COGNITION

Cognition 64 (1997) 231–248

False categories in cognition: the Not-The-Liver fallacy

Felice L. Bedford*

Department of Psychology and Program in Cognitive Science, University of Arizona, Tucson AZ 85721, USA

Received 11 June 1995; accepted 5 May 1997

Abstract

This paper reports on an increasingly frequent error committed in cognition research that at best slows progress, and at worse leads to self-perpetuating false claims and misguided research. The error involves how we identify meaningful processes and categories on the basis of data. Examples are given from three areas of cognition: (1) memory, where the misconception has fueled the popular implicit/explicit categories, (2) perception, where the misconception is used to re-evaluate the classic what/where division, and (3) motor skills, where it is used to draw conclusions from patients with Huntington's disease. Reasons for the prevalence of this error, how it relates to double dissociations, and what it suggests about scientific reasoning are offered. © 1997 Elsevier Science B.V.

1. Introduction

The year is 1750. The place: An Unknown Medical school. You are in the audience of a well attended medical conference, about to listen to the highly regarded Dr. Fright. "I have a new discovery", Dr. Fright begins. "I have isolated the organ system which removes toxins from the blood - I call it the 'liver'". Amidst oohs and ahs, Dr. Fright provides some evidence for his discovery: "When this organ is removed from a rat, toxins quickly build up, and the rat dies" Applause from the crowd. "But that is not all", the lecture continues, "I have discovered a second organ. This organ circulates the blood, absorbs nutrients, expels waste products from the body, and attacks foreign invaders." "For when the liver is removed", he argues, "the body is still able to do all these things and more, until such time as the toxin buildup is fatal. I suggest we call this second organ 'Not-the-Liver'". Although there are one

* e-mail: bedford@u.arizona.edu

PII S0010-0277(97)00019-X

or two ‘oohs’ and ‘aahs’ remaining, most of the audience is baffled. While you admire the doctor’s discovery of the functions of the liver, you realize that the doctor has not discovered a second organ at all. He has merely shown that the liver is not the ONLY organ that is present in the body.

The medical example just presented is fictitious and clearly outrageous. Yet I suggest that this error in logic, where one discovery is unwittingly turned into two, occurs often in cognitive psychology. This error, which I hereafter will refer to as the ‘Not-the-Liver Fallacy’, contributes to faulty categories and theories that do not reflect meaningful natural divisions of the mind.

2. Memory

If we replace the fictitious year 1750 with the year 1987, then we can see an analogous error in one of the most influential recent trends in cognition: the dissociation of implicit and explicit memory. I suggest that the category of ‘implicit memory’ is not a category at all. There is no evidence that implicit memory is a meaningful concept that has any psychological or biological reality. It is instead a construct founded on the Not-the-Liver Fallacy.

The meanings of the terms themselves provide a clue that something is amiss. Schacter (1987) used the term *explicit memory* to refer to ‘...conscious recollection of recently presented information, as expressed on traditional tests of free recall, cued recall, and recognition’ (p. 501). Synonyms for ‘conscious’ have included ‘intentional’ or ‘deliberate’, and synonyms for ‘recently presented information’ have included ‘a specific learning episode’. Thus, ‘the intentional or deliberate recollection of a specific learning episode’ also helps characterizes what is meant by explicit memory. This traditional type of memory, previously known only as ‘memory’, is contrasted with a new type: implicit memory. Namely, ‘...information that was encoded during a particular episode is subsequently expressed without conscious or deliberate recollection’ (Schacter, 1987, p. 501). Or, ‘...unintentional, non-conscious form of retention that can be contrasted with explicit memory’ (Schacter, 1992, p. 559). Note that while explicit memory is described in terms of positive attributes, implicit is described on the basis of what it is not. Based on the descriptions, implicit memory begins to sound a lot like ‘not explicit memory’.

The argument is not a semantic one. The seminal inspiration for the distinction between implicit and explicit processing comes from amnesic patients. Patients with lesions of the hippocampus and related structures have trouble encoding new memories, as assessed by the failure on standard tests of recall and recognition. These patients, however, are nonetheless able to learn new things (see Schacter, 1987 for examples; e.g. Milner, 1966; Johnson et al., 1985; Glisky et al., 1986). HM, for instance, improved over a series of sessions involving practice in mirror tracing, while having no memory of the sessions themselves. An amnesic patient can be taught to successfully use a computer, yet have no memory of the instruction. A melody played every few minutes will be reported as novel each time it is played, yet a preference for the melody may develop. The idea that these abilities can occur in a person without conscious recollection has captivated many investigators. Indeed, anyone who does not find the phenomenon of some interest is probably as dull as lint[1]. But precisely what should one conclude from these findings?

Damage to the hippocampal system produces a selective deficit in the intentional and conscious recollection of recently presented material, while leaving the patient otherwise unimpaired. This pattern of damage and selective impairment of (explicit) memory does provide evidence that (explicit) memory itself is a candidate for a meaningful independent process. It does not, however, suggest that everything ELSE that the patient can do is a meaningful independent process. The patient can also brush his teeth, sing songs, and draw pictures, as well as acquire fast mirror tracing, computer skills, and musical sophistication. Thus, explicit memory is dissociated from OTHER STUFF, but the data give us no way to understand what remains. Is there a brain disorder which eliminates *all* ability to implicitly acquire knowledge, while sparing the ability to do so explicitly? Such a disorder would provide one piece of evidence that implicit processing may be a meaningful unit. To my knowledge, no such disorder has ever been reported. The single finding of damage and its consequences tells us only that explicit memory *dissociates from other stuff*. (See also Section 6 later in this article.) To return to the medical example, the preservation of some bodily functions with destruction of the liver in no way implicates a 'Not-the-Liver' organ in the body. Analogously, the fact that amnesiacs are still competent at other things does not implicate an 'implicit memory' organ of the mind.

