outlined above may provide a natural solution (see Discussion) but which in the case of the 2-4-6 problem proves maladaptive.

## Narrowing a Hypothesis

Although the iterative counterfactual strategy may be a necessary *component* of a discovery procedure, it is not sufficient. This can be seen from considering how the strategy copes with the four possible relationships between a hypothesis H and the target rule T, identified by Klayman and Ha (1987; 1989) (we will reserve a discussion of the case of H and T overlapping until later). Most trivially, H and T will never be disjoint, since H is always generated either from the seed triple which is guaranteed to be in T, or from the list of enumerated instances L, including the seed triple, which are also guaranteed to be in T. The iterative counterfactual strategy is best suited to the situation where H is included in T (i.e. an embedded H), since the accumulation of partially incompatible instances will ensure that the hypotheses generated will be of ever increasing generality. However, this strategy is of no assistance when T is included in H (i.e. a surrounding H). If H surrounds T, then on the iterative counterfactual strategy, "No" feedback will always be received at (5), indicating H is possible. This is because when H surrounds T, H′ (the complement of H) and T are necessarily disjoint. Thus whatever happens in the positive sub-loop, the procedure will always iterate again. However, *no* instances will accumulate to suggest ways of narrowing H in the same way that they accumulate to suggest ways of expanding H. Thus, whether a narrower hypothesis is selected is left to the unexplicated procedures of hypothesis generation in box (2).

However, the qualification response provides a natural way of narrowing a hypothesis.[10] This is especially so because "No" feedback, either in the positive sub-loop or on rule announcement, does *not* exclude H *plus a qualification* from being a good contender for T [i.e. before box (13) in Fig. 2, an option to add qualifications should be included]. The primary risk involved, as we saw above, is that an embedded hypothesis (i.e. H included in T) will result. However, if the iterative counterfactual strategy is then adopted, a hypothesis tester may also be able to recover from this situation. When subjects begin with overlapping hypotheses (as they most frequently do; Klayman & Ha, 1989), which of the situations Klayman and Ha identify they move to next will depend on the particular hypothesis H and the instances they choose to test. It may be that H′ overlaps T to such an extent that "Yes" feedback is highly likely, in which case H and H′ will be rejected quite rapidly. Alternatively, there may be little or no overlap between H′ and T, in which case H may be accepted as plausible. This could lead to qualifications and an embedded hypothesis.

The qualification procedure is not depicted in Fig. 2. This is because the additional complexity would defeat the diagram's function to depict the iterative counterfactual strategy clearly. The basis of the strategy is to repeat the process of extracting common properties from L to be conjoined with H. It is then a matter of conjecture whether the counterfactual strategy is rejoined or the positive sub-loop. If the latter, and an embedded hypothesis has resulted, then it won't be discovered until once again the criterion $x$ has been reached and the complex rule announced. If, however, the counterfactual strategy is rejoined, then subjects confront the problem of whether to look for a joint complement or just the complement of the new property in order to construct H′. Nonetheless, they still have the possibility of rejecting the new H, which they do not if they rejoin the positive sub-loop.

The qualification response is, however, maladaptive in the 2-4-6 task because T is usually *a unitary* feature of the seed triple, e.g. "ascending". Therefore, when narrowing, a unitary hypothesis needs to be selected

---

[10]Like Klayman and Ha (1989), Farris and Revlin (1989b) also suggest that hypotheses must be narrowed. However, unlike Klayman and Ha, Farris and Revlin are not specific about how this is to be achieved. The qualification response may, however, be what Farris and Revlin (1989b) intend when they suggest that "slight alteration of the original hypothesis is introduced" (but they provide no example of such an alteration, so it is difficult to tell).

which is consonant with the seed triple and which applies to fewer number triples in the domain of all triples than the previous hypothesis. To achieve this, accurate estimates are required of the frequencies of the two hypotheses in the domain of triples. This is equivalent to possessing knowledge of the base rates of two unitary hypotheses in the population. And this is information which subjects are unlikely to possess (Kahneman & Tversky, 1973; but see also Gigerenzer, Hell, & Blank, 1988).[11] Thus a principled way of discovering appropriate unitary hypotheses when T is included in H may be beyond most subjects' reach. Perhaps surprisingly, although maladaptive in the 2-4-6 task, the qualification response, which would be condemned as *ad hoc* by Popper (1959), is ubiquitous in the sciences, as we shall discuss further below.

