

Research

Do different tests of episodic memory produce consistent results in human adults?

Lucy G. Cheke and Nicola S. Clayton¹

Department of Psychology, University of Cambridge, Cambridge CB2 3EB, United Kingdom

A number of different philosophical, theoretical, and empirical perspectives on episodic memory have led to the development of very different tests with which to assess it. Although these tests putatively assess the same psychological capacity, they have rarely been directly compared. Here, a sample of undergraduates was tested on three different putative tests of episodic memory (What-Where-When, Unexpected Question/Source Memory, and Free Recall). It was predicted that to the extent to which these different tests are assessing the same psychological process, performance across the various tests should be positively correlated. It was found that not all tests were related and those relationships that did exist were not always linear. Instead, two tests showed a quadratic relationship, suggesting the contribution of multiple psychological processes. It is concluded that not all putative tests of episodic cognition are necessarily testing the same thing.

Episodic memory is the ability to mentally relive one's own past events. Most psychologists would agree on what episodic memory is, and that when they discuss this ability they are talking about the same phenomenon. As Suddendorf and Corballis (2007) put it, "... we know what [episodic memory] is because we can introspectively observe ourselves doing it and because people spend so much time talking about their recollections." However, perhaps because of such reference to the personal experience of remembering, people tend to have slightly different working definitions of episodic memory. Such small differences in working definition have produced large differences in the behavioral criteria used to study episodic memory, and lead to significant inconsistency in how this ability is tested.

As well as being studied in its own right, episodic memory is investigated within a number of distinct yet overlapping memory literatures (e.g., source memory, autobiographical memory, long-term memory), each of whose definitions are in turn distinct yet overlapping. Furthermore, various literatures within the field may use the same name for two potentially different processes, or different names for what may be a single process. On top of this, psychological tests are sometimes labeled for the way they assess memory (e.g., free recall [FR], cued recall), sometimes for the nature of what is remembered (e.g., source memory), and sometimes for the memory that they putatively test (episodic memory test, semantic memory test), though different researchers disagree on which of the former correspond to which of the latter. Because of all this, the internal consistency of the episodic memory literature is limited, with very different tasks being used to assess the same ability. Unfortunately many of these tasks have never been tested in the same species, let alone within the same subjects, and consequently we have no idea to what extent performance across the various tasks might be related. The present study aims to address this problem by assessing the extent to which some of these different putative tests of episodic cognition assess the same or related cognitive processes in adult humans. To the extent that these different tests do test the same underlying cognitive ability, the prediction is that they will be positively correlated when tested in the same subjects.

Methods of assessing episodic memory

Adult humans are able to verbally report both the content of their memory and their subjective experience of remembering. As such, many researchers have investigated episodic cognition using interview or self-report (e.g., Crovitz and Schiffman 1974; Kopelman et al. 1989; Hashtroudi et al. 1990; Levine et al. 2002; Buckner and Carroll 2007; Hassabis and Maguire 2007; Schacter and Addis 2007). There is little controversy that measurement of accuracy and vividness of mental experience through interview can effectively tap episodic processes (e.g., Barr et al. 1990; Hokkanen et al. 1995; Kapur et al. 1997; Maguire and Mummery 1999; Tanaka et al. 1999; Buckner and Carroll 2007; Hassabis et al. 2007; Addis and Schacter 2008) and these methods are an excellent way in which to assess the nature of episodic cognition, and its neural basis. However, in terms of assessing the existence and/or functionality of episodic cognition in target groups (e.g., patients, children, animals), tests that rely heavily on both receptive and productive verbal competence are not always appropriate. Thus there is a need for behavioral tests of episodic cognition that are less reliant on verbal capabilities.

Laboratory tests of a cognitive process essentially define a behavioral criterion: If a subject demonstrates behavior X, then it can be said to be using mental process Y. To define such a criterion one must have an extremely precise definition of what mental process Y is and how it differs from other processes. However, consensus on a true behavioral criterion for episodic memory has proved difficult. In the first instance, different definitions of episodic memory highlight different defining features. As such, the behavioral criteria born out of these definitions emphasize different features. A related problem is that there is little agreement (yet little explicit debate) over which behavioral criteria can be said to test episodic memory itself, and which test related, yet distinct, processes such as semantic memory, rule-learning, or familiarity. Small differences in definition lead to large differences in behavioral criteria. So what are the features that different theorists have argued to be definitive of episodic memory?

Episodic memory must ... code spatiotemporal relations: The What-Where-When test

Clayton and Dickinson (1998) have argued that a behavioral criterion for episodic memory could be derived from Tulving's original

¹Corresponding author
E-mail nsc22@cam.ac.uk

Article is online at <http://www.learnmem.org/cgi/doi/10.1101/lm.030502.113>.

definition of episodic memory as a system that “receives and stores information about temporally dated episodes or events, and temporal-spatial relations among these events” (Tulving et al. 1972); in other words, as a memory for what happened, where, and when relating to a single past experience (Clayton and Dickinson 1998). In later writings, Clayton and colleagues emphasize that for a What-Where-When (WWW) paradigm to be considered to be testing something resembling episodic memory, the memory demonstrated must be integrated (i.e., the what, where, and when elements must form a single representation) and should be able to be used flexibly to guide behavior (i.e., the subject should represent the event, not a fixed response rule or intention based on the event [Clayton et al. 2003a,b]). Over the past 15 years, the WWW test has been used extensively to assess “episodic-like” memory in the animal cognition literature (e.g., Clayton and Dickinson 1998; Hampton et al. 2005; Babb and Crystal 2006; Naqshbandi et al. 2007; Ferkin et al. 2008; Zinkivskay et al. 2009; Martin-Ordas et al. 2010).

Episodic memory must . . . be internally cued: The Free Recall test

Memory retrieval can occur in response to external cues that trigger the retrieval of a memory (cued recall) or uncued/in response to an internally generated cue (free recall [FR]). Tulving (1985) argued that the contribution of episodic cognition (or rather, the self-knowing “autonoetic” consciousness that he argues accompanies episodic cognition) to memory for items on a list can be assessed by asking subjects if they remembered an item’s occurrence on the list, or whether they simply knew “on some other basis” (Tulving 1985) that the item was on the list. He found that subjects were more likely to report “remembering” items from a word list during a FR task than during a cued recall task (although this difference was small, 88% compared with 75%). Thus he concluded that if the memory was internally cued by some cognitive process then it was more likely to be remembered episodically than if it was externally cued by some information in the environment (such as the first letter of the to-be-remembered word). This finding has led to the conclusion by some theorists (e.g., Perner and Ruffman 1995) that FR tasks tap episodic cognition to a greater extent than cued recall tasks. FR tests are extremely common in the human episodic memory literature (e.g., Tulving 1985; Wheeler et al. 1995; Herlitz et al. 1997; Backman et al. 2001; Howard and Kahana 2002) and are included in many neuropsychological evaluations (e.g., Cognitive Drug Research Battery [CDR], Consortium to Establish a Registry for Alzheimer’s Disease Neuropsychological Test Battery [CERAD]).

