Human Learning and Memory

Advances in Theory and Application

Edited by

Chizuko Izawa • Nobuo Ohta

With a Foreword by William K. Estes

HUMAN LEARNING AND MEMORY:

ADVANCES IN THEORY AND APPLICATION

The 4th Tsukuba International Conference On Memory

筑波大学

Edited by

CHIZUKO IZAWA

Tulane University

NOBUO OHTA

University of Tsukuba

LEA LAWRENCE ERLBAUM ASSOCIATES, PUBLISHERS

2005 **Mahwah, New Jersey** **London**

Lawrence Erlbaum Associates, Inc., Publishers
10 Industrial Avenue
Mahwah, New Jersey 07430

Library of Congress Cataloging-in-Publication Data

Tsukuba International Conference on Memory (4th : 2003: Tsukuba Diagaku)
Human learning and memory : advances in theory and application : the 4th Tsukuba International Conference on Memory / edited by Chizuko Izawa, Nobuo Ohta.
p. cm.
Includes bibliographical references and index.
ISBN 0-8058-4788-X (alk. paper)
1. Memory–Congresses. 2. Learning, Psychology of–Congresses. I. Izawa, Chizuko. II. Ohta, Nobuo. III. Title.
BF371.T78 2003
153.1'2–dc22 2004054103

Books published by Lawrence Erlbaum Associates are printed on acid-free paper, and their bindings are chosen for strength and durability.

Printed in the United States of America
10 9 8 7 6 5 4 3 2 1

Contents

Invited Participants to Tic4 (The 4th Tsukuba International Conference on Memory)
Human Learning and Memory: Advances in Theory and Application

Available for the Photo Session
at Noon 12 January 2003

From Left to Right:

Row 1: Nobuo Ohta, Chizuko Izawa, Alice Healy, Vicki Schneider, Lynn Hasher;
Row 2: Nelson Cowen, Rich Shiffrin, A. Mayers;
Row 3: Mike Humphreys, Doug Nelson, Cathy McEvoy, Bill Hockley, Sal Soraci;
Row 4: Mike Wenger, M. Meeter, Jeroen Raaijmakers, E.-J. Wagenmakers, T. Mantyla, Debra Pate; and
Row 5: Jun Kawaguchi

This page intentionally left blank

Foreword

William K. Estes
Indiana University, USA

This Foreword to the proceedings of the fourth Tsukuba International Conference on Memory (Tic4) will not serve the usual function of sketching the setting of the Conference and introducing the participants with the customary laudatory comments about their contributions. These tasks have been accomplished with high enthusiasm by Professor Chizuko Izawa, the American co-organizer of the Conference and co-editor of this volume. What is left for me, evidently, is to comment briefly on the accomplishments of the Conference series and projections of its possible futures from the viewpoint of potential readers and users of the products of Tic4 who, like me, have had no first-hand acquaintance with this enterprise.

To assess the accomplishments of this Conference series, one needs to know its objectives, which are nowhere explicitly stated in the proceedings. In view of the fact that the series has been supported by the Ministry of Education of Japan, one might surmise that a major purpose is to inform attendees about developments in research that bear directly on problems of education. In this connection, one notes that the proceedings of Tic4 include just three chapters that deal explicitly with the application to education of products of research on memory, those of Izawa, Hayden, and Franklin (Chap. 2); Izawa, Maxwell, Hayden, Matrana, and Izawa-Hayden (Chap. 5), and Hasher, Goldstein, and May (Chap. 9). Each of these could offer guidance in the scheduling of study and testing experiences if followed up by research designed to extend the findings from single experimental sessions to the longer durations of time characteristic of school learning. A fourth chapter, by Healy and colleagues, bears on education in the limited sense of personnel training in the military.

It might seem that the potential contributions to education are meager; but perhaps that judgment would be superficial. The concept of memory, the common thread running through all of the chapters, is being treated, not in isolation, but as a component of learning, speech and language processing, problem solving, and decision making. With that thought in mind, one can see how the timely reviews of research and theory in the Tic4 proceedings might be quite relevant to those who conduct research in educational settings and to those who educate both the researchers and the educators.

A disturbing note in co-organizer Nobuo Ohta's opening remarks to the Conference, summarized in an appendix to the Preface in this volume, is that recently Japan has been undergoing extensive changes in education-- reduction in the length of the school year and a shift of focus from the acquisition of knowledge to motivational aspects of learning. The new focus is on the abilities, skills, and habits children need in order to cope with problems of everyday living. Does this development presage an end to the series of Tsukuba conferences on memory and learning? Perhaps not, for there are reasons to believe that the shift away from emphasis on acquisition of knowledge may be transitory.

Whatever may be the problems faced by Japanese children in their local neighborhoods, the children will grow up to face ever escalating problems of global scope: depletion of natural resources, degradation of the environment, viral pandemics, terrorism, the complexities of the world economy. Preparation of Japan, as of the U. S. and other countries, to cope with these problems, will require comparably escalating efforts to arm their citizenries with knowledge and understanding of the problems and possible solutions.

In this connection, I am reminded of a communication from the British Royal Society recently reported in the *Manchester Guardian* under the heading: Warning on Science Education, and beginning, "The crisis in science education [too few students studying science] may inflict huge damage on Britain's prosperity and quality of life." Similar warnings have come in recent years from the U. S. National Academy of Sciences and National Science Foundation.

Efforts to meet this crisis will depend on enhanced, not diminished emphasis by educators on the acquisition of knowledge and on communicating new developments in the cognitive sciences to those engaged in relevant educational research and its applications. This communicative process has been furthered by the Tsukuba conference series, and perhaps it will continue to be in the future.

This page intentionally left blank

Preface

Welcome to the 4th Tsukuba International Conference on Memory (Tic4) — Human Learning and Memory: Advances in Theory and Application

Chizuko Izawa, Tulane University, USA and Nobuo Ohta, University of Tsukuba, Japan

On Sunday, 19 November 2000 at the Psychonomic Society Meeting in New Orleans, Louisiana, USA, Chizuko Izawa accepted the honor and responsibility as a co-organizer, overseas-organizer, co-chair, and associated duties of the 4th Tsukuba International Conference on Memory (Tic4) to be held at the University of Tsukuba in Japan during Academic Year 2002-2003 under the auspice of the Government of Japan.

Her responsibilities included inviting all overseas speakers and poster presenters, writing a grant proposal in English that addressed a coherent and significant theme, editing the English texts of all oral presentations, which were to appear in the volume of the conference proceedings for Tic4. Nobuo Ohta was to take charge of all financial matters, including the acquisition of the necessary funds from the *Monbusho* [the Japanese Ministry of Education, Science, Sports, and Culture], translating the proposal by the Overseas Chair into Japanese, as well as essential travel arrangements for overseas invitees. He was also responsible for local arrangements, daytime scientific presentations, and the opening night hospitality/banquet as well as the Japanese-cultural productions. In addition, Professor Ohta arranged for all oral presentations to be translated into Japanese and he collated all individual chapter indexes prepared by the authors into one for the entire Tic4 volume.

The University of Tsukuba (Japan's most modern and highly prestigious national university located about 80 km northeast of Tokyo) was built in 1972 by the then Prime Minister Kakuei Tanaka with an impressively healthy budget for "Science City" at Tsukuba, a community specifically created to support this new comprehensive university. In addition to absorbing all departments of the former Tokyo University of Education, it has created many other disciplinary venues that are the lifeblood of cutting edge science education and technologies.

This spacious and appealing campus features beautiful multi-level landscaping, which includes a scenic river and creeks, running among diverse stands of pine. It is serviced by a modern but non-intrusive superhighway along with the national rail and city bus nets. Inclusivity and openness to international exchanges exemplify the University of Tsukuba. Overseas scientists working in every field enjoy modern, high-tech facilities. Substantial student/scholar discounts are available for room, board, and campus purchases (as is true in all universities/colleges in Japan). A bicycle is a must, because most campus roads between buildings are off limits for automobiles. Those who do not use bicycles may ride city or campus buses, which continuously circle the campus perimeter on an hourly basis.

International conferences are held year round in Tsukuba at on- or off-campus conference centers equipped with up-to-date technology and Western style hotels. Ohta *Sensei* [Professor Ohta] (*Sensei* = Professor, a person of great wisdom. Overseas Organizer Izawa taught all Tsukuba participants this word "*Sensei*" before they left for Japan!) helped to get the Tsukuba International Conference on Memory (Tic1) started in 1998 after successfully acquiring the first grant for this series from the *Monbusho*. The current conference is its fourth, Tic4.