## Summary

The iterative counterfactual strategy resolves the logical inconsistencies in Farris and Revlin's (1989a; 1989b) account. On receiving "Yes" feedback to a complement instance, *both* H and H′ are rejected. The counterfactual strategy is attempted falsification, but via the argument form of *reductio ad absurdum* rather than *modus tollens* as in the standard disconfirmation strategy. It is this innovation that accounts for the predominance of positive instances of hypotheses proposed in the 2-4-6 task, although subjects are actually falsifying. On our analysis of the iterative counterfactual strategy, it also provides a component of a discovery procedure for arriving at new hypotheses to test. On the one hand, it provides a range of inputs to the hypothesis generation process that enables a hypothesis to be expanded. On the other hand, the qualification response provides a natural mechanism for narrowing a hypothesis, although this is maladaptive in the 2-4-6 task. Thus an efficient eliminative procedure for discounting hypotheses simultaneously allows the enumeration of instances to constrain hypothesis generation. Taken together these factors offer some resolution of the paradox of why subjects succeed on this task while on the face of it adopting an irrational strategy.

The iterative counterfactual strategy makes several empirical predictions. First, for subjects who adopt this strategy, a common observation should be cases of "Yes" feedback being followed immediately by a change

---

[11] However, all that is required is *relative* base rate information, not *absolute* base rate information. Thus, without knowledge of the absolute base rates of *black things* and *ravens*, my knowledge of many black non-ravens provides good evidence that *P*(black thing) > *P*(raven). However, even this much background knowledge is likely to be beyond most subjects for the number theoretic domains involved in the 2-4-6 task.

of hypothesis. This should show up in the next number triple being incompatible with the triple that just elicited a "Yes" response. Second, on this strategy, "No" feedback indicates the plausibility of H which is then tested in the positive sub-loop. This leads to the expectation that "No" feedback should frequently be followed by a long string of "Yes" responses (especially if H is included in T). Third, as also pointed out by Farris and Revlin (1989a), on the counterfactual strategy, success should not be dependent on "No" feedback. Therefore, we should not expect the percentage of "No" feedback to distinguish between solvers and non-solvers on this strategy. Further information could be obtained by adopting the proposals made by Gorman (1991). If subjects were asked to indicate their current best guess hypothesis at each trial, then the logical relations between triple and hypothesis could be ascertained. On the counterfactual strategy, there should be more frequent mismatches between the triples proposed and the current best guess hypothesis. In testing the utility of L—the list of partially incompatible instances of T—one would expect that the best guess hypotheses should be compatible with all triples for which "Yes" feedback was received. Distinguishing between strategies may also be facilitated by the use of training studies where subjects are given explicit instruction concerning which strategy to adopt (see also Gorman & Gorman, 1984).

The components of the iterative counterfactual strategy may not only be reflected in the data, but may also be at work in actual scientific practice. We discuss this possibility in the next section.

## DISCUSSION

There are two main issues that require further discussion. First, we argue that most of the processes we have introduced are well documented in the history of science. Actual scientific practice provides our yardstick of rationality (Brown, 1989). Thus if we show that the processes we propose are regularly found in scientific practice, then we have good grounds for assuming that they are rational. Second, we wish to highlight some of the unrealities of the 2-4-6 problem. There are aspects of this task which may not invoke strategies normally engaged by scientists or everyday hypothesis testers. To the extent that this is so, the validity of the task as a measure of people's hypothesis testing abilities is called into question.

### Rationality

We will look at three subprocesses in the iterative counterfactual strategy: the qualification response, the positive sub-loop, and the enumerative input to the processes of hypothesis generation. For each we will argue that there is a parallel in the history of science which indicates the rationality of these procedures.

Qualifying a rule or a scientific law is equivalent to restricting its range of application to particular contexts. Thus, the *ascending **even** number* rule we discussed above can be interpreted as restricting the *ascending number* rule to the domain or context provided by even numbers. Popper (1959) argues that such *ad hoc* hypotheses are to be avoided; however, such restrictions are found in many of science's most successful and enduring laws (Cartwright, 1983). In arguing that most scientific laws which can lay claim to describing real causal processes are similarly context-sensitive, Cartwright (1983) employs the example of Snell's law. This is the familiar optical law which states that the ratio of the angle of incidence and the angle of refraction is a constant. However, Cartwright observes, this only holds in *isotropic* mediums (in *anisotropic* mediums there are *two* refracted rays). The precise meaning of "isotropic" need not concern us, the point is that here is a law which is useful enough to be retained even when it is known to be strictly false, *because* we know its appropriately restricted domain of application. Many other examples abound—all swans *native to the northern hemisphere* are white, all *genetically normal* ravens are black, and so on. It seems that in the face of falsifying evidence, it is not always rational to wholly abandon a hypothesis, since some minor adjustment in our system of scientific beliefs may save it from outright refutation (Quine, 1953). In consequence, the qualification response is *normally* a rational procedure.[12] In the standard 2-4-6 task, rather than qualifying the old rule, the seed triple requires *re-classification* (see Holland, Holyoak, Nisbett, & Thagard, 1986) under a different atomic or unitary concept and thus the qualification response proves maladaptive. However, there are versions of the 2-4-6 task where the possibility of error is introduced and where rules are employed that allow qualifications (Gorman, 1989). When this is done, subjects naturally qualify their hypotheses even in abstract tasks.