2.1. Castles in the air.

What about other evidence for implicit memory? The second cornerstone of the revolution came from normal subjects rather than amnesiacs and appeared to show that implicit and explicit memory have different rules. For instance, implicit memory appears to be largely modality specific, and explicit memory largely modality independent (e.g. Jacoby and Dallas, 1981; Graf et al., 1985; Roediger and Blaxton, 1987). In a typical modality shift experiment, subjects hear a list of words, but are tested for memory of the words visually. Explicit memory is assessed by recall and recognition, and implicit memory assessed by tests such as stem completion and lexical decision. In stem completion, subjects are given only a few letters of a word, and asked to fill in the remaining letters with the first word that comes to mind. Increased prevalence of words that came from the study list suggest implicit knowledge of the words from the study list. In lexical decision, subjects have to respond as quickly as possible as to whether or not a letter string constitutes an actual word of english. Decreased reaction time for words from the study phase suggests implicit knowledge of the words from the study phase. Modality shift experiments find

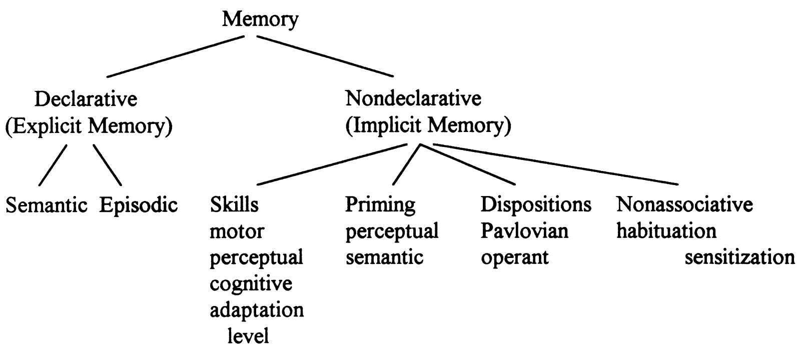
[1]It is not easy to capture why these types of phenomena captivate our interest. The bulk of complex unconscious processes such as digestion, circulation, and the workings of the immune systems do not similarly captivate most psychologists. It is as if we first narrow down the field to those processes that can occur through conscious mental thought, and then are intrigued if these processes can also occur unconsciously.

explicit tests of memory unaffected by the shift from audition to vision, but performance on implicit tests of memory impaired.

Note, however, the finding has only shown that *repetition priming* appears to be modality specific. Repetition priming refers to the phenomenon of facilitation in the processing of a stimulus as a function of a recent encounter with the same stimulus. In fact, nearly all of the modern empirical evidence on normal subjects that formed the basis for the explicit–implicit distinction comes from some type of priming study[2]. I have no objection to the phenomena of priming. Nor do I have an objection to the claim that priming and (explicit) memory appear to have different rules. However, data suggesting that repetition priming and (explicit) memory may dissociate from one another in no way implicates the existence of a general category of ‘implicit memory’. Rather than conclude that ‘implicit memory’ has different rules than (explicit) memory, the studies on priming should have led to the conclusion that *priming* has different rules than explicit memory–not as newsworthy, but more accurate[3]. While this in and of itself is not an example of the Not-the-Liver Fallacy, it illustrates that the establishment of a false category based on the Not-the-Liver Fallacy may lead to a misleading overgeneralization from data that otherwise would not have been made.

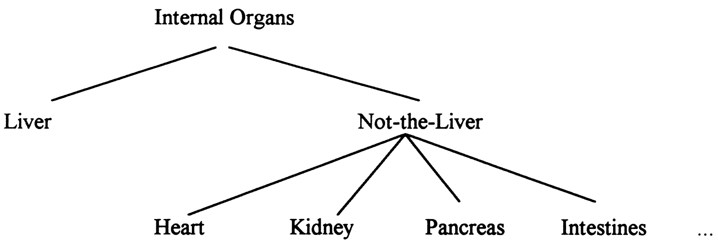
2.2. False taxonomies.

It has become increasingly more common to talk about implicit memory as ‘heterogeneous’, perhaps as more and more researchers realize that the many rules gathered about priming may not really generalize to Pavlovian conditioning, instrumental learning, mirror tracing, prism adaptation, and language acquisition to name just a few learning processes that are Not Explicit Memory. Squire (1992), for instance, has offered a categorization of processes shown below:

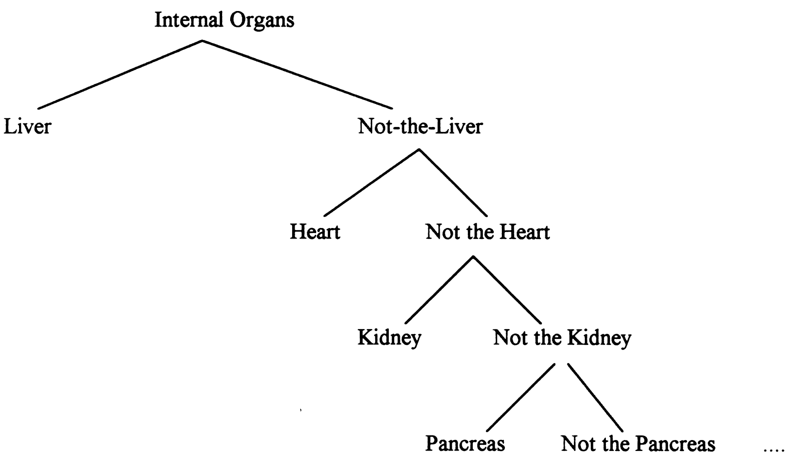


[2]When the distinction was being developed, in 1986, 1987, and 1988, there were 17 articles on ‘implicit memory’, as determined by a search in Psychological Abstracts for ‘implicit memory’. Of these articles, the majority were on priming save for one that was difficult to classify, and 2 broader theoretical articles on implicit memory by Schacter. The following year, 1989, there were 37 articles. Of these, two were difficult to classify, two were on the priming of pictures, and one was on tactile exploration. The remaining empirical studies were all on priming involving words.

