MEMORY, 2002, *10* (5/6), 397–403

Levels of processing, transfer-appropriate processing, and the concept of robust encoding

Robert S. Lockhart

University of Toronto, Canada

The theoretical status of levels of processing and its relation to the general principle of transfer-appropriate processing is discussed. One possible description of this relationship is that levels of processing has its effect by influencing the likelihood that the processing will prove to be transfer-appropriate. This transfer account of the levels effect is discussed in terms of the concept of robust encoding (Lockhart & Craik, 1990). Available evidence provides little support for any simple form of this concept, but a modified version is suggested as a possibility.

At the age of 30, levels of processing would seem to be still very much alive. However, recent years, have seen it undergo an apparent change of status. Once a focal subject in the introduction and general discussion of experimental reports, it is now more often to be found in the method section: levels of processing has become an experimental manipulation, a paradigm. To some, this shift in status from conceptual to methodological may seem to make being alive at 30 a dubious achievement, but a consideration of the original ideas (Craik & Lockhart, 1972) justifies a more positive view. It must be remembered that levels of processing was presented, not as a theory of memory, but as a framework for memory research. The claim was that our understanding of memory would be greatly enhanced by research that examined the consequences for remembering of qualitatively different forms of processing.

Recent research using the "levels paradigm" can be thought of as doing precisely this. Exploring the "levels effect" has done much to clarify our understanding of the relations among memory systems and their distinctive roles in traditional memory tasks. The experiments reported by Challis, Velichkovsky and Craik (1996) provide a good example. More specifically, the presence or absence of levels effects has become a benchmark for the distinction between implicit and explicit memory (e.g., Roediger, Weldon, Stadler, & Riegler, 1992). Similarly, manipulating levels of processing has substantially added to our understanding of the difference between remembering and knowing as distinct components of recognition (Gardiner, 1988; Rajaram, 1993).

IS LEVELS OF PROCESSING A TAUTOLOGY?

These examples of the pragmatic success of the levels-of-processing framework is a partial answer to those critics who interpreted it as a theory, the truth or falsity of which might be decided by a suitable *experimentum crucis*. The answer is only partial, however, in that it sidesteps the basic concerns that motivated the criticism in the first place. This is the claim that the concepts underlying levels of processing are untestable and unfalsifiable.

The accusation of being "circular" or tautological has been one of the most persistent criticisms

Requests for reprints should be sent to Robert S. Lockhart, Department of Psychology, University of Toronto, Toronto, Ontario, Canada M5S 3G3. Email: lockhart@psych.utoronto.ca

I thank F.I.M. Craik for helpful comments on an earlier version of this paper.

http://www.tandf.co.uk/journals/pp/09658211.html DOI:10.1080/09658210244000225

of levels of processing. The claim is that because the concept of depth lacks an independent definition, levels of processing involves a tautology in which the putative cause (deep processing) is nothing more than a restatement of the effect (better remembering). By their very nature tautologies are unfalsifiable, thus rendering the concept of depth devoid of empirical substance (Nelson, 1977). This criticism has been discussed briefly in a number of places (e.g., Lockhart & Craik, 1978, 1990), and most of the points made there will not be repeated here. However, apart from attempting to put the circularity argument finally to rest, there are aspects of this issue that warrant further attention because they lead to a number of other matters. The most important of these is the relation between levels of processing and the closely related concepts of transfer-appropriate processing (Morris, Bransford, & Franks, 1977) and the encoding specificity principle (Tulving & Thomson, 1973).

In standing accused of tautology, levels of processing is in distinguished company. Darwinian natural selection has frequently been similarly criticised and defended (see, for example, Gould, 1977; Stamos, 1996) and it is instructive to compare the two cases. With natural selection, the focus of criticism is usually the concept of adaptation, and in particular Herbert Spencer's phrase "the survival of the fittest", a phrase that Darwin adopted in later editions of the *Origin of Species* where it was added to the heading to Chapter 4. The circularity argument is that there is no independent definition of fitness: the fittest are defined in terms of those that survive. Fitness is thus a cause that is nothing more than a restatement of its effect. Evolution by natural selection is therefore a tautology.