The counterfactual strategy is based on the methodology of *crucial experiments* or *strong inference* (Platt, 1964). That is, an experiment is devised such that two mutually incompatible hypotheses make diverse predictions concerning its outcome. However, the role of such experiments in accepting a novel hypothesis may be minimal without a demonstration of some successful novel predictions, as we have suggested in the *positive*

---

[12]However, as Oaksford and Chater (1991) discuss, rules with many qualifications—which seems characteristic of the bulk of commonsense knowledge—cause problems of their own, especially with regard to the tractability of the inferential processes which need to be employed with such a knowledge base. For further discussion in the context of the processes of scientific inquiry, see Oaksford and Chater (forthcoming).

*sub-loop* in the flow diagram outlined above. The Michelson–Morley experiment (see Hacking, 1983; Lakatos, 1970) is often cited as crucial in deciding against the classical theory of the aether (the medium through which light waves were hypothesised to travel) and for relativity theory. However, this was largely a *post hoc* rationalisation. Many physicists continued to adhere to the aether theory even after Einstein published his theory of special relativity in 1905 (Lakatos, 1970). Indeed, this response continued unabated at least until 1925 when a paper by Miller appeared in *Science* supporting the aether theory via a critique of the Michelson–Morley experiment (see Lakatos, 1970, p. 165). Only a long period of sustained predictive success led to the adoption of relativity theory and the rejection of the classical framework. The new theory was accepted because of its remarkable propensity for making novel predictions, not as a result of strong inference (although the simplicity and elegance of Einstein's theory also strongly influenced its adoption by the scientific community). Thus making novel predictions in the positive sub-loop is indeed a rational strategy.

We have suggested that the iterative counterfactual strategy provides a range of data which serves to constrain possible hypotheses. We now observe that constraining possible hypotheses by considering a range of data is also the norm in the natural sciences. As an example domain we return to physical optics. The development of optical theory has had to deal with a whole range of empirical observations from diverse areas, e.g. reflection, refraction, double refraction, interference and diffraction effects, colour spectra, absorption spectra, and so on. Arriving at any hypothesis capable of covering some of this data was a Herculean intellectual task. Moreover, any hypothesis that could cover this data, albeit *post hoc*, would be strongly favoured, simply in virtue of its coverage. It is notable that in almost 500 years of research, only two hypotheses have ever been serious contenders to explain these optical phenomena: the corpuscular theory of Newton and the wave theory of Huyghens. Thus, as we suggested above, subjects may, perhaps prematurely, announce a rule without sufficient testing, simply because of the range of data it covers. Moreover, the iterative counterfactual strategy generates a body of diverse data constraining the hypotheses generated for its description.

## Unrealities

In this section, we will look at four unrealities in the 2-4-6 task: the plethora of possible hypotheses available, the lack of constraint provided by a single seed triple, that fact that subjects know there is a law to be discovered, and the implicit acceptance of the view that only *true* laws or theories have any utility.

The counterfactual strategy is based on the logic of crucial experiments (see above). It is an eliminative strategy, like falsification, which relies on the existence of a wealth of hypotheses to test and reject. However, such a strategy is not prevalent in a mature science. As we have observed, there is usually a dearth rather than an abundance of reasonable hypotheses in mature scientific domains. Moreover, genuine crucial experiments are either rare or a philosopher's *post hoc* rationalisation. In contrast, the plethora of number theoretic predicates (and their corresponding "opposites") means that the counterfactual strategy is highly appropriate in the 2-4-6 task.