This is a reasonable attempt to categorize learning processes, especially since it is one of the very few to attempts to do so. However, it is simply making the best of a premise which is false. Consider what the fictitious Dr. Fright's categorization would look like if we accepted his false premise:



or worse:



The problem is not that implicit memory is a fractionable category rather than a unitary category. Implicit memory is not a *heterogeneous* category. It is not a *category*. At least no evidence has been presented that it is any more a category than the construct Not-the Liver.

So here we have an influential example of the Not-the-Liver fallacy, where one discovery, that memory is a candidate for a meaningful category which dissociates

[3]Many of the studies from 1986–1989 investigated the rules of priming, but advertised these as rules of implicit memory. In addition to modality specificity, these included the effects of elaboration on 'implicit memory' and explicit memory (i.e. really priming and explicit memory), the effects of cognitive rigidity, effects of diazepam, effects of anxiety states, differences in learning disabled children, differences among adults of different ages, and comparisons of different 'implicit memory' tasks (i.e. different priming measures).

from other stuff is mistakenly turned into two–that we have two meaningful processes, explicit and implicit. But psychologists have long known about explicit memory, so unlike the fictitious Dr. Fright's liver discovery, there is nothing new there. Should psychologists be surprised to discover that memory is not the ONLY way to learn? Only if you're a cognitive psychologist trained only on the works of other cognitive psychologists. The field of cognitive psychology in part grew out of a rejection of behaviorism. But the great behaviorists were also the great learning theorists, and the study of learning was rejected along with the behaviorist philosophies behind the research (see e.g. Adams, 1987; Shanks and Dickinson, 1987). Memory was spared, perhaps because nothing could be further from S-R muscle twitching than traditional memory, which so clearly required deliberate mental control to work. Any researcher who happened to continue to work on any problem such as Pavlovian conditioning, instrumental conditioning, prism adaptation or motor skills has long known that there are ways to learn other than explicit memory. But cognitive psychologists working on explicit memory perhaps became excited by the 'discovery' that there are other ways to learn. The wealth of attention given to implicit memory by cognitive psychologists may reflect excitement over the reintroduction of learning issues into the domain of cognitive psychology. The issue of how the mind is changed by experience should be a central question in cognition but until recently has been largely ignored by cognitive psychologists. It is possible that a fallacy in logic contributed to renewed interest in learning. The question is whether implicit memory is a construct that will disappear when cognitive psycholgists know more; I believe that it will, and hopefully I will be around to see it.

3. Vision

The same type of inference is found in other areas. An influential theory in visual perception has been the separation of 'what' vs. 'where' visual systems. The 'where system', which enables one to localize objects in space, is thought to involve the dorsal stream of projections from the striate cortex to the posterior parietal region; the 'what system', which enables object recognition, appears to involve the ventral stream of projections from the striate to the infero temporal cortex (see Mishkin et al., 1983). Recently, Goodale and Milner (1992) have suggested that the two separate visual information processing systems may be better thought of as enabling judgments about objects (ventral), and the control of action (dorsal). Thus, while these two separate cortical pathways are usually thought to subserve separate 'what' vs. 'where' systems, Goodale and Milner argue that they may instead be for perception vs. action or 'what' vs. 'how'. Others have also argued for a perception vs. action division of labor (e.g. Bridgeman et al., 1981), but without the physiological instantiation.

At least one type of evidence for the distinction between visually guided action and visual object judgments is problematic. In a paper entitled 'A neurological dissociation between perceiving objects and grasping them', Goodale et al. (1991) report a case study of a person with brain damage who has lost the ability to judge the perceptual property of orientation. When she is asked to judge the orientation of a slot in a disk, through such tasks as verbal reporting, or picking the matching orientation from a set differently oriented lines, she cannot do so. However, when she is asked to put a card through the slot, the position of her hand changes in accordance with the orientation of the slot very early in the movement, just as it does with normal subjects. An analogous dissociation was found for size in the same patient. The selective deficit reported doesn't really show a dissociation between traditional visual judgments about objects and action to objects, but between those visual judgments and OTHER STUFF, which may turn out to be precisely action, or may include action, or may include just a piece of action, but which we don't really know how to parse.

While this may sound like hair-splitting, consider an experiment which makes good on the concern. In a later study (where Goodale was, in fact, one of the authors; Humphrey et al., 1991), the same patient was tested to see if she would be subject to a phenomenon known as the McCollough effect. The McCollough effect is a visual illusion produced by repeatedly alternating two striped gratings, one oriented vertically with alternate green and black stripes and the other oriented horizontally with alternate magenta and black stripes. This induction procedure leads to a long-lasting color aftereffect that is contingent on orientation. After a few minutes of induction, the white stripes of a vertical black and white test grating appear pink, but the white stripes of a horizontal grating appear green (see e.g. McCollough, 1965; Bedford and Reinke, 1993). Can the patient, who cannot see the difference between vertical and horizontal lines, see the illusory pink and green colors which depend on those very orientations? Yes! Intriguing, but note that the McCollough Effect does not require any *action.* The OTHER STUFF which is preserved includes visual information which serves some function other than action. Therefore, the finding of a normal McCollough Effect visual illusion is contrary to a straight-forward dissociation between visual judgments and visually guided *action*.

Goodale and Milner do have valid arguments against the belief that the dorsal pathway is just for the spatial localization of objects as this function is usually described (e.g. Goodale and Milner, 1992), but their reinterpretation may be premature. The full pattern of data invites other possible reinterpretations. For instance, an alternative to 'action' may be more generally mappings between perceptual dimensions. That would include action, because there are many visual-motor mappings of dimensions involving action, and would also include effects such as the McCollough Effect, which involves a mapping between dimensions of orientation and opponent color (see Bedford, 1995). Whether this is true or not, the general point is that the Not-the-Liver fallacy contributes to premature conclusions and categorizing that may subsequently have to be undone.