For those who were unable to come to Tsukuba for Tic4, Chapter 1 and Postscript will portray as closely as possible how this distinguished conference proceeded speech by speech. For those of us who were there, this will be a chance to revisit the fond memory that was Tic4. Furthermore, for most overseas participants, this conference led to their first visit of Japan. Many took advantage of the opportunity to experience much of the more than 2,100 year old Japanese cultural heritages that characterize Kyoto, Nara, and Tokyo; but more of that later (Chapter 1 and Postscript Appendix PS.6).

Other "firsts" in the Tic series include a Program Summary in Appendix PS.1, all poster presenters from Japan and overseas are seen in Appendix PS.2, and all officers of Tic4 are acknowledged in Appendix PS.7 in the Postscript. In addition, we invite you, especially many of those from overseas who are so fascinated by Japan and its enduring traditions and cultures, to visit Postscript Appendix PS.6.

Tic4, as compared to the previous conferences, was most blessed by the inclusion of William K. Estes' foreword! He is the recipient of the United States National Medal of Science, whose name was one of the most frequently cited during Tic4 presentations.

Let us now recreate the unforgettable journey to the 4th Tsukuba International Conference on Memory, *Human Leaning and Memory: Advances in Theory and Application!*

APPENDIX PRF.1

The following was an excerpt of Ohta Sensei's Opening Remarks on 11 January 2003 at Tic4 (cf. Chapter 1 and Appendix PS.1):

> ". . . In the 21st century, we are pressed to solve difficult conflicts among nations, violations of human rights, incurable diseases and aging, failings of education, increases in terrorism, crime, and acts of war, and so on. Psychologists need to address these problems with greater urgency and contribute to the well being of humankind. The starting point for problem-oriented research ought to be real world problems. Theory-based basic research is valuable especially when it has potential for solving real life problems. In the 21st century our global situation is becoming worse and worse. In my opinion, the impetus for psychologi-cal research must incorporate real life problems and their solutions.
>
> One example of real life problems in urgent need of research revolves around those of education in Japan. Recently, Japan has been undergoing major changes in education. It has adopted a 5-day school week and cut as much as 30% of the school curriculum. Evaluations of students have shifted from a relative evaluation system to an absolute evaluation system. Although knowledge aspects of learning were previously regarded as

important, emphasis has shifted toward motivational aspects of learning. Amidst all these changes a new curricular focus has emerged: comprehensive learning. Its purpose is coping skills for daily living — skills to surmount obstacles and to actualize one's potential everyday. Although these educational reforms have been introduced only recently, they raise many important issues. Central questions: What are the abilities for daily living? How do scholastic achievements relate to those abilities required for daily life? While various definitions of abilities for living had been proposed, they oftentimes bound reflecting contemporary concerns. Recently, people have become very apprehensive about declines in scholastic achievement scores due to a shorter school week, significant cuts to the curriculum, and the shift from knowledge-oriented learning to more motivational aspects of learning. However, we believe that such concerns reflect a narrow understanding of learning. It is important for cognitive psychologists to identify and foster abilities required by the world we live in and find efficient means for their acquisition.

Japanese psychologists should focus on tackling the problems of real life. In short, we hope that the psychology of learning will improve individual lives as well as society."

VOLUME ACKNOWLEDGMENTS

The Editors are deeply indebted to devoted editorial assistants, Althea J. E. K. Izawa-Hayden, Salvador Iloreta, and Leith Edgar who gave their precious time to help complete this volume without compensation, as interested and concerned volunteers who greatly cared about the subject matter of this enterprise and the scientific discoveries and contributions reported in this volume. They were "quick studies" who did quality work with dispatch seasoned by early professional experiences.

At New Orleans and Tsukuba
Spring, 2004

Chapter 1

Introduction and Contributors to the 4th Tsukuba International Conference on Memory (Tic4) — Human Learning and Memory: Advances in Theory and Application

Chizuko Izawa
Tulane University, USA

"I am enough of an artist to draw freely upon my imagination. Imagination is more important than knowledge. Knowledge is limited. Imagination encircles the world." (Albert Einstein, 1929)

INTRODUCTION

Tsukuba International Conferences on Memory (Tics)

Memory is central to practically all cognitive processes, if not all psychological processes. Indeed, memory processes approach omnipotence, when learning phenomena are considered at both the level of theory and application, especially in education. It is well recognized that child rearing and education reflect local traditions and cultures. Indeed, how the young are educated may differ from country to country, from culture to culture, albeit the governing principles may be the same everywhere.

Thus, both theoretical and applicational memory research may greatly benefit from the exchange of ideas, and comparison of psychological research worldwide. An ideal place for such cross-national and cross-cultural conferences is the University of Tsukuba (cf. Preface); where Nobuo Ohta *Sensei* [Professor, a person of great wisdom] initiated the first of the Tsukuba International Conference Series on Memory in 1998.

Given the tremendous scope of memory issues within psychology, it made sense to assign each Tic conference a theme to be narrowed down to manageable and productive proportions, amenable to discussions/ presentations. As for the frequency of conferences, an interval of 1-3 years was viewed as useful for keeping up with the rapidly evolving knowledge base.

Consequently, Tic1, held in March 1998, addressed *Memory and Consciousness*, boasting of such speakers (in the alphabetical order) as: J. R. Anderson, J. Engelkamp, J. Gardiner, L. Jacoby, M. Masson, L.-G. Nilsson, P. Perruchet, L. Reder, and H. L. Roediger.

Tic2 explored *Lifespan Memory Development* (December, 1999) with: L. Bäckman, D. F. Bjorklund, N. Cowan, A. de Ribaupierre, R. Fivush, P. Graf, G. Hitch, E. A. Maylor, L.-G. Nilsson, D. C. Rubin, T. A. Salthouse, H. Tajika, and H. Sakata.

Tic3 examined *Memory and Society* (March, 2002), with: D. Albert, E. Bjork, R. Bjork, M. Conway, F. Craik, D. Herrmann, Y. Itsukushima, S. Lindsay, R. Logie, I. Lundberg, M. Mimura, L.-G. Nilsson, K. Pezdek, J. Schooler, and B. Wilson.

The 4th Tsukuba International Conference on Memory (Tic4) — *Human Learning and Memory: Advances in Theory and Application*

In the present unsettling and often-violent times, it is essential that the good citizens of the world foster as much normalcy in daily activities as is possible. Come what may, war, pestilence, or peace, science must advance, and enhance human learning and education. This is the key to ultimate world peace.

The Tic4 theme was both critically important and most timely. Indeed, almost nothing is more important in the world today than enhancing efficient learning at school, work, home, and everywhere else. An ever changing/developing technology demands a process of life-long-learning, and therefore a better understanding of memory processes. Hence, the current theme, "*Human learning and memory: Advances in* *theory and application.*"

Furthermore, Tic4 strove to present a number of sophisticated formal mathematical/quantitative models, not well represented in the first three Tic sessions. (Because Japan's mathematics education is one of the best in the world, developments in formal quantitative-theoretical approaches in learning and memory should be more pervasive. What seems needed in Japan is more encouragement. Indeed, Tic4 provided this sorely needed stimulus.)

Learning is a major component of all education, and in today's world, every country has the essential task of enhancing learning levels. This includes attention to all types of learning in the population, not only children, but also young and older adults involving all of their five senses: For example, learning the meaning of verbal and non-verbal materials such as: letters, characters, symbols, words, numbers, sentences, colors, pictures, faces, music, odors, and tastes, as well as learning to read. Learning motor skills is also essential to most of what is required in daily life, as is the manipulation of all sorts of devices (computers included). Indeed, learning to use our own body in certain ways as in sports and health giving exercises is part of essential learning.

Of crucial import are the effective applications of learning for rehabilitating the disabled, those prone to mental disturbances, or leading misdirected lives (e.g., criminals and terrorists). For the latter, fundamental redirections toward the acquisition of civility and decency must take place. Thus, Tic4 had the imperative task to address a great human need.

Both Professors Ohta and Izawa, the organizers for the 4th Tsukuba International Conference on Memory (Tic4), sought to achieve the difficult goal of surpassing the previous three Tic conferences by emphasizing memory research which was not substantively addressed during the earlier three Tic sessions. They strove to survey as many diverse facets of memory as possible within each selected theme. Tic4 was held on 11- 13 January 2003.

