ESSAYS ON THE CONCEPT OF INHIBITION

REVISED VERSIONS TO APPEAR IN:

Roediger, H.L., Dudai, Y. & Fitzpatrick, S.M. (Eds), *Science of Memory: Concepts*. Oxford: Oxford University Press.

Inhibition in Long-Term Memory
Michael C. Anderson
University of Oregon

Diversity of Cortical Functions is Secured by Inhibitory Mechanisms
György Buzsáki
Rutgers, The State University

Inhibition: An Attentional Control Mechanism
Lynn Hasher
University of Toronto

Cognitive Inhibition: Elusive or Illusion?
Colin M. MacLeod
University of Waterloo

Inhibition as an Essential and Contentious Concept in Memory
Robert A. Bjork
University of California, Los Angeles

Inhibition as an Essential and Contentious Concept in Memory

Robert A. Bjork
University of California, Los Angeles

In a 1989 essay on "Retrieval inhibition as an adaptive mechanism in human memory," written for a volume honoring Endel Tulving, I argued that inhibitory processes played little or no role in then-current theories of human memory (Bjork, 1989). I viewed that fact as puzzling—given that inhibitory processes had long been acknowledged by brain and behavioral scientists to be critical at the neural, sensory, attentional, and motor levels, and in the ontogeny of brain development. Towards explaining *why* inhibition seemed out of favor as an explanatory concept, I suggested that two aspects of the prevailing research zeitgeist played a role:

> *"First, notions of inhibition or suppression in human memory have an unappealing association to certain poorly understood clinical phenomena, such as repression. Second, the information-processing approach, grounded as it is in the computer metaphor, leads us to think in term of processes like storing, scanning, grouping, erasing, and so forth. Notions like inhibition, suppression, unlearning, and spontaneous recovery are not easily compatible with the computer metaphor. (p. 310)*

My own history of research on directed forgetting, tracing back to my graduate-school days, illustrates the influence of such factors. I spent the first 15-20 years saying and writing that research on directed forgetting was important *not* because it had anything to do with clinical phenomena, such as repression, but because it could shed light on how our memories are kept current, how rehearsal and encoding resources are allocated, and how competing items are segregated and differentiated in memory. In papers I authored or co-authored during what Colin MacLeod (1998), in a remarkable review of the literature, referred to as the "golden age" of research on directed forgetting (1968-74), which began with a paper by Bjork, LaBerge, and Legrand (1968), and through much of what he refers to as the "silver age" (1975-85), I attempted to explain directed-forgetting findings via non-inhibitory processes such as selective rehearsal and set differentiation. It took an accumulation of directed-forgetting findings that proved hard to interpret without reference to inhibition, especially those reported by Geiselman, Bjork, and Fishman (1983), to convince me that retrieval inhibition played a key role in directed forgetting.

Historical Perspective on the Reluctance to Postulate Inhibition

Taking a broader historical view, however, the hesitancy to postulate inhibitory mechanisms in learning and memory theories clearly pre-dates the emergence of computer metaphor and the information-processing approach. In fact, concerns about the necessity of assuming inhibitory processes go back to the early decades of controlled research on human and animal learning and memory. In the experimental and theoretical analysis of extinction phenomena, for example, questions and debates arose that are very reminiscent of present-day issues that have emerged in the context of blocking-versus-suppression accounts of retrieval-induced forgetting phenomena (cf. the Anderson's and

MacLeod's essays in this volume). In research on conditioning, for example, explanations of extinction in terms of inhibitory processes (e.g., Pavlov, 1927, and Hull, 1951) did battle with interference explanations (e.g., Guthrie, 1935) that asserted, in essence, that extinction is merely the learning of a new response to the old conditional stimulus.