Theorists are normally confronted with an array of results that for some reason require a unified explanation, for example the range of optical data alluded to above. Existing computer programs, which discover general laws (e.g. BACON: Langley, Simon, Bradshaw, & Zytkow, 1987), also assume a data set and attempt to uncover general qualitative or quantitative laws for its description. Such procedures would be insufficiently constrained by a single instance, as in the 2-4-6 task. A virtue of the iterative counterfactual strategy is that it effectively expands the data set to where reasonable hypotheses may be generated. This may suggest that the 2-4-6 task is more analogous to the experimental procedures that generate the scientist's data set in the first place. These procedures are indeed enumerative, very often fortuitous and highly unlikely to be formalisable. Nonetheless, in the highly constrained domain of *reasonable* number predicates and number triples, subjects do appear to adopt a very sensible enumerative strategy.

A further unreality of the task is the fact that the subjects are told that there is a law governing the generation of the seed triple for them to discover. In inducing general laws, the scientist is again never in the privileged position of knowing whether a law-like relation exists or not, prior to trying to discover what it is. [However, recently, Gorman (1989) has introduced procedures that may more closely mimic the scientist's normal situation in this respect.]

A final unreality of the task is that however close the announced rule comes to covering the same domain of triples as T, it will always receive "No" feedback unless it is correct. This underestimates the utility of good as opposed to true theories. Theory or hypothesis construction begins by covering easy cases and progressively attempts to bring more aspects of the data under the explanatory umbrella of the theory. However, a theory is not regarded as without utility even if it fails to cover some aspects of the data. For example, as they acknowledged, Rescorla and Wagner's (1972) theory of Classical Conditioning failed to explain several well-documented learning phenomena (see Dickinson, 1979, for a review). Nonetheless, this theory was and remains crucially important in the development of theories of conditioning.

## CONCLUSIONS

Only the counterfactual strategy of Farris and Revlin (1989a; 1989b) would appear extendible to account for why subjects are so successful at the 2-4-6 task. This is because it naturally enumerates a set of inputs to the processes of hypothesis generation in a way that the positive test heuristic does not. Moreover, in combination with the qualification response, a strategy that mirrors some of the patterns of reasoning involved in actual scientific practice can be defined. The phases of the iterative counterfactual strategy are consistent with Tukey's (1986) observation that subjects do not describe themselves as engaging in a single strategy. Rather, particular triples may be proposed for different purposes.

There are important aspects of the data on the 2-4-6 task that the iterative counterfactual strategy fails to address. Most importantly, we have not addressed any of the manipulations that facilitate an eliminative strategy (e.g. Gorman, 1986; Gorman & Gorman, 1984; Tweney et al., 1980; Wetherick, 1962), such as proposing that all triples fall into one of two mutually exclusive classes, one named DAX the other named MED (Tweney et al., 1980). While aware of the problem, we remain unable to offer anything more than an unprincipled and *ad hoc* account of why such a procedure should make a difference. It should be borne in mind, however, that the unrealities of the 2-4-6 task *may* indicate that while important to this laboratory task, this manipulation may have no analogue in real scientific inquiry. Its efficacy is often discussed (Tweney et al., 1980) in terms of the role of crucial experiments and strong inference (Platt, 1964). Naming serves to identify the mutually exclusive hypotheses under test. However, as we have pointed out, the history of science reveals that such episodes are in fact rare and usually inconclusive. Unfortunately, of course, this also suggests that the counterfactual strategy itself could not be expected to be much in evidence in real scientific inquiry.

While the actual processes of hypothesis generation still remain to be elucidated, it is certain that world knowledge will play a central role. No strategy which subjects could adopt is ever *guaranteed* to find target rule T. The space of possibilities is too large. The set of all number triples is itself uncountably infinite, let alone the members of the power set of that set, each of which would, in extension, define a property. Experimenter and subject alike make massive assumptions about what is *reasonable*. The principle assumption is that T is a rule describable using the number theoretic predicates of which the normal, intelligent adult is likely to be aware. These shared assumptions reveal the reliance which hypothesis generation and test places on prior world knowledge. The choice of predicates to incorporate in hypotheses *and* target rules has been constrained by the subject's and experimenter's respective world knowledge.

An understanding of human reasoning and the 2-4-6 task in particular will rely on the development of accounts of hypothesis generation that integrate with theories of hypothesis testing (Fodor, 1983; Klayman & Ha, 1989). However, despite some notable progress (e.g. Holland et al., 1986; Langley et al., 1987; Sternberg, 1988), it must be conceded that the processes of hypothesis generation and thus of scientific creativity in general remain profoundly obscure (Fodor, 1983). Given the central role of these processes, it is tempting to recall Fodor's pessimistic conclusion in discussing this issue that, "In this respect, cognitive science hasn't even started. . ." (Fodor, 1983, p. 129), and to reflect that little has changed in the intervening years.