The organizers noted: The success of Tic4 was primarily a function of the outstanding scholars who gathered there. They included a recent recipient of the David Rumelhart Prize (Psychology's version of the Nobel Prize), Richard M. Shiffrin, the keynote speaker. A major Tic4 innovation in this series was the translations of all oral presentations from English into Japanese.

Eleven overseas speakers and one Japanese speaker were selected from five nations in four continents. They were (in alphabetical order):

Charles J. Brainerd, University of Arizona, USA
Nelson Cowan, University of Missouri, USA
Alice F. Healy, University of Colorado, USA
Lynn Hasher, University of Toronto, Canada
Michael S. Humphreys, University of Queensland, Australia
Chizuko Izawa, Tulane University, USA
Jun Kawaguchi, Nagoya University, Japan
Douglas L. Nelson, University of South Florida, USA
Jeroen G. W. Raaijmakers, University of Amsterdam, the Netherlands
Valerie Reyna, University of Arizona, USA
Richard M. Shiffrin, Indiana University, USA (keynoter), and
Robert L. Solso, University of Nevada, USA.

Professor Solso could not join us, due to serious illness. Also missed were Chuck Brainerd and Val Reyna who reported that the first leg of their flights to Japan was grounded due to bad weather. Fortunately, Takafumi Terasawa (Japan) and Terry Joyce (Japan/UK) *Sensei*s graciously accepted the difficult job of presenting, practically without notice, the papers by the two absent authors. In spite of these struggles on the afternoon of Day 1, all those present did marvelous jobs scientifically and otherwise during all three days, breaking a variety of records (see the Volume's Postscript).

JAPANESE CULTURAL PRESENTATIONS AT Tic4

Given the international aspects of Tic4, participants and audience enjoyed unique experiences only Japan can afford. For an in-depth historical background of the Japanese cultural heritage relevant to Tic4, see Postscript Appendix PS.6. Here (Chap. 1), we touch on only a few of the "cultural" experiences at Tic4, experiences deriving from cultural presentations, in addition to the conference's cutting edge scientific discourse.

Hina Dolls and *Nihon Odori* [Classic Japanese Dances]

Japanese *Hina* dolls are attribute of the "Girls'/Flower Festival," celebrated on 3 March for many centuries, having originated at the *Genji* "Court," described in Lady Murasaki's classic (970s-1020s), the *Tale of Genji* [*Genji Monogatari*], the first great novel in world literature (cf. Seidensticker, 1983). These dolls magnificently robed are featured in a traditional red-carpeted multi-shelf display for weeks during the spring flower/girls' festival, celebrated in every Japanese household.

A small sample of *Hina* Dolls was displayed opposite to the check-in counter in the lobby of the Headquarters Hotel, Okura. When next you witness Japanese Imperial Household Festivities, you will see the dazzling life-size *Heian* imperial robes, designed along the lines of *Hina* dolls attire.

On opening night, 11 January 2003, those at Tic4 heartily enjoyed a sumptuous Japanese feast featuring rare, fresh seafoods and lavish libations (courtesy of the University of Tsukuba). It was accompanied by a fabulous Japanese dance performance.

Each Japanese dance, performed by three virtuosos, Daichi, Hyogo, and Wakaya Kotobuki (a male and two female dancers), depicted centuries old traditions. Each motion of the dancer has meanings, as explained prior to the performances. Indeed, the performance reflects decades of dedication and training. The first was entitled, *Echigo Jishi* [Lion from Echigo], an adaptation to Japanese Dancing, in which a male dancer energetically and at times acrobatically danced, mimicking a playful young lion.

The second number was a well-known classic, *Yoimachigusa* [An Evening Primrose] which blooms quietly at night without being noticed, representing a Japanese girl of bygone days, patiently waiting for her lover's visit. The final number was another classic, *Fuji-Musume* [A Young Maiden with a Wisteria Vine], a youthful dance, inspired by the wisteria flower's (angel) spirit.

All dancers wore beautiful *Kimonos*; some made of *Nishijin* fabrics, one of the best silks showing gold and silver threads, originating in the *Heian* Period (794-1185), Lady Murasaki's era, although the origin of the *Kimono* itself dates back to the *Nara* period (710-794, then the capital). Both *Kimonos* and Japanese dances continue to thrive today as part of Japan's enduring cultural heritages.

In honor of Lady Murasaki's monumental accomplishments (cf. Postscript Appendix PS.6), true to her spirit and name (*murasaki* = purple), Co-Editor Izawa selected purple for the cover of our Tic4 Volume. Purple is also the color of mystical wisteria flowers after the grand finale of the opening night's Japanese classic dances, *Fuji-Musume* [A Maiden with a Wisteria Vine]. Imperial purple is frequently used for dignified occasions at the Imperial Court as well. Purple is also the color of the University of Tsukuba. Indeed, there is no better color than purple to honor the host of Tic4.

Moreover, "*San-Go-no-Kiri*" [Three-five Blossomed Pauloconia], the crest of the University of Tsukuba, is on the cover and constitutes the decorative theme of the Tic4 volume. In accordance with Japan's enduring cultural and historical traditions, our volume distinguishes itself by being appropriately dignified, respectful, and of high quality both in substance and appearance.

Japanese Poems [*Wakas*] and the National Anthem of Japan

The first official collection of Japanese poems published in A.D. 905, *Kokin Wakashu* [*the Collection of Japanese Poems from Ancient and Modern Times*] (*Kokinshu* for short; see Postscript Appendix PS.6 for details) includes "*Kimigayo*," the most famous of all *Waka*s. Both *Kokinshu* and *Genji Monogatari* [the *Tale of Genji*], represent one of the most creative periods of the Japanese court. Their enduring significance over the past 1,100 years may be attributable to the authors' artistic imagination and their depictions of Nature and the world in terms of human attributes (sentiments, emotions, joys, happiness, and destiny). These works were never confined to the artificial boundaries of nations, cultures, or races.

The traditions established during the *Heian* Period have been maintained in the Japanese Imperial Court ever since. For example, in early January, to celebrate the New Year, *Waka/Tanka* recitals take place annually at the Imperial Palace in Tokyo today, just as was done during the *Heian* Period in Kyoto, or much earlier. Everybody continues to be invited to submit *Waka*s annually. The winning Japanese poems are recited to the sound of traditional Imperial Court musical instruments before all winners of the year are honored at the dazzling pageantry, in the presence of Their Majesties, the Emperor and Empress of Japan and their family. At times, some winners are overseas poets, including

Americans and Europeans.

The Japanese National Anthem:
Kimigayo: His Majesty's Reign

Kimi ga yo wa	Thousands of years of happy reign be thine;
Chiyo ni yachiyo ni	Rule on, my lord, till what are pebbles now
Sazare ishi no	By age united to mighty rocks shall grow
Iwao to nari te	Whose venerable sides the moss doth line.
Koke no musu made.	

Words: by an anonymous Japanese poet in ca AD 905
English translation: by Basil H. Chamberlain.

FIG. 1.1. *Kimigayo*: The national anthem of Japan.

As for *Waka Kimigayo,* in 1880 the Japanese Imperial Household Ministry invited the public to submit musical scores for the poem to be sung uniformly (rather than a variety of different ways as was the case for the previous 10 centuries). The melody submitted by Hiromori Hayashi was selected (after an English military music teacher, John W. Fenton's did not produce satisfactory results). *Kimigayo* in Fig. 1.1 by Hayashi was played for Emperor Meiji (great-grand father of the current Emperor) for his birthday, on 3 November 1880, and became an enormous public success.

From that point forward, the *Kimigayo* has become the National Anthem of Japan and is played for practically all events/functions, public or private, especially school and sports events. Fig. 1.1 also gives the poem in Romanized (phonetically expressed) Japanese, along with an English translation by Basil H. Chamberlain (1850-1935, the First Professor of Japanese and Philology at the (then Imperial) University of Tokyo[1]) as "His majesty's Reign." Indeed, his English translation is very well done, and is most fitting to the longest lasting imperial family in the world. Currently, His Majesty, the 125th Emperor of Japan, is Emperor Akihito.

Yet, Chamberlain's or anyone's translation cannot convey all the subtleties and allusions embedded in the Japanese language, since it is inherently impossible to convey everything from one language to another. That is, of course, also true of Lady Murasaki in translation, since she does not, to this day, have an equal among Japanese writers. Her exquisite use of the language's many shadings and exceedingly subtle evocative devices present daunting challenges to even the very best translator.