4. Motor Skill

A recent development concerning skilled motor performance is the claim that there are at least two separate, dissociable processes of motor skill that can operate independently (Willingham and Koroshetz, 1993). They argue that one process involves the ability to learn a repeating sequence of movements, and the other the ability to learn a new stimulus-response mapping. Yet as we will see, they only provide evidence for one independent process.

The evidence for the dissociation comes from a series of experiments using patients with Huntington's disease, who have atrophy to the striatum. Previously, Huntington's patients have been said to be simply impaired in learning new motor skills, but Willingham and Koroshetz (1993) show the impairment is more selective. Specifically, patients with Huntington's disease were first given a task which did not require a new stimulus response mapping, but did require that a repetitive sequence be learned. An asterisk appeared on a computer screen in one of several locations and subjects were required to push a key directly below the asterisk as quickly as possible. Critically, the asterisks did not appear randomly, but were programmed to occur in a particular repetitive sequence. Normal controls were able to take advantage of the repeating motor pattern of pressing keys, and reaction time fell quickly. Huntington's patients, however, were greatly impaired at learning this task. Next, the task demands were reversed. A different set of subjects were required to learn a new stimulus-response mapping, but did not have to learn a repeating sequence of motor actions. In this study, the task required pressing a key which was immediately to the right of the asterisk's position (the new mapping), rather than the key directly below. The asterisks appeared randomly rather than in a repeating pattern. In contrast to performance on the first task, Huntington's patients did relatively well; they do not appear specifically impaired at learning a new simple motor response to a visual stimulus. The investigators suggest that these results provide evidence for a dissociation between sequences of movements and learning new perceptual-motor mappings, and that these two processes are two components of motor skill.

But do the data really suggest that learning a new perceptual-motor mapping is a component of motor skill, or that it is a meaningful independent component? What Willingham and Koroshetz have shown is that motor skill is likely not unitary, that people with Huntington's disease are likely not equally impaired at all aspects of motor skill, and that mastering repetitive sequences of movements is a likely candidate for one of those subcomponents. Their experiments successfully isolate sequencing from other motor skill task demands, and they find that Huntington's patients are selectively impaired at that function. Beyond that, the data show that these patients are spared the ability to do at least some OTHER MOTOR LEARNING STUFF. But the experiments do not show how the other motor learning stuff parse into meaningful components. The investigators have made several interesting contributions, but they do not extend to visual-motor mappings. They may have other reasons for singling out visual-motor mappings, but it would be incorrect to argue that *these data* provide evidence for that hypothesis. One can safely say that the data do not provide evidence *against* the idea that learning new visual-motor mappings is a distinct separable component of motor skill, but that is not a very strong statement. In fact, there is some reason to believe that the acquisition of visual-motor mappings often has little to do with motor skills, but instead tap into perceptual learning, which may be a different type of learning process than motor skill learning (see Bedford, 1993a, also Bedford, 1993b). Whether the acquisition of visual-motor mappings is or is not a meaningful independent component of motor skill acquisition is a matter of empirical and theoretical debate, but the Not-the-Liver Fallacy slows progress and adds confusion.

5. What the error is not and related arguments

Two explicit clarifications of the what the present argument entails may be useful to avoid any misunderstandings. The Not-the-Liver argument is not an argument against the existence of *heterogeneous* categories or *fractionable* categories or *non-unitary* categories or *multi-level* categories. For instance, the claim is not that the problem with implicit memory is that too many researchers assumed it was a unitary category, nor is the claim that the problem with implicit memory is that it would have to subdivided into too many further categories; rather, the claim is that implicit memory is *itself* not a category, and therefore questions as to whether it is unitary or needs to be subdivided are not meaningful questions. Analogously, it is not meaningful to ask whether Not-the-Liver should be further subdivided, although one could do what looks like a subdivision into heart and kidney. The set of all things consisting of both the number 7 plus the color red can be subdivided into '7' and 'red'–one could even present compelling evidence that '7' and 'red' are independent of one another, but that does not make their concatenation a meaningful category.

In contrast, consider a candidate for an actual multi-level category, visual perception. Visual perception can be subdivided into color vision and depth perception, to name two processes at the next level. Not only is there evidence that color vision and depth perception are each dissociable and independent, but there is reason to believe that the entity at the superordinate level, visual perception, is *itself* a meaningful category. A number of different criteria are useful for parsing the mind, e.g. unique *inputs*, *internal states*, *output*, *brain states*, and *function*; the more criteria that can uniquely distinguish one piece from remaining processing, the more likely that piece is to be a meaningful unit. Vision meets several criteria. For instance, the *input* to vision is electromagnetic radiation with wavelengths between 350 and 750 nanometers, largely distinguishable from the input to other abilities such as audition; removing the input to vision, such as through closing the eyes, carves away vision but little else. Vision has unique *outputs* in part because the experience of seeing feels different than anything else we do, such as hearing. Many of the *internal states* of vision are unique, autonomous, and distinguishable from the internal states of other processes; one type of evidence for this is that the inner workings of vision are largely impenetrable to other abilities, such as to conscious thought and to audition (cf. modularity and Fodor, 1983). The *brain states* serving vision are believed to largely unique and isolable (beginning with doctrine of specific nerve energy and Mueller). At the next level, vision can be subdivided into pieces that in turn appear to meet several of these criteria. For instance, color vision takes different input (wavelength) than other visual processes (e.g. spatially extended arrays). Another example of a system with meaningful units at multiple levels may be language processing, with syntax and semantics as two of its subprocesses.

