A

A Stability Bias in Human Memory

Nate Kornell
Department of Psychology, Williams College,
Williamstown, MA, USA

Definition

Human memory is anything but stable: We constantly add knowledge to our memories as we learn and lose access to knowledge as we forget. Yet people often make judgments and predictions about their memories that do not reflect this instability. The term *stability bias* refers to the human tendency to act as though one's memory will remain stable in the future. For example, people fail to predict that they will learn from future study opportunities; they also fail to predict that they will forget in the future with the passage of time. The stability bias appears to be rooted in a failure to appreciate external influences on memory, coupled with a lack of sensitivity to how the conditions present during learning will differ from the conditions present during a test.

Theoretical Background

All memories are not created equal. Some memories feel strong, vivid, and familiar; others feel shakier and less reliable. People are generally confident in the first type of memory but unsure about the second. Behavior reflects this difference; for example, most people only volunteer to answer a question in class if they feel confident about their response.

The term *metacognition* refers to the process of making judgments about one's cognition and, frequently, about one's memory (Dunlosky and Bjork 2008). Metacognitive processes are used to distinguish accurate memories from inaccurate ones. A memory is only valuable to the degree that we can trust it, which makes metacognition vital in our day-to-day use of memory. Moreover, virtually all memory retrievals are associated with a feeling of certainty (or lack thereof). Thus, metacognition is a critical, and omnipresent, component of human memory.

Metacognitive judgments are often accurate. For example, your memory of what you ate for breakfast today is probably more accurate than your memory of what you ate for breakfast on this date 11 years ago, and it probably feels more accurate as well. It would be natural to assume that metacognitive judgments are made on the basis of the memory being judged – that is, that when confidence is low, it is *because* a memory is weak. The empirical evidence suggests otherwise.

Instead of being made based on memories themselves, metacognitive judgments appear to be made based on inferences about those memories. For example, if an answer comes to mind quickly and easily, people tend to judge that they know that answer well. This inference is usually correct. But it is an inference all the same, and when conditions are created that reverse this relationship – when answers that come to mind quickly are *less* memorable – people give high *judgment of learning* ratings to information that comes to mind quickly, *not* to information that is highly memorable (Benjamin et al. 1998).

If metacognitive judgments are inferential, what is the basis of the inferences? Koriat (1997) put forward a highly influential framework that has successfully accounted for a great deal of subsequent data. He proposed that three categories of cues influence metacognitive judgments. *Intrinsic cues* were defined as information intrinsic to the information being judged (e.g., the semantic relatedness of a question and its answer). *Mnemonic cues* were defined as information related to the learner's experience (e.g., the fluency with which an answer comes to mind). *Extrinsic cues* were defined as information extrinsic to the learner and the to-be-learned material (e.g., the number of times an item was studied).

A second key distinction, related to Koriat's (1997) framework, is between judgments based on direct experience and judgments based on analytical processes (Kelly and Jacoby 1996). Intrinsic cues and mnemonic cues tend to elicit *experience-based* judgments. That is, these cues (e.g., how easily one thinks of an answer) are part of the learner's experience at the time of the judgment. Metacognitive judgments are usually highly sensitive to a person's current experience. Thus, experience-based judgments often occur automatically.

Norbert Seel (ed.), *Encyclopedia of the Sciences of Learning*, DOI 10.1007/978-1-4419-1428-6,

Extrinsic cues, by contrast, tend to elicit more analytical *belief-based* judgments. For example, the number of times an item will be studied is not a salient part of the learner's experience while studying. Instead, responding to an extrinsic cue often requires applying one's beliefs about memory (e.g., I will do better on items I study more). Doing so does not tend to happen automatically. As a result, people regularly fail to make belief-based judgments, even when they should. Thus, people tend to be sensitive to experience-based cues but not belief-based cues.

It is important to be able to predict how future events will affect one's memory. For example, a student may need to predict the value of spending the rest of the day studying. Future events are extrinsic cues – they are external to the learner's current experience – and, as such, they require belief-based judgments. Thus, people should exhibit a stability bias: They should be relatively insensitive to the impact of future events on their memories.