It is instructive to pursue this parallel between levels of processing and natural selection. Concepts such as "fitness" (or any other description of adaptability) and "depth of processing" (or any other description of the memory trace) have an important feature in common. Neither is capable of fully predicting its relevant outcome—survival and retrievability respectively. At best they can speak probabilistically: certain characteristics of the organism might increase the likelihood of survival, and certain features of a memory trace may make retrieval more probable. It is this inability to predict deterministically that gives both concepts their post hoc quality, thus making them vulnerable to the accusation of circularity. Outside of the laboratory, a full account of a particular instance of survival or retrieval is possible only after the event. The reason for this state of affairs is basically the same in both cases: Adaptability and retrievability are concepts that are intrinsically relational. Survival is not fully predictable on the basis of some a priori definition of fitness, or any other property of the organism, because survival is a relationship between such properties and environment conditions. Analogously, retrievability is not fully predicted by some a priori definition of depth of processing, or any other property of the memory trace (such as strength or distinctiveness) because retrievability is a relationship between such properties and conditions under which retrieval is required.

This post hoc quality is not damaging to the scientific status of either, and it should not be confused with circularity or a tautology. The important point is that mechanisms of survival can be studied and subjected to experimental evaluation. Moreover, some degree of predictability is possible to the extent that future environmental conditions can be specified, or controlled in the laboratory. Similarly, mechanisms of memory retrieval can be studied and subject to experimental evaluation and predictions made contingent on specified retrieval conditions. The scientific integrity of natural selection or levels of processing is not compromised by the inability to specify in advance the conditions in which they will operate.

TRANSFER-APPROPRIATE PROCESSING

The relational property between the memory trace and the conditions of retrieval is, of course, the fundamental claim underlying the concepts of transfer-appropriate processing (Morris et al., 1977) and the encoding specificity principle (Tulving & Thomson, 1973). The evidence supporting the general validity of these conceptual twins is overwhelming. Notice, however, that transfer-appropriate processing is vulnerable to the same inappropriate accusation of circularity as levels of processing, and for exactly the same reason: it, too, possesses a certain post-hoc quality and lacks a precise definition. Under circumstances in which retrieval conditions cannot be specified in advance, processing can be identified as "transfer-appropriate" only after retrieval has occurred. Its defence is the same as that made for levels of processing in the preceding paragraph. Moreover, a lack of precision in defining concepts such as levels of processing or transfer-appropriate processing is not as great a weakness as is often claimed. Many important scientific ideas begin as vague concepts and their clarification is the goal of research, not its starting point.

Granted the general validity of transfer-appropriate processing and the encoding specificity principle, the question that then arises is whether these concepts render levels of processing irrelevant or redundant. Is transfer-appropriate processing, when suitably elaborated, sufficient, or is there still a place for concepts such as levels of processing which claim that some forms of processing are more likely than others to afford subsequent retrieval, and, if so, in what sense can such a claim be justified? Moreover, even if this claim is granted, fundamental questions remained unanswered. Roediger and Gallo (2001) put the matter succinctly when they point out that after 30 years of research, we do not know why or how we get the typical levels-of-processing effect.

In addressing this question it is helpful to push the analogy with natural selection one step further. A possible line of argument runs as follows. Certain species are fitter than others in the limited sense that the species could survive under a wider range of future environmental conditions, or survive under conditions that are the most likely to obtain in the future. According to this view, a rodent such as the rabbit might be described as fitter than a marsupial such as the koala. The rabbit has qualities (such as fecundity and dietary flexibility) that would support survival across a wide range of environments whereas the koala has worked itself into an extraordinarily narrow ecological niche. It is possible, of course, to specify environments in which the koala would survive and the rabbit become extinct; but the former are highly specific, intolerant of even slight variation, and unlikely to arise in practice. In betting on survival in an uncertain future environment, smart money will take the rabbit.

The claim, then, is that the typical levels-of-processing effect reflects the greater "survival" value of deep processing for those retrieval conditions most likely to occur. Again, it is possible to arrange retrieval conditions such that shallow processing outperforms deep processing, a point demonstrated by Morris et al. (1977) when they showed that for rhyme-recognition, rhyme encoding yielded superior performance to semantic encoding. Granted the theoretical importance of this result in establishing the validity of the principle of transfer-appropriate processing, the result obtains only under highly specific retrieval conditions of rhyme recognition. It is reasonable to predict that with almost any other retrieval task (such as "standard" recognition) the usual superiority of semantic processing would be observed. In betting on retrieval in an uncertain future retrieval environment, smart money will take deep processing.

