Article:

Baddeley, Alan David orcid.org/0000-0002-9163-0643 and Hitch, Graham James (2017) Is the Levels of Processing effect language-limited? Journal of Memory and Language. pp. 1-13. ISSN 0749-596X

Is the levels of processing effect language-limited?

Alan D. Baddeley and Graham J. Hitch

University of York

Correspondence concerning this article should be addressed to A Baddeley, Department of Psychology, University of York, YO10 5DD, United Kingdom. Email: ab50@york.ac.uk

Abstract

The concept of Levels of Processing (LOP), proposing that deep coding enhances retention, has played a central role in the study of episodic memory. Evidence has however been based almost entirely on retention of individual words. Across five experiments, we compare LOP effects between visual and verbal stimuli, using judgments of pleasantness as a method of inducing deep encoding and a range of shallow encoding judgments selected so as to be applicable to both verbal and visual stimuli. LOP effects were consistent but modest across the visual stimuli (mean effect size 0.5). In contrast, LOP effects for verbal stimuli varied widely, from modest for people's names and unfamiliar animals (mean effect size 0.6) to large for familiar animals and household items (mean effect size 1.4), typical of the dramatic LOP effects that characterize the existing verbal literature. We interpret our data through the Gibsonian concept of "affordance", proposing that visual and verbal stimuli vary in the number and richness of features they afford, and that access to such features will in turn depend on encoding strategy. Our hypothesis links readily with Nairne's feature model of long-term memory.

Keywords: levels of processing; long-term memory; affordance; visual memory; verbal memory; encoding

Craik and Lockhart (1972) proposed that memory is a by-product of processing, the deeper the processing the better the retention. Their paper is one of the most highly cited in the history of cognitive psychology (Roediger & Gallo, 2001), and "one of the most influential systematic conceptual frameworks within which problems of memory can be raised and investigated" (Tulving, 2001, p24). While the assumption of a series of levels leading from perceptual to semantic was subsequently abandoned (Craik & Tulving, 1975), Levels of Processing (LOP) has continued to serve as a broad theoretical framework, accounting for a wide range of data within the field of human memory and potentially providing a fruitful basis for further investigation (Conway, 2002). Furthermore, the principle underlying the levels approach is of considerable practical relevance, providing an important and valuable means of improving learning, in contrast to the common tendency for learners to rely on rote rehearsal.

On the other hand, despite many replications and the magnitude of the effects shown (a series of studies by Hyde and Jenkins (1969) and Walsh and Jenkins (1973) yielded an average effect size based on Cohen's d of 2.27), the use of the framework to broaden our knowledge of human memory has been somewhat limited. One exception to this comparative lack of development comes from the demonstration by Tulving and Thomson (1973) of the importance of the match between encoding and retrieval in determining memory performance. This point was further developed with the introduction of the concept of Transfer Appropriate Processing (TAP), as proposed by Morris, Bransford and Franks (1977). They showed that shallow phonological coding led to better performance than deeper semantic coding when rhyming words were used as retrieval cues for the items to be recalled, again demonstrating that memory performance depends crucially on conditions at retrieval, as well as at encoding. The concept of TAP is an important reminder that retrieval needs to be considered, but leaves open the question of how to determine transfer appropriateness.

In an attempt to develop the concept of TAP, Roediger (Roediger & Blaxton, 1987; Roediger, Weldon & Challis, 1989) proposed to link it to the distinction between explicit episodic memory and more automatic implicit memory. Most explicit memory tasks involve processing in terms of meaning, hence benefiting from deeper encoding while implicit tasks tend to be perceptually based, depending more on the exact replication of shallower encoding cues. However, although there were many examples in the literature that fitted this pattern, it is not always possible to make a clear distinction between perceptual data-driven levels of analysis and analysis at a more conceptual or semantic level. Roediger, Srinivas and Weldon (1989) proposed that any given situation could have components involving both levels of analysis which might or might not trade off against each other. While plausible, this compounds the problem of measuring transfer appropriateness. Furthermore, data began to appear suggesting that dissociations occurred within the proposed perceptual and conceptual paradigms (Hunt & Toth, 1990) presenting further difficulties in using TAP as a way of developing the original LOP approach, and leading Roediger (2002, p321) to conclude "we suggest that the field in general has not yet been able to develop an adequate characterization of procedures that account for memory phenomena despite efforts in this direction".

One important question to be asked of any theoretical framework concerns its breadth of application. As Roediger and Gallo (2001 p42) observe, LOP can be regarded as "a special case of transfer-appropriate processing that applies to memory for words in meaning-based tests". However, although language is clearly important, it is only part of our capacity to experience and remember the world, suggesting a need for LOP studies of non-verbal memory. We describe a series of experiments that began with the question of whether reliable LOP effects could be demonstrated using visual material. As relatively little is known we adopted an exploratory approach of comparing LOP effects for a range of visual and verbal materials. Our results show that different types of visual materials all yield modest LOP effects whereas verbal materials give a wider range such that the dramatic advantage to deep encoding typically found depends crucially on the nature of the material. These findings led us to propose a modified explanation of LOP effects that takes into account the "affordances" of a stimulus (Gibson, 1977) and applies to both verbal and non-verbal material.

An early critique of the LOP concept (Baddeley 1978) noted the lack of evidence for LOP effects using visual stimuli. Although subsequent research on LOP has also been dominated by use of verbal stimuli, a number of studies have been performed across a range of other modalities, though largely using implicit memory measures for which LOP effects were, unsurprisingly found not to apply (Graf & Mandler, 1984; Jacoby & Dallas, 1981). There appears to be very little investigation of the LOP effect in studies of explicit episodic memory using nonverbal stimuli. Some exceptions to this generalization do however occur.

In the case of music, Halpern and Bartlett (2010) comment on a paucity of LOP studies in the literature, reporting only one positive result. Peretz, Gaudreau and Bonnel (1998), found that judgments of the familiarity of a tune led to better subsequent recognition than judging the instrument playing the tune, commenting however that " the current authors failed to find LOP effects for unfamiliar music on numerous occasions (some published, some languishing in bottom drawers)" (Halpern & Bartlett, 2010 p 234).

Attempts have also been made to study LOP effects in olfactory memory. Lyman and McDaniel (1986) varied encoding instructions in a study involving recognition of 30 odors after a one week delay. No difference in hit rate was found, but an advantage on a d' measure suggested that attempting to name and define each odor or linking it to a life episode led to better performance than forming a visual image or simply trying to memorize each stimulus. A subsequent replication by Zucco (2003) again found a significant effect for d' but not hit rate, with only the life episode condition showing a significant advantage. These results suggest a modest overall effect of deeper processing, operating mainly through reducing false alarm rate, far from the robust effects typical of verbal material.

There have been rather more attempts to detect LOP effects in visual memory, reflected largely in studies of memory for faces. Warrington and Ackroyd (1975) report better face recognition following pleasantness judgments than from estimation of the person's height, a somewhat challenging task from a portrait photograph. A much easier "shallow" task was used by Bower and Karlin (1974), judging the sex of the person portrayed. This proved less effective in facilitating subsequent recognition than did judgments of likeableness or honesty. This could however simply reflect the need to scan the face more intently in order to make these "deeper" judgments, as proposed by Winograd (1981) who found that an instruction to identify the most distinctive facial feature of a given face was more effective than the apparently deeper task of making a personality judgment. On the other hand, a study by Patterson and Baddeley (1977) which compared categorization on physical dimensions such as nose size and thickness of lips found these to be slightly less effective than judgments of pleasantness or intelligence. An attempt to increase depth of processing by providing a semantic context for each face by adding a description of the unfamiliar person's occupation, background and habits however, proved ineffectual (Baddeley, 1982; Baddeley & Woodhead, 1982). An attempt to maximize TAP by presenting the contextual information at both encoding and recognition did increase rate of detection, but this proved to be entirely attributable to inducing a positive response bias (Baddeley & Woodhead, 1982), with participants also more likely to erroneously say yes to a novel face, if accompanied by a previously presented description. Once again therefore, although it would be unwise to rule out the possibility of an LOP effect for faces, any such effects are clearly far weaker than those routinely found for verbal materials.