In the majority of tests putatively assessing episodic memory, there are focal elements (e.g., words on a word list) that are the center of attention and the subjects are explicitly informed of the impending memory test. These features set those memory tests apart from the majority of scenarios in which episodic memories are encoded and retrieved in everyday life. In reality, memories are often encoded without the encoder’s knowledge that they are later to be tested, and memories must often be retrieved about events that were not the focal element at the time they occurred. Some schools of thought suggest that these features of episodic memory can be considered defining. That is, episodic memory is unique in its catch-all recording of entire events, whether or not they are deliberately memorized or the central focus of attention (e.g., Morris and Frey 1997).

This “automatic encoding” feature of episodic memory has been investigated in two ways. Some researchers (mostly in the animal cognition literature) have emphasized nondeliberate encoding as indicative of episodic memory. In these experiments, the to-be-remembered item is the center of attention at the time of en-

coding, but the subject is not aware that their memory will be tested. Others (mostly in the human literature) have investigated memories for contextual details of a learning event, rather than the “target” item. Here, the to-be-remembered detail is incidental at the time of encoding, but the subject is (usually) aware that a memory test will take place.

Episodic memory must . . . be memory for context: Source memory

Fact memory for focal elements and source memory for contexts have been established as independent (e.g., Johnson and Raye 1981; Shimamura and Squire 1987) suggesting that they represent differential memory systems and may be equivalent to semantic and episodic memory, respectively. Some patients with amnesia have demonstrated relatively preserved memory for facts learned during an experimental session, but severely impaired memory for how and when those facts were learned (e.g., Schacter et al. 1984; Shimamura and Squire 1987). This inability to remember the source of a remembered fact is known as source amnesia. Source memory tasks are now extremely widely used in the human adult (e.g., Shimamura and Squire 1987; Johnson et al. 1993; Simons et al. 2002; Davachi et al. 2003; Lundstrom et al. 2005) as well as the developmental episodic memory literature (Gopnik and Graff 1988; Wimmer et al. 1988; O’Neill and Gopnik 1991; Taylor et al. 1994; Whitcombe and Robinson 2000; Drummey and Newcombe 2002). It could be argued that the difference in memory for focal and contextual features of an event may be due to the deliberateness of encoding, in that focal items are deliberately encoded, while contexts are not.

Episodic memory must . . . not be deliberately encoded: The unexpected question

Zentall and colleagues (2001, 2008) argue that deliberate encoding of stimuli may lead them to be stored as semantic rather than episodic memories: “The critical aspect of the question is that at the time of encoding, there should be no expectation that one would be asked to retrieve the information” (Zentall et al. 2008). However, while there is evidence that deliberate encoding of information may improve recall performance (e.g., Paivio 1971; Bower and Reitman 1972; Craik and Lockhart 1972; Carlson et al. 1976; Greene 1986; Neill et al. 1990) there is little evidence to suggest that deliberate encoding reduces the contribution of episodic memory to recall performance. Shimamura and Squire (1987) demonstrated that amnesic patients had comparable deficits in fact and source memory, regardless of whether the fact learning and testing situations were explicit (i.e., subjects were taught a fact and then tested for the fact and when they had learned it) or incidental. This finding suggests that knowledge of the impending memory tests did not reduce the impairment of the amnesic patients on either fact or source memory, nor did it increase the relative performance of the controls. Nonetheless, the unexpected question paradigm has been regularly used by some researchers in animal cognition (Zentall et al. 2001, 2008; Singer and Zentall 2007; Fujita et al. 2012; Zhou et al. 2012).

Different ways of testing the same thing?

The behavioral tests reviewed above differ in the criteria they set with which to measure episodic memory, and thus it is difficult to compare the results they produce. Nonetheless all of these tasks have been widely used for many years, and many literature reviews discuss studies using different methodologies as if they were equivalent. Only recently have researchers begun to assess the extent to which these tests assess the same ability.

Plancher and colleagues (2010) used memories of a tour around a virtual town to assess What-Where-When (WWW) memory in young and elderly adults, and compared performance on this test to Free Recall (FR) and self-reported memory complaints in everyday life (i.e., a tendency to lose keys, forget names, etc.). The authors found that WWW performance did not correlate with FR, and that levels of memory complaint in everyday life correlated significantly with WWW binding ability, but not with individual elements (what/where/when) or with FR. This emphasis on the binding of components corresponds with Clayton and colleagues' (2003b) arguments that it is not the content but the integration of the what, where, and when elements that makes the memory episodic. Furthermore, the discovery that it is the binding of contextual features that is the feature most correlated with experience of memory problems in everyday life is consistent with the finding that episodic amnesics are impaired on object-location, object-order, and object-person-binding memory, but not on object recognition (Burgess et al. 2002).

Holland and Smulders (2011) found that people's accuracy in remembering where they had hidden two different types of coin (what) on two consecutive days (when) was related to their ability to remember incidental features of the hiding episodes (i.e., an Unexpected Question/Source Memory [UQ/SM] test). They also found that subjects were generally more likely to associate their WWW memories with the experience of "remembering" rather than the feeling of "knowing." This pattern was replicated by Easton and colleagues (2012), who found that subjects were significantly more likely to report "remembering" than "knowing" when they had previously seen a specific object–location–context combination.

The aim of this study is to investigate the relationship between performance on three putative tests of episodic memory (What-Where-When, Unexpected Question/Source Memory, and Free Recall) in order to establish whether they tap into the same underlying process. The WWW task was a computerized task in which subjects were asked to hide coins of different types at different times and were later asked to indicate the location of a particular type of coin which they hid at a particular time. The FR task followed the standard procedure of an aurally presented word list that must be repeated after a delay. The UQ/SM test involved a series of questions relating to contextual/incidental features of the other experiments and was administered unexpectedly at the end of the study. To investigate the potential contribution of episodic memory to performance on any or all of the tests, subjects were asked to report on what "strategies" they used in performing these memory tests. It was predicted that those individuals who performed better on one test would also perform better on the others, and that the extent of the relationship would be predicted by the subjects' self-reported use of "episodic" strategies.