While some may take issue with the criteria or the claims that visual perception or language processing are meaningful categories, the examples are discussed to illustrate that there is no general bias against entertaining the existence of multi-level hierarchical categories. There is nothing about the Not-the-Liver argument that precludes likely fractionable categories such as these. The claim has been that 'implicit memory' (for instance) is more like the non-organ 'not the liver' or the concatenation '7 plus the color red', than like the (likely) categories visual perception or language processing. And that the creation of the false categories is caused by the Not-the-Liver Fallacy.

A more serious potential problem concerns the fact that the present thesis has been about 'not-X'–but what about 'X?' That is, if you successfully isolate a process, or mechanism, or function X from everything else, what remains is just undifferentiated amorphous stuff–not itself a meaningful category and with no evidence from which to extract the meaningful categories contained within. But whereas you have isolated X, how can you be so sure that X really reflects a meaningful category any more than Not-X? The answer is that we can't, but the uncertainty surrounding that conclusion is of a very different sort than the Not-the-Liver Fallacy.

Consider a specific line of research discussed in this paper, motor skills. Huntington's patients are selectively impaired on a task which requires detecting and predicting a repeating sequence for efficient performance (a sequence of positions of an asterisk is repeated), but not on a task that requires them to learn a new visual-motor mapping (each position of an asterisk is associated with a new position for a key-press). In the last section, it was argued that the research did not successfully isolate visual-motor mappings as an independent component of motor skill as the authors claimed because this remained an ability of the patient, which was not isolated from the many remaining things the patient can still do. However, the authors also claimed that learning repetitive sequences was an independent component of motor skill, and this claim went unchallenged. How can we be so sure that sequencing is a coherent subsystem of motor skill?

We can't: It may turn out that there are impairments on tasks other than sequencing, and that sequencing cannot be isolated from these other tasks. In this case, it would not be correct to argue that sequencing is an isolable component of motor skills. Dissociation data require caution. In some sense, you get out what you put in. To say there is selective impairment on one task implies that one has tried every conceivable task there could be and found all but one intact - clearly impossible. Thus what one does choose to test is often guided by preconceived notions of what the parts already are. This necessarily introduces bias and makes it more likely that you find evidence for that which you already believe is a part. However, the difference is that if X goes away with damage, X is a candidate for a meaningful category, even if it doesn't always turn out to be, whereas if X goes away, there is no evidence whatsoever that not X is a category. You have learned that not X can work without piece X, which can be useful in some situations, but that doesn't suggest that not X is meaningful.

There are further cautions concerning a conclusion that X is a subsystem. What if you didn't damage only one 'thing'–then what goes away could look like one process but really be two independent processes that have nothing to do with one another. If you remove the liver *and* the gallbladder, you could mistakenly believe that bile is produced and stored by the same system. Also, if X goes away with damage it is tricky to conclude that X is an independently functioning subsystem, because X is no longer functioning–it's damaged! You don't have evidence that it can function by itself. Consequently, one needs to believe in the dropped watch model–that things come apart at the joints when they're broken–to make conclusions about mental subsystems from dissociation data. It may turn out that there are so many assumptions and cautions involved in using brain dissociations, that other criteria should be preferred for inferring natural kinds, such as inputs, functions and so on mentioned earlier.

The point of this paper, however, is not to examine the logic of brain dissociation methodology or to examine the criteria of identifying natural kinds. It is clearly not an easy question to know how to parse the mind into meaningful divisions; however it is easier to know how *not* to do so, and one error of this sort is presented in this article. The Not-the Liver Fallacy can be viewed as one part of a larger set of issues that concerns how to identify natural kinds; one part that can be show to be occurring now in cognition and leading to erroneous conclusions.

6. What about Double Dissociations?

One final digression may be useful. Readers particularly knowledgeable about, or uninterested in, double dissociations can skip to the next section. In cognitive neuroscience, a single dissociation takes the form that damage to an area of the brain leads to selective impairment of a function while preserving other functions. A double dissociation is said to occur when the reciprocal pattern is found: some other damage now destroys function B, while preserving A. The present examples of the Not-the-Liver Fallacy have largely come from findings that use the logic of the single dissociation. Yet it is the double dissociation which has become the standard to achieve in cognitive neuroscience; claims of separation of processing subsystems are usually not taken seriously until the double dissociation is discovered. Is the Not-the-Liver fallacy prevented whenever there exist double dissociations?

Unfortunately, the typical double dissociation does not remove the potential X, not X, error in reasoning. A typical 'double dissociation' usually reflects what is essentially 2 single dissociations, which presents the opportunity for the Not-the-Liver fallacy times 2 (X, not X; Y, not Y). The double dissociation, while widely used and accepted, is not the magic bullet for all problems that occur in the interpretation of deficit data. Its existence solves one particular problem with interpreting a single dissociation, that of relative task difficulty (cf. Teuber, 1955, see e.g. McCarthy and Warrington, 1990). That is, if damage causes function A to go away but not function B it is tempting to conclude that A and B are carried out by separate processing systems. But Teuber pointed out that they could be carried out by the same system, such that A goes away with damage and B does not only because A is harder than B. The discovery of a reciprocal pattern where B can go away but not A rules out the hypothesis that A will always goes away first simply because is harder than B. For instance, McCarthy and Warrington (1990) use the example of Broca's aphasia. Patients with damage to Broca's area are unable to talk but can still understand language, which suggests a dissociation between producing and comprehending language. However, it could be that they are part of the same processing system, but production is harder than comprehension. The discovery that damage to Wernicke's areas produces the opposite pattern of impaired comprehension with preservation of production argues against an interpretation that language production is simply harder than comprehension.