It could be argued of course, that despite the obvious importance of faces, they are a rather special form of visual stimulus, with their own specific anatomical processing area (Kanwisher, McDermott & Chun, 1997), possibly also associated with a relatively automatic link to emotional coding (Öhman, 2009). For that reason, it is important to extend the study of LOP effects in visual memory to other stimuli. Unfortunately, the small number of studies that have attempted this previously have used different methods and given contradictory results with D'Agostino, O'Neill and Paivio, (1977) finding a positive effect using readily nameable line drawings while Intraub and Nicklos (1985), found a negative effect for some of their cued recall conditions, suggesting the need for a more systematic approach.

The starting point for our investigation was the observation that any study of the role of LOP must deal with three variables, the nature of the initial encoding, deep versus shallow, the nature of the retrieval test, bearing in the mind the importance of TAP, and the characteristics of the material to be remembered. Neither the method of encoding nor the range of materials involves a simple binary choice, hence the range of possible experiments becomes very large indeed. For that reason we fixed our deep encoding method, basing it on judgments of pleasantness, and always used a four-alternative forced-choice recognition retrieval measure. Holding constant the method of ensuring deep encoding and the testing procedure then allowed us to manipulate the variable central to our enquiry, the nature of the material, allowing comparison between visual and verbal memory, and importantly, of variations in material within each modality. This approach raises a number of further issues which will be discussed next.

The first concerns our selection of judgments of pleasantness as our deep encoding procedure. We did this because we needed a semantic judgment that is readily applicable to a wide range of materials. In his attempt to develop a measure of meaning that extended beyond verbal material Osgood developed a complex rating scale, the semantic differential which factor analysis suggested yielded three factors of which the strongest was consistently the hedonically evaluative good-bad dimension (Osgood, May & Miron, 1975). In the case of words, encoding on this dimension has been shown to produce a particularly powerful LOP effect (Hyde & Jenkins, 1969) and indeed Packman and Battig (1978) found that judgments of pleasantness were substantially more effective than other "deep" judgments such as concreteness or meaningfulness. Furthermore, the widespread use of pleasantness judgments in clinical assessments such as the Warrington (1984) recognition test involving words and faces, reflects the fact that it is a task that participants find natural and relatively easy to use for both verbal and visual stimuli.

Choosing shallow encoding tasks is less straightforward given that they need to be applicable to both visual and verbal material and to ensure that participants process the stimuli at the required level. Finally, to avoid the risk of basing our conclusions on a single atypical task, we use a range of different "shallow" processing instructions. Our earlier research concerned with developing a clinical test of visual memory opted to use door scenes as they are familiar, allowing a range of degrees of similarity and resulting difficulty. Two lists of 12 doors tested using four-alternative forced choice proved both sensitive to memory deficit and patient friendly (Baddeley et al., 1994).

Photographing doors subsequently proved addictive to A.B., resulting in a data base of over 2000 visual stimuli. In order to increase their experimental usability we classified each item along a range of dimensions, thus making it relatively easy to select sets of differing levels of inter-item similarity (Baddeley, Hitch, Quinlan, Bowes & Stone, in press). In addition to our having a very large readily available set, doors have the advantage that, unlike faces they almost certainly do not have a specific brain area devoted to their processing and are unlikely to have atypically strong links to emotional and social processing (Öhman, 2009).

Having established that a LOP effect can be obtained using door scenes in pilot work, we continued to include door stimuli as a baseline against which other types of visual and verbal stimulus material could be compared. This led to the question of what other type of material. In this essentially exploratory study, rather than setting up and testing precise hypotheses, we used pragmatic constraints to select our material. We opted for lists of 24-30 items per condition, choosing four-choice recognition rather than two-alternative or yes/no recognition reduced baseline guessing to obviate the need for longer lists. We wanted to maintain certain characteristics of our doors test, namely that the items should come from a single broad semantic category, and that there should be sufficient similarity between items to allow a level of recognition approximately equal across materials. It is worth noting at this point that simply selecting visual recognition items from a wide range of categories, with distractors chosen at random tends to lead to levels of performance of 90% or more, even with very long lists (Brady, Konkle, Alvarez & Oliva, 2008; Konkle, Brady, Alvarez & Oliva, 2010a; Nickerson, 1965; Standing, Conezio & Haber, 1970). The experiments that follow reflect these constraints.

Experiment 1 therefore compares recognition memory for door scenes or concrete words processed either “deeply” in terms of pleasantness, or more shallowly in terms of stimulus color. Experiment 2 attempts to replicate this with different sets of stimuli and a different shallow processing task, while Experiments 3a, b and c explore the generality of our initial findings by extending them to a broader range of visual and verbal materials, using the method of converging operations to determine which aspects of the material are crucial

Experiment 1

Design and Procedure

A 2 x 2 within participants design combined two types of material, doors and words and two types of encoding instruction involving judgments of pleasantness and color. All participants were tested on each of the four conditions in counterbalanced order. A total of 20 student volunteers were tested.[1] They and all participants in the remaining studies were University of York students who gave informed consent and were rewarded either by course credit or a small honorarium.

Material comprised two lists of 24 doors selected to have a range of colors, with half the door scenes being predominantly brown. There were also two lists of 24 words, presented in Courier New 18pt font. Half the words contained a majority of brown letters with the rightmost n letters displayed in PC color "saddle brown". The remaining letters were chosen at random from "red, white, blue and green", all being the same color with a minority of brown letters. In all conditions the background screen was silver and other textual materials were presented in black. In the deep conditions, participants judged whether they found each door or word pleasant or unpleasant. In the shallow condition they judged whether the stimulus was predominantly brown.

Each presented stimulus was tested using a 4AFC recognition procedure in which a previously presented item occurred together with three distractors arranged in a square formation. In the case of the doors, these were again selected so as to be from the same category, for example all church doors or garage doors. The verbal stimuli comprised one list of animals (e.g. lion, leopard, puma, tiger), and one of household objects (e.g. plate, cup, saucer, mug), the reason being that these stimulus categories did not have sufficient similar distractors to create two lists of 24 items with three broadly similar foils.

Figure 1 shows monochrome examples of the door scenes used.

Figure 1 about here

Each learning trial involved a central fixation point presented for 1s followed by a 1s stimulus presentation after which two response choices were displayed for 2 s, involving computer keys 1 and 2. In the shallow conditions "brown-1, not brown-2" was displayed, and in the deep conditions "pleasant-1, unpleasant-2". Once a response was made the next trial fixation cross was presented. In this and all subsequent experiments, each list was followed by a four alternative recognition test in which each of the target items was presented together with three "new" foils. Each test array was preceded by a fixation cross, presented for 500ms and followed immediately by the four test stimuli, one in each of the screen's quadrants. Target location within each test set was random. Participants were told that one of the four had been shown previously and to report the "old" item by pressing one of four response keys on a computer keyboard in which the keys corresponded to the spatial layout to the stimuli on the screen (keys 1, 3, 7 and 9). The targets and distractors were always matched in predominant color.

Each of the four conditions began with a practice block comprising four study items followed by four test items, leading to the experimental block which involved 24 study items immediately followed by 24 test trials. The order of the four conditions was balanced across participants using a Latin square.

Results

Figure 2 shows the mean probabilities of correct responses across the four conditions. A 2 x 2 repeated measures ANOVA that combined stimulus type (words versus picture) and LOP (shallow versus deep) showed main effects of stimulus type (F (1,19) = 47.57, $p < .001$, $\eta^2 = .72$) and LOP (F (1,19) = 29.38, $p < .001$, $\eta^2 = .61$) together with a stimulus type x LOP interaction (F (1,19) = 12.06, $p < .01$, $\eta^2 = .39$). Further analysis of the interaction showed a significant effect of LOP for both types of stimuli (words: $t(19) = 5.77$, $p<.01$, $d = 1.29$; doors: $t(19) = 2.88$, $p<.01$, $d = 0.64$) although the size of the effect was significantly larger for the word stimuli than for pictures ($t(19) = 3.47$, $p < .01$).