ACCOUNTING FOR THE MATCHED-TRANSFER LEVELS EFFECT

However, even if we grant that typical retrieval environments are more likely to favour deep processing, this appeal to the demographics of memory sidesteps more fundamental questions about the relationship between levels of processing and transfer-appropriate processing. In this matter, the typical levels effect, such as the superiority in free recall of a semantic over a non-semantic orienting task, is not the most critical evidence favouring the generally beneficial effects of deep processing. Rather, it is the well-documented finding (e.g., Fisher & Craik, 1977; Morris et al., 1977) that even under conditions aimed at maximising transfer-appropriateness, shallow encoding yields poorer performance than deep encoding. Rhyme recognition following rhyme encoding is poorer than standard recognition following semantic encoding. Because the following discussion will make continued reference to this particular finding it is convenient to label it. It will be referred to as the *matched-transfer* levels effect. More recently, Marmurek (1995) has shown that this effect cannot be attributed to differences in the difficulty of retrieval tasks, thereby clarifying the rather careless use of the term "main effect" by Lockhart and Craik (1990).

Conceptually, it is possible to distinguish two potential sources of the matched-transfer levels effect. The first source consists of characteristics of the memory trace itself, qualities that could be specified independently of any particular retrieval environment. Such qualities might be, for example, distinctiveness, resistance to interference, or even something as traditional as strength. We will refer to this account as the *trace* explanation of the matched-transfer levels effect. On the basis of their experimental findings Moscovitch and Craik (1976) and Fisher and Craik (1977) argued for such an explanation, emphasising the role of trace-distinctiveness.

The second possible influence on the matched-transfer levels effect is transfer-appropriate processing itself. Such a possibility may seem unlikely, given the claim that transfer-appropriateness has been matched. However, despite the nominal equating of the level of transfer-appropriate processing for the two encoding conditions, it is conceivable that semantic processing in conjunction with semantic cueing nevertheless represents a greater degree of transfer appropriateness than the corresponding match for shallower processing such as rhyme. Assuming that processing is never fully transfer-appropriate—that the retrieval environment never achieves a perfect recapitulation of encoding processes—shallow processing may be more vulnerable to a mismatch produced by small variations in the retrieval environment. It presents, as it were, a smaller target for retrieval processing and, to pursue the metaphor, is thus vulnerable to the slightest misdirection. We will refer to this account as the *transfer* explanation of the matched-transfer levels effect. This view does not deny the general advantage of deep processing, but claims levels of processing is a phenomenon that can be subsumed under transfer-appropriate processing; transfer-appropriate processing is the mechanism through which it operates. Levels of processing is simply an indirect way of influencing the degree to which processing is transfer-appropriate.

The transfer and trace accounts of the matched-transfer levels effect are not, of course, mutually exclusive. It is possible that both are operative either additively or interactively. The evidence for the trace account as a contributing factor is quite strong (e.g., Fisher & Craik, 1977; Marmurek, 1995). The contribution of the transfer factor, if any, is less clear and therefore warrants further attention.

THE CONCEPT OF ROBUST ENCODING

According to the transfer account, for an unspecified future retrieval context, levels of processing has its effect by influencing the likelihood that the processing will prove to be transfer-appropriate. Using the language of the encoding specificity principle, another way of expressing this claim is to say that deep processing expands the domain of elements that can interact ecphorically with the retrieval environment. Lockhart and Craik (1990) introduced the term *robust encoding* to capture this concept. Schacter (1996, p.63) expresses a similar idea when he suggests that a major reason why a semantic orienting task typically works better than a non-semantic one might be because the semantic processing yields a trace that is accessible to a broader range of retrieval cues. A robust trace, then, is one more likely to survive variations in the subsequent retrieval context because it presents a broader target to subsequent retrieval processing.