While double dissociations of A from B and B from A rule out the 'ease of processing' alternative interpretation of the single dissociation A from B, they still not do not ensure that A and B are separate subsystems. For instance, consider the example just provided, Broca's and Wernicke's aphasias. Others have argued that the same data really suggest a dissociation between closed class items (e.g. 'and', 'the') which are especially difficulty for Broca's patients and open class items ('truck',) which Wernicke's patients have trouble with, rather than implying a dissociation between production and comprehension. Another example of the less than straight-forward interpretation of double dissociations comes from recent work on imagery. Behrmann et al. (1994) first review evidence that mental imagery and visual perception draw from the same representations and may share common neural mechanisms. They also discuss the single dissociation that others have reported in which patients have lost imagery, but perception remains intact. The bulk of the paper reports the discovery of the reciprocal pattern, a clean case where a patient has lost perception, but not imagery. Thus, they have found the coveted double dissociation and as they point out, this rules out the possibility that the single dissociation had been found simply because imagery was harder than perception; they further point out that the way double dissociations are usually interpreted implies that the double dissociation between imagery and perception suggests that the two functions are separate and independent. They go on to say: 'The paradox, then, is how to reconcile the findings of separation with the overwhelming evidence supporting the shared substrate for imagery and perception' (p. 1083).

Finally, consider an example discussed earlier in this paper, the issue of separate visual processing systems mediated by the ventral and dorsal pathways. Goodale and Milner (1992) have challenged the suggestion that the two pathways correspond to separate identification ('what') and localization ('where') functions (cf. Mishkin et al., 1983). Yet there exist double dissociations both in human patients and animals such that damage ventrally causes loss of 'what' but not 'where', and damage dorsally causes the reverse. For instance, patients with visual agnosia are impaired at identifying objects, but not at reaching for objects or navigating in the world, whereas patients with optic ataxia cannot reach in the correct direction for objects, but have no trouble identifying the objects. What went wrong then? Goodale and Milner argue that 'closer examination' of the deficits leads to a different conclusion;

patients with ataxia not only can't point to the correct location, but also can't do other things with their hands, such as adjust the size of the opening between thumb and forefinger to match the size of an object to be grasped, or adjust their hand to an appropriate orientation. This they argue points to an 'action' ('how') subsystem, rather than a 'space' ('where') subsystem.

As these examples demonstrate, the presence of a double dissociation between A and B do not always imply that A and B reflect separate independent meaningful categories of processing.

The Not-the-Liver Fallacy can cause biased interpretations of a double dissociation. For instance, to continue the 'what' vs. 'where' or 'how' story, Goodale and colleagues, as discussed earlier, have a patient who cannot perceive the orientation of a line, yet can adjust her hand appropriately to post a card through an oriented slot. Also as discussed earlier, this does not provide any evidence that what she can still do in this task, automatic visually guided action, is the meaningful independent category, since other things are preserved as well. But if one now erroneously believes one has evidence for 'action' as a category, one will search for a seeming double dissociation where 'action' is eliminated, and not look further for what else might be eliminated along with it, or not look further for the best characterization of what is eliminated. Yet this identification vs. action interpretation of the double dissociation is subject to the very same objections which caused Goodale and Milner (1992) to challenge the earlier interpretation of double dissociation data: not having performed a 'closer examination' of the true deficit. (In fact, to end the convoluted tale, at least the animals with damage dorsally don't just have problems with action, but with other non-action spatial tasks as well. 'Visually guided action' does not seem to provide a full characterization of the deficit any more than the initial 'space–where' interpretation.)

In general, when A is dissociated from B, and B is dissociated from A, the two cases appear to be complements, but only rarely are they exact complements. If they are not exact complements then one is really dealing with 2 single dissociations, A from not A and B from not B, which are then subject to two Not the Liver problems. Exact complements would be helpful at preventing false conclusions, but they are rare. If A and B were the only 2 processes/tasks/functions in existence, then the double dissociation would be a genuine double dissociation where the two deficits are exact complements. (Not A = B and Not B = A). If A and B are not the only 2 functions, a genuine double dissociation is also possible. For instance, in the domain of memory, if a deficit was reported where *all* not-explicit memory were eliminated, while preserving explicit memory–the exact complement of the current deficit–this genuine double dissociation would provide evidence that not-explicit memory (implicit memory) was a meaningful category. (The hypothetical single dissociation where all implicit memory functioning goes away, would provide evidence that not implicit memory was a candidate for a meaningful process and the existence of the reciprocal dissociation would add that implicit and explicit weren't part of the same system with one or the other simply being easier). Otherwise, without the exact-complement genuine double dissociation, the presence of a typical double dissociation doesn't prevent the Not-the-Liver fallacy.

7. When does the error occur?

The Not-the-Liver fallacy is becoming common. The frequency of the error may be increasing along with the increases in attempts to dissociate learning processes or dissociate any modules of the mind more generally. The temptation to commit the error depends in part on one's intuition about the size of the underlying domain of investigation.

If a process or mechanism 'X' has been successfully identified and isolated, then what should be done with what remains?; i.e. 'not X'. The chances that 'Not X' is a single coherent process, 'Y', go up as the size of the space 'Not X' goes down. Thus, if one believes that X is a very large piece of a very small pie, then one is more likely to suspect that what remains may also be a meaningful process. If one believes X is carved from a big pie, then one is less likely to conclude much about 'Not X' from the original extraction of X. For instance, in 'implicit memory', if one had the intuition that the amount of processing that could occur without intentional conscious direction was limited to a very small subset of learning, than one is more likely to have the intuition that the subset might well be a meaningful independent process. If one instead had the intuition that many processes could occur without conscious intervention, and that deliberate recall and recognition were the exception rather than the rule, then one is less likely to believe that what remains, not explicit memory, is itself a meaningful entity. Unless we know ahead of time that the underlying space consists of only two items, 'X' and 'Y'–which is usually the very thing we are trying to discover–we cannot conclude anything about the coherence of Not X as a process from the isolation of X.

