Volume 22: No. 4, April 2008

Memory Attribution: Deliberating About Memories

by Ian Dobbins

Ian Dobbins, currently an Associate Professor of Psychology at Washington University in Saint Louis, received his PhD in Psychology at The University of California, Davis under Dr. Neal E. A. Kroll. His laboratory studies the cognitive processes and neural mechanisms underlying how people both deliberately and automatically recover memories. His research particularly focuses on how regions within the prefrontal cortex contribute to the deliberate retrieval of memories and how regions in other parts of the brain may instead regulate more automatic expressions of memory. Tools used in the laboratory include behavioral experiments, decision modeling, and brain imaging with functional magnetic resonance imaging (fMRI).

We have all experienced the slight social panic that occurs when we are unsure whether a quickly approaching person is known or unknown. Questions such as “Do I wave or is a brief smile a safer bet?” arise, and sometimes the memory status of the individual isn’t concluded until embarrassingly late. Such examples illustrate that memory judgments often involve complex and rushed decisions (Johnson, M.K., Hashtroudi, S, & Lindsay, D, 1993; Schacter D.L., Norman, K. A., & Koutstaal, W., 1998). For example, one’s willingness to wave, given a certain level of memory evidence, may depend upon the rapid weighing of numerous contextual factors. If one is at an occasion where there is a high likelihood of knowing most of the attendees, say a small close-knit conference, then minimal evidence may be deemed sufficient to initiate an effusive greeting. In contrast, contexts where one is likely to be a relative stranger may lead to greater caution when initiating greetings. The mechanisms that influence the mapping between memory evidence and actions or judgments are often termed “decision criteria,” and to understand the significance of decision criteria it is useful to examine a simple model of recognition judgments termed Signal Detection Theory (SDT) (Figure 1).

Under SDT, recognition judgments are assumed to rely upon a unidimensional memory evidence value often referred to as memory strength or familiarity. During the discrimination of recently studied from novel items, it is assumed that prior study has increased the baseline evidence of the items yielding two overlapping normal evidence distributions (Figure 1). The observer must parse this evidence into two discrete response options, “old” and “new,” and does so by establishing a fixed value or criterion (c). Items generating greater memory evidence than the criterion are labeled old whereas those eliciting evidence values below the criterion are classified as new (Macmillan & Creelman, 1991). The criterion then, represents the subjects’ best guess as to how much evidence is likely, in that situation, for items that were studied. The distance between the distributions defines the resolution or accuracy of the observer and is termed d’. By way of analogy, consider how one would judge the gender of individuals if the only information provided were height. Males and females would be expected to differ on average but not categorically so, such that there would be overlapping distributions of height values. Here the observer would have to pick a value as a cutoff and the chosen location of that cutoff may be affected by factors such as whether the sample was being drawn from Scandinavia or East Asia, or whether one was judging pre-schoolers or adults.

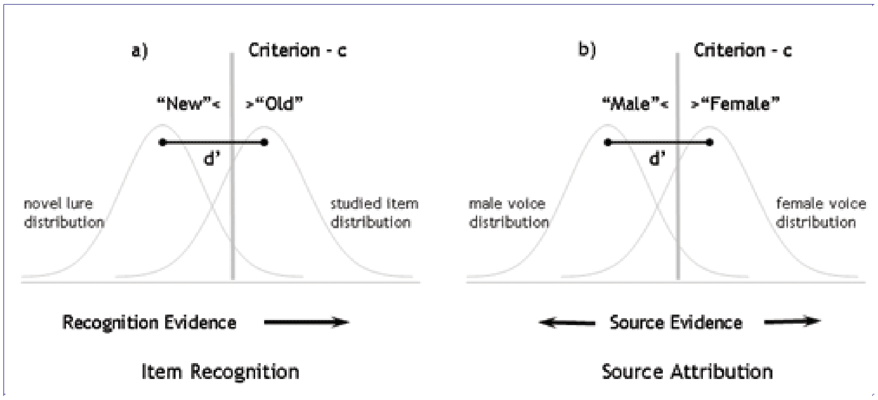

Figure 1. Signal Detection Theory (SDT) decision models of item recognition (panel a) and source memory judgment (panel b).

The SDT decision model is likely an oversimplification of the way explicit memory judgments are rendered; nonetheless, the model illustrates the importance of considering decision criteria. Given two observers with different resolutions (d'), it is not necessarily the case the observer with the higher resolution will be the most successful in rendering correct memory judgments. This is because an observer who is sensitive to situational factors (such as the event contexts mentioned above) may be able to strategically place the criterion in a manner that increases his or her success relative to an observer who is insensitive and does not adjust the criterion adaptively. In short, while resolution is important, so is criterion placement, and nature is presumably neutral as to how the correct or most rewarding judgments are reached. Surprisingly, although the SDT decision model has been applied to recognition memory since the mid 1960s (Parks, 1966), relatively little work has examined how or if subjects adaptively position decision criterion.