Results

Performance for each of the tests (WWW, UQ/SM, and FR) is reported, followed by the relationships between performances on the different tests.

Gender effects

Men and women did not differ in their performance on any of the tests (FR, $t_{(75)} = -1.061$, $P = 0.292$; UQ/SM, $t_{(75)} = 1.441$, $P = 0.154$; WWW, $t_{(65)} = 1.471$, $P = 0.145$). As such, all of the following analyses were conducted with data from men and women combined.

What-Where-When

Subjects had an average integrated WWW score of 0.73 (± 0.242)—equivalent to correctly locating 3.6 of the five target coins. In terms of the individual what, where, and when elements, subjects on average correctly identified gold coins ("what") with a score of 0.8 (± 0.217), correctly identified coins from the first hiding episode ("when") with a score of 0.91 (± 0.188), and correctly located coins ("where") with a score of 0.9 (± 0.164). Thus subjects were generally very successful at the individual elements of the WWW test, but less successful at integrating this information. Subjects were differentially successful at the different elements of the test (Friedman's ANOVA, $F_{(2)} = 16.39$, $P < 0.001$). Planned contrasts (related samples Wilcoxon) revealed that subjects were significantly worse at identifying the type of coin ("what") than at identifying a coin from the correct hiding period ("when," $W = 487$, $P = 0.001$) or identifying coin location ("where," $W = 486$, $P = 0.015$), but that these latter two elements did not differ from each other ($W = 245$, $P = 0.322$).

Subjects reported their strategies for remembering the location of the treasure, although it was often difficult to categorize these reports, given that many of them could be either semantic or episodic. Conservatively, only "re-experiencing" and "narrative" strategies were termed episodic; 30% of subjects reported using such methods. Overall, 48% of subjects reported hiding coins in a specific order and then using the order as a retrieval cue, 35% reported using landmarks, 26% reported "re-experiencing" the hiding event, 19% reported matching the color of the background to the color of the coin when hiding, and thus identifying location–identify combinations according to color, 4% reported creating a story out of the locations, and 4% reported using the geometry of the screen to identify locations. There was no significant difference in performance between those reporting "episodic" strategies (re-experiencing or story creation) and those not ($t_{(69)} = 0.887$, $P = 0.378$).

Unexpected Question/Source Memory

Subjects scored an average of 0.61 (± 0.119) in the UQ/SM test, which is equivalent to 6.7 correct answers out of the 11 questions.

In the debrief, 87% of subjects reported using an episodic strategy in the UQ/SM test, while only 27% reported using a semantic strategy. Specifically, 80% of subjects reported visualizing the context they were attempting to remember, 40% reported guessing, 19% "just knew" the answers, 10% reported using logical inference to infer the answers, and 6% reported attempting to visualize, but being unable to. Because only four people did not report at least one episodic strategy, it was not possible to compare performances of those reporting episodic strategies and those reporting only semantic strategies or guessing.

Free Recall

On average, the subjects achieved a score of 0.389 (± 0.146) on the Free Recall test. This is equivalent to recalling approximately 11 words out of a possible 28.

In the debrief, 53% of subjects reported using an episodic strategy while 48% reported using a semantic strategy (subjects could report as many strategies as they felt they used, so the same subjects contribute to multiple categories). Specifically, 40% reported visualizing the words, 35% reported linking the words together, 20% of people reported mentally hearing the experimenter's voice, 14% reported that they "just knew" that the words had appeared on the list, 10% reported creating a narrative out of the words, 5% reported sorting the words into categories

and using these as retrieval cues, and 5% reported remembering their own thoughts in response to hearing the words. There was no significant difference in performance between people reporting (any) episodic and those reporting (only) semantic strategies ($t_{(38)} = -0.230, P = 0.819$).

Memory tests overall

The subjects' scores on all three memory tests were averaged into a single "memory score," influenced equally by performance on all three tests. The subjects' self-reported strategies were grouped across the memory tests and scored according to the number of tests in which subjects reported using episodic strategies (1, 2, or 3). This was used as a metric of "tendency to use episodic cognition." There was no "0" score because only two subjects never reported using an episodic strategy in any of the tests. There was no effect of the tendency to use episodic strategies on the overall memory score ($F_{(3)} = 0.845, P = 0.474$). This suggests that those subjects reporting episodic strategies were neither more nor less successful overall in the memory tests.

The relationship between memory tests

Figure 1 illustrates the relationships between performances on the different memory tests. The graphs appear to illustrate a linear correlation between WWW and FR, and a quadratic relationship between WWW and UQ/SM. The graph indicates no clear relationship between FR and UQ/SM. Statistical analysis supports this interpretation. Relationships between the memory tests were assessed using Pearson's correlations to assess linear relationships and using univariate general linear models (GLMs) to assess non-linear relationships. There was a significant positive correlation between WWW and FR ($R = 0.233, P = 0.041$), which was driven by subjects who reported episodic strategies in the FR test (episodic, $R = 0.514, P = 0.014$; semantic, $R = -0.344, P = 0.163$) (see Fig. 1). There was no correlation between UQ/SM and either FR or WWW (FR-UQ/SM, $R = -0.124, P = 0.281$; WWW-UQ/SM, $R = -0.024, P = 0.833$).

The lack of correlation between the performance in the WWW and UQ/SM tests may have been due to a nonlinear relationship. Univariate GLM revealed a quadratic relationship between WWW and UQ/SM scores (UQ/SM, $F_{(1,74)} = 3.733, P =$

0.057 ; $UQ/SM^2, F_{(1,74)} = 4.343, P = 0.041$). This suggests that subjects who performed very well or very badly on the WWW test perform poorly on the UQ/SM test, while subjects who performed moderately on the WWW test perform well on the UQ/SM test. By contrast, there was no quadratic relationship between FR and UQ/SM (UQ/SM, $F_{(1,74)} = 0.660, P = 0.419$; $UQ/SM^2, F_{(1,74)} = 0.713, P = 0.401$) overall, or when only subjects who reported episodic strategies in the FR test were included (UQ/SM, $F_{(1,74)} = 0.023, P = 0.881$; $UQ/SM^2, F_{(1,74)} = 0.022, P = 0.884$).