Aside from the whether inhibition or interference accounts of extinction could provide a better account of extinction phenomena, such as "disinhibition" effects, or that massing of extinction trials often resulted in more effective extinction than did spacing those trials, whereas the opposite was true for acquisition—both of which posed difficulties for interference/new-learning accounts—reviews from that period clearly reflect the hesitancy to postulate inhibitory processes. In referring to Pavlov's appeal to inhibitory mechanisms in accounting for extinction effects, for example, Woodworth and Schlosberg's (1954) say "Pavlov's rather speculative ideas of what goes on in the brain may be of little importance. Some psychologists go so far as to reject the concept of inhibition, although it seems to be a necessary—and respectable—concept in physiology" (p. 559). And they later, in discussing the reaction of the field to Pavlov's interpretation of disinhibition, comment that "It is not surprising that psychologists who disliked inhibition regarded this suggestion as adding insult to injury" (p. 561). It became common, in fact, and for mostly good reasons, to avoid using the term *inhibition* in labeling empirical effects, but that convention may have been spurred on, in part, by the desire to avoid the concept altogether. In his experimental psychology book, for example, Osgood (1953), for example, in a preamble to his discussion of transfer and retroaction in his experimental psychology book, provides the following caution:

> *"Although the term "retroactive facilitation is commonly and acceptably used for positive retroaction, the term 'retroactive inhibition' has unfortunately been applied when negative retroaction is found. What is referred to here is simply an observed decrement in performance, not a process—the decrement may or may not be due to some inhibitory process—so henceforth we shall use the more neutral term, retroactive interference. (p. 520)*

Roots of the Reluctance.

Why, historically, has there been a reluctance to postulate inhibitory mechanisms? Beyond any unsavory association to poorly understood clinical dynamics, or any effect of the computer metaphor, I think two other factors may play a role. One is a kind of parsimony consideration: If effects can be explained in terms of cognitive processes that are better understood, such as selective rehearsal and interference/blocking dynamics, then why appeal to inhibitory mechanisms, which are more poorly understood? A second and related consideration has to do with our subjective experience. We all have the conscious experience of selecting some items to rehearse or encode or retrieve—and the experience of some items in memory interfering or blocking our access to other items in memory—but inhibitory processes are not accompanied by the same volitional and conscious experience. Even in cases where inhibitory effects are powerful and undeniable, such as in dichotic listening, when attending to one ear is accompanied by a gating out of input from the other ear, what is available to consciousness is the decision

to allocate attention to a given ear coupled with the processing of input from that ear, not the inhibitory operations that suppress input from the other ear.

Inhibition in the Current Research Zeitgeist

At the end of my 1989 essay, I predicted that in the "near future" there would be consensus that inhibitory processes play a critical role in the overall functioning of human memory. I made that prediction, in part, because I thought the combination of two developments—the emergence of new techniques to examine neural and structure dynamics in the human brain, and the emergence of neural/connectionist modeling of human memory processes—would lead researchers towards, rather than away from, theories that incorporated inhibitory processes. That proved to be a good prediction—at least as indexed by books devoted to the role of inhibition that appeared shortly thereafter (e.g., Dagenbach & Carr, 1994; Dempster and Brainerd, 1995; Smith, 1992), by the subsequent proliferation of the word "inhibition" in the titles of articles, and by the keen interest in experimental tasks thought to instigate inhibitory processes, such as the retrieval-induced forgetting, think/no-think, and directed-forgetting tasks—but concerns about the necessity of assuming inhibitory processes have been resurrected, too, as exemplified by Colin MacLeod's essay in this volume.

Comments and Perspective on the Present Essays

In the sections that follow, I comment—in the context of the splendid essays by Michael Anderson, György Buzáski, Lynn Hasher and Colin MacLeod—on the definition of inhibition; the adaptive nature of inhibitory mechanisms; and what a see as the remaining key issues and points of contention.

Defining Inhibition

Colin MacLeod cites Brunton's (1983) definition of inhibition, repeated below, and Michael Anderson provides a useful and detailed characterization of the attributes of inhibition that goes beyond, but is consistent with, Brunton's definition:

> *"the arrest of the function of a structure or organ, by the action upon it of another, while the power to execute those functions is still retained, and can be manifested as soon as the restraining power is lifted" (Brunton, 1883)*

I endorse that definition, but with respect to *retrieval* inhibition in particular, I have tended to adopt an empirical, rather than conceptual, definition—one that focuses on the impairment of the *recall* of inhibited memory representations. One result that implicates retrieval inhibition is a violation of what might be called the "law of forgetting"; that is, when something that is not recallable after a shorter delay becomes recallable at a longer delay. An example of such an effect is when the earlier learned of two competing memory representations becomes more recallable, in absolute terms, as time passes. Such "regression" effects (Bjork, 1978) are common in both motor and verbal learning, they occur on time scales ranging from seconds to months and years, and they occur across species (for a review, see Bjork, 2001). A second result that implicates