Figure 2 about here

Discussion

Experiment 1succeeds in showing a positive effect of LOP for our visual stimuli. It is however significantly less than that found for words where the magnitude of the LOP effect is almost certainly constrained by a ceiling effect with many of the participants scoring at or near the maximum of 24 in the deep processing condition. A second problem concerns the tendency for our door stimuli to be harder than the verbal material even under shallow encoding conditions. This difference occurred despite our attempt to make the three distractor words as similar as possible to the target. Category size limitations, even using the two large semantic categories, of animals and household objects, proved insufficient to keep performance down to an equivalent level, probably because of the difficulty in finding enough similar distractors. This then leaves open the possibility that the interaction between material and encoding strategy stems not from material per se, but from the difference between high and low performance. Such a result would be interesting in placing constraints on the generality of the LOP effect, but would itself require further explanation. Fortunately, verbal materials matched in difficulty to the doors test were already available from the Doors and People Test (Baddeley et al., 1994) based on people's names, where level of difficulty can readily be manipulated by varying distinctiveness. Hence, an easy item such as Graham Cammiss with distractors Graham Clayton, Graham Crosby and Graham Coles can be contrasted with a more difficult item such as John Robertson with distractors John Robinson, John Roberson, and John Robson. Names have the advantage that although obviously verbal, they tend not to have useful semantic associations, hence Mrs Black is no more likely to have dark hair than Mrs White, or indeed Mrs Green. Finally, a third condition was added, namely occupations, providing a large category of semantically rich items, where difficulty can be introduced by making relatively fine distinctions between the target and distractors as in the case for example of the target television director against distractors television engineer, television presenter, television producer. We expected this category to give a level of performance somewhat higher than names but lower than our previous categories of animals and household objects. Finally, we increased list length from 24 to 30 in order to reduce the risk of ceiling effects.

Experiment 2

Design

A total of 24 participants, were each tested on three types of stimuli, doors, names and occupations, in each case processed at two levels, shallow and deep. Each encoded list was followed by an immediate four alternative forced-choice test. All participants completed all six conditions, half beginning with the shallow condition and half with deep. The order of stimulus presentation within each encoding condition was counterbalanced using a 3x3 Latin square. Half began with the three deep encoding instructions and half with the shallow. Order of material type was the same for each encoding depth, hence if the participant began with deep coding of doors, they would also begin with the deep encoding of names and occupations.

Materials

Two sets of 30 target doors were prepared, each followed by 30 test sets comprising the target and three distractors. Two sets of 30 names were generated using the Doors and People Test, supplemented by the local telephone directory, but avoiding names of famous people. Each name was presented with both a given and family name, with the given name remaining constant across the recognition set. Two sets of 30 occupations were constructed. In order to increase level of difficulty and avoid ceiling effects, targets and distractors were typically selected to be similar for example the target <u>dressmaker</u> was tested against distractors <u>tailor</u>, <u>seamstress</u>, and <u>embroiderer</u>.

The doors were presented in full color and the verbal stimuli were broadly equivalently colored <u>red</u>, <u>green</u>, <u>brown</u>, <u>blue</u> or <u>black</u>. In the shallow condition, participants were asked to name the color of the door or word. When tested, both the target and distractors were in the original target color. In the deep condition, participants were required to make a verbal judgment of pleasant or unpleasant for each of the items.

Procedure

Stimuli were presented at a rate one second per item, with a 1.5 second interval during which participants were required to make their judgments. They were informed that they would subsequently be required to remember the items using a four-alternative forced choice procedure. Each 30 item list was preceded by a six item practice list using the relevant encoding procedure.

Each memory block was followed by an immediate test in which 30 sets of four items were presented in a square array of four locations labeled A to D. Participants responded by writing the relevant letter on an answer sheet. They were allowed a maximum of eight seconds to respond but were encouraged to do so at their own pace. Location of the target was randomized across the locations. After the first three lists, there followed a three minute interval during which they completed a maze distractor task, before going on to the alternative encoding procedure.

Results

Mean performance across the three material types and two levels of processing is shown in Figure 3. A 3 x 2 ANOVA was performed involving material (doors, names and occupations), and LOP (deep and shallow). It showed a main effect of material type $F(2,46) = 16.14$, $p < .001$, $\eta_p^2 = .41$, together with a clear effect of processing level $F(1,23) = 38.53$, $p < .001$, $\eta_p^2 = .63$, with an overall advantage to deep processing, together with a significant interaction between material and level $F(2,46) = 8.62$, $p = .001$, $\eta_p^2 = .27$. A series of t-tests showed a significant LOP effect for each type of material, with the greatest being for occupations $t(23) = 5.13$, $p < .001$, $d = 1.05$, a somewhat smaller effect for names $t(23) =$

3.46, $p = .002$, $d = 0.71$ and a smaller but significant effect for doors, $t(23) = 2.60$, $p = .016$, $d = 0.53$.

Figure 3 about here

Further 2 x 2 ANOVAs were performed, using the interaction as an indicator of the magnitude of the levels effect across types of material. Analysis of occupation and names, the two verbal conditions, showed a significant interaction, $F(1,23) = 8.15$, $p = .009$, $\eta_p^2 = .26$, indicating a more pronounced levels effect for occupations. When occupations and doors were compared in a 2 x 2 ANOVA, a somewhat more substantial interaction resulted $F(1,23) = 14.27$, $p = .001$, $\eta_p^2 = .38$. However, when names and doors were compared, the interaction failed to reach significance $F(1,23) = 1.19$, $p = .287$, $\eta_p^2 = .05$.

Discussion

We can draw three clear conclusions from our results. First, we replicate the impact of depth of processing on visual stimuli, using a different set of doors. Second, we were successful in finding a verbal task that was at least as difficult as our visual condition by choosing names, finding a LOP effect comparable to that with doors. A significantly larger effect of levels of processing was found in the case of occupations where overall level of performance was equivalent to that of the door stimuli. We can therefore rule out a simple association between overall level of performance and impact of LOP. This pattern of results thus appears to rule out two simple hypotheses, namely that LOP effects are always more substantial for verbal than visual material, and that the magnitude of the levels effect simply reflects the overall level of performance. Importantly, they also point to differences within verbal memory of susceptibility to the effect of LOP.

While it is tempting to propose specific explanations at this point, it would clearly be unwise to do so on the basis of only a limited range of materials and shallow encoding operations. Our next step therefore was to test the generality of our findings in a three part study, each involving three different types of material and two levels of processing. In each case we include a baseline condition involving our original doors task, together with one visual and one verbal condition. In each case, the visual and verbal stimuli were determined by the practical constraints of ensuring a sufficiently large sample of material, in which degree of similarity allowed broadly comparable levels of overall performance. The design in each of the three was broadly similar to that in Experiment 2, with lists comprising 30 items tested by 4AFC. In each case deep processing involved judgments of pleasantness, while the shallow processing depended on the nature of the stimulus material, but was always the same for all three types of material.

Experiment 3a

The materials selected were as follows:

1) The 240 doors used in Experiment 2
2) A total of 240 clocks. These were selected from the internet using Google Search under five subcategories: circle clocks, square clocks, pendulum clocks, alarm clocks, and street clocks. All words on the pictures were removed using Adobe Photoshop CS 2.256.
3) A total of 240 verbal items came from food menus, again selected from the internet using search terms: Chinese food menu, English food menu, Japanese food menu, Dessert menu and Drinks menu. All food names consisted of either three or four words (e.g. Cheddar and tomato soup, Pea and bacon soup, Creamy sweet corn soup, Hot and sour soup) and were displayed in Times New Roman in font size 40.

All stimuli were presented in central vision on a 14 inch flat screen monitor. Within each set, 240 items were used in the test and 16 during the practice session. For each type of material, two lists of 30 target items were selected, together with three distractors, selected using target-distractor similarity to avoid ceiling and floor effects and to balance lists in approximate level of difficulty.

Items were presented at encoding for one second, with a further two seconds allowed for a verbal judgment. In each case, the deep judgment was pleasant or unpleasant, while the shallow involved judged size. In the case of doors participants judged whether each item was larger than a typical domestic door. For clocks, the judgment was larger or smaller than a domestic mantelshelf clock, and in the case of food items whether the name comprised more than three words. Stimuli were exposed for 1s, with a further 2s allowed for a verbal response. Recognition was self paced, within an 8s maximum.

A total of 48 participants (15 males), were tested. All were student volunteers from the University of York reporting normal color vision and reading competence. Again a within-subject design was used with appropriate counterbalancing of order and materials.