Discussion

A sample of undergraduates was tested on three putative episodic memory tests (WWW, UQ/SM, and FR). The only significant correlation was between WWW and FR. There was a possible quadratic relationship between WWW and UQ/SM, suggesting that those who performed very well or very badly on WWW were less accurate on UQ/SM than those who performed moderately. There was no relationship between UQ/SM and FR. The finding that WWW did not correlate with UQ/SM is in contrast to previous findings that WWW performance correlated positively with UQ/SM (Holland and Smulders 2011). However, there are a number of methodological differences between that study and the one presented here that may have resulted in this contrasting finding (for example, the unexpected questions in Holland and Smulders's study were in a yes/no format, while in the present study they were open-ended).

According to subjects' self-reports, they used an almost equal mix of episodic and semantic strategies in the FR test. This is in contrast to UQ/SM where almost all subjects reported using an episodic strategy. In none of the memory tests was there a significant effect of reporting using an episodic strategy on performance. This result is similar to that demonstrated by Easton and colleagues (2012), who found that when subjects were asked to recall when they had previously seen a particular object–location combination, their answers were significantly more likely to be "remembered" than "known." However, whether subjects reported "remembering" or "knowing" the answer did not have any effect on their performance. Interestingly, reports of "knowing" in the current study were relatively common in the FR and UQ/SM tests (14% and 19%, respectively) but not in the WWW,

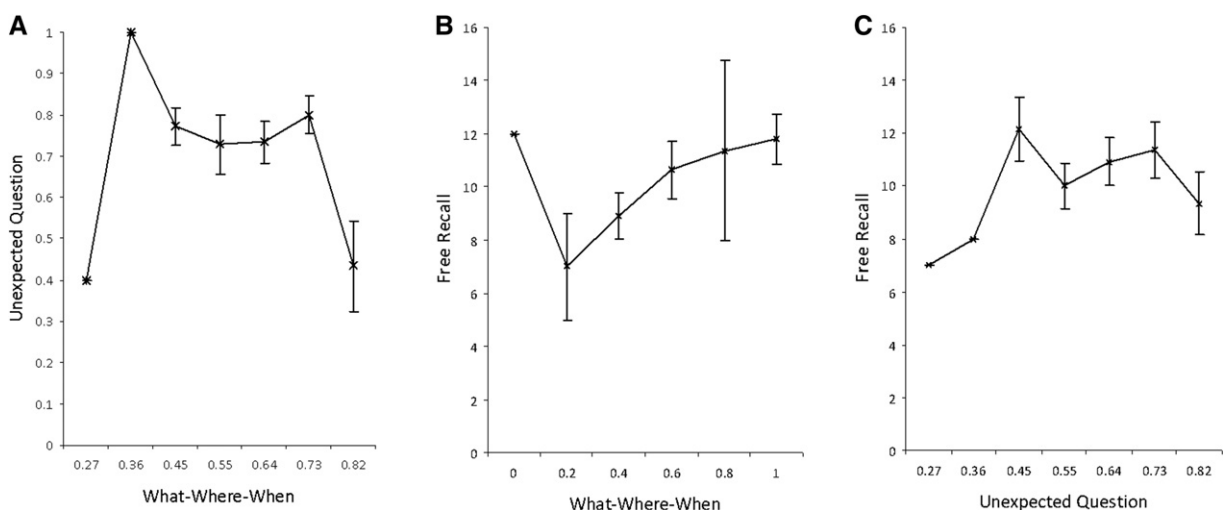


Figure 1. Relationship between memory tests. (A) Relationship between What-Where-When (WWW) and Unexplained Question/Source Memory (UQ/SM). (B) Relationship between WWW and Free Recall (FR). (C) Relationship between UQ/SM and FR. Error bars represent standard error.

in which “knowing” was not reported by any subject. This supports the findings of Holland and Smulders (2011) and Easton and colleagues (2012) that WWW tests are overwhelmingly reported as “remembered” rather than “known” (when given the option only between these two reports). However, this may also suggest that the type of memories tapped by WWW tests do not lend themselves to being reported using the word “known,” but that this does not necessarily mean they are episodically “remembered”; simply that “known” is not an appropriate word to describe the experience.

It is interesting to note that the relationship between WWW and UQ/SM was U-shaped. Such a pattern of results may indicate the contribution of multiple factors to performance. Episodic memories are constrained and specified (“scaffolded”) by known facts about the world. For example, if someone knows that there were 28 words on a word list, they are more likely to attempt to recall all 28 words—potentially resulting in recall of more words than if they were unaware of the number of words on the list. In both the WWW and FR tests (but not the UQ/SM test) subjects had the opportunity to use encoding strategies to aid their memories. The use of deliberate encoding strategies may have increased the contribution of such semantic “scaffolding.” For example, subjects may have mentally categorized the words, leading them to know that there were two words relating to body parts and attempt to remember both. In contrast, those engaged in answering an unexpected question on a single unusual event (hiding virtual treasure on a computer-generated island) have very little semantic scaffold (that is, few facts) to support their memory. Specifically, there is both no opportunity to deliberately encode semantic scaffolds (such as rules about where certain items are located) nor (in this study, at least) is there much of a knowledge base to rely upon (there is not, for example, a “usual” number of birds in a virtual beach scene). This test, however, is far more likely to be affected by subjects not encoding items in the first place, for example because they didn’t pay attention to the item. The varying degree of contribution from semantic knowledge to test performance is likely to affect the degree of relationship between different tasks and contribute to the U-shaped relationship we see between WWW and UQ/SM.

One possible reason for the low level of correlations between tests is that there may have been low intra-test consistency (i.e., low reliability). This is a recurring problem with tests of memory (e.g., Dikmen et al. 1999), suggesting that even different items within the same memory test may be testing the same thing theoretically, but not empirically. Unfortunately, reliability measures could not be taken for the tests used in this study due to low item-numbers and single-block testing. However, even if we assume that each test had a level of reliability comparable to standard verbal FR tests (i.e., around 0.6–0.7 [Lowe and Rabbitt 1998]) this would be in the range of “poor” to “questionable” by standard rules of thumb (Kline 1999). Thus weak and absent correlations between memory tests may be the result of weak or absent correlations within memory tests. Such a possibility does not weaken, but rather adds to the argument that focus is needed to improve the consistency of episodic memory measures, both within and between tests.