One relevant factor appears to be the subjects' expectations with regard to memory availability. For example, Dobbins & Kroll (2005) examined recognition judgments for studied photos drawn from the participants' local environment (e.g., the local coffee shop) versus similarly constructed photos drawn from the surroundings of a different campus. Under self-paced recognition, subjects were less likely to incorrectly endorse unstudied items drawn from their local environment compared to those drawn from an unfamiliar environment. Critically, if only familiarity governed response tendencies, the opposite pattern should have emerged, namely, a higher intrusion rate for familiar lures drawn from the local environment. This suggested that subjects used a higher or more rigorous standard for photos that they anticipated would have yielded rich or robust recognition had they in fact just been studied. This strategy has been termed a memorability heuristic (Brown, J., Lewis, V.J., & Monk, A. F., 1977), and estimating the memorability of stimuli is presumably a controlled reasoning process. Supporting this interpretation, when subjects were forced to respond extremely rapidly, the pattern reversed (Dobbins & Kroll, 2005). That is, lures drawn from the local environment were now incorrectly endorsed at a higher rate than those chosen from an unknown location, suggesting subjects were unable to evaluate memorability in the limited time available, and were instead relying solely on item familiarity. Although the neural substrates supporting subjective memorability heuristics have not been directly investigated, neuropsychological casework (Curran *et al*., 1997) and functional magnetic resonance imaging (fMRI) work on the inhibition of automatic learned responses (Dobbins, I.G., Schnyer, D.M., Verfaellie, M., & Schacter, D.L. 2004) suggest that dorsolateral prefrontal cortex (PFC) regions may be critical for exerting executive control during recognition.

Subjective memorability heuristics reflect a type of metamemory or meta-awareness in which subjects adjust criteria based on what they know about the general operating characteristics of their own memory; for example, the knowledge that memory for personally distinctive materials, such as photos of their favorite coffee shop, should be vivid or rich if recently encountered (at least when encountered in a laboratory) (Schwartz, B.L., Benjamin, A.S., & Bjork, R.A., 1997). Although sensitive to these individual probe characteristics, subjects often appear strikingly insensitive to general trends in the quality of memory evidence and their concurrent performance during testing. For example, if studied items become generally hard to detect during the second half of a test list, subjects do not spontaneously adjust the criterion downwards in an attempt to improve detection rates (Verde & Rotello, 2007). Such criterion rigidity appears to reflect an ambiguity inherent in the design of standard recognition tasks that do not provide performance feedback. More specifically, memory evidence may begin to routinely fall below one's criterion because the criterion is too stringent, or because one is simply encountering a large run of lures and the criterion is appropriate. To investigate subjects' ability to strategically adjust the criterion in the absence of changes to the memoranda, Han & Dobbins (in press) used a novel biasing feedback procedure in which subjects were correctly informed about successful responses (correct "old" and "new") responses, but were tacitly misinformed about incorrect responses of a given type. Whereas half the subjects were informed that incorrect endorsement of new items (false alarms) were in fact correct, the other half were informed that incorrect rejection of old items (misses) were in fact correct. Thus each group was given positive feedback in response to a particular type of memory error and correct negative feedback in response to another type of error. Two experiments demonstrated that this procedure led to prominent and durable shifts in the relative decision criterion across groups. Interestingly, subjects appeared to be largely unaware of the skewed or biased nature of the biased feedback manipulation. Follow-up research using these types of biased feedback manipulations further suggests that the effect may rely upon reinforcement learning mechanisms implicated in the learning of novel categories, and that magnitude of the induced effect covaries with personality traits linked to reward-seeking outside the laboratory (Han & Dobbins, under review). Such data suggest the exciting possibility that explicit memory judgments may be influenced not only by conscious strategies and heuristics, but also by incrementally acquired response tendencies linked to reinforcement history and reward sensitivity (Wixted & Gaitan, 2002).