8. Final thoughts

If an error is so common and so appealing, then perhaps it is pointing to a deeper problem in the way we do science. In all science, we would like our theories to be capable of generating new predictions, rather than simply explaining the past. Robyn Dawes (1993) argues that in the social sciences there is an asymmetry between predicting the future and 'predicting' the past. For instance, if a person is diagnosed as schizophrenic, we are able to look back and identify his risk factors, such as a biological mother with schizophrenia. However, our knowledge of schizophrenia does not allow us to do what we would like–to predict who will become schizophrenic. If a child's mother is schizophrenic, we cannot predict whether the child will or will not become schizophrenic. This type of asymmetry between predicting the past vs. the future will occur, Dawes argues, when certain characteristics of the underlying problem prevail (Dawes, 1993). For instance, multiple antecedents for a single effect will contribute to the asymmetry. All the characteristics are those which are more commonly found in the social sciences than in the physical sciences. Does our knowledge of schizophrenia allow us to make any predictions at all? Yes: We can predict that: I*f you do not have a mother who is schizophrenic you are likely NOT to become schizophrenic.*

What does this have to do with the Not-the-Liver fallacy? The above statement about schizophrenia is not one that anyone would happily claim or publish as a prediction concerning schizophrenia. Predictions in science do not take that form. But if we had a category of 'not-schizophrenia', then we might be able to turn the negative statement about schizophrenia into a more acceptable positive statement about non-schizophrenia. Then we can predict who will become 'non-schizophrenic', rather than predicting who will not become schizophrenic. The Not-the Liver fallacy may be related to the desirability to make predictions while keeping with the (often implicit) rules involving theorizing in science. Our notions of what it means to be a theory and what types of predictions are considered meaningful are based on the physical sciences. If Dawes is correct, they may be based on assumptions that do not hold in the social sciences. Perhaps the Not-the-Liver Fallacy points to a more general premise about the way we do science that needs to be reinterpreted.

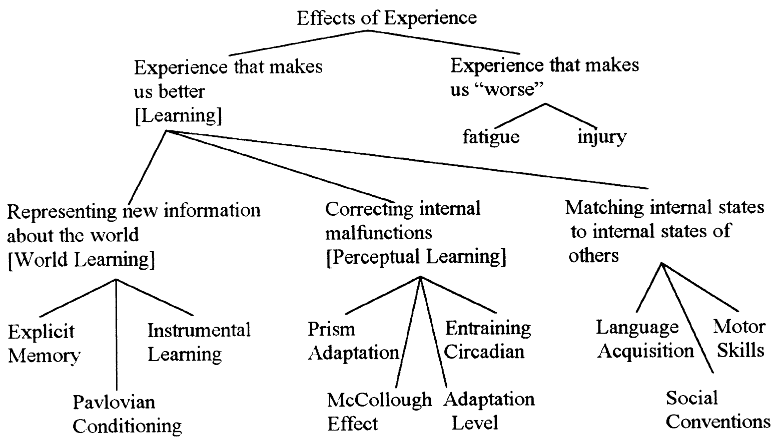
Either way, we should all be on our guard against committing the Not-the-Liver Fallacy.

9. Acknowledgements

Thanks to colleagues Thomas Bever, Paul Bloom and William Ittelson for helpful comments and discussions and to an anonymous reviewer for a very thoughtful review that lead to Sections 5 and 6.

Appendix A. Functional Categorisation of learning and memory

I describe here a functional categorization of learning and memory processing that was suggested recently (Bedford, 1993a). Note that the analysis attempts to avoid the Not-the-Liver Fallacy. The categorization is shown first followed by an explanation of each level.



Level 1. The top of the hierarchy is a very large set consisting of all possible effects of experience on psychological processing. Yet not all of these effects involve what we would consider to be learning.

Level 2. The relevant subset concerns ‘those experiences that make us better’. To paraphrase Rozin and Schull (1988): Isn’t it curious that learning always makes us better? It is precisely those changes that make us better rather than worse that characterize what we mean by ‘learning’. Older attempts to define learning with statements such as ‘a permanent or relatively permanent change in behavior caused by experience except for effects such as fatigue and injury’, can be replaced with a working definition such as ‘Changes in psychological processing caused by experience that lead to adaptive improvement’. One of the dissatisfactions in the past with defining learning at all was having to exclude phenomena by listing them, such as ‘except injury’. Such qualifiers seemed inelegant and lacking in theoretical justification. Yet intuitively, some changes brought about by experience did not seem to belong in the study of learning. Changes from experience that make us better captures the intuition about what is meant by learning, and makes explicit something that had been implicit in older definitions of learning. (Note that what is left, ‘things that do not make us better’, is a non-category of entities that awaits those interested in effects of injury on psychological processing and other topics to sift through and identify meaningful processes).

Level 3. Next, Learning can be broken down further into three major categories. The most familiar of these, World learning, includes processes whose function it is to apprehend new information about the world. A much less familiar category is Perceptual learning, which includes processes that correct internal errors or otherwise improve sensory systems. Perceptual learning mechanisms do not lead to improvement by representing facts about the world the way World learning mechanisms do, but rather by updating the sensory systems themselves (see Bedford, 1993a; Bedford, 1995). Accurate sensory systems are essential for engaging in new and accurate world learning. Finally, there is also likely a third general category. The first category, world learning, involves matching ones own internal representations to the world; the second category, perceptual learning, involves matching ones own internal representations to other of ones own internal representations. (e.g. updating sensory systems on location involves aligning internal visual representations of location with other internal representations of location). Many processes do not fit clearly in either of these categories. The third category involves matching ones own internal representations to the *internal representations of other people*. Many adaptive changes are not about acquiring new information about the world, or about correcting sensory systems, but instead serve this third purpose to mimic what others do.

Level 4. (a) World learning subdivides into processes including memory. By memory, I mean what is currently called explicit memory, or what used to be called simply memory. Another process in this category is Pavlovian conditioning, a primitive associative learning mechanism that allows an organism to represent the structure of the world (see Rescorla, 1988).