Important Scientific Research and Open Questions

Koriat et al. (2004) investigated how sensitive people are to future forgetting. After studying a list of word pairs, their participants were asked to predict their likelihood of recalling the pairs on a cued-recall test (i.e., their ability to recall the second word in the pair when shown the first word). There were three groups of participants, who were told, respectively, that their test would take place immediately, a day later, or a week later.

Actual recall performance dropped off precipitously as the delay between study and test increased. Shockingly, predictions hardly changed at all. In other words, the participants demonstrated a stability bias: They acted as though they would remember just as much in a week as they would remember immediately. The predictions were highly sensitive to the degree of association between the pairs, which is an experience-based, intrinsic cue. But they were insensitive to retention interval, an extrinsic cue. In one extreme case, tests that would take place immediately and in one *year* elicited the same predictions.

A key change in the procedure greatly altered participants' predictions. When a single participant was told about all three retention intervals, their predictions became sensitive to retention interval. It appeared as though the participants believed that they would forget over time, but that they did not apply that belief in the first experiment. When they were told about all of the retention intervals, they began to apply belief-based judgments. Phrasing the question in terms of forgetting had a similar effect: Apparently, making the idea of forgetting salient was enough to make judgments sensitive to retention interval.

One potential implication of ignoring retention interval is extreme overconfidence. People tend to be overconfident in their memories in general. But when someone is overconfident about an immediate test, and is not sensitive to retention interval, their overconfidence is destined to grow. For example, assume you have a 70% chance of recalling a fact from your textbook if you are tested in 10 min. If you are tested in a week, that chance might decrease to 20%. If you judge that you have an 80% chance of recalling the fact at either retention interval, you will be overconfident immediately, but only by 10% points. A week later, you will be overconfident by 60% points. This increase in overconfidence with time has been referred to as *long-term overconfidence* (Kornell 2010).

The stability bias is not limited to forgetting. Kornell and Bjork (2009) investigated predictions about another seemingly obvious principle of memory, namely, that people learn by studying. Their participants were told that they would be allowed to study a list of word pairs between one and four times. They were asked to predict how they would do when they took a test on the pairs. The predictions were almost entirely insensitive to the number of study repetitions, again demonstrating a stability bias. The stability bias did not go both ways; people recognized the value of *past* studying, but underestimated the value of *future* studying. Like with forgetting, when the concept of learning was made salient, in a within-participants design, the predictions became more sensitive. Unlike forgetting, however, the predictions continued to underestimate the value of studying. As a result, across a number of different experiments, participants were overconfident in their current knowledge, but simultaneously *underconfident* in their learning ability.

One potential implication of undervaluing future study opportunities is that people might underestimate their own learning potential. For example, a student might look at a set of challenging course materials and decide to drop out of a class, assuming that he or she cannot learn all of the material. If this student is underconfident in his or her learning, he or she might be giving up in the face of a challenge that could be overcome.

Cross-References

► Confidence Judgments in Learning
► Cued Recall
► Metacognition and Learning
► Metacognitive Learning: The Effect of Item-Specific Experiences
► Overconfidence

► Self-confidence and Learning
► The Role of Stability in the Dynamics of Learning, Memorizing, and Forgetting

References

Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (1998). The mismeasure of memory: When retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General, 127*, 55–68.

Dunlosky, J., & Bjork, R. A. (Eds.). (2008). *A handbook of metamemory and memory*. Hillsdale: Psychology Press.

Kelly, C. M., & Jacoby, L. L. (1996). Adult egocentrism: Subjective experience versus analytic bases for judgment. *Journal of Memory and Language, 35*, 157–175.

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 126*, 349–370.

Koriat, A., Bjork, R. A., Sheffer, L., & Bar, S. K. (2004). Predicting one's own forgetting: The role of experience-based and theory-based processes. *Journal of Experimental Psychology: General, 133*, 643–656.

Kornell, N., & Bjork, R. A. (2009). A stability bias in human memory: Overestimating remembering and underestimating learning. *Journal of Experimental Psychology: General, 138*, 449–468.

Kornell, N. (2010). Failing to predict future changes in memory: A stability bias yields long-term overconfidence. In A. S. Benjamin (Ed.), *Successful Remembering and Successful Forgetting: A Festschrift in Honor of Robert A. Bjork* (pp. 365–386). New York, NY: Psychology Press.