Results

These are shown in Figure 4. A 3 x 2 ANOVA revealed a significant main effect of LOP on performance, $F(1,47) = 100.47$, $p < .001$, $\eta_p^2 = .68$, with deeper processing leading to better performance. There was however no overall effect of stimulus type, $F(1,72) = 1.7$, $p = .192$, $\eta_p^2 = .01$, indicating that the attempt during piloting to equate overall level of difficulty had been successful. There was however a significant interaction between LOP and type of stimulus material $F(2,94) = 23.47$, $p < .001$, $\eta_p^2 = .33$. The LOP effect was significant and large for food names, $t(47) = 10.59$, $p < .001$, $d = 1.52$, was significant for clocks $t(47) = 6.1$, $p < .001$, $d = 0.88$, and was significant but much smaller for doors $t(47) = 1.87$, $p < .05$, $d = 0.27$.

Figure 4 about here

A series of 2 x 2 ANOVAs involving material and level of processing were performed in order to compare the processing effect across conditions. The relevant interaction proved significant when comparing menu names with clocks, $F(1,47) = 31.93$, $p < .001$, $\eta_p^2 = .41$, and with doors $F(1,47) = 34.77$, $p < .001$, $\eta_p^2 = .43$. However the interaction failed to reach significance when doors were compared with clocks $F(1,47) = 2.4$, $p = .128$, $\eta_p^2 = .05$.

Discussion

Once again deeper processing enhances memory for doors but the effect is relatively small. Clocks give a similarly modest LOP effect; however, the most substantial effect is for the verbal material of food names. In this case we have succeeded in matching overall levels of performance, again ruling out a simple association between difficulty and the impact of deeper processing. Further discussion will be postponed until after Experiments 3b and c.

Experiment 3b

The overall design is identical to 3a with the exception that different materials were used, and different shallow processing judgments required. In this case the shallow judgment was to report the dominant color of each stimulus.

Materials Again the same 30 doors were used followed by 30 test items each comprising one target and 3 distractors. The second visual set comprised 240 pictures of mobile phones together with a further 8 items used for practice. These were downloaded from the web using four categories (flip, straightboard, touch-monitor and slide) in colors of white, pink, red, silver, green, blue, purple and black from Google and Yahoo using search terms "Japanese mobile phone images" and "hand phone images". Any words, brand names or other non-pictorial cues were removed to minimize the possibility of verbal coding.

The verbal stimuli comprised a total of 240 names of London Underground stations, (e.g. Waterloo, Watford, Wansted, Warren Street), again sourced from Google together with two invented names. They were printed with first letters in uppercase in colors green, red, black, purple, blue or pink. For each condition, sets of 4 items were selected so as to be broadly similar to the target items, again located at random within the test sets. Since piloting suggested these sets of material might be somewhat more difficult than earlier stimuli, exposure time was increased from 1s to 2s, allowing 2s for the judgment. Again, recognition was unpaced with the exception that a maximum of 8 seconds was allowed, with participants marking the location of the target on a response sheet. A total of 36 undergraduate students, 18 male, from the University of York served as unpaid participants. All reported an absence of color blindness.

Results

These are shown in Figure 5 and were again analyzed using a 3 x 2 ANOVA. This showed significant main effects of stimulus material $F(2,70) = 26.88$, $p < .001$, $\eta_p^2 = .43$, and of depth of processing $F(1,35) = 39.16$, $p < .001$, $\eta_p^2 = .53$. There was also a significant interaction between material and depth of processing $F(2, 70) = 3.64$, $p < .05$, $\eta_p^2 = .09$ suggesting that the material differed in susceptibility to LOP. There was however a significant depth effect for all three types of material, for underground station names, $t(35) = 6.29$, $p<.001$, $d = 1.05$, for mobile phones, $t(35) = 3.47$, $p<.001$, $d = 0.58$ and for doors, $t(35) = 2.53$, $p<.01$, $d = 0.42$.

Figure 5 about here

Once again the magnitude of the levels effect was compared between conditions using a series of 2 x 2 ANOVAs and testing for an interaction. The interaction proved significant when comparing phones and underground station names $F(1,35) = 7.54$, $p = .009$, $\eta_p^2 = .18$ suggesting a greater levels effect for station names. When station names were compared to doors the interaction was marginally significant, $F(1,35) = 4.18$, $p = .049$, $\eta_p^2 = .11$. However no significant interaction occurred when the two visual stimulus sets were compared $F(1,35) = 0.27$, $p = .871$, $\eta_p^2 = .00$.

Discussion

We again find modest but significant effects of LOP using visual stimuli, but once again the effects are smaller than found with words, despite the fact that the verbal material almost certainly contained many names that were unfamiliar to the participants, very few of whom would be familiar with the London area. We postpone discussion until describing our final study, which again samples a different set of material.

Experiment 3c

This used the same overall design as 3a and 3b, using a pleasantness judgment for deep processing, this time compared with a shallow judgment of whether the stimulus had one dominant color or was multi-colored. Three types of material were used, one comprising the same 30 doors, a second stimulus set comprising scenes as used by Konkle, Brady, Alvarez and Oliva (2010b) at http://cvcl.mit.edu/MM/sceneCategories.html (Appendix). We used 240 scene images from 10 categories (streams, libraries, snowy mountains, deserts, dining rooms, restaurants, forests, icebergs, bars, and canyons). Images with people or atypical distinguishing features in the background were excluded. A total of 30 images were selected as stimuli and the remainder used as distractors, with distractors chosen to be from the same category, for example four desert scenes or four mountain scenes. A further 16 scenes were used during the practice session.

The verbal stimuli comprised animal names collected using a Google search engine under categories marine mammals, land mammals, reptiles, amphibians and insects. They were arranged into sets with the distractors broadly similar to the targets. In order to avoid the very high level of performance that occurs when familiar animals are used as in Experiment 1, subcategories were allowed for example, striped owl, spotted owl, snowy owl, barn owl or field ants, carpenter ants, fire ants, leaf-cutter ants. These were displayed using Times New Roman font size 44, presented either in black ink or using two or more colors for different letters. When tested, the target and distractors were reproduced using the same door or letter colors.

Again the pleasant/unpleasant judgment was used for deep processing while the shallow condition involved judgment of whether the stimulus had one predominant overall color or comprised multiple colors. The encoding phase involved 2s stimulus presentation followed by a 2s interval during which a verbal encoding decision was to be made. Again the relevant variables were counterbalanced using a Latin Square. A total of 48 undergraduate students, 26 male, were tested, all reporting normal color vision.

Results

Figure 6 shows the mean level of performance across the six conditions. A 3 x 2 ANOVA found no effect of material, suggesting that levels had been successfully matched. There was a significant effect of LOP $F(1,47) = 24.49$, $p < .001$, $\eta_p^2 = .34$, but no significant interaction between stimulus material and processing level $F(2,94) = 0.96$, $p = .386$, $\eta_p^2 = .02$. Further analysis showed significant LOP effects for doors, $t(47) = 3.33$, $p < .001$, $d = 0.48$, for scenes, $t(47) = 3.45$, $p < .001$, $d = 0.50$, and for words $t(47) = 3.52$, $p < .001$, $d = 0.51$.

Figure 6 about here

Once again we find reliable but modest effects of levels for visual stimuli, although in this case the trend for a more prominent effect with verbal material was absent.

Discussion

This is the only occasion in which we find no difference in LOP magnitude across materials, all of which show a significant effect of processing depth. The most obvious interpretation for the lack of an enhanced effect for verbal stimuli is that participants knew insufficient about the animals named to be able to set up a precise encoding; how, for instance does a field ant differ from a fire ant? Taken in isolation, this result could therefore be regarded as obvious and hence uninteresting. In the broader context of our range of LOP effects however, it can be regarded as a potentially informative if relatively extreme case. Also of interest is the clear but relatively modest LOP effect for scenes, extending our results beyond our previous range of pictures of objects or limited door scenes. We return to this in the general discussion.

General Discussion

We will begin by summarizing our results before going on to suggest an interpretation. This will then be applied to the studies of the effect of LOP on verbal and nonverbal material more generally, as summarized in the introduction before concluding with a discussion of the potential significance of our results for other recent studies of visual LTM.

We set out with a broad question; is the positive effect of deep processing limited to language-based materials? We compared the effect of judgments of pleasantness with shallow encoding conditions involving judgments of size and color and emerged with the conclusion that LOP effects can clearly be found with non-verbal materials, hence placing earlier, sometimes equivocal evidence for this, on a broader and more robust base. More interesting than this simple conclusion however, is the variation in size of the effects, and the way in which this depends on the nature of the remembered material.