The different tests explored in this study have arisen out of different perspectives on what the “defining features” of episodic memory are, and as such they explicitly test different mnemonic skills. The WWW test explicitly requires the ability to bind spatio-temporal features. This binding process has been shown to be specifically affected by the aging process and is predictive of self-reported everyday memory impairment in a way that individual what, where, and when elements, as well as word-list FR, are not (Plancher et al. 2010). By contrast, the UQ/SM task explicitly requires the ability to reanalyze previous experiences for new in-

formation, something not required by the other tasks. Finally, the FR task assesses the ability to mentally initiate and guide retrieval in the absence of external cueing (as is present in both the WWW and UQ/SM tasks). While these different elements are by no means the only important features of episodic memory (others include, for example, reality monitoring and emotion memory), they are all integral elements of episodic memory as it is used in everyday life. They also cannot, by definition, all be assessed by the same task. As such, while the small number and potential nonlinearity of relationships between these different tasks may be considered a weakness of the literature at present, it has the potential to become a great strength. If these tasks can be developed, along with others, into a battery of episodic cognition tests, then this might have great potential for assessing the functionality of an extremely multifaceted cognitive ability. Such a battery would be capable of assessing performance on different types or components of episodic memory and might therefore have significantly more ability to predict deficits in everyday functioning than any one test alone. Such a prospect relies on the assumption that these tests are all, in their own way, measures of some “core” episodic memory ability. The self-report judgments described here, as well as the remember/know judgments in previous studies (e.g., Tulving 1985; Holland and Smulders 2011; Easton et al. 2012; Mickes et al. 2013) provide hope that this may, indeed, be the case, although further research is needed.

In summary, the findings presented here demonstrate some relationships between the different types of episodic memory test (WWW, UQ/SM, and FR). Figure 1 indicates that FR was positively correlated with WWW, and WWW had a quadratic relationship with UQ/SM, but that there was no relationship between FR and UQ/SM. The presence of quadratic relationships may be indicative of the contribution of multiple factors. Such a small degree of relationship between these tasks does not encourage confidence that these tests are assessing the exact same underlying cognitive process. Given these results, researchers of episodic cognition should proceed with caution when directly comparing literature using different methodologies. Which test is the “true” test of episodic cognition is yet to be determined. Indeed, it is likely that such a thing does not and cannot exist. Thus it is possible that, as in the field of IQ, research emphasis should be placed not on a particular “true” test of this multifaceted cognitive ability, but on developing a battery of tests that encompass all its diverse qualities. Such a battery may make it possible for non/less verbal tests of episodic memory to provide a similar level of assessment as interview/self-report-based assessments.

Materials and Methods

The experiments took place in a room in the Department of Experimental Psychology, Cambridge, in the spring of 2011. The department was renamed Department of Psychology in October 2012.

Subjects

The sample consisted of 77 subjects, of whom 42 were male. Subjects were undergraduates at the University of Cambridge, aged between 18 and 23 years.

Memory tests

What–Where–When

The WWW test was conducted in the form of a computer game. Participants took on the character of “Swashbuckle,” a pirate

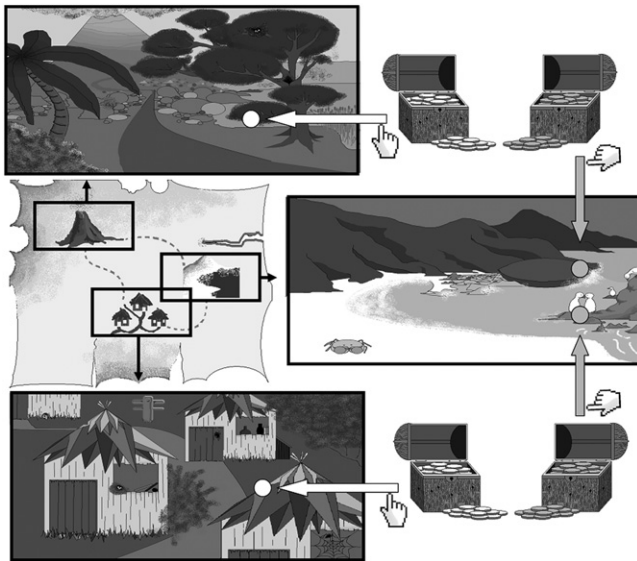


Figure 2. Schematic of WWW pirate game hiding phase. Clicking on any of the symbols on the map takes the subject to one of the three scenes: mountain (*top*), beach (*right*), or village (*bottom*). At the *bottom* of each scene were two treasure chests from which subjects could drag and drop gold or silver coins into the scene to hide them.

who has run aground on a desert island and must hide all his treasure before his evil rival “Pinkbeard” arrives to steal it. To hide the treasure, subjects could navigate between three locations on a “treasure map”: the beach, the village, and the mountain. Clicking on any of these locations took the subject to the appropriate “scene,” where they could then hide the treasure. The treasure was presented in two virtual treasure chests, one containing gold coins and one containing silver coins, from which subjects could drag and drop individual coins into virtual scenes, where they would disappear (see Fig. 2). Before beginning to hide the treasure, subjects were informed that Swashbuckle would have to find the treasure again, and thus to try to remember where it was hidden. After the subjects had hidden half of their treasure (five gold and five silver coins) they were informed that the sun was going down and they would have to continue “in the morning.” There followed a delay in which subjects undertook a planning task (reported elsewhere). They were then told that it was now morning and that they should finish hiding the treasure. When all the treasure was hidden, it was revealed that Swashbuckle’s treasure was, in fact, stolen from a leprechaun, meaning that the gold coins would disappear if left outside overnight. The subjects were then asked to identify on each scene the locations (“where”) gold coins (“what”) had been hidden before Swashbuckle slept (“when”) (i.e., the treasure that would have disappeared). They did so by clicking on the area of the map that they thought they had hidden it, thus producing an “X” to mark the spot. Subjects made five “X marks the spot” judgments in total.

The accuracy of subjects’ coin location was measured in terms both of integrated WWW and individual what, where, and when elements. The integrated WWW score was calculated by assessing whether there was a “target” coin (i.e., a gold coin from the first hiding period) within 60 pixels (two coin widths) of the subject’s “X marks the spot.” The individual “What” and “When” elements were coded in terms of the identity (what) and origin (when) of the nearest coin to the subject’s “X marks the spot.” Thus if the nearest coin was gold, subjects were coded as correct for “what” and if the nearest coin was hidden during the first hiding session they were coded as correct for “when.” The “where” element was calculated in terms of absolute distance from any coin. If there was a coin within 60 pixels of the subject’s

Table 1. “Unexpected” questions asked at the end of the session

1. What color was Swashbuckle’s shirt?
2. Which of Pinkbeard’s shoulders was his parrot sitting on?
3. Was the beach on the east or west of the island?
4. Of the three scenes, how many contained a bird’s nest?
5. How many trees were there in the volcano scene?
6. Was the lake on the left or the right of the screen in the volcano scene?
7. Which animal was there in the village?
8. What color were the roofs of the huts in the village?
9. How many birds were there in the beach scene?
10. What color were the birds in the beach scene?
11. How many starfish were there in the beach scene?