Thus, the Chamberlain translation represents only one interpretation, albeit overwhelmingly accurate. "Kimigayo," however, can also be simply interpreted as "Your life/world" or "The generations of your family/household" for anyone. This is no surprise: After all, the Japanese Emperor/Empress is the head of the very large family (the nation) of Japan. This may explain, in part, why this poem has been repeatedly sung for the last 1,100 years! (See Postscript Appendix PS.6 for details.)

It is unique, indeed, that a poem written by an anonymous author in the year 905 (or earlier) became the national anthem, some 10 centuries later. It was most fitting, indeed, that our international conference on memory, Tic4, underwritten by the Japanese Ministry of Education, Science, Sports, and Culture commenced with the *Kimigayo*, at its opening session on the morning of 11 January 2003.

---

[1]The author is grateful to Historian Charles D. Chamberlain, III for his assistance in accessing this information.

BRIEF PERSONAL INTRODUCTIONS/COMMENTARIES TO CONTRIBUTORS AND CHAPTERS

This section highlights what transpired at Tic4 at the Tsukuba Conference Center from 9:00 AM on Saturday, 11 January 2003 through 4:45 PM on Monday, 13 January 2003. The program is summarized in Appendix PS.1 at the end of this volume. Co-Chair Izawa introduced all scientific presenters throughout the three conference days as well as Co-Organizer Ohta for his opening remarks. In turn Co-Chair Ohta introduced Co-Organizer Izawa for her closing statement. Co-Chair Yama, as timekeeper, kept each speaker on schedule throughout the three day Tic4 gathering.

During these three days, over 225 well-known experts in memory presented 11 incisive talks and 79 excellent posters (of the latter, 13 represented the efforts of 29 memory specialists invited from abroad). Their data derived from many thousands of human participants (See Appendix PS.2 for all posters on the Tic4 program). This in itself is no small achievement: They came from 10 nations in four continents: Europe, Australia/New Zealand, North America, and Asia.

What follows now summarizes both attributes of the contributors and their papers. Extensive career histories are available from the usual sources. However, before everything else, we salute the most eminent psychologist of all:

WILLIAM K. ESTES (Foreword)

Notwithstanding his physical absence from Tsukuba, Bill Estes' work was ubiquitous at Tic4. We are indeed most thrilled and honored, to have Bill Estes, perhaps the world's foremost cognitive psychologist, a mentor and friend to many of us, grace this volume with his foreword!

Bill Estes earned his Ph. D. from the University of Minnesota (B. F. Skinner, an early (1968) recipient of the United States National Medal of Science, as his dissertation chair). Thereafter Estes was on the faculties of Indiana, Stanford, Rockefeller, and Harvard Universities. Subsequent to his retirement from Harvard, he returned to Indiana in 1999 as a Distinguished Scholar in Psychology and Cognitive Science (history does repeat itself!). The quality of his scientific creativity, insightfulness, and scale of contributions are simply unparalleled, as has been true of his commitments to students and willingness to share his talents with colleagues, friends, and all those committed to exploring experimental psychology and cognitive science.

Estes pioneered the foundation of modern quantitative psychology and clearly demonstrated the productive consequences of mathematical reasoning for such fields of cognitive psychology as learning, memory, visual perception, attention, categorization, conceptualization, and many others, with 13 weighty volumes, about 200+ quality papers (he never had the time to count them!) and the more numerous influential oral presentations at all manners of scientific and academic conferences/ gatherings both at home and abroad (probably exceed other contributions in number but here also his creative activities have left him little time for counting and sorting).

His productivity after retirement from Harvard in 1989 is equally impressive: Two books and a score of major articles, plus many speeches of importance, including his key-note address at the Psychonomic Society meeting in 2001!

Bill Estes has also championed the sound development of the science of psychology as editor of major journals including *Journal of Comparative and Physiological Psychology*, *Psychological Review*, and *Psychological Science*, and as head of such organizations as the Midwestern Psychological Association, American Psychological Association (APA) Division 3, Psychonomic Society, Society for Mathematical Psychology, etc.

Also as one views the world today, his generous work for world peace deserves major emphasis. The National Research Council's Committee on Prevention of Nuclear War was chaired by Bill Estes! As a small child who directly experienced and is still haunted by the absolutely horrific terrors of "strategic" bombing during WWII, it is my position that no child should ever be killed, wounded, or terrorized by acts of violence. I must, therefore, urge everyone to commit to and work together for the prevention of *all* wars, and the positive promotion of world peace!

It is no surprise that Bill Estes has collected numerous impressive honors and awards. He was a very early recipient of the Distinguished Scientific Contribution Award (1962), and later the Gold Medal for Lifetime Achievement in Psychological Science (1992) from the American Psychological Association. He also was awarded the Warren Medal by the Society of Experimental Psychologists, elected to the National Academy of Sciences, and is a Fellow of the American

Academy of Arts and Sciences, as well as Honorary Life Member of the New York Academy of Sciences, among many other awards/honors.

In recognition of all this, President Clinton in 1997 selected Bill Estes to receive the United States National Medal of Science. This recognized merit is of far broader scope and depth than the Nobel Prize, because of the latter's much narrower gauged selection process and the number of disciplines considered is much smaller and therefore the competition less intense. He was only the 12th psychologist so honored in the history of this award (established by the United States Congress in 1959).

INTRODUCTIONS of the Tic4 SPEAKERS by Co-Chair IZAWA

For all three-days of Tic4 proceedings all spoken introductions and comments by Bilingual Co-Chair Izawa are shown between quotation marks (" ") usually in the first segment/paragraph of each speaker's presentation. The English part Co-Chair Izawa's introduction was immediately followed by her translation into Japanese. Her brief introduction of each speaker and chapter in this volume are seen below. However, the order in which speeches are presented in this written rendition differs from the sequence of meeting presentations below.

Tic4 DAY ONE (Saturday, 11 January 2003)

"Good morning, fellow memory psychologists of the world, especially those of Japan, our host country and the country of my birth. I am Chizuko Izawa, Tulane University, USA. I am truly honored to be co-organizer and co-chair of the entire 3-day proceedings. I am in charge of invited overseas speakers and poster-presenters for the 4th Tsukuba International Conference on Memory, Tic4 for short. Its theme is *Human Learning and Memory: Advances in Theory and Application.* My immediate mission is to set the stage for the founding father of the Tsukuba International Conference on Memory to officially open this outstanding 4th conference on memory, the best of its kind on the planet."

"Out of respect for our host country, we begin the 4th Tsukuba International Conference on Memory with the national anthem of Japan, *Kimigayo.* I have been asked to distribute the English translation, and musical score to all our guests from abroad. It will also be seen on the auditorium screen (Fig. 1.1, Page 7 of this volume)."

"As per custom, I shall lead it by singing the first notes. Then, I will signal with my conductor's baton for you to join in singing *Kimigayo* from the beginning. Japanese colleagues, please sing with your pride and vigor so that not only your Prime Minister and Emperor, but especially your Departmental Chairs can hear you."

"Please rise. <Japanese National Anthem in Fig. 1.1 was sung very well by all.>
Please be seated."

NOBUO OHTA (Opening Statement, Preface Appendix PRF.1)

"Now, the moment we have waited for has arrived. Please welcome the founding father of the Tsukuba International Conference on Memory, Tsukuba's shining star, Professor of Psychology, Nobuo Ohta *Sensei*!"

Ohta *Sensei* received his Ph. D. and MA in educational psychology at Nagoya University, and is currently a Full Professor (since 1992) at the University of Tsukuba. He is a prolific writer with 110 publications to his credit singly or jointly with others, including 21 books, mostly in Japanese, but some in English. Books include textbooks, reviews-commentaries, research, and translations into Japanese (including Tulving's work). As a committed scholar, Ohta *Sensei* participates in or attends almost every International Congress of Psychology, International Conferences, the meetings of the Psychonomic Society, and the Japanese Psychological Association.

He welcomed all participants and attendees around the world, and shared his thoughts as in Appendix PRF.1 (Preface).

Tic4 Ground Rules for Questions and Answers

"Prior to the first presentation, let me indicate how 'Questions and Answers' sessions will be conducted. After each speech, about 5-10 min are allocated for questions and answers. Please ask your questions in English. It is my understanding that the Japanese audience is anxious to use English, as they often lack such opportunities in Japan. Here you have a rare chance! I encourage my Japanese colleagues to take advantage of it and ask your questions in English, followed by your own rendering into Japanese. But, if you prefer Japanese, please ask your questions accordingly, and I will handle it."