Although the simple SDT model shown in Figure 1a is most frequently applied to recognition data, it has also been argued to apply to more contextually specific memory discrimination tasks referred to as source memory (Figure 1b). In a source memory task subjects encounter items in two different study situations, for example, words read by a male or female presenter. Later, they are required to attribute the items to the particular prior sources, in this case the male or female presenter. The SDT model of source memory is formally identical to that for item recognition. Items are assumed to evoke normally distributed continuous evidence values favoring one or the other source and subjects parse the evidence axis into two discrete decisions using a criterion value. However, the formal equivalence of the recognition and source SDT models belies the fact that the two judgments likely depend on different decision processes or combinations of processes. For example, early neuropsychological evidence suggests that source memory judgments may be particularly dependent upon the integrity of prefrontal cortex (Janowsky, J.S., Shimamura, A. P., and Squire, L.R.,1989). This hypothesis has received considerable confirmation from fMRI studies of source memory demonstrating much greater recruitment of PFC regions during source memory compared to item recognition judgments (Dobbins, I. G., Foley, H., Schacter, D. L., & Wagner, A. D., 2002; Mitchell, K.J., Johnson, M. K., Raye, C. L., and Greene, E. J., 2004; Rugg, M.D., Fletcher, P.C., Chua, P. M., & Dolan, R..J., 1999), and this increased recruitment does not simply reflect the relative behavioral difficulty of the two tasks.

One factor thought to differentiate source memory from item recognition is the degree to which the former requires increased planning or contextual reinstatement (Norman & Bobrow, 1979). Whereas item recognition can be based on a sense of familiarity or novelty, source tasks typically require the active consideration of finer grained aspects of prior experiences and subjects must appreciate that not only may items potentially differ in terms of their old/new status, but that they may also differ in their ability to evoke contextual memories diagnostic of one versus the other source. Thus whereas judging items as old or new requires little in the way of considering specifics about ones' prior experiences, judging an item as from a particular source requires keeping in mind a template or description of the types of memory characteristics that will aid in identifying particular sources. Supporting this distinction, recent fMRI research suggests that source memory tasks impose greater processing demands early in the trial than item memory tasks. To examine this, Dobbins and Han (2006) presented subjects with memory triplets composed of one new item, and two old items encountered in two different prior sources (pictures previously rated for pleasantness or realism). On some trials subjects were tasked to select the item arising from a particular prior source (e.g., earlier pleasantness rating task) whereas on other trials they were instructed to select the new item. Critically, the memory queries preceded the appearance of the probe items by varying time periods. This enabled the isolation of the neural response to the question (cue activity) from that which occurred in response to the later arrival of the memoranda (cue + probe activity). A key finding was that in relation to item recognition questions, source memory questions led to the early recruitment of left PFC regions, even before the appearance of the actual memoranda. Prior research suggests that the left pre-central gyrus and medial prefrontal regions implicated in this contrast are involved in verbal working memory maintenance. Thus the evidence suggests that source memory tasks place a greater working memory demand upon subjects early in the train of processing consistent with the hypothesis that subjects must bring to, and hold in mind a much more complex task characterization when attempting source attribution (Dobbins & Han, 2006).

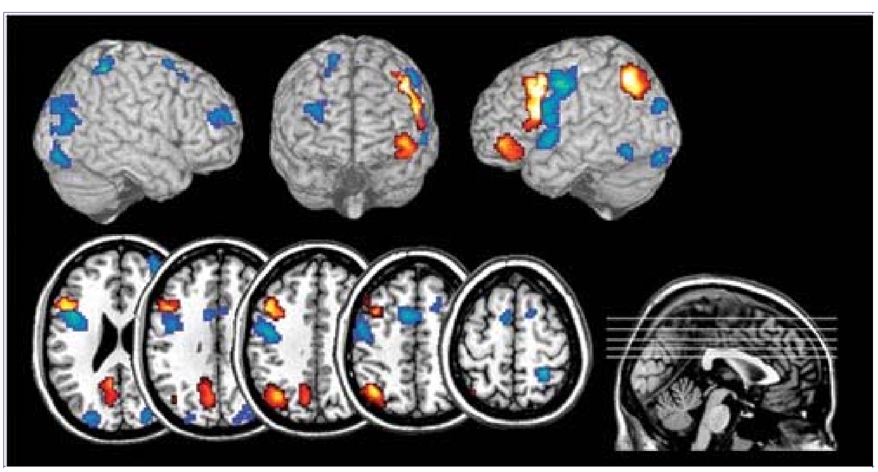

Figure 2. Regions demonstrating greater activity for source memory trials compared to item recognition trials in Dobbins & Han (2006). Regions in blue demonstrated greater activity during the initial cue/question period, when the question indicated the upcoming task would require source attribution. Thus greater recruitment occurred even though memoranda were not yet visible, reflecting the planning of contextual retrieval. Many of these regions have been implicated in verbal working memory maintenance. Regions in orange demonstrated greater activity for

source versus item memory trials as well, although this difference was not apparent until the memoranda were actually present suggesting these responses reflect processes involved in evaluating the memoranda or the memory content they evoke.