“X marks the spot,” they were coded as correct for “where.” To prevent subjects from reporting the same coin multiple times, individual coins that had contributed to the score (either integrated or individual) on a previous “X marks the spot” judgment it was discounted from subsequent judgments and the next nearest coin was considered.

Unexpected Question/Source Memory

Participants were unexpectedly asked 11 questions about aspects of the WWW test at the end of the computer session (see Table 1). Questions were relatively evenly distributed between asking identity-related questions (e.g., “What animal was there in the village scene?”), spatial questions (e.g., “Which of the pirate’s shoulders was the parrot sitting on?”), and number questions (e.g., “How many birds were there in the beach scene?”). This was designed such that the questions did not differ in content from the WWW questions, but simply differed in the fact that they (1) were unexpected, (2) did not explicitly require integration of memory for what, where, and when, and (3) were assessing memory for nonfocal features of the event. For our purposes, contextual or “nonfocal” elements were those elements of the scene which were not central to the task that the subject was undergoing at the time of encoding (WWW). Any of these “contextual” items could have been removed from the WWW task altogether and the subject could still have completed it. In contrast, a “focal item” was defined as something the subject had to attend to in order to solve the WWW task, for example, the color of coin, where they hid it, or when. The questions were open-ended; subjects typed their answers into blank boxes. To make sure all subjects answered all questions, participants were not permitted to continue until they had typed an answer in every box.

Table 2. Fourteen high-concreteness ($m = 6.49$), high-imagery ($m = 6.39$) words and 14 low-concreteness ($m = 2.44$), low-imagery ($m = 1.72$) words from Paivio et al. (1968)

High-concreteness, high-imagery words	Low-concreteness, low-imagery words
Umbrella	Truth
Mountain	Tendency
Orchestra	Thought
Kettle	Misconception
Magazine	Intellect
Ink	Idea
Hotel	Gist
Forehead	Essence
Garden	Fate
Elbow	Ego
Engine	Criterion
Book	Disposition
Car	Concept
Clock	Attitude

Table 3. Self-report questions asked to subjects for each memory test

Test	Question
What-Where-When	"How do you think you went about identifying where you hid the coins in the pirate game?"
Unexpected Question/ Source Memory	"How do you think you went about answering the questions presented on the screen?"
Free Recall	"How do you think you recalled the words from the word list?"

Free Recall

A list of 28 words was selected from a list of 925 nouns that were rated for abstractness and imagery (Paivio et al. 1968) (see Table 2). Fourteen words were selected for scoring highly in "imagery" and "concreteness" while 14 words were selected for having low "imagery" and "concreteness" scores.

Participants were read the word list and then immediately asked to recall it. They were then read the list in a different order and informed that they would be asked to recall the list again at the end of the experiment. The retention interval lasted ~30 min, in which subjects undertook the computerized tests (WWW, UQ/SM, and a planning task).

Procedure

The WWW and UQ/SM were computer based and in the form of video games. The FR test was administered orally by the experimenter.

The procedure was a nested design in which the phases of the different tests formed the retention interval for the other tests. Thus every subject received the tests in a set order. Some of the retention intervals were spent undergoing a future-planning task (reported elsewhere). The subjects underwent the encoding phase of the FR test before starting the computer-based tests. They then undertook the WWW test. Between the first and second encoding episode of the WWW test, subjects undertook the planning task. The WWW test phase occurred immediately after the second encoding phase. The final section of the computer game was the UQ/SM test. Finally, the subjects were asked to verbally recall the words that were read to them at the beginning of the ses-

sion. The subjects were then debriefed about the nature and purpose of the experiment, during which they were asked to give their consent for the use of their data and to report on how they went about the different memory tests (see Table 3). Their answers were coded into categories that captured the essence of the report (Table 4). For FR and UQ/SM, these were then further coded into episodic and semantic strategies. For the WWW test, many of the different strategies could be interpreted as either episodic or semantic. Interestingly, many strategies offered by subjects involved methods both of encoding and retrieval, and often the encoding and retrieval elements of the strategies could not be separated.

As such the strategies are not subdivided into encoding and retrieval strategies, but coded as episodic or semantic on the basis of whether it involved the phenomenology of re-experiencing or added spatiotemporal information to the memory at encoding (such as creating a "narrative" from the hiding period; this included memory palaces and method of loci techniques).

Analyses

To make them comparable, scores on all memory tests were recalculated as a proportion, such that the maximum score was 1 and the minimum score was 0. Where assumptions of normality were met, data were analyzed using paired and independent samples *t*-tests and Pearson's correlation. Where the data did not meet assumptions of normality, Friedman's ANOVA, related-sample Wilcoxon, and Kendall's τ correlations were used. Nonlinear relationships between tests were assessed using a univariate general linear model (GLM).

Acknowledgments

We thank the research students who were involved in this project, Mathilda Hay and Stephanie Bailey. We also thank Netta Chachuma and Francesca Lewis for help on an early version of this study, and James Thom, Ljerka Ostojic, Jon Simons, and Tom Smulders for commenting on versions of the manuscript, and Daniel Booth and Anthony Cheke for proofreading. Special thanks to Chris Stephenson for creating, maintaining, and updating the computer-based task. L.G.C. was funded by an MRC doctoral training studentship, an MRC Centenary Early Career Award, and a Junior Research Fellowship at Girton College, Cambridge.

Table 4. Categories of self-reported memory strategies

Free Recall	Unexpected Question	What-Where-When
Heard	Visualized	Order
Felt that they could hear experimenters voice in their head	Felt that they could see the scene in their head	Hid the coins in a specific order and used this rule to infer "what" and "when"
Visualized	Failed visualization	Color-matching
Felt that they could see the words in their head	Attempted to see the scene in their head, but couldn't	Hid coins in areas of a similar color and thus used color as a retrieval cue for "what" and "where"
Categorized	Guessed	Landmark
Categorized words into different types	Aware that answer was total guess	Used salient landmarks when hiding and retrieving
Linked	Semantics	Revisited
Linked words together	Used logical inference to reach answer	Mentally "went back" to the hiding event
Knew	Knew	Story/memory palace
"Just knew"	"Just knew"	Made the hiding event into a narrative
Made a story		Geometry
Turned the word list into a narrative		Used the geometry of the screen when hiding and retrieving
Remembered own thoughts		
Remembered what they were thinking at the time		

These were categorized by two coders with 91% agreement. Those presented in pale gray were coded as episodic strategies and those presented in darker gray were coded as semantic strategies, while those presented in white were coded as either/neither.