"For questions from overseas guests, not yet proficient in the local language, I will provide instant Japanese translations of your questions. Similarly, I suggest Japanese speakers answer the questions in English first, followed by their own Japanese translations. I will translate the English answers by overseas speakers into Japanese, unless they want to do it themselves."

"Let us now begin the scientific portion of Tic4 for which you have traveled so far."

Speaker 1: JEROEN G. W. RAAIJMAKERS (Chapter 4)

"I have chosen this first speaker for two reasons: One is the historical tie between Japan and the speaker's home country, the Netherlands. Japan is said to have shut its doors to foreign countries, especially to those in the West for nearly 300 years, commencing in about 1603, exactly 400 years ago. That is, however, only partially correct, because the Dutch were exempted. Thus, we Japanese have had closer ties to the Dutch people for about 3 centuries longer than with other western countries. In fact, one would not be surprised to run into descendants of Dutch export-import traders on the streets of Nagasaki. Indeed, our three Dutch participants in Tic4 may look much like their countrymen traders who did business here centuries ago. Now one of them is about to speak."

"A second reason for this speaker's attendance was our need for a high caliber memory specialist to set the tone of this International Conference. The person who fulfills this awesome responsibility is Professor Jeroen Raaijmakers of the University of Amsterdam. The translator for the first talk is Yayoi Miyaji *Sensei* from Kobe College. Professor Jeroen Raaijmakers will discuss '*Modeling Implicit and Explicit Memory.*' Please welcome Raaijmakers *Sensei*, *Dozo*! [= Please! This is another Japanese word that Overseas Organizer Izawa required of all overseas participants to learn before arriving in Japan]."

A native son of Nijmegen of the Netherlands, Jeroen Raaijmakers completed all graduate degrees, cum laude, at the University of Nijmegen, and is currently Professor of Cognitive Psychology and was until 2002 Director of Research at the Department of Psychology of the prestigious University of Amsterdam. From his graduate years forward, his close associations with Rich Shiffrin at Indiana (our keynote speaker) led to highly productive outputs as exemplified by such notable theoretical and empirical work as the SAM model (Search of Associative Memory, Raaijmakers & Shiffrin, 1981). Raaijmakers is a prolific writer in both Dutch and English having produced about 100 papers, of which about half are major publications, which include an acclaimed graduate textbook on human learning in the Netherlands. Numerous presentations at scientific meetings/conferences, and universities reinforce his enviable stature even further.

Raaijmakers, in Chapter 4, elucidates the power of modeling, especially mathematical models (based on explicit memory), and addresses the difficult task to infuse more precision into mathematical modeling of implicit memory which can only be indirectly detected at this juncture. No doubt more difficulties are ahead. Yet, his review and evaluation of previous studies suggest a promising direction for future research. Given that REM (Retrieving from Memory, Shiffrin and Steyvers, 1997) provides a theoretical account for both episodic and semantic memory, and for both explicit and implicit memory, he sets a new standard for future mathematical modeling. He has a clear vision and workable plan: An extension of the REM model toward the formulation of a unified theory for implicit and explicit memory as well as for episodic and semantic memory. We look forward to his and his group's accomplishments in years to come!

Speaker 2: CHIZUKO IZAWA (Chapters 2 and 5)

Co-Chair Ohta introduced the second speaker, Chizuko Izawa and the title of her presentation.

Bill Estes' first Ph. D. at Stanford University; Post-Doctoral Fellow at the University of California at Berkeley, Institute of Human Learning; BA, the University of Tokyo was Chizuko Izawa, later a Full Professor of Psychology at Tulane University (1980). She has published 3 books (plus 1 in press, 3 others pending) and 75 papers. She has also presented 120+ papers worldwide. These contributions deal with the theoretical, empirical, and clinical components of human cognitive processes especially learning and memory, as these relate to cross-cultural issues, gender, efficient learning/behavior, the effect of mass media on children, and health. Honors include Fellow of APA's Divisions 1 and 3, APS (American Psychological Society), and WPA (Western Psychological Association). She serviced on the Executive Committee of both SWPA (Southwestern Psychological Association, 1978-80) and SEPA (Southeastern Psychological Association, 1998-2001), and chair of SWIM (Southeastern Workers in Memory), and is a charter member of SEPA's Commission on the Status of Women, and is continuing her work with CEPO (Committee on Equal Professional Opportunity) by heading the Minority Interest Group and other SEPA Committees.

Chizuko Izawa has overcome many obstacles such as the initial language handicap as a Stanford graduate student, and later difficulties pertaining to her minority status. But these experiences forged an unyielding determination to insure equal rights for all who work to advance science.

In Chapter 2, "*Newly Uncovered Psychophysiological Processes and Study-Test-Rest (S-T-R) Presentation Program Effects for Optimal Learning: Empirical and Theoretical Investigations*," addressed at Tic 4, Izawa, Hayden, and Franklin note previously unexamined new phenomena in their original, innovative experiment. It measured heart rate (HR) and galvanic skin response (GSR), during both S (study) and T (test) trials. Tested by these newly procured data were traditional general learning theories that postulate learning is a function of the number of S trials, Bugelski's total time hypotheses (original and modified), Witten's test trial learning (S = T) hypothesis, the hypermnesia view, and Izawa's study-test-rest (S-T-R) presentation program hypothesis.

New empirical discoveries by Izawa, Hayden, et al. include: (**a**) the effects of S and T trials were distinctively different for HR and GSR, respectively, and (**b**) the amount and direction of these differences were controlled by the S-T-R presentation programs. (**c**) Overall, the S-T-R program effects strongly interact with S and T trial differentials and are larger for GSR in magnitude than for HR. (**d**) The greater the T trial density, the more efficient learning per S trial, the higher the attention/alertness activation (HR), and the greater, the learners' comfort (GSR). (**e**) the S-T-R presentation program hypothesis is well supported by these brand new data, at the expense of the four alternatives.

How do HR and GSR differ for learned items vs. unlearned items? This question will be Izawa and Hayden's next project as well as reviewing the all-or-none vs. incremental learning theory debate from novel perspectives. Is there any resolution forthcoming from psychophysiological data? Does history repeat itself? Stay tuned!

In Chapter 5 entitled, "*Optimal Foreign Language Learning and Retention: Theoretical and Applied Investigations on the Effects of Presentation Repetition Programs,*" Izawa, Maxwell, Hayden, Matrana, and Izawa-Hayden conducted experimental research in the interest of enhancing efficient education. The experiment utilized a natural German language learning situation, under the Study plus Test (S+T/S'+T') list-repetition program (traditional), item-repetition program (Izawa's innovation), and S/T alternation program (standard study-test method). Results from that original research permitted the assessment of seven types of learning theories: In addition to examining the replicability of results re the five theories dealt with in Chapter 2 (this volume), the authors newly investigated the study plus test (S+T, S'+T', S/T) presentation duration and frequency hypotheses under list- and item-repetition and S/T alternation presentation programs, respectively.

The new findings by Izawa, Maxwell, et al. clearly support the S-T-R presentation program hypothesis. Overall, large differences were observed for conditions in both short-term acquisition performances (STM) and long-term retention (LTM) data. The results favored the S/T alternation program suggesting that students program their time to include tests as in STSTST ... in moderately challenging learning situations at school and home.

Chapter 5, in mid volume specifically honors Bob Solso, a scheduled speaker for Tic4. Regrettably, he suffered a serious illness just before the meeting and was thus unable to attend.

Speaker 3 in Absentia: Charles. J. BRAINERD (Chapter 10)

"Good afternoon, and welcome back! We are most sorry to report that our first speaker this afternoon, Professor Charles J. Brainerd, an outstanding psychologist at the Informatics and Decision Making Laboratory, College of Medicine at the University of Arizona, USA, is unable to be with us this afternoon. He reported that bad weather had cancelled his first flight connecting the international airlines to Japan. Fortunately, however, one of his interpreters, Takefumi Terasawa *Sensei* is with us, and he will be able to present both the written paper by Brainerd *Sensei* and his translation with his colleague, Tazuko Aoki *Sensei* from Okayama University. The topic is, '*Fuzzy-Trace Theory and Memory.*' Terasawa *Sensei* would like to add a few words on his own research, at the end of the presentation. Terasawa *Sensei, Dozo*!"