We can answer our initial question with some confidence. Modest but consistent LOP effects were shown for all, the visual materials we studied, namely doors, clocks, mobile phones and scenes. In contrast, the data for verbal materials are considerably more varied than found with visual stimuli. This is shown in Figure 7 which lists, for each broad type of material, the LOP effect size measured in terms of Cohen's *d*. For purposes of comparison, studies run by Walsh and Jenkins (1973) and Hyde and Jenkins (1969) using judgments of pleasantness with familiar word stimuli report seven LOP experiments with effect sizes ranging from 1.78 to 2.65 with a mean of 2.27. However it should be borne in mind that these studies typically used free recall, rather than experimental conditions that are matched with our own. Furthermore, many studies in the literature have used an incidental learning paradigm, not possible in our case where participants were tested across conditions so as to allow a more powerful within participant comparison. Learning would hence have been incidental on only the first condition. A more appropriate comparison would be with the recognition conditions tested by Craik and Tulving (1975) but unfortunately it is not possible to calculate effect sizes from their data.

Figure 7 about here

As the previous summary would suggest, the smallest effect sizes in Figure 7 tend to be found for doors, with similar effect sizes for the other visual stimuli, while effect sizes for verbal material are widely spread ranging from a small effect for names, broadly comparable to that found for visual stimuli, through to a relatively substantial effect for familiar animals and household goods. This is still lower than those typically reported by Jenkins and colleagues, but was almost certainly constrained by a ceiling effect.

How might we explain this pattern of data? In an informal interchange, Craik agreed that work on LOP had principally focused on studies involving individual words, speculating that perhaps "pictures drive a 'deep' encoding regardless of orienting instructions, whereas words can be processed in a variety of ways" (Craik, personal communication 7th May 2014). However, while this could give an account of a visual-verbal difference, it does not give an adequate account of the wide range of results found within our verbal conditions. For this, we need an explanation that incorporates the characteristics of the material to be remembered and the fact that material and encoding strategy interact, with the result that the same encoding instructions can have markedly different effects across different types of material.

One way of thinking about this is offered by the concept of "affordance" proposed by J.J. Gibson (1977) and subsequently used more widely to refer to the relation between an object and an organism, whereby the object provides, or "affords" the opportunity to perform an action (Norman, 1988). An example might be a simple chair that affords the action of sitting, but also of using it to stand and reach a high shelf, or indeed less probably as a weapon in a bar room brawl.

Craik and Lockhart's (1972) LOP concept could be seen as capitalizing on the affordances offered by printed words. Each word affords encoding as a simple visual stimulus, as a pronounceable word, or as providing access to a rich and potentially complex set of associations through its meaning. Following the abandonment of the initial simple processing depth hypothesis the tendency has been to interpret the enhanced memory for "deeply" processed items in terms of the richness of encoding, with more encoded features leading to greater probability of subsequent recall or recognition (Craik, 2002). Viewed in this way, it seems likely that a judgment of pleasantness will call upon a wide range of potential features which have to be sampled, implicitly at least, in order to come up with an overall pleasantness judgment.

Suppose we consider a picture of a domestic door in the same way. Unless it is a familiar door with rich associations, pleasantness will presumably be judged using relatively superficial visual cues such as shape, size and physical condition, features that are likely to be encoded rapidly and incidentally, even when making other judgments such as the color of the door. We would therefore modify Craik's suggestion of visual stimuli affording immediate "deep" processing, proposing instead that a potential lack of depth characterizes many such stimuli including our door scenes. A similar case can be made for the other visual materials used, clocks, mobile phones and visual scenes.

It is important to note however, that differences are more likely to occur for verbal materials that will vary in the degree and usefulness of the elaboration any word affords. Hence, a familiar animal name such as "leopard", will afford the visual representation of a spotted animal, of speed and fierceness perhaps also linked to its habitat, possibly also including indirect associations, for example to the title of a novel about Sicily. This will allow a clear discrimination between its elaborated encoding and that for example of a lion, different color, different shape, a mane and perhaps associated with the musical “The Lion King” (cf Experiment 1). Similarly menu items are likely to have a range of affordances (Experiment 2). Such richness of detail will not however be available in comparing relatively unfamiliar creatures such as different varieties of ant (Experiment 3c). A similar problem will occur in the case of peoples’ names, which are essentially arbitrary, with potential semantic links, between the name and the person being relatively unhelpful (Experiment 2). We suggest therefore that rather than emphasizing modality as the critical factor, we use the more general concept of affordance. This allows us to integrate results from verbal and visual memory. Thus, in contrast to the rich set of semantic associations afforded by a familiar concrete word, a picture of a domestic door affords rather little in terms of potential semantic elaboration. A door may of course yield a wide array of visual features, any one of which might provide the feeling of familiarity necessary for a positive memory response. However, such features will have little impact on success in recognition unless they differentiate between the target and potential distractors, a situation minimized in our own studies by selecting distractors that are broadly similar in category and visual appearance, for example four church doors.

We suggest a distinction between two broad classes of affordances, perceptual and semantic. Any stimulus, whether visual or auditory will typically provide relatively automatic perceptual encoding that may potentially involve a large range of features. Visually presented words on the other hand, afford a relatively impoverished perceptual stimulus, but one that can be elaborated by making use of a complex network of semantic associations, a process that will depend to some extent at least on the strategy adopted by the participant. It is important to stress that we use the term “semantic” in the broad sense of elaborated and structured knowledge rather than in a purely linguistic sense. This is of particular relevance where differences in levels of expertise are involved as proposed by Bransford et al. (1979) and is well illustrated in the case of memory for visually presented chess positions where performance is strongly influenced by level of expertise (De Groot 1965) in a task that shows no dependence on verbal coding (Robbins, Anderson, Barker, Bradley et al 1996).

Rich and elaborate encoding will not itself guarantee recognition however, unless the encoded features can distinguish between targets and distractors, a central issue addressed by Nairne (2002), in what he terms a “simple model”. This feature-based model provides a useful bridge between our concept of affordance and more classic approaches to theories of verbal LTM[2]. Nairne’s model follows his critique of the concept of transfer appropriate processing in which he criticizes what he describes as the “myth of the encoding-retrieval match” assumption underlying TAP. He proposes instead a correlational link between the number of features present at both encoding and retrieval. Such a correlation is strongly influenced by what he terms the “diagnosticity” of the matching cues, the extent to which they allow a distinction between the correct response and incorrect distractors. Here he introduces the concept of cue overload whereby a potentially useful retrieval cue becomes much less so, if it is also associated with a range of incorrect responses (Earhard, 1967; Watkins & Watkins, 1975). This can arise either because of similarity, whereby many of the features are common to both target and distractor as in the case of our church doors, or because of a relative paucity of cues due to lack of information available at encoding, as in the case of our study using relatively rare animals such as several varieties of ant. In short, what is necessary for successful retrieval is not determined by the number of cues in common between encoding and retrieval, but by their “diagnosticity” based on their capability of distinguishing between targets and non targets.

Nairne's "simple model" provides a plausible overall framework encompassing the processes involved in recognition, while leaving open the question of material and its potential interaction with LOP. The model could be adapted to explain our results by incorporating the concept of affordance. This would potentially provide a simple model that is able to give an account of standard learning phenomena (see Nairne, 2002), LOP, TAP and their interaction with material type.

If the proposed framework is to be developed further we need to identify and measure the relevant features. Identifying the nature of the semantic features of a given verbal stimulus has already been addressed as part of an attempt to develop semantic theory (Lambon Ralph, Lowe & Rogers, 2007; Patterson, Nestor & Rogers, 2007) and to compare the various theories proposed to explain category-specific semantic deficits reported in neuropsychological patients (Cree & McRae, 2003) The method used is simple if somewhat laborious, namely to present each stimulus item (e.g. the word dog) and require participants to produce as many brief associated statements as they can in a specified time. These can be either generic e.g. "has four legs, is a mammal, barks", or specific "called Fido", "bit me last week" etc. We would expect recognition memory for a given item to be a joint function of the number of such responses, together with their distinctiveness-based diagnosticity. Such potentially diagnostic features are likely to be substantially more numerous for familiar concrete nouns such as leopard than for a less familiar creature such as a leaf-cutting ant. It should, in principle, be possible to apply the same method to visual stimuli presenting each item and asking for associated features to be listed. Here, we suspect the discriminative features within a limited category such as doors will be less rich and consequently less likely to be diagnostic in separating targets from distractors although within each list of items such as doors or natural scenes, we would expect some items to be more distinctive and hence more memorable than others, hence offering a potential validation of the application of the method to episodic as well as semantic memory.