References

- Addis DR, Schacter DL. 2008. Constructive episodic simulation: Temporal distance and detail of past and future events modulate hippocampal engagement. *Hippocampus* **18**: 227–237.
- Babb SJ, Crystal JD. 2006. Episodic-like memory in the rat. *Curr Biol* **16**: 1317–1321.
- Backman L, Small BJ, Fratiglioni L. 2001. Stability of the preclinical episodic memory deficit in Alzheimer's disease. *Brain* **124**: 96–102.
- Barr WB, Goldberg E, Wasserstein J, Novelly RA. 1990. Retrograde-amnesia following unilateral temporal lobectomy. *Neuropsychologia* **28**: 243–255.
- Bower GH, Reitman JS. 1972. Mnemonic elaboration in multilist learning. *J Verb Learn Verb Behav* **11**: 478–485.
- Buckner RL, Carroll DC. 2007. Self-projection and the brain. *Trends Cogn Sci* **11**: 49–57.
- Burgess N, Maguire EA, O'Keefe J. 2002. The human hippocampus and spatial and episodic memory. *Neuron* **35**: 625–641.
- Carlson RF, Kincaid JP, Lance S, Hodgson T. 1976. Spontaneous use of mnemonics and grade-point average. *J Psychol* **92**: 117–122.
- Clayton NS, Dickinson A. 1998. Episodic-like memory during cache recovery by scrub jays. *Nature* **395**: 272–274.
- Clayton NS, Bussey TJ, Dickinson A. 2003a. Can animals recall the past and plan for the future? *Nat Rev Neurosci* **4**: 685–691.
- Clayton NS, Bussey TJ, Emery NJ, Dickinson A. 2003b. Prometheus to Proust: The case for behavioural criteria for 'mental time travel'. *Trends Cogn Sci* **7**: 436–437.
- Craik FIM, Lockhart RS. 1972. Levels of processing—framework for memory research. *J Verb Learn Verb Behav* **11**: 671–684.
- Crovitz HF, Schiffman H. 1974. Frequency of episodic memories as a function of their Age. *B Psychonomic Soc* **4**: 517–518.
- Davachi L, Mitchell JP, Wagner AD. 2003. Multiple routes to memory: Distinct medial temporal lobe processes build item and source memories. *Proc Natl Acad Sci* **100**: 2157–2162.
- Dikmen SS, Heaton RK, Grant I, Temkin NR. 1999. Test–retest reliability and practice effects of expanded Halstead–Reitan neuropsychological test battery. *J Int Neuropsychol Soc* **5**: 346–356.
- Drumme AB, Newcombe NS. 2002. Developmental changes in source memory. *Developmental Sci* **5**: 502–513.
- Easton A, Webster LA, Eacott MJ. 2012. The episodic nature of episodic-like memories. *Learn Mem* **19**: 146–150.
- Ferkin MH, Combs A, delBarco-Trillo J, Pierce AA, Franklin S. 2008. Meadow voles, *Microtus pennsylvanicus*, have the capacity to recall the "what", "where", and "when" of a single past event. *Anim Cogn* **11**: 147–159.
- Fujita K, Morisaki A, Takaoka A, Maeda T, Hori Y. 2012. Incidental memory in dogs (*Canis familiaris*): Adaptive behavioral solution at an unexpected memory test. *Anim Cogn* **15**: 1055–1063.
- Gopnik A, Graff P. 1988. Knowing how you know: Young children's ability to identify and remember the sources of their beliefs. *Child Dev* **59**: 1366–1371.
- Greene RL. 1986. Word stems as cues in recall and completion tasks. *Q J Exp Psychol-A* **38**: 663–673.
- Hampton RR, Hampstead BM, Murray EA. 2005. Rhesus monkeys (*Macaca mulatta*) demonstrate robust memory for what and where, but not when, in an open-field test of memory. *Learn Motiv* **36**: 245–259.
- Hashtroudi S, Chrosniak LD, Johnson MK. 1990. Aging and qualitative characteristics of memories for perceived and imagined complex events. *Psychol Aging* **5**: 119–126.
- Hassabis D, Maguire EA. 2007. Deconstructing episodic memory with construction. *Trends Cogn Sci* **11**: 299–306.
- Hassabis D, Kumaran D, Vann SD, Maguire EA. 2007. Patients with hippocampal amnesia cannot imagine new experiences. *Proc Natl Acad Sci* **104**: 1726–1731.
- Herlitz A, Nilsson LG, Backman L. 1997. Gender differences in episodic memory. *Mem Cognit* **25**: 801–811.
- Hokkanen L, Launes J, Vataja R, Valanne L, Iivanainen M. 1995. Isolated retrograde-amnesia for autobiographical material associated with acute left temporal-lobe encephalitis. *Psychol Med* **25**: 203–208.
- Holland SM, Smulders TV. 2011. Do humans use episodic memory to solve a What-Where-When memory task? *Anim Cogn* **14**: 95–102.
- Howard MW, Kahana MJ. 2002. When does semantic similarity help episodic retrieval? *J Mem Lang* **46**: 85–98.
- Johnson MK, Raye CL. 1981. Reality monitoring. *Psychol Rev* **88**: 67–85.
- Johnson MK, Hashtroudi S, Lindsay DS. 1993. Source monitoring. *Psychol Bull* **114**: 3–28.
- Kapur N, Millar J, Colbourn C, Abbott P, Kennedy P, Docherty T. 1997. Very long-term amnesia in association with temporal lobe epilepsy: Evidence for multiple-stage consolidation processes. *Brain Cogn* **35**: 58–70.
- Kline P. 1999. *Handbook of psychological testing*. Routledge, London.
- Kopelman MD, Wilson BA, Baddeley AD. 1989. The autobiographical memory interview—a new assessment of autobiographical and personal semantic memory in amnesic patients. *J Clin Exp Neuropsychol* **11**: 724–744.
- Levine B, Svoboda E, Hay JF, Winocur G, Moscovitch M. 2002. Aging and autobiographical memory: Dissociating episodic from semantic retrieval. *Psychol Aging* **17**: 677–689.
- Lowe C, Rabbitt P. 1998. Test/re-test reliability of the CANTAB and ISPOCD neuropsychological batteries: Theoretical and practical issues. Cambridge Neuropsychological Test Automated Battery. International Study of Post-Operative Cognitive Dysfunction. *Neuropsychologia* **36**: 915–923.