Encoding variability

The term "robust encoding" is in need of clarification, not to mention supporting evidence. Note first that the general idea is by no means new. For example, it will no doubt seem familiar to those who have followed the fortunes of the concept of encoding variability. This idea predates levels of processing (e.g., Madigan, 1969; Martin, 1968; Melton, 1970), and continues to have some currency (e.g., Soraci et al., 1999). Thus one interpretation of robust encoding is that it represents a special form of encoding variability in which deep processing (such as semantic elaboration) has its effect by laying down a greater range of potential retrieval routes. General support for the beneficial effects of encoding variability would therefore seem essential to the plausibility of the robust encoding concept. Two lines of experimental evidence are relevant. First, if robust encoding is a form of encoding variability, then experimentally eliminating encoding variability from processing should eliminate any benefit that deep processing might impart to subsequent retrieval. Second, the experimental enhancement of encoding variability should benefit retrieval. We will briefly consider each of these points before a consideration of more direct evidence on the concept of robust encoding.

The concept of maintenance (or type 1) rehearsal as described by Craik and Lockhart (1972) can be thought of as a form of processing designed to eliminate, or at least minimise, the kind of encoding variability envisaged in the concept of robust encoding. Experimental support for the claim that maintenance rehearsal yields no additional benefit to retrieval would therefore seem essential to the plausibility of robust encoding. On the other hand, although such evidence is necessary, it provides only weak positive support in that it also supports (and is demanded by) virtually any interpretation of the levels-of-processing effect.

The evidence relating to the effects of maintenance rehearsal has been thoroughly reviewed by Greene (1987). The data offer support for the general conclusion that in so far as maintenance rehearsal facilitates retrieval its effects are largely limited to recognition memory. The evidence is quite strong, especially when one takes into account the fact that pure maintenance processing is an ideal that actual experimental conditions can only approximate. Poor approximations to the ideal may well account for the occasional finding that maintenance rehearsal facilitates free recall. Greene attributes the effect of maintenance rehearsal on recognition memory to "self-coding". Naveh-Benjamin and Jonides (1984) discuss a similar explanation in terms of intra-item integration. A more specific version of this interpretation, making use of dual-process theories of recognition, is that maintenance rehearsal has its effects via the non-recollective component of recognition. Such an account meshes nicely with the finding that this non-recollective component is unaffected by level of processing (Gardiner, 1988; Rajaram, 1993) leading to the conclusion that, although maintenance rehearsal may facilitate some aspects of retrieval, its effects are qualitatively distinct from those of levels of processing. Thus the data would seem to leave intact (and provide weak support for) the initial claim that the experimental elimination of encoding variability from processing should remove any benefit that deep processing might impart to subsequent retrieval.

However, the evidence for the effects of increased encoding variability is anything but clear-cut. Despite some success in accounting for the spacing effect in free recall (Glenberg, 1979; Madigan, 1969) and the effect on final recall of different forms of initial retrieval (McDaniel & Masson, 1985) direct evidence is mixed. Postman and Knecht (1983), for example, performed a direct test by comparing the recall of words repeated within the same sentence with words repeated in different sentences. They found no support for the encoding variability hypothesis. McDaniel and Masson (1985) suggest that this failure may be the consequence of an experimental manipulation that produced insufficient variation in encoding. This may be a valid explanation of the Postman and Knecht results, but if so, it does little to help the interpretation of robust encoding as a form of encoding variability. It is surely implausible to claim that the typical semantic orienting task such as judging synonymity yields greater encoding variability than the manipulation used by Postman and Knecht (1983). Moreover, it is possible to make the opposite claim and argue that enhanced encoding variation of the type used by Postman and Knecht adds no further benefit because the semantic analysis achieved in a single presentation is sufficient to produce maximum effects of encoding variability, at least of the kind envisaged by the concept of robust encoding. It would seem that encoding variability is a slippery concept.