(b) Examples of Perceptual learning include the process responsible for the phenomenon of prism adaptation, which updates sensory systems in order to produce accurate location percepts, the process responsible for the phenomenon called the McCollough Effect, which updates sensory systems in order to produce consistent color perception, and the entrainment of circadian rhythms, which updates internal clocks.

(c) An example in the third category is language acquisition. Languages are not out in the world, nor is their acquisition a matter of correcting erroneous sensory systems. Instead, what matters in language acquisition is to match your language to someone else's. Another example is the acquisition of motor skills. Motor skills may be best thought of as a reduction in the degrees of freedom of movement in order to match what someone else can do.

This taxonomy is novel in many ways, including the inclusion of phenomena often omitted because no one knew what to do with them (e.g. prism adaptation), the compilation of processes usually regarded as unrelated into the same category (e.g. language acquisition and motor skill learning), and the division of processes sometimes regarded as similar into different categories (e.g. motor skill acquisition and prism adaptation). This new type of categorization may point to new research directions. For instance, might all processes in the third category, and only in the third category, be subject to critical periods?

References

Adams, J.A., 1987. Historical review and appraisal of research on the learning, retention, and transfer of human motor skills. Psychological Bulletin 101, 41–74.

Bedford, F.L., 1995. Constraints on perceptual learning. Objects and dimensions. Cognition 54, 253–297.

Bedford, F., 1993a. Perceptual learning. In: Medin, D. (Ed.), The Psychology of Learning and Motivation, Vol. 30. Academic Press, New York, pp. 1–60.

Bedford, F.L., 1993b. Perceptual and cognitive spatial learning. Journal of Experimental Psychology: Human Perception and Performance 19, 517-530.

Bedford, F.L., Reinke, K.S., 1993. The McCollough Effect: dissociating retinal from spatial coordinates. Perception and Psychophysics 54, 515–526.

Behrmann, M., Moscovitch, M., Winocur, G., 1994. Intact visual imagery and impaired visual perception in a patient with visual agnosia. Journal of Experimental Psychology: Human Perception and Performance 20, 1068–1087.

Bridgeman, B., Kirch, M., Sperling, A., 1981. Segregation of cognitive and motor aspects of visual function using induced motion. Perception and Psychophysics 29, 336–342.

Dawes, R.M., 1993. Prediction of the future versus an understanding of the past: a basic asymmetry. American Journal of Psychology 106, 1–24.

Fodor, J.A., 1983. Modularity of mind. MIT Press, Cambridge, MA.

Goodale, M.A., Milner, A.D., 1992. Separate visual pathways for perception and action. Trends in Neuroscience 15, 20–25.

Goodale, M.A., Milner, A.D., Jakobson, L.S., Carey, D.P., 1991. A neurological dissociation between perceiving objects and grasping them. Nature 349, 154–156.

Glisky, E.L., Schacter, D.L., Tulving, E., 1986. Computer learning by memory-impaired patients: acquisition and retention of complex knowledge. Neuropsychologia 24, 313–328.

Graf, P., Shimamura, A.P., Squire, L.R., 1985. Priming across modalities and priming across category levels: Extending the domain of preserved function in amnesia. Journal of Experimental Psychology: Learning, Memory, and Cognition 11, 385–395.

Humphrey, G.K., Goodale, M.A., Gurnsey, R., 1991. Orientation discrimination in a visual form agnosic: evidence from the McCollough Effect. Psychological Science 2, 331–335.

Jacoby, L.L., Dallas, M., 1981. On the relationship between autobiographical memory and perceptual learning. Journal of Experimental Psychology: General 110, 306–340.

Johnson, M.K., Kim, J.K., Risse, G., 1985. Do alcoholic Korsakoff's syndrome patients acquire affective reactions? Journal of Experimental Psychology: Learning, Memory, and Cognition 11, 27–36.

McCarthy, R.A., Warrington, E.K., 1990. Cognitive Neuropsychology. New York: Academic Press.

McCollough, C., 1965. Color adaptation of edge detectors in the human visual system. Science 149, 1115–1116.

Milner, B., 1966. Amnesia following operation on the temporal lobe. In: Whitty, C.W.M., Zangwill, O.L. (Eds.), Amnesia. Butterworth and Co., London, pp. 109–133.

Mishkin, M., Ungerleider, L.G., Macko, K.A., 1983. Object vision and spatial vision: two cortical pathways. Trends in Neurosciences 6, 414–417.

Rescorla, R.A., 1988. Pavlovian conditioning: it's not what you think it is. American Psychologist 43, 151–160.

Roediger, H.L. III, Blaxton, T.A., 1987. Retrieval modes produce dissociations in memory for surface information. In: Gorfein, D.S., Hoffman R.R. (Eds.), Memory and Cognitive Processes: The Ebbinghaus Cocentennial Conference. Erlbaum, Hillsdale, NJ, pp. 349–379.

Rozin, P. and Schull, J., 1988. The adaptive-evolutionary point of view in experimental psychology. In: Atkinson, R.C., Herrnstein, R.J. , Lindzey, G., Luce, R.D. (Eds.), Handbook of Experimental Psychology. Wiley-Interscience, New York.

Schacter, D.L., 1987. Implicit memory: History and current status. Journal of Experimental Psychology: Learning, Memory, and Cognition 13, 501–518.

Schacter, D.L., 1992. Understanding implicit memory: a cognitive neuroscience approach. American Psychologist 47, 559–569.

Shanks, D.R. and Dickinson, A. 1987. Associative accounts of causality judgement. In: Bower, G. (Ed.), The Psychology of Learning and Motivation, Vol. 21. Academic Press, San Diego, CA, pp. 229–261.

Squire, L.R., 1992. Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. Psychological Review 99, 195–231.

Teuber, H.L., 1955. Physiological psychology. Annual Review of Psychology 6, 267–296.

Willingham, D.B., Koroshetz, W.J., 1993. Evidence for dissociable motor skills in Huntington's disease patients. Psychobiology 21, 173–182.