Despite the tendency for source attribution tasks to recruit PFC to a greater degree than item recognition tasks, it is important not to over-simplify this characterization. Increasingly, fMRI research designs are using individual differences to facilitate the functional interpretation of regional activity. Although not often stressed, subjects differ considerably not only in estimates of recognition accuracy, but also in the relative placement of decision criteria. That is, whereas some subjects are inherently very cautious in their approach to recognition endorsements, others appear very lax (Dobbins, 2001). As part of a broader study examining item recognition and reinforcement, Han, Huettel and Dobbins (in preparation) conducted an initial fMRI scan in which observers engaged in simple old/new recognition judgments for serially presented words. As expected, scoring outside the scanner suggested considerable individual variability in criterion levels. When these criterion estimates were used to interrogate regions showing greater activation for old versus new items, left frontopolar cortex demonstrated a particularly strong relationship with criterion variability (Han, Huettel, & Dobbins et al., in prep). More specifically, this region demonstrated more activation for subjects who were more conservative or strict in their adopted criterion. Although there is considerable controversy regarding the function of frontopolar cortex, it may play a critical role in the integration or joint consideration of evidence of different types during reasoning. In the case of recognition, this characterization would suggest that observers who demonstrate conservative criteria consider multiple converging pieces of memory evidence before endorsing an item as old. This has the benefit of reducing incorrect endorsement of new items since these are unlikely to yield multiple spurious indicators of prior study, however, it comes at the cost of also reducing the number of old items that are detected.

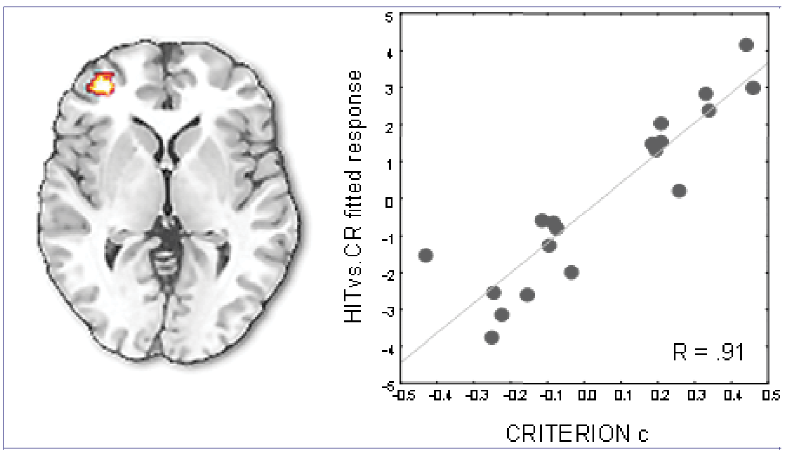

Figure 3. A region of left frontopolar cortex that demonstrated sensitivity to individual differences in criterion placement. During recognition testing this region demonstrated greater fMRI activation for the correct identification of old items (hits) than new items (correct rejections - CR). Additionally, those subjects who where more conservative or strict when responding, as estimated by the SDT measure c (x - axis), demonstrated greater recruitment of this region. Thus recruitment of this region, known to participate in complex reasoning and planning, may be an indicator of cautious or thorough evaluation of memory content.

Although the study of the decision mechanisms guiding memory judgments is in its early stages, several new and exciting possibilities have arisen from recent investigations. What has generally become clear is that memory judgments are complex acts, even when they are confined to judgments about individual stimuli. They appear to be mediated by a host of factors such as explicit strategies, reinforcement histories, personality, and the contextual specificity of the memory demand (e.g., source versus item recognition judgments). Furthermore, even simple recognition judgments can recruit regions of PFC implicated in complex reasoning abilities. While simple decision models such as SDT are extremely useful, they are perhaps best thought of as basic measurement tools or data summaries. For example, there is nothing inherent in the two models illustrated in Figure 1 that would lead one to predict that compared to item recognition, source queries would preferentially recruit PFC regions linked to working memory earlier in the trial (Dobbins & Han, 2006). Similarly, the models are agnostic with respect to whether individual differences in conservative versus liberal responding should lead to different patterns of cortical activation, yet the data of Han, Huettel & Dobbins (in preparation) demonstrate that conservative individuals demonstrate increased recruitment of left frontopolar cortex (Figure 4), a region also recruited during higher order reasoning. Finally, the models in Figure 1 are not useful in explaining why certain manipulations of recognition test items influence the criterion position (Dobbins & Kroll, 2005) whereas others do not (Verde & Rotello, 2007), and work in this area is beginning to suggest that shifts in the measured criterion might result from fundamentally different mechanisms. For example, the shift that results from biased feedback may be fundamentally different from that which occurs in response to explicit warnings by the experimenter to avoid certain types of errors. Cataloguing and explaining these multiple and perhaps interactive mechanisms promises to be a fruitful area of research over the coming years and may shed light not only the impressive feats of human memory, but also its extreme vulnerability to neuro-degenerative diseases and psychological and physical trauma.