A Michigan State Ph. D. in Developmental/Experimental Psychology, Charles J. Brainerd is well known for the fuzzy-trace theory and works closely with Val Reyna. Prolific Chuck Brainerd has published over 130 papers singly or jointly with others (some in press), 19 monographs/books, 26 commentaries/reviews, and 76 presentations at meetings on cognitive development, false memory, forgetting, judgment/decision making, mathematical psychology, and psychology of law among others.

Given that the concept of false memories has aroused substantial interest in recent years, we must comment on aspects of it: A good theory is the fuzzy-trace theory (FTT). In Chapter 10, Brainerd effectively explains how FTT involves dual or opponent processes model which excels over older one process models such as constructivism and resource-monitoring models, or other dual process models to predict and account for complex results of true and false memories, including: the developmental aspects of children, adolescents, and adults. Brainerd maintains that FTT interfaces with memory's higher reasoning processes, and sees promise in this interface for integrative theorizing via research on developmental relations between memory and reasoning. An excellent viewpoint, indeed.

Speaker 4 in Absentia: VALERIE REYNA (Chapter 11)

"We are again very sorry to inform you that our next speaker, Valerie Reyna *Sensei*, Professor of Surgery and Medicine from the University of Arizona, USA was also booked on the cancelled flight from Tucson, Arizona and cannot be with us this afternoon. However, we are fortunate that Terry Joyce *Sensei* from the UK/Japan, graciously offered to read Val Reyna's text for us. She reports that she currently serves as a Senior Research Advisor to the Office of Educational Research and Improvement in the US Department of Education. Her paper is on *'Fuzzy Trace Theory, Judgment, and Decision-Making,'* translated by Shigeru Ono *Sensei* of Tokyo Metropolitan University. Joyce *Sensei* and Ono *Sensei, Dozo!*"

A 1981 Rockefeller Ph. D., returning from a Bush Administration appointment to her teaching post as Professor of Psychology, now at the University of Texas at Arlington, Val Reyna is a versatile and prolific writer. In close association with Chuck Brainerd, she has authored over 95 publications singly or jointly with others (a few more in press) including 2 books, and over 95 presentations on medical judgment, decision-making in children, adolescents, and adults, learning, memory, suggestibility, language, reasoning and cognitive development, among others.

In Chapter 11, Reyna presents a thoughtful review and critique of traditional unitary process assumptions about relationships between memory and higher cognitive functions such as reasoning, judgment, and decision-making, and advances rationales for the advantages of the dual processes models, especially those of FTT that can account for opposing processes. Children and adults acquiring expertise demonstrated developmental progress/evolution from verbatim-analytical processing to gist-intuitive processing. Her illustrative examples and error analyses are both enlightening and fascinating. A very interesting chapter!

Tic4 DAY TWO (Sunday, 12 January 2003)

"Good morning Fellow Memory Psychologists. Welcome to the second, or premier day of Tic4 including the keynote speech this morning and a plethora of poster presentations in the afternoon."

Speaker 5: MICHAEL S. HUMPHREYS (Chapter 3)

"The first speaker of this important day is a former Head of the Psychology Department, and current Director of the Key Centre for Human Factors and Applied Cognitive Psychology of the University of Queensland, Australia. He is Professor Michael S. Humphreys, an ex-classmate of mine at Stanford. He presents '*Recollection and Familiarity.*' Ryuta Iseki *Sensei* from the University of Tsukuba will interpret this presentation. Humphreys *Sensei* and Iseki *Sensei*, *Dozo*!"

Another Estes Ph. D. from Stanford, an excellent psychologist and productive author, publishing in excess of 87 papers, responsible for numerous presentations at scientific meetings/conferences in diverse fields of human memory and its processes, including the theoretical and application domains, in recognition and recall, the role of context in human memory, similarities and differences between direct and indirect memory tests, lexical access and the use of cues and codes in short-term memory, among many others. He has worked in and acquired grants from the United States, Canada, and Australia, and is currently among the most eminent psychologists in Australia.

Chapter 3 ponders similarities and differences between recollection and familiarity. Initially, if we emphasize the differences between these two psychological processes, we are led to the independence hypothesis as in, for example, Jacoby and Dallas, (1981), Mandler (1989), Tulving (1985), or Yonelinas (2002). If, however, emphasis is placed on similarities, the redundancy hypothesis beckons, as may be seen in many studies by Humphreys and his colleagues and those of Gillund and Shiffrin (1984) among many others.

However, things are far more complex as "context" and other complex experimental variables are considered. Humphreys and Maguire have painstakingly reviewed, analyzed, evaluated, interpreted, and synthesized many relevant studies in detail for and against these hypotheses, as well as hybrids of these two.

Notwithstanding complexities (e.g., Macken, 2002), in pursuit of clarification, Humphreys and McGuire conducted a new experiment, and fostered significant insights into future research requirements.

Speaker 6, Keynoter: RICHARD M. SHIFFRIN (Appendixes PS.3 and PS.4)

"After much waiting, I now present our keynote speaker to you. We are very honored to have one of the most eminent psychologists in the world today, Distinguished Professor Richard M. Shiffrin from Indiana University, USA. He is the recipient of last year's David Rumelhart Prize (Psychology's version of the Nobel Prize), and 2004 Distinguished Scientific Contribution Award of the American Psychological Association in recognition of his many enduring innovative theories and models, including the Atkinson-Shiffrin Rehearsal Buffer, SAM (Search in Associative Memory), and REM (Retrieving Effectively from Memory) Models."

"Rich Shiffrin's accolades include appointment to the Luther Dana Waterman Chair and Professorship at Indiana. He is a former editor of the *Journal of Experimental Psychology: Learning, Memory, and Cognition*, and co-edited many volumes. Today, Shiffrin *Sensei* will present '*Modeling of memory and perception*'" translated by Hideaki Shimada *Sensei* of the University of Tsukuba. Shiffrin *Sensei*, and Shimada *Sensei*, the stage is yours. *Dozo!*"

Because of previous time commitments demanded by our scientific community, Rich Shiffrin is regrettably unable to write a chapter at this time for our Tic4 volume. His star talent as the world's foremost psychologist will be, however, shown at the next opportunity. We are most fortunate that he was able to grace the Tsukuba International

Conference on Memory in January 2003 in Japan. For those who were unable to attend Tic4, Appendix PS.3 in the Postscript section gives an abstract of Shiffrin's presentation. For those interested in more details, his contact information is: Distinguished Professor Richard M. Shiffrin, Cognitive Science Program, Department of Psychology Indiana University, Bloomington, IN 47405 USA. E-mail: shiffrin@indiana.edu.

In addition, for the time being, the readers of our volume can get a glimpse of Shiffrin's writing talents via Appendix PS.4, which Co-Editor Izawa accepted for publication earlier. This piece is highly valuable, for only he can write it! Enjoy!

Speaker 7: JUN KAWAGUCHI (Postscript Appendix PS.5)

"I find special pleasure in welcoming a delegation from the host country, Japan. The psychologist to represent this land of the rising sun is Professor Jun Kawaguchi from Nagoya University, located east of Kyoto, west of Nagano, a recent Winter Olympics site at the very center of Japan in Honshu, the largest of Japan's four main islands. He will discuss '*Interaction between Memory and Environment: Automatic and Intentional Processes.*' Being very good at Japanese, he presents his talk both in English and Japanese. Please welcome Kawaguchi *Sensei*!"

Most regrettably, Kawaguchi *Sensei* is unable to participate in our Tic4 volume. However, we are happy to include the talk's abstract in Appendix PS.3 under Postscript. For those interested in his approach, please contact him as follows: Professor Jun Kawaguchi, Department of Psychology, Graduate School of Environmental Studies, Nagoya University, Furo-cho, Chigusa-ku, Nagoya-shi, Aichi-ken, 464-8601, Japan; phone & fax: +81-52-789-3505; e-mail: jun@info.human.nagoya-u.ac.jp.

Speaker 8: ALICE F. HEALY (Chapter 6)

"I am especially happy to present our next speaker, an outstanding psychologist with a prolific record of publications and many other professional accomplishments, including the editorship of *Memory & Cognition*, and associate editorship of *Journal of Experimental Psychology: Learning, Memory, and Cognition*. She is Professor Alice

F. Healy from the University of Colorado, USA who enlightens us about '*Optimizing the Speed, Durability, and Transferability of Training*'. Hama Watanabe *Sensei* of Nagoya University will translate her speech. Please welcome Healy *Sensei* and Watanabe *Sensei*. Dozo!"