This leads on to the question of whether there are likely to be differences between typical visual and verbal stimuli. We suggest that this is likely to be the case simply because verbal stimuli will frequently lend themselves more readily to more general semantic elaboration, particularly if encouraged by “deep” encoding instructions. This seems much less likely for most visual stimuli, for which the relevant diagnostic cues may be less semantically rich, and more dependent on purely perceptual features.

We began with a simple question concerning the applicability of the concept of LOP to visual LTM. We arrived at the answer that modest but consistent effects could be shown but in doing so discovered that the classic substantial verbal effect depends crucially on the verbal material employed. In explaining our pattern of data we imported the concept of affordances from the field of perception and action combining it with the feature based approach widely used in theories of verbal LTM. The resulting broad theoretical framework incorporates the concepts of LOP and TAP, together with the roles of perceptual and semantic coding and the potential effects of expertise. Further development is likely to benefit from capitalizing on currently developing methods and concepts in the areas of semantic memory and picture processing.

Author Note

We are grateful to Philip Quinlan for his help and advice, and to Stephanie Chung, Po Fu, Stephanie Motley, Susan Oei, Stephen Rhodes, Natalie Whitehead and Suet Wong for their contribution to the development of material and to testing and to Fergus Craik and James Nairne for their constructive comments on an earlier draft.

References

Baddeley, A. D. (1978). The trouble with levels: A re-examination of Craik and Lockhart framework for memory research. *Psychological Review, 85*, 139-152.

Baddeley, A. D. (1982). Domains of recollection. *Psychological Review, 89*, 708-729. doi: 10.1037/0033-295X.89.6.708

Baddeley, A. D., Emslie, H., & Nimmo-Smith, I. (1994). Doors and people: A test of visual and verbal recall and recognition. Bury St Edmunds, Suffolk: Thames Valley Test Company.

Baddeley, A. D., Hitch, G. J., Quinlan, P., Bowes, L. & Stone, R. (in press). Doors for memory: A searchable database. *Quarterly Journal of Experiment Psychology*.

Baddeley, A., & Woodhead, M. (1982). Depth of processing, context, and face recognition. *Canadian Journal of Psychology, 36*, 148-164. doi: 10.1037/h0080635

Bower, G. H., & Karlin, M. B. (1974). Depth of processing pictures of faces and recognition memory. *Journal of Experimental Psychology, 103*, 751-757. doi: 10.1037/h0037190

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of National Academy of Sciences U S A, 105,* 14325-14329.

Bransford, J. D., Franks, J. J., Morris, C. D., & Stein, B. S. (1979). Some general constraints on learning in memory research. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels of Processing in Human Memory.* (pp. 331-354). Hillsdale, NJ: Erlbaum.

Conway, M. A. (2002). Sensory-perceptual episodic memory and its context: autobiographical memory. In A. Baddeley, M. Conway & J. Aggleton (Eds.), *Episodic Memory* (pp. 53-70). Oxford: Oxford University Press.

Craik, F. I. M. (2002). Levels of processing: Past, present ... and future? *Memory, 10,* 305-318. doi: 10.1080/09658210244000135

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing. A framework for memory research. *Journal of Verbal Learning & Verbal Behavior, 11,* 671-684. doi: 10.1016/S0022-5371(72)80001-X

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104(3)*, 268-294. doi: 10.1037/0096-3445.104.3.268

Cree, G. S., & McRae, K. (2003). Analyzing the factors underlying the structure and computation of the meaning of chipmunk, cherry, chisel, cheese, and cello (and many other such concrete nouns). *Journal of Experimental Psychology: General, 132*, 163-201.

D'Agostino, P. R., O'Neill, B. J., & Paivio, A. (1977). Memory for pictures and words as a function of level of processing: Depth or dual coding? *Memory and Cognition, 5,* 252-256. doi: 10.3758/BF03197370

De Groot, A. D. (1965). *Thought and choice in chess*. New York: Basic Books.

Earhard, M. (1967). Subjective organization and list organization as determinants of free-recall and serial-recall memorization. *Journal of Verbal Learning & Verbal Behavior, 6*, 501-507. doi: 10.1016/S0022-5371(67)80007-0

Gibson, J. J. (1977). The theory of affordances. In R. Shaw & J. Bransford (Eds.), Perceiving, Acting and Knowing (pp. 67-82). Hillsdale, NJ: Erlbaum.

Graf, P., & Mandler, G. (1984). Activation makes words more accessible, but not necessarily more retrievable. Journal of Verbal Learning & Verbal Behavior, 23, 553-568.

Halpern, A. R., & Bartlett, J. C. (2010). Memory for melodies. In M. R. Jones, R. R. Fay & A. N. Popper (Eds.), *Music Perception. Springer Handbook of Auditory Research.* (Vol. 36, pp. 233-258). doi: 10.1007/978-1-4419-6114-3_8

Hunt, R. R., & Toth, J. P. (1990). Perceptual identification, fragment completion, and free recall: Concepts and data. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 16,* 282-290. doi: 10.1037/0278-7393.16.2.282

Hyde, T. S., & Jenkins, J. J. (1969). Differential effects of incidental tasks on the organization of recall of a list of highly associated words. *Journal of Experimental Psychology, 82,* 472-481. doi: 10.1037/h0028372

Intraub, H., & Nicklos, S. (1985). Levels of processing and picture memory: The physical superiority effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 284-298. doi: 10.1037/0278-7393.11.2.284

Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General, 110,* 306-340. doi: 10.1037/0096-3445.110.3.306

Kanwisher, N., McDermott, J., & Chun, M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience, 17*, 4302-4311.

Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010a). Conceptual distinctiveness supports detailed visual long-term memory for real-world objects. *Journal of Experimental Psychology: General, 139,* 558-578. doi: 10.1037/a0019165

Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010b). Scene memory is more detailed than you think: The role of categories in visual long-term memory. *Psychological Science, 21,* 1551-1556. doi: 10.1177/095679610385359

Lambon Ralph, M. A., Lowe, C., & Rogers, T. T. (2007). Neural basis of category-specific semantic deficits for living things: Evidence from semantic dementia, HSVE and a neural network model, *Brain, 130*, 1127-1137. doi: 10.1093/brain/awm025

Lyman, B., & McDaniel, M. (1986). Effective encoding strategies on long-term memory for odours. *Quarterly Journal of Experiment Psychology, 38A*, 753-765. Doi: 10.1080/14640748608401624

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning & Verbal Behavior, 16,* 519-533. doi: 10.1016/S0022-5371(77)80016-9

Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition, 18,* 251-269. doi: 10.3758/BF03213879

Nairne, J. S. (2002). The myth of the encoding-retrieval match. *Memory, 10*, 389-395. doi: 10.1080/09658210244000216

Nickerson, R. S. (1965). Short-term memory for complex meaningful visual configurations: A demonstration of capacity. Canadian Journal of Psychology, 19, 155–160

Norman, D. A. (1988). *The psychology of everyday things*. New York: Basic Books.

Öhman, A. (2009). Of snakes and faces: An evolutionary perspective on the psychology of fear. *Scandinavian Journal of Psychology, 50,* 543-552. doi: 10.1111/j.1467-9450.2009.00784.x

Osgood, C. E., May, W. H., & Miron, M. S. (1975). *Cross cultural universals of affective meaning*. Urbana, IL.: University of Illinois Press.

Packman, J. L., & Battig, W. F. (1978). Effect of different kinds of semantic processing on memory for words. *Memory & Cognition, 6*, 502-508. doi: 10.3758/BF03198238

Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience, 8,* 976-987.

Patterson, K. E., & Baddeley, A. D. (1977). When face recognition fails. *Journal of Experimental Psychology: Human Learning and Memory, 3,* 406-417. doi: 0.1038/nrn2277

Peretz, I., Gaudreau, D., & Bonnel, A.-M. (1998). Exposure effects on music preference and recognition. *Memory & Cognition, 26*, 884-902. doi: 10.3758/BF03201171

Robbins, T., Anderson, E., Barker, D., Bradley, A., Fearneyhough, C., Henson, R., . . . Baddeley, A. (1996). Working memory in chess. *Memory and Cognition, 24*(1), 83-93.