Acknowledgments

Preparation of this article, and some of the research reported, was supported in part by grant RO1-MH073982 from the National Institute of Mental Health.

References

Brown, J., Lewis, V. J., & Monk, A. F. (1977). Memorability, word frequency and negative recognition. *The Quarterly Journal of Experimental Psychology, 29*(3), 461-473.

Curran, T., Schacter, D. L., Norman, K. A., & Galluccio, L. (1997). False recognition after a right frontal lobe infarction: Memory for general and specific information. *Neuropsychologia Cortical and subcortical circuits for perception and memory., 35*(7), 1035-1049.

Dobbins, I. G. (2001). The systematic discrepancy between a for overall recognition and remembering: A dual-process account. *Psychonomic Bulletin & Review, 8*(3), 587-599.

Dobbins, I. G., Foley, H., Schacter, D. L., & Wagner, A. D. (2002). Executive control during episodic retrieval: Multiple prefrontal processes subserve source memory. *Neuron, 35*(5), 989-996.

Dobbins, I. G., & Han, S. (2006). Cue- versus probe-dependent prefrontal cortex activity during contextual remembering. *J Cogn Neurosci, 18*(9), 1439-1452.

Dobbins, I. G., & Kroll, N. E. (2005). Distinctiveness and the recognition mirror effect: Evidence for an item-based criterion placement heuristic. *J Exp Psychol Learn Mem Cogn, 31*(6), 1186-1198.

Dobbins, I. G., Schnyer, D. M., Verfaellie, M., & Schacter, D. L. (2004). Cortical activity reductions during repetition priming can result from rapid response learning. *Nature, 428*(6980), 316-319.

Han, S., & Dobbins, I. (under review). Regulating explicit recognition decisions through implicit reinforcement learning.

Han, S., Huettel, S. A., & Dobbins, I. G. (in prep). Reward and recognition: Recruitment of striatal and frontopolar regions during recogition decisions.

Janowsky, J. S., Shimamura, A. P., & Squire, L. R. (1989). Source memory impairment in patients with frontal lobe lesions. *Neuropsychologia, 27*(8), 1043-1056.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. (1993). Source monitoring. *Psychological Bulletin, 114*(1), 3-28.

Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide:* (1991).

Mitchell, K. J., Johnson, M. K., Raye, C. L., & Greene, E. J. (2004). Prefrontal cortex activity associated with source monitoring in a working memory task. *J Cogn Neurosci, 16*(6), 921-934.

Norman, D. A., & Bobrow, D. G. (1979). Descriptions: An intermediate stage in memory retrieval. *Cognitive Psychology, 11*(1), 107-123.

Parks, T. E. (1966). Signal-detectability theory of recognition-memory performance. *Psychological Review, 73*(1), 44-58.

Rugg, M. D., Fletcher, P. C., Chua, P. M., & Dolan, R. J. (1999). The role of the prefrontal cortex in recognition memory and memory for source: An fmri study. *Neuroimage, 10*(5), 520-529.

Schacter, D. L., Norman, K. A., & Koutstaal, W. (1998). The cognitive neuroscience of constructive memory. *Annu Rev Psychol, 49*, 289-318.

Schwartz, B. L., Benjamin, A. S., & Bjork, R. A. (1997). The inferential and experiential bases of metamemory. *Current Directions in Psychological Science, 6*(5), 132-137.

Verde, M. F., & Rotello, C. M. (2007). Memory strength and the decision process in recognition memory. *Mem Cognit, 35*(2), 254-262.

Wixted, J. T., & Gaitan, S. C. (2002). Cognitive theories as reinforcement history surrogates: The case of likelihood ratio models of human recognition memory. *Anim Learn Behav, 30*(4), 289-305.

Science Directorate
750 First Street, NE • Washington, DC • 20002-4242
Phone: 202-336-6000 • TDD/TTY: 202-336-6123
Fax: 202-336-5953 • E-mail