- Lundstrom BN, Ingvar M, Petersson KM. 2005. The role of precuneus and left inferior frontal cortex during source memory episodic retrieval. *Neuroimage* **27**: 824–834.
- Maguire EA, Mummery CJ. 1999. Differential modulation of a common memory retrieval network revealed by positron emission tomography. *Hippocampus* **9**: 54–61.
- Martin-Ordas G, Haun D, Colmenares F, Call J. 2010. Keeping track of time: Evidence for episodic-like memory in great apes. *Anim Cogn* **13**: 331–340.
- Mickes L, Seale-Carlisle TM, Wixted JT. 2013. Rethinking familiarity: Remember/know judgments in free recall. *J Mem Lang* **68**: 333–349.
- Morris RGM, Frey U. 1997. Hippocampal synaptic plasticity: Role in spatial learning or the automatic recording of attended experience? *Philos Trans R Soc B* **352**: 1489–1503.
- Naqshbani M, Feeney MC, McKenzie TL, Roberts WA. 2007. Testing for episodic-like memory in rats in the absence of time of day cues: Replication of Babb and Crystal. *Behav Processes* **74**: 217–225.
- Neill WT, Beck JL, Bottalico KS, Molloy RD. 1990. Effects of intentional versus incidental-learning on explicit and implicit tests of memory. *J Exp Psychol Learn* **16**: 457–463.
- O'Neill DK, Gopnik A. 1991. Young children's ability to identify the sources of their beliefs. *Dev Psychol* **27**: 390–397.
- Paivio A. 1971. *Imagery and verbal processes*. Holt, New York.
- Paivio A, Yuille JC, Madigan SA. 1968. Concreteness, imagery, and meaningfulness values for 925 nouns. *J Exp Psychol* **76**: 1–25.
- Perner J, Ruffman T. 1995. Episodic memory and autoegetic consciousness: Developmental evidence and a theory of childhood amnesia. *J Exp Child Psychol* **59**: 516–548.
- Plancher G, Gyselinck V, Nicolas S, Piolino P. 2010. Age effect on components of episodic memory and feature binding: A virtual reality study. *Neuropsychology* **24**: 379–390.
- Schacter DL, Addis DR. 2007. The cognitive neuroscience of constructive memory: Remembering the past and imagining the future. *Philos Trans R Soc Lond B Biol Sci* **362**: 773–786.
- Schacter DL, Harbluk JL, Mclachlan DR. 1984. Retrieval without recollection—an experimental analysis of source amnesia. *J Verb Learn Verb Behav* **23**: 593–611.
- Shimamura AP, Squire LR. 1987. A neuropsychological study of fact memory and source amnesia. *J Exp Psychol Learn* **13**: 464–473.
- Simons JS, Verfaellie M, Galton CJ, Miller BL, Hodges JR, Graham KS. 2002. Recollection-based memory in frontotemporal dementia: Implications for theories of long-term memory. *Brain* **125**: 2523–2536.
- Singer RA, Zentall TR. 2007. Pigeons learn to answer the question "where did you just peck?" and can report peck location when unexpectedly asked. *Learn Behav* **35**: 184–189.
- Suddendorf T, Corballis MC. 2007. The evolution of foresight: What is mental time travel, and is it unique to humans? *Behav Brain Sci* **30**: 299–312.
- Tanaka Y, Miyazawa Y, Hashimoto R, Nakano I, Obayashi T. 1999. Postencephalitic focal retrograde amnesia after bilateral anterior temporal lobe damage. *Neurology* **53**: 344–350.
- Taylor M, Esbensen BM, Bennet RT. 1994. Children's understanding of knowledge acquisition: The tendency for children to report that they have always known what they have just learned. *Child Dev* **65**: 1581–1604.
- Tulving E. 1985. Memory and consciousness. *Can Psychol* **26**: 1–12.
- Tulving E, Donaldson W, Bower GH, United States Office of Naval Research. 1972. *Organization of memory*. Academic Press, New York.
- Wheeler MA, Stuss DT, Tulving E. 1995. Frontal lobe damage produces episodic memory impairment. *J Int Neuropsychol Soc* **1**: 525–536.
- Whitcombe EL, Robinson EJ. 2000. Children's decisions about what to believe and their ability to report the source of their belief. *Cogn Dev* **15**: 329–346.
- Wimmer H, Hogrefe J, Perner J. 1988. Children's understanding of informational access as source of knowledge. *Child Dev* **59**: 386–396.
- Zentall TR, Clement TS, Bhatt RS, Allen J. 2001. Episodic-like memory in pigeons. *Psychon Bull Rev* **8**: 685–690.
- Zentall TR, Singer RA, Stagner JP. 2008. Episodic-like memory: Pigeons can report location pecked when unexpectedly asked. *Behav Processes* **79**: 93–98.
- Zhou W, Hohmann AG, Crystal JD. 2012. Rats answer an unexpected question after incidental encoding. *Curr Biol* **22**: 1149–1153.
- Zinkivskay A, Nazir F, Smulders TV. 2009. What-Where-When memory in magpies (*Pica pica*). *Anim Cogn* **12**: 119–125.

Received January 23, 2013; accepted in revised form June 12, 2013.



Do different tests of episodic memory produce consistent results in human adults?

Lucy G. Cheke and Nicola S. Clayton

Learn. Mem. 2013 20: 491-498

Access the most recent version at doi:[10.1101/lm.030502.113](https://doi.org/10.1101/lm.030502.113)

References

This article cites 62 articles, 7 of which can be accessed free at:
<http://learnmem.cshlp.org/content/20/9/491.full.html#ref-list-1>

Creative Commons License

This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first 12 months after the full-issue publication date (see <http://learnmem.cshlp.org/site/misc/terms.xhtml>). After 12 months, it is available under a Creative Commons License (Attribution-NonCommercial 3.0 Unported), as described at <http://creativecommons.org/licenses/by-nc/3.0/>.

Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

To subscribe to *Learning & Memory* go to:
<http://learnmem.cshlp.org/subscriptions>