retrieval inhibition is when something becomes non-recallable, but remains at full strength by other measures. An example is when an instruction to forget impairs subsequent recall of the to-be-forgotten materials, but not the subsequent recognition or relearning of those materials, or the effects of those materials on indirect measures of retention, such as priming (see, e.g., Bjork & Bjork, 2003). Again, such effects are very general and appear in research on animal learning as well as human learning. As Bouton (e.g., 1994) and Rescorla (e.g., 2001) have emphasized, based on research with animals, new associations do not over-write or destroy old associations and post-treatment "return of fear," which is common in the clinical treatment of phobias (see Lang, Bjork, & Craske, 1999), is an especially salient example that the same is true for humans as well.

Inhibition as an Adaptive Mechanism

A thread that runs through the essays by Anderson, Buzsaki, and Hasher is that inhibitory processes play a key and adaptive role in how our brains function and in how we contend with the learning, memory, decision, and emotional challenges that are part of living, learning, and managing ourselves. Gyorgy Buzsake, in remarkably few words, provides a clear and compelling description of how inhibitory networks and inhibitory interneurons multiply and refine the computational power of principal cells. That "brain systems with 'simple' computational demands," such as the basal ganglia, thalamus, or cerebellum, are characterized by only a few neuron types, whereas systems that support conscious memory functioning are characterized not only by five principal cell types, but also by "numerous classes" of inhibitory neurons, is an interesting and provocative aspect of the brain's structure. Speculatively—perhaps *too* speculatively on my part—that property of the brain's neural organization seems to link to Hasher's argument that familiar stimuli "activate their representations automatically" and that in many circumstances "this activation (and its spread to associated representations) can and must be down regulated in order for organized behavior to achieve an individual's long and short-term goals." The basic idea is that such down regulation is accomplished by "inhibitory mechanisms that operate in the service of goals."

The emphasis, in Hasher's treatment, is on attentional control and the ability to have one's thoughts and actions be guided by goals and plans, not by the activation triggered by environmental and other stimuli. Individuals, in her view, who have poor inhibitory abilities will also have difficulty "in stopping thoughts and actions that were recently relevant, but no longer are." Such a stopping function maps to one of the two adaptive functions Michael Anderson attributes to inhibition, the other being to resolve competition in the use of our memories, but the emphasis is a bit different in Anderson's and Hasher's frameworks. Anderson's emphasis is on "memory stopping"—that is, on stopping the retrieval of information that is stored in memory, but, when recalled, is a source of emotional discomfort or "undermine performance on some task"—whereas Hasher's emphasis is on inhibiting activations that are stimulus driven, automatic, and perhaps unaccompanied by awareness.

With respect to resolving competition among items in memory, Anderson argues that selective retrieval—that is, selecting a target item from memory from among all the

items in memory that might be associated to a given retrieval cue or cues—engages an inhibitory mechanism that suppresses the non-target/competing items and, thereby, enhances access to the target representation. This suppression, though, can persist and result in *retrieval-induced forgetting* of non-target items (Anderson, Bjork, & Bjork, 1994)—at least for some time—should the recall of those items be required. Anderson and Spellman (1995) have argued that selective retrieval has properties that are parallel to selective attention; in both cases inhibitory mechanisms act to enhance access to the external or internal target by suppressing competition from unattended external or internal non-targets. Bjork, Bjork, and Anderson (1998) have pushed that argument further by suggesting that such a selection-plus-suppression mechanism may be "*the* primary solution in the functional architecture of the human as an information-processing device to the problem of avoiding interference and competition at various levels of cognitive processing … in a broad range of motor and cognitive activities, *selecting* appears to involve *inhibiting*." (p. 133)

In my own case, I have argued that retrieval inhibition is a uniquely human and adaptive solution to the problem of keeping one's memory current. In contrast to man-made memory devices, in which—without special precautions—the storage of new information replaces or erases old, out-of-date, information, learning and using new information (such as a new home phone number) does not destroy the representation of the information it replaces, but, rather, renders it non-recallable overtime. The old phone number (or street address, software procedure, maiden name, locker number, etc.) remains in memory, can often be recognized, and—should it again become relevant—can be relearned with great rapidity and savings.