A thorough review of the experimental findings surrounding encoding variability shows an intricate pattern of interactions that makes it difficult to speak in general terms about the effects of encoding variability. Experiments reported by McDaniel and Masson (1985), or Soraci et al. (1999) provide a good illustration of this point. Such complexity should not surprise anyone who takes seriously the underlying encoding specificity principle or transfer-appropriate processing. According to these principles, encoding variability is not intrinsically or universally beneficial, but should be effective only in so far as the variability encompasses elements that are recapitulated at retrieval. The effect of variability not so recapitulated would at best be neutral or, by exerting some form of interference or lessening discriminability, the effect could quite possibly be negative. The relevant question with respect to the concept of robust encoding is this: Does deep processing result in form of encoding variability (robust encoding) that is more likely to transfer appropriately to the retrieval conditions typical of free recall and the recollective component of recognition? We turn now to a consideration of experiments that attempt to manipulate directly the relationship between levels of processing and variations in the conditions of retrieval.

Direct tests of robust encoding

A straightforward interpretation of robust encoding leads to a simple prediction. Variations in the retrieval environment should have a smaller effect for items that have been deeply processed. Results reported by Hannon and Craik (2001) support this prediction. They argue that deeper processing makes successful recognition less dependent on the reinstatement of the original encoding context. However, most of the evidence relevant to the prediction contradicts it. Marmurek (1995) points this out with respect to the data from Morris et al. (1977) as well as his own data, and experiments by Fisher and Craik (1977) and Moscovitch and Craik (1976) show exactly the same result. When the retrieval environment is varied by providing either semantic or non-semantic retrieval cues, the effect of this cue variation is much greater for items that were deeply processed. Such results indicate that the concept of robust encoding is either badly mistaken or in need of some very serious refinement. Evidence, as it stands, would seem to favour some version of the trace account, but clearly the issue needs more research.

Refinement of the concept of robust encoding may well be possible, but it is beyond the scope of this paper to attempt such a task. However, a few points can be made. First, there is no reason to suppose that the conjectured robustness resulting from deep encoding should extend beyond the semantic domain. It seems more appropriate to consider it a range of semantic elaboration that can be taken advantage of by semantic cueing. If this is true, then the robustness associated with deep encoding should not enhance transfer-appropriateness in relation to a non-semantic cue. It is to be expected, therefore, that despite semantic encoding, retrieval will be poor under conditions of non-semantic cueing, the result obtained by Morris et al. (1977) and Fisher and Craik (1977). Given the relatively high level of performance of semantic cueing following semantic encoding, a large difference between the two cueing conditions is inevitable and, it could be argued, irrelevant to any plausible realisation of robust encoding. Defenders of robust encoding might further point out that the low level of performance for rhyme encoding, even when the task is rhyme recognition, allows little room for downward variation for other sub-optimal cueing conditions. In brief, such experiments constitute too blunt an instrument to provide a critical test.

CONCLUSION

The challenge facing the concept of robust encoding (or any other transfer-based concept) as an account of the matched-transfer levels effect is to show that there exist more sharply focused cueing conditions that can elevate performance for non-semantic encoding. That challenge seems formidable. A more promising alternative is to qualify the sharp distinction made between trace- and transfer-based accounts. Whereas this distinction can be drawn at a conceptual level, empirically the two may be largely inseparable. That is to say, the features of the memory trace envisaged by trace accounts of deep processing may be precisely those that, at the same time, impart the robustness envisaged by transfer accounts. If this is so, then no refinement of cueing conditions alone will elevate performance for non-semantic encoding to that of semantic encoding paired with semantic cueing. The only way in which to increase transfer-appropriateness would be tantamount to introducing greater depth of processing. This interactive interpretation of the relationship between the trace and transfer accounts is entirely compatible with that given by Fisher and Craik (1977) in their general discussion.

These speculations clearly need to be spelled out in ways that make them amenable to experimental investigation. However, the questions raised by concepts such as robust encoding are central to any thorough understanding of human memory. The fundamental issues remain what they have been for 30 years: the nature of the interactive relationship between encoding and retrieval processes. Broad concepts such as levels of processing and transfer-appropriate processing have generated a large body of research that has greatly increased our understanding, although much work remains to be done.

REFERENCES

Challis, B.H., Velichkovsky, B.M., & Craik, F.I.M. (1996). Levels-of-processing effects on a variety of memory tasks: New findings and theoretical implications. *Consciousness & Cognition: An International Journal*, *5*, 142–164.

Craik, F.I.M., & Lockhart, R.S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*, 671–684.