Alice Healy is the first Estes' Ph. D. at Rockefeller University (1973), and an extensive contributor to psychology, in addition to serving the profession in many other important capacities. During her three-decade career, she has attracted numerous academic collaborators to produce over 150 publications (including 6 books), not counting those currently in press, and made over 200 presentations at professional meetings/conferences, universities, and scientific gatherings.

Her group of authors, mostly her students and colleagues, six collaborators in all, consolidated efforts to produce Chapter 6. It overlaps with Izawa's Chapters 2 and 5 on optimizing learning and retention. Healy and her colleagues pursue optimization of learning speed, durability, and transferability in applied settings including skill learning. Healy, Kole, Wohldmann, Buck-Gengler, Parker, and Bourne summarize a series of several recently published and unpublished studies that aimed at optimizing and balancing three processes of training and its utilization: (**a**) speed of acquisition, (**b**) retention, and (**c**) transferability.

Most interestingly, optimizing the totality of these processes may not necessarily coincide with optimizing any one of them. Indeed, the Gestalt principle still stands the test of time (nearly a century since Wertheimer, 1912): The whole still may not be a sum of its parts! According to the authors, practice for applications/tests should include all elements in order to be optimal. Very informative!

Tic4 DAY THREE (Monday, 13 January 2003)

Speaker 9: NELSON COWAN (Chapter 7)

"Good morning, fellow memory psychologists of the world! On this final day of Tic4, the first of today's high-powered speakers is Professor Nelson Cowan from the University of Missouri at Columbia, representing the United States' mid-west. Middlebush Professor of the Social Sciences at Missouri, Cowan *Sensei* discusses '*Working-Memory Capacity Limits in a Theoretical Context.*' Satoru Saito *Sensei* of the University of Kyoto serves as his interpreter. Cowan *Sensei* and Saito *Sensei, Dozo*!"

A 1980 Wisconsin Ph. D., Nelson Cowan's versatility and productivity is abundantly clear through approximately 110 publications singly or jointly with others, including 2 books, plus several more papers currently either in press or pending. His interests lie in various aspects of both child and adult short-term memory (STM) or working memory (WM), and especially relationships between WM and selective attention in information processing. Cowan is a charismatic speaker, and thus over a half of his presentations at symposiums, universities, and professional conferences/meetings are by invitation.

It has been 30 long years since a seminal work, "working memory" by Baddeley and Hitch (1974), stimulated a large number of investigations. At the same time, this field continues to be marked by controversies, not only between those for and against WM approaches but also among WM theorists themselves. A significant source of this chaotic situation is that no two researchers on WM can agree upon the definition of it! Specialists (e.g., Miyake & Shaw 1999) often ask, "What is working memory anyway?" This indicates the depth of the problem. To help resolve this definitional chaos, Izawa (2001, 2002) proposed a more comprehensive concept, working cognition (WC, inclusive of attention) that is not limited to the memory component of this phenomenon (e.g., Mizuno, 2002 supports Izawa's WC approach).

WM controversies are abundant, not to mention debates regarding WM capacities and/or how to measure them, attention factors, and just about everything involved. We are lucky to have Nelson Cowan discuss the limits of WM capacity in a theoretical context in Chapter 7, which focuses on attention. His theoretical contributions are well recognized; his creative summary of the current state of affairs, inclusive of new information only available from his unpublished data and that of his associates, broadens the horizon of the attentive reader. His challenge is the difficult task of measuring WM that he approaches with ideas both interesting and enlightening.

Speaker 10: DOUGLAS L. NELSON (Chapter 8)

"The next accomplished speaker is Distinguished Research Professor of Psychology at the University of South Florida at Tampa, representing the warm Southeastern part of the United States. Douglas Nelson *Sensei* talks about '*Implicit Activated Memories, the Missing Links of Remembering,*' and his interpreter is Kazuo Mori *Sensei* from Shinshu University. Please welcome Nelson *Sensei* and Mori *Sensei*. *Dozo*!"

Doug Nelson is among the most active and energetic memory specialists, often collaborating with current coauthor Cathy L. McEvoy and others. He is also a Wisconsin Ph. D., and has authored over 100 papers, a book, three websites, as well a few in press. He is an enthusiastic speaker, as evidenced by over 54 presentations both at home and abroad.

At the glorious dawn of the then new field of human learning and memory in psychology, Ebbinghaus' inaugural volume, *Über das Gedächtnis: Untersuchungen zur experimentallen Psychologie* [On memory] (1885), presented the reasons why he created 2,300 artificial words, commonly known as nonsense syllables for use as rigorously controlled learning materials. With these, learners then were presumed to all start at the same base line for each nonsense-syllable/artificial-word, devoid of differential pre-experimental knowledge or experiences. His concern here was, of course, that pre-experimental experiences might bias current or episodic learning in unknown ways.

Over a century later, Nelson, McEvoy, and associates have proven, in a sense, the legitimacy of Ebbinghaus' concern. Pre-experimental knowledge-experience-information significantly interacts with current episodic learning. By means of free associations, often used in this field to obtain norms (already utilized several decades earlier (e.g., Glaze, 1928 to establish association values for nonsense syllables), Nelson et al. established interrelationships between and among over 5,000 words, and developed PIER2 (Processing Implicit and Explicit Representations 2), to account for how episodic learning was influenced by implicit pre-experimental associations each word elicits. Izawa will be able to add to their efforts utilizing free association data being compiled for numbers/number words, for a volume publication (contract signed).

In Chapter 8, Nelson and McEvoy summarize the current status of PIER2 with its strengths, deficits, and some comparative discussion of other models, inclusive of those not published as yet, and ongoing efforts. Theoretical and applicational outlooks are quite good. They appear to involve interrelationships and linkages among words in map form and word associative structures that vary systematically in respect to resonance, connectivity, and set size. Very fascinating!

Speaker 11: LYNN HASHER (Chapter 9)

"We have now reached the grand finale of the 2003 Tic4. As in the world of Kabuki, the most accomplished speaker, next to the keynote speaker, makes the final presentation. That is the main reason for asking Professor Lynn Hasher, the Chairperson of Psychology at the University of Toronto, Canada to do this most difficult job of bringing the 4th Tsukuba International Conference on Memory to a successful and satisfying conclusion. Translated by Etsuko Harada *Sensei* of Hosei University, Lynn Hasher *Sensei* proclaims '*It's about Time: Circadian Rhythms, Memory, and Aging.*' Hasher *Sensei* and Harada *Sensei*, please go for it. *Dozo*!"

Lynn Hasher earned her Ph. D. from the University of California, Berkeley in 1970 while Leo Postman was running Berkeley's Institute of Human Learning. Overseas Organizer Izawa was a postdoctoral fellow there shortly before Lynn received her doctorate. Hasher is a productive and prolific scientist, publishing more than 91 papers singly or jointly with others, and delivering more than 118 invited presentations, and over 83 poster presentations at various conferences on human memory, with a special focus on age-related, gerontological phenomena.

Please note here that Izawa's long-term projects indicated that objectively the more tests, the better the learning (e.g., Chaps. 2 & 5 of this volume). Yet, it is hard to find students who love tests! That is, participants' subjective preference may not necessarily correlate positively with their objective performances.

Quite interestingly, however, at the time of testing, things are quite different. Hasher, Goldstein, and May in Chapter 9 demonstrated that personal characteristics, such as "Morningness" and "Eveningness," are positively correlated with the preferred time of being tested. In turn, the participants' subjective preferences correlate positively with objective measures of their performances.

Hasher et al. present very intriguing and important facts about Circadian rhythms' impact on memory tasks in comparisons of young and old adult participants. Generally, younger adults prefer afternoons, older ones, mornings. Although well-established strong knowledge may not be subject to the synchrony effect (testing during a subjectively and a somewhat narrowly defined peak time), performance that requires intricate cognitive tasks that include all sorts of memory elements and interferences are highly susceptible to the time of day. When old and young participants were compared in the early morning, only small, often non-significant differences were found. However, in the late afternoon, differences dramatically increased to the disadvantage of the older group. The authors rightfully pointed out that these comparative studies without regard of circadian rhythms are likely to artificially inflate the differentials between the age groups with the lower scores attributable to the older group.

It is must reading, especially for those of us, 50+ who are really kicking along just as young as ever! Likewise, college students who are failing morning examinations may be able to request accommodations for an afternoon test? Clear expositions, summarizing extant findings plus projects currently in progress as well as the latest developments from Hasher's laboratory serve to enhance knowledge about this field.

SPECIFIC REFERENCES

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. H. Bower (Ed.). *The psychology of learning and motivation: Advances in research and theory.* (pp. 47-89). New York: Academic Press.