Remaining Issues and Points of Contention

In his current essay, and in earlier review (MacLeod, Dodd, Sheard, Wilson & Bibi, 2003), Colin MacLeod provides a broad critique of the readiness of today's researchers to attribute various empirical findings to inhibitory mechanisms. The central arguments reiterate, update, and embellish the arguments that emerged decades ago in accounting for phenomena such as extinction, as alluded to above, and retroactive interference. With respect to his two "start vectors, " that there is (a) no denying that there is inhibition at the neural/brain level, but that (b) the evidence of such inhibition cannot, by itself, be offered as evidence of inhibition at the cognitive/memory level, I believe there can be no serious disagreement. Nor should there be any disagreement with his reiteration of the time-honored caution in discussions of interference versus interference: "inhibition is not an outcome; it is a theory about the cause of the outcome" (p. XXX).

At the risk of over-simplifying MacLeod's arguments, I interpret his critique as saying, first, that to attribute some empirical effect to inhibition, versus some alternative process, requires meeting rigorous criteria and, second, that alternative interference-based explanations must be ruled out before inhibition is inferred. With respect to the first point, he suggests that we need evidence of dissociations analogous to those that have implicated inhibition in research on attention, such as the effect, in research on inhibition

of return, of a pre-cue changing from a benefit to a cost with cue-target delay. I believe that results of that kind, if not exactly of that kind, already exist. In A-B, A-D paired-associate list-learning experiments, for example, one type of evidence supporting response-set-suppression hypothesis (Postman, Stark, & Fraser, 1968) is that the retroactive effects of list-two on list one are greater than the proactive effects of list one on list two, given a short retention interval, but the converse is true at a long delay—consistent with the proposal that B responses are suppressed during A-D learning, but then recover.

A recent finding from research on retrieval-induced forgetting can also, in my view, be seen as a kind of dissociation that implicates inhibition. Items that are most recallable if tested—by virtue, for example, of being the strongest associates to a category cue (Anderson et al., 1994), or by virtue of their valence (Storm, Bjork, & Bjork, 2005)—turn out to be the most, not least, subject to retrieval-induced forgetting. This result, one of the three properties of retrieval-induced forgetting Anderson cites as favoring inhibition, is especially difficult to accommodate from within an interference/blocking interpretation. Another recent finding—that retrieval success is not necessary for retrieval-induced forgetting to happen (Storm, Bjork, Bjork, & Nestojko, 2006)—also poses a major problem for interference/blocking theories, given that no competitor is strengthened when retrieval practice fails, but is readily interpretable in terms of a suppression mechanism.

There is, I think, an implicit—and subtle—assumption in MacLeod's analysis, one that is widely shared, even by those of us who are convinced that inhibitory processes play a key role in human memory. The assumption is that a kind of theoretical pecking order is justified: non-inhibitory ideas are to be preferred, unless totally untenable, and the burden of proof is on inhibition theorists. Before an inhibition interpretation is to be believed, interference mechanisms must be refuted, whereas the converse requirement does not seem, typically, to be placed in interference interpretations. This subtle assumption, which might even be justified via a kind of Bayesian prior-odds reasoning, constitutes, I think, another reflection of the reluctance to assume inhibitory processes.

Finally, even if the evidence of inhibition is considered convincing, a key question remains: Is inhibition a by-product of other activities, such as selective retrieval, selective attention, or self distraction, or does what might be termed "pure" inhibition also characterize human memory and cognition? Stated in the context of Anderson and Green's () think/no-think paradigm, can we respond to instruction not to think of a target item in a direct, unmediated, way that inhibits subsequent access to the target, or can such suppression only be achieved via, say, self distraction or active retrieval of arbitrary non-target items? In the context of research on directed forgetting, research by Gelfand and Bjork (1985; described in Bjork, 1989) suggests that the subsequent inhibited access to to-be-forgotten items is a by-product of new learning coupled with the mental set to replace to-be-forgotten items with upcoming to-be-remembered items, but the question remains, as do many other questions about exactly how and if inhibition is involved in human cognition.

References

Anderson, M.C., Bjork, R.A., & Bjork, E.L. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 1063-1087.

Anderson, M..C., & Green, C. (2001). Suppressing unwanted memories by executive control. *Nature, 410(6826),* pp 131-134.

Bjork, E. L., & Bjork, R. A. (2003). Intentional forgetting can increase, not decrease, the residual influences of to-be-forgotten information. *Journal of Experimental Psychology: Learning, Memory,Cognition, 29,* 524-531.