Fisher, R.P., & Craik, F.I.M. (1977). Interaction between encoding and retrieval operations in cued recall. *Journal of Experimental Psychology: Human Learning and Memory*, *3*, 701–711.

Gardiner, J.M. (1988). Functional aspects of recollective experience. *Memory & Cognition*, *16*, 309–313.

Glenberg, A.M. (1979). Component-levels theory of the effects of spacing of repetitions on recall and recognition. *Memory & Cognition*, *7*, 95–112.

Gould, S.J. (1977). Darwin's untimely burial. In *Ever Since Darwin* (pp. 39–45). New York: Norton.

Greene, R.L. (1987). Effects of maintenance rehearsal on human memory. *Psychological Bulletin, 102*, 403–413.

Hannon, B., & Craik, F.I.M (2001). Encoding specificity revisited: The role of semantics. *Canadian Journal of Experimental Psychology, 55*, 231–243.

Lockhart, R.S., & Craik, F.I.M. (1978). Levels of processing: A reply to Eysenck. *British Journal of Psychology, 69*, 171–175.

Lockhart, R.S., & Craik, F.I.M. (1990). Levels of processing: A retrospective analysis of a framework for memory research. *Canadian Journal of Psychology, 44*, 87–112.

Madigan, S.A. (1969). Intraserial repetition and coding processes in free recall. *Journal of Verbal Learning and Verbal Behavior, 8*, 828–835.

Marmurek, H.H. (1995). Encoding, retrieval, main effects and interactions: were Lockhart and Craik (1990) on the level? *Canadian Journal of Experimental Psychology, 49*, 174–192.

Martin, E. (1968). Stimulus meaningfulness and paired-associate transfer: An encoding variability hypothesis *Psychological Review, 75*, 421–441.

McDaniel, M.A., & Masson, M.E.J. (1985). Altering memory representations through retrieval. *Journal of Experimental Psychology: Human Learning and Memory, 11*, 371–385.

Melton, A.W. (1970) The situation with respect to the spacing of repetitions and memory. *Journal of Verbal Learning & Verbal Behavior, 9*, 596–606.

Morris, C.D., Bransford, J.D., & Franks, J.J. (1977). Levels of processing versus transfer-appropriate processing. *Journal of Verbal Learning and Verbal Behavior, 16*, 519–533.

Moscovitch, M., & Craik, F.I.M. (1976). Depth of processing, retrieval cues, and uniqueness of encoding as factors in recall. *Journal of Verbal Learning and Verbal Behavior, 15*, 447–458.

Naveh-Benjamin, M., & Jonides, J. (1984). Maintenance rehearsal: A two-component analysis. *Journal of Experimental Psychology: Human Learning and Memory, 10*, 369–385.

Nelson, T.O. (1977). Repetition and levels of processing. *Journal of Verbal Learning & Verbal Behavior, 16*, 151–171.

Postman, L., & Knecht, K. (1983). Encoding variability and retention. *Journal of Verbal Learning & Verbal Behavior, 22*, 133–152.

Rajaram, S. (1993). Remembering and knowing: Two means of access to the personal past. *Memory & Cognition, 21*, 89–102.

Roediger, H.L.III, & Gallo, D.A. (2001). Levels of processing: Some unanswered questions. In M. Naveh-Benjamin, M. Moscovitch, & H.L. Roediger III (Eds.), *Perspectives on human memory and cognitive aging: Essays in honour of Fergus Craik* (pp. 28–47). Philadelphia, PA: Psychology Press.

Roediger, H.L.III, Weldon, M.S., Stadler, M.L., & Riegler, G.L. (1992). Direct comparison of two implicit memory tests: Word fragment and word stem completion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 1251–1269.

Schacter, D.L. (1996). *Searching for memory*. New York: Basic Books.

Soraci, S.A., Carlin, M.T., Chechile, R.A., Franks, J.J., Wills, T., & Watanabe, T. (1999). Encoding variability and cuing in generative processing. *Journal of Memory & Language, 41*, 541–559.

Stamos, J. (1996). Popper, falsifiability, and evolutionary biology. *Biology and Philosophy, 11*, 161–191.

Tulving, E., & Thomson, D.M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80*, 359–380.