Roediger, H. L. (2002). Reconstructive memory. In N. J. Smelser & P. B. Baltes (Eds.), *International encyclopedia of the social and behavioral sciences.* Oxford, England Elsevier.

Roediger, H. L., & Blaxton, T. A. (1987). Retrieval modes produce dissociations in memory for surface information. In D. S. Gorfein & R. R. Hoffman (Eds.), Memory and learning: The Ebbinghaus centennial conference (pp. 349-379). Hillsdale, NJ: Erlbaum.

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

Roediger, H. L., Srinivas, K., & Weldon, M. S. (1989). Dissociations between implicit measures of retention. In S. Lewandowsky, J. C. Dunn & K. Kirsner (Eds.), *Implicit memory: Theoretical issues* (pp. 67-84). Hillsdale, NJ: Erlbaum.

Roediger, H. L., Weldon, M. S., & Challis, B. H. (1989). Explaining dissociations between implicit and explicit measures of retention: A processing account. In H. L. Roediger & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving (*pp. 3-39). Hillsdale, NJ: Erlbaum.

Standing, L., Conezio, J., & Haber, R. N. (1970). Perception and memory for pictures: Single-trial learning of 2500 visual stimuli. *Psychonomic Science, 19,* 73-74. doi: 10.3758/BF03337426

Tulving, E. (2001). Does memory coding exist? In M. Naveh-Benjamin, M. Moscovitch & H. L. Roediger III (Eds.), *Perspectives on human memory and cognitive aging: Essays in honor of Fergus Craik* (pp. 6-27). New York: Psychology Press.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80*, 352-373.

Walsh, D. A., & Jenkins, J. J. (1973). Effects of orienting tasks on free recall in incidental learning: "difficulty," "effort," and "process" explanations. *Journal of Verbal Learning & Verbal Behavior, 12*, 481-488. doi: 10.1016/S0022-5371(73)80028-3

Warrington, E. (1984). Recognition memory test. Windsor UK: NFER-Nelson.

Warrington, E. K., & Ackroyd, C. (1975). The effect of orienting tasks on recognition memory. *Memory & Cognition, 3*, 140-142. doi: 10.3758/BF03212890

Watkins, O. C., & Watkins, M. J. (1975). Buildup of proactive inhibitions as a cue-overload effect. *Journal of Experimental Psychology: Human Learning and Memory, 1,* 442-452. doi:

Winograd, E. (1981). Elaboration and distinctiveness in memory for faces. *Journal of Experimental Psychology: Human Learning and Memory, 7,* 181-190.

Zucco, G. M. (2003). Anomalies in cognition: Olfactory memory. *European Psychologist, 8*, 77-86. doi: 10.1027//1016-9040.8.2.77

Footnotes

[1] Unfortunately gender balance for Experiments 1 and 2 was not recorded but was likely to contain a higher proportion of female participants.

[2] Although we find this version of feature theory helpful, our enthusiasm does not extend to Nairne's (1990) attempt to extend feature theory to STM, which involves strong assumptions about the role of modality-specific features.

Figure 1

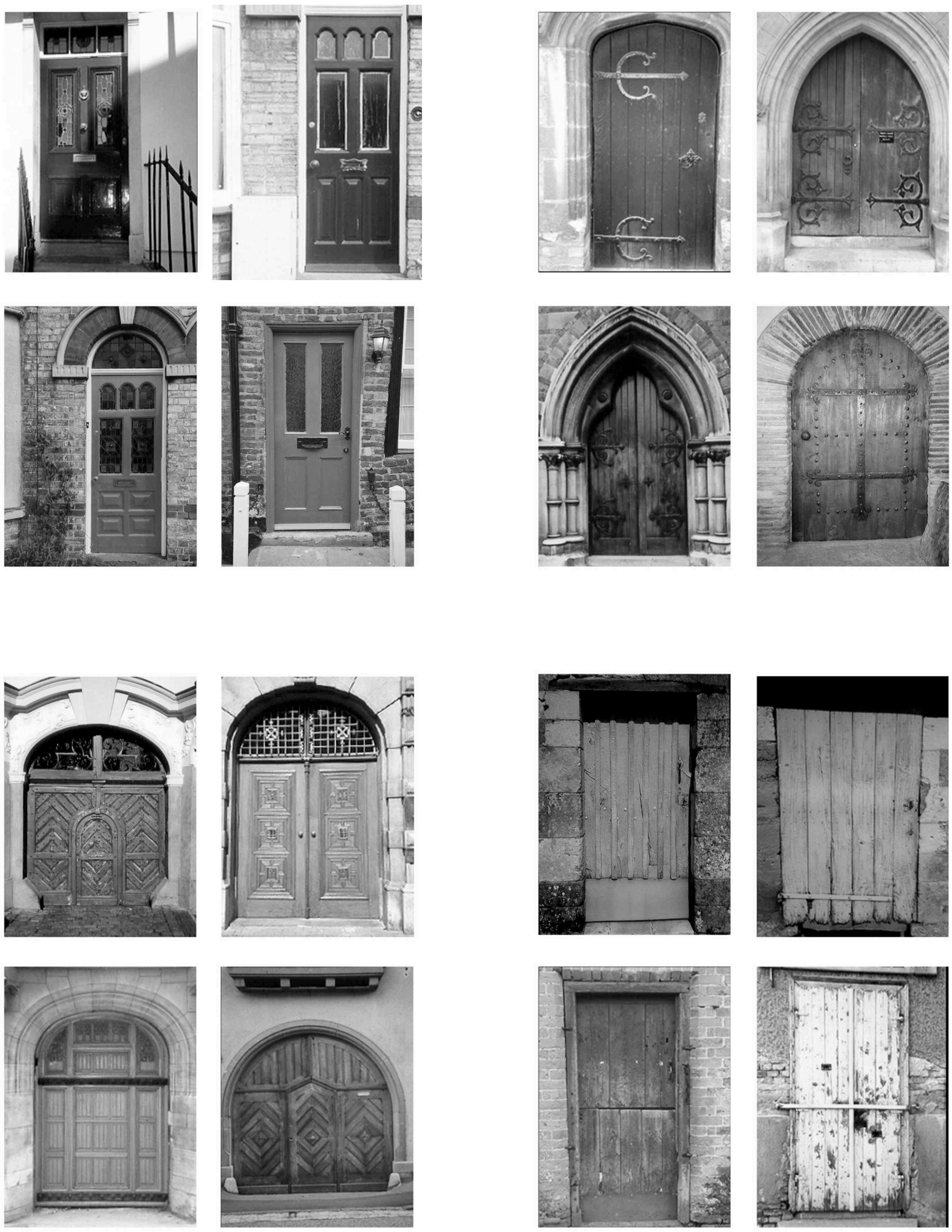

Figure 2

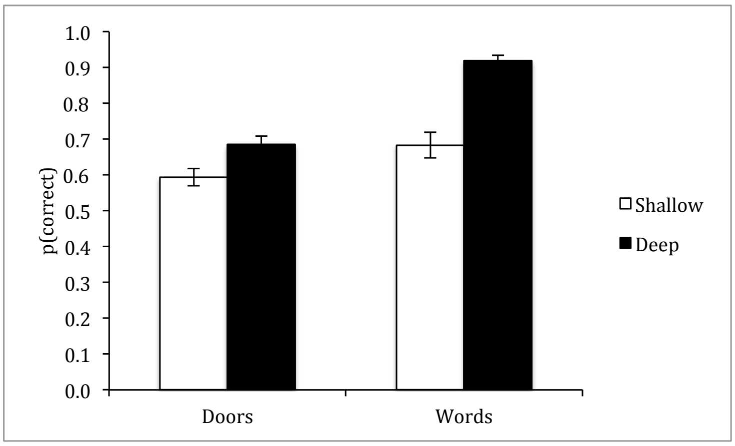
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
p(correct)
Doors
Words
Shallow
Deep