Bjork, R. A., LaBerge, D., & LeGrande, R. (1968). The modification of short-term memory through instructions to forget. *Psychonomic Science, 10,* 55-56.

Bjork, R. A. (1978). The updating of human memory. In G. H. Bower (Ed.), *The psychology of learning and motivation.* (Vol. 12., pp. 235-259). New York: Academic Press.

Bjork, R. A. (1989). Retrieval inhibition as an adaptive mechanism in human memory. In H. L. Roediger and F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 309-330). Hillsdale, NJ: Erlbaum.

Bjork, E.L., Bjork, R.A., & Anderson, M. C. (1998). Varieties of goal-directed forgetting. In J. M. Golding and C. MacLeod (Eds.), *Intentional forgetting: Interdisciplinary approaches* (pp.103-137). Hillsdale, NJ: Erlbaum.

Bjork, R. A. (2001). Recency and recovery in human memory. In H. L. Roediger, J. S. Nairne, I Neath, & A. M. Surprenant (Eds.), *The nature of remembering: Essays in honor of Robert G. Crowder* (pp. 211-232). Washington, DC: American Psychological Association Press.

Bouton, M. E. (1994). Context, ambiguity, and classical conditioning. *Current Directions in Psychological Science, 3,* 49-53.

Dagenbach, D., & Carr, T. H. (1994). *Inhibitory processes in attention, memory, and language.* Orlando, FL: Academic Press.

Dempster, F. N., & Brainerd, C. J. (1995). *Interference and inhibition in cognition.* San Diego, CA: Academic Press.

Geiselman, R. E., Bjork, R. A., & Fishman, D. (1983). Disrupted retrieval in directed forgetting: A link with posthypnotic amnesia. *Journal of Experimental Psychology: General, 112,* 58-72.

Gelfand, H. & Bjork, R. A. (1985; November). *On the locus of retrieval inhibition in directed forgetting.* Meetings of the Psychonomic Society, Boston, MA.

Hull, C. L. (1951). *Essentials of behavior.* New Haven: Yale University Press.

Lang, A. J., Craske, M. G., & Bjork, R. A. (1999). Implications of a new theory of disuse for the treatment of emotional disorders. *Clinical Psychology: Science and Practice, 6,* 80-94.

MacLeod, C. M. (1998). Directed forgetting. In J. M. Golding and C. MacLeod (Eds.), *Intentional forgetting: Interdisciplinary approaches* (pp. 1-57). Hillsdale, NJ: Erlbaum.

MacLeod, C. M., Dodd, M. D., Sheard, E. D., Wilson, D. E., & Bibi, U. (2003). In opposition to inhibition. In B. H. Ross (Ed.), *The Psychology of Learning and Motivation, Vol. 43* (pp. 163-214). San Diego, CA: Academic Press.

Osgood, C. E. (1953). *Method and theory in experimental psychology*. New York: Oxford University Press.

Pavlov, I. P. (1927). *Conditioned reflexes.* Translated by G. V. Anrep. New York: Oxford University Press.

Postman, L., Stark, K., & Fraser, J. (1968). Temporal changes in interference. *Journal of Verbal Learning and Verbal Behavior, 7,* 672-694.

Rescorla, R. A. (2001). Experimental Extinction, In R. R. Mowrer & S. Klein (Eds.) *Handbook of contemporary learning theories* (pp. 119-154). Hillsdale, N.J.: Erlbaum.

Smith, R. (1992). *Inhibition: History and meaning in the sciences of mind and brain*. Berkeley, CA: University of California Press.

Storm, B. C., Bjork, E. L., & Bjork, R. A. (2005). Social metacognitive judgments: The role of retrieval-induced forgetting in person memory and impressions. *Journal of Memory and Language, 52,* 535-550.

Storm, B. C., Bjork, E. L., Bjork, R. A, & Nestojko, J. (2006). Is retrieval success a necessary condition for retrieval-induced forgetting? (Unpublished paper; accepted, pending revisions, *Psychonomic Bulletin and Review*).

Wegner, D.M., Eich, E., & Bjork, R.A. (1994). Thought suppression. In D. Druckman and R.A.Bjork (Eds.), *Learning, remembering, believing: Enhancing human performance* (pp.277-293). Washington, DC: National Academy Press.

Woodworth, R. S. & Schlosberg, H. (1954). *Experimental psychology.* New York: Henry Holt and Company.