Manuscript received February 1993
Revised manuscript received October 1993

## REFERENCES

Brown, R. (1989). *Rationality*. London: Routledge and Kegan Paul.

Cartwright, N. (1983). *How the laws of physics lie*. Oxford: Oxford University Press.

Colin, A.J.T. (1980). *Fundamentals of computer science*. London: Macmillan.

Dickinson, A. (1979). *Contemporary animal learning theory*. Cambridge: Cambridge University Press.

Farris, H.H., & Revlin, R. (1989a). Sensible reasoning in two tasks: Rule discovery and hypothesis evaluation. *Memory and Cognition*, *17*, 221–232.

Farris, H.H., & Revlin, R. (1989b). The discovery process: A counterfactual strategy. *Social Studies of Science*, *19*, 497–513.

Fodor, J.A. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.

Gigerenzer, G., Hell, W., & Blank, H. (1988). Presentation and content: The use of base rates as a continuous variable. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *14*, 513–525.

Gorman, M.E. (1986). How the possibility of error affects falsification on a task that models scientific problem solving. *British Journal of Psychology*, *77*, 85–96.

Gorman, M.E. (1989). Error, falsification and scientific inference: An experimental investigation. *Quarterly Journal of Experimental Psychology*, *41A*, 385–412.

Gorman, M.E. (1991). Counterfactual simulations of science: A response to Farris and Revlin. *Social Studies of Science*, *21*, 561–564.

Gorman, M.E., & Gorman, M.E. (1984). A comparison of disconfirmatory, confirmatory and control strategies on Wason's 2-4-6 task. *Quarterly Journal of Experimental Psychology*, *36A*, 629–648.

Hacking, I. (1983). *Representing and intervening*. Cambridge: Cambridge University Press.

Holland, J.H., Holyoak, K.J., Nisbett, R.E., & Thagard, P. (1986). *Induction: Processes of inference, learning and discovery*. Cambridge, MA: MIT Press.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, *80*, 237–251.

Klayman, J., & Ha, Y. (1987). Confirmation, disconfirmation and information in hypothesis testing. *Psychological Review*, *94*, 211–228.

Klayman, J., & Ha, Y. (1989). Hypothesis testing in rule discovery: Strategy, structure and content. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *15*, 596–604.

Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In I. Lakatos & A. Musgrave (Eds), *Criticism and the growth of knowledge*, pp. 91–196. Cambridge: Cambridge University Press.

Langley, P., Simon, H.A., Bradshaw, G.L., & Zytkow, J.M. (1987). *Scientific discovery: Computational explorations of the creative processes*. Cambridge, MA: MIT Press.

Oaksford, M., & Chater, N. (1991). Against logicist cognitive science. *Mind and Language*, *6*, 1–38.

Oaksford, M., & Chater, N. (forthcoming). *Cognition and inquiry*. London: Academic Press.

Oaksford, M., & Stenning, K. (1992). Reasoning with conditionals containing negated constituents. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *18*, 835–854.

Platt, J.R. (1964). Strong inference. *Science*, *146*, 347–353.

Popper, K.R. (1959). *The logic of scientific discovery*. London: Hutchinson.

Quine, W.V.O. (1953). *From a logical point of view*. Cambridge, MA: Harvard University Press.

Rescorla, R.A., & Wagner, A.R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A.H. Black & W.F. Prokasy (Eds), *Classical conditioning II*. New York: Appleton-Century-Crofts.

Sternberg, R.J. (Ed.) (1988). *The nature of creativity*. Cambridge: Cambridge University Press.

Tukey, D.D. (1988). A philosophical and empirical analysis of subjects' modes of inquiry in Wason's 2-4-6 task. *Quarterly Journal of Experimental Psychology*, *38A*, 5–33.

Tweney, R.D., Doherty, M.E., Worner, W.J., Pliske, D.B., Mynatt, C.R., Gross, K.A., & Arkkelin, D.L. (1980). Strategies of rule discovery in an inference task. *Quarterly Journal of Experimental Psychology*, *32*, 109–123.

Wason, P.C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology*, *12*, 129–140.

Wason, P.C. (1966). Reasoning. In B. Foss (Ed.), *New horizons in psychology*. Harmondsworth: Penguin.

Wetherick, N.E. (1962). Eliminative and enumerative behaviour in a conceptual task. *Quarterly Journal of Experimental Psychology*, *14*, 246–249.