Figure 3

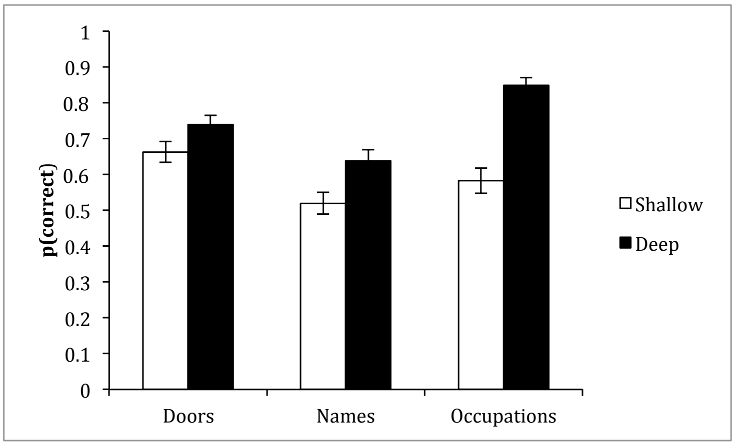
p(correct)
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Doors
Names
Occupations
Shallow
Deep

Figure 4

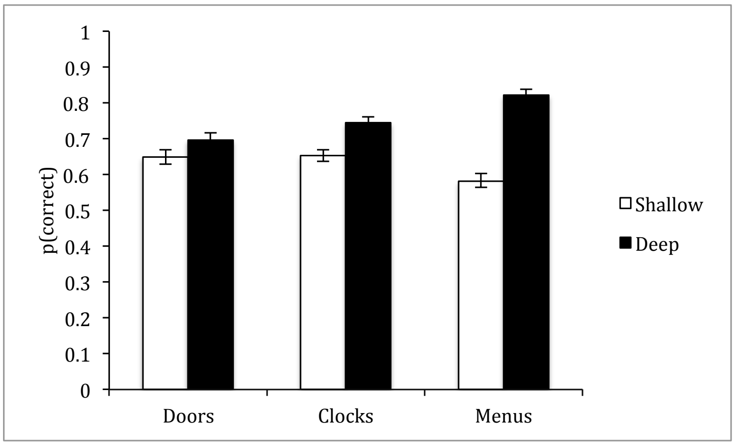
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
p(correct)
Doors
Clocks
Menus
Shallow
Deep

Figure 5

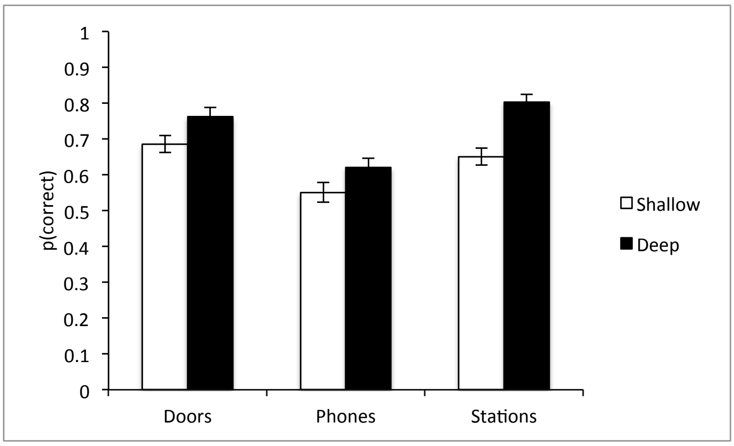
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
p(correct)
Doors
Phones
Stations
Shallow
Deep

Figure 6

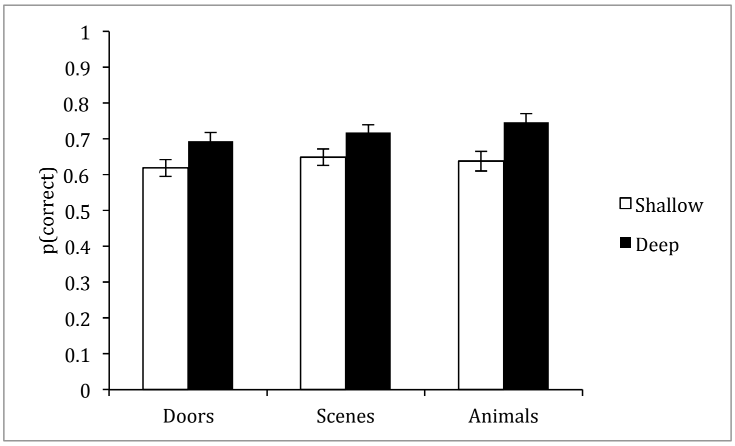
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
p(correct)
Doors
Scenes
Animals
Shallow
Deep

Figure 7

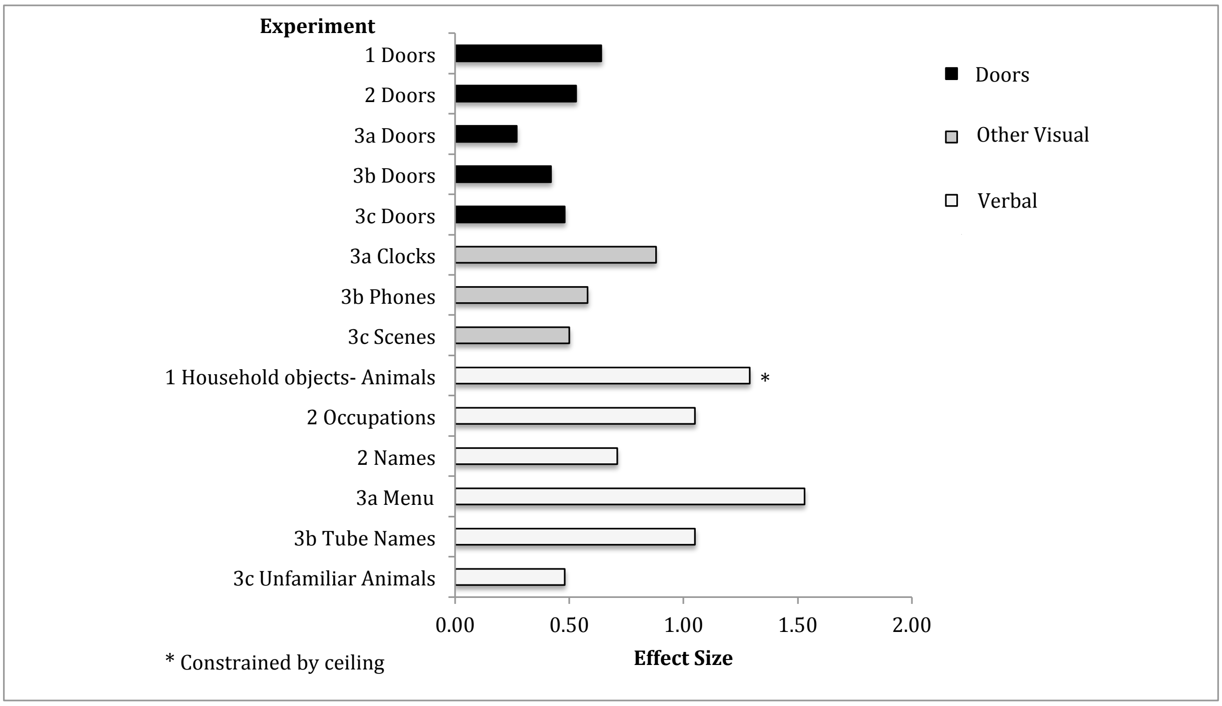
Experiment
1 Doors
2 Doors
3a Doors
3b Doors
3c Doors
3a Clocks
3b Phones
3c Scenes
1 Household objects- Animals
2 Occupations
2 Names
3a Menu
3b Tube Names
3c Unfamiliar Animals
*
0.00
0.50
1.00
1.50
2.00
Effect Size
Doors
Other Visual
Verbal
* Constrained by ceiling

Figure captions

Figure 1

Examples of the four alternative forced choice door scenes used in Experiment 1. The full database is accessible via Baddeley et al. (in press).

Figure 2

Influence of levels of processing on door scenes, words selected from the categories of household goods and familiar animals.

Figure 3

Influence of depth of processing on door scenes, people's names and occupations.

Figure 4

Influence of processing depth on recognition of door scenes, clocks and menu items.

Figure 5

Influence of processing depth on recognition of door scenes, mobile phones and London underground station names.

Figure 6

Influence of processing depth on recognition of door scenes, outdoor scenes and unfamiliar animals.

Figure 7

Effect sizes of the influence of depth of processing across our three experiments.