Bugelski, B. R. (1962). Presentation time, total time, and mediation in paired associate learning. *Journal of Experimental Psychology, 63*(4), 409-412.

Ebbinghaus, H. (1885). *Über das Gedächtnis: Untersuchungen zur experimentellen Psychologie.* [On memory] Leipzig, Germany: Duncker & Humbolt.

Einstein, A. (1929). A quotation from *the Saturday Evening Post,* 26 October 1929.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91,* 1-67.

Glaze, J. A. (1928). The association value of nonsense syllables. *Journal of General Psychology, 35,* 255-267.

Izawa, C. (1999). Efficiency in short-term memory: Study-test-rest presentation program and learning difficulty. In C. Izawa (Ed.), *On human memory: Evolution, progress, and reflections on the 30th Anniversary of the Atkinson-Shiffrin model* (pp. 165-213). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Izawa, C. (2000). Total time and efficient time management: In search of optimal learning and retention via study-test-rest presentation programs. *American Journal of Psychology, 113*(2), 221-248.

Izawa, C. (2001). A new temporary maintenance system, working cognition: A construct to incorporate working memory in unified theories of cognition. [*Review of the book: Models of working memory: Mechanisms of active maintenance and executive control*]. *American Journal of Psychology, 114* (1), 115-126.

Izawa, C. (2002). Strengths of working cognition, a temporary maintenance system that incorporates working memory. *The Annual Report of Educational Psychology in Japan, 41,* 194-204. (In Japanese).

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

Macken, W. J. (2002). Environmental context and recognition: The role of recollection and familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 153-161.

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review, 87,* 252-271.

Miyake, A., & Shaw, P. (1999). (Eds.). *Models of working memory: Mechanisms of active maintenance and executive control.* New York: Cambridge University Press.

Mizuno, R. (2002). The durability of the effect of an optimal spaced learning method, the modified low-first method: Demonstration and development of a predictive model. *Cognitive Studies, 9(4),* 532-542.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88,* 93-134.

Seidensticker, E. G. (1983). Tale of Genji. *Kodansha Encyclopedia of Japan, 7,* 328-331. Tokyo: Kodansha.

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM: Retrieving effectively from memory. *Psychonomic Bulletin and Review, 4,* 145-166.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychologist, 26,* 1-12.

Wertheimer, M. (1912). *Exprimentelle Studien ueber das Sehen von Bewegungen, 61,* 161-165.

Whitten, W., & Leonard, J. (1980). Learning from tests: Facilitation of delayed recall by initial recognition. *Journal of Experimental Psychology: Human Learning and Memory,* 6(2), 127-134.

Yonelinas, A. P., & Jacoby, L. L. (1996). Noncriterial recollection: Familiarity as automatic, irrelevant recollection. *Consciousness and Cognition, 5,* 131-141.

GENERAL REFERENCES

Parts of general backgrounds/sources of the detailed historical facts for the first section of Chapter 1 came from the nine volumes of the *Kodansha Encyclopedia of Japan*, Tokyo: Kodansha (1983), including some of the sections by E. Minor, E. G. Seidensticker, Y. Suzuki, and S. Thompson, among many others.

Chapter 2

Newly Uncovered Psychophysiological Processes and Study-Test-Rest (S-T-R) Presentation Program Effects for Optimal Learning: Empirical and Theoretical Investigations

Chizuko Izawa, Robert G. Hayden, and Michael Franklin
Tulane University, U. S. A.

> *"To see with one's own eyes, to feel and judge without succumbing to the suggestive power of the fashion of the day, to be able to express what one has seen and felt ... –is that not glorious?" (Albert Einstein, The World as I See it, 1949).*

Are there better ways to learn? We focus in this chapter on novel perspectives of study (S) and test (T) presentation programs. An S event in cued-recall/paired-associate learning (PAL) is defined as the presentation of both cue/stimulus and target/response terms of the pair, and a T event, as the presentation of the cue alone for the participants to respond with the target, without feedback.

Since 1885, the great majority of theoretical and empirical investigations in learning, including those classics of major importance in its history by Ebbinghaus (1885), Hull (1943), and later modern mathematical approaches beginning with Estes (1950), concentrated primarily on S (study) trials. They made the general assumption that learning increments are a positive function of the number of S trials, N. The larger the N, the greater the learning, L (i.e., $N+1$ S trials produce higher L than N S trials), as for example in Equation 2.1:

$$L = \mathrm{f}(N). \qquad (2.1)$$

Let us refer to the largest "family" of learning theories among the most influential ones as general learning theories (the more S trials, the greater the learning). The popularity of countless numbers of theoretical positions expressable by Equation 2.1 comes, in part, from its intuitive and common sense appeal, backed by daily experience: Increasing S (study) sessions raises acquisition and retention!

T (test) trials, on the other hand, have long been regarded as nothing but a way to measure acquisition (L) on S trials, and thus have long been neglected, notwithstanding early advice by Gates (1917) to the contrary. To make matters worse, no adequate experimental methods were developed to research T trials. For example, in cued-recall/PAL situations, under the traditional/standard anticipation method, extant since the days of Ebbinghaus/Pilzecker, was accepted as fitting Skinnerian teaching machine principles (Appendix 2.1). Here, T and S trials for each item came alternately as T-S, T-S, ... in any cycle, making it impossible to separate T effects from S effects.

However, under the standard study-test method since about the 1960s, S and T trials are administered in separate cycles as in Appendix 2.2, but S and T cycles still alternate as S_1T_1 S_2T_2... Thus, it is also impossible to isolate T effects with this study-test method. Therefore, a new method had to be invented.

To examine the function of Ts per se, it is necessary to *administer T cycles successively*. Izawa, consequently, developed an innovative method for learning and teaching by programming the learner's time for S, T, and R (rest) cycles. Equation 2.2 illustrates one of Izawa's many general S-T-R presentation programs (e.g., Izawa, 1966, 1992, 1999, 2000, 2003). It specifies a unit/pattern of repetitions or *replications*:

$$S_1 \ldots S_i R_1 \ldots R_j T_1 \ldots T_k R_1 \ldots R_l, \qquad (2.2)$$

where the subscripts $i > 1$, $j \geq 0$, $k \geq 0$, and $l \geq 0$. Under the list-design, each S, T, or R stands for a study, test, or rest cycle, respectively. An S or T cycle is defined as a presentation of the n-items within a list, once each, one at a time (either as an S event or a T event, respectively) in random order from cycle to cycle. The replication units recur until the end of the experiment.

By varying subscripts, a large number of presentation programs can be generated. For example, $j = l = 0$, $i = 1$, and $k = 7$, produced Presentation Program STTTTTTT. We used this as Condition 5 of the present experiment. Interchanging subscript values of i and k alone, while holding the others constant, led to Program SSSSSSST for

Condition 1. Similarly, by letting Subscripts $j = l = 0$, $i = 3$ or 1, and $k =$ 1 or 3, Program SSST or STTT (Cond. 2 or 4) resulted. Likewise, Program ST (Cond. 3) emerged from subscript $j = l = 0$ and $i = k = 1$. The latter is the standard study-test method (Appendix 2.2), and is a special case of Izawa's more general S-T-R presentation programs in Equation 2.2. Table 2.1 summarizes the current experimental design.

Aided by this (variations of j and l generate rest/neutral trials, or spaced practice effects, e.g., Izawa, 1971, 1988, 1992) and other new systematic programming methods, free from the alternating S and T trial/cycle constraints of the standard methods, Izawa discovered several *positive* effects of T trials (e.g., Appendix 2.3).

TABLE 2.1
Experimental Design: Five Study-Test (S-T) Presentation Programs

Program/ Condition	Cycle 1 2 3 4 5 6 7 8 9	23 24 25 26
1 (SSSSSSST)	S_1 S_2 S_3 S_4 S_5 S_6 S_7 T_1 S_8 S_9 S_{10} S_{11} S_{12} S_{13} S_{14} T_2 . .	S_{21} T_3
2 (SSST)	S_1 S_2 S_3 T_1 S_4 S_5 S_6 T_2	S_{18} T_6
3 (ST)	S_1 T_1 S_2 T_2 S_3 T_3 S_4 T_4 S_5 T_5	S_{12} T_{12}
4 (STTT)	S_1 T_1 T_2 T_3 S_2 T_4 T_5 T_6	T_{17} T_{18} S_7 T_{19}
5 (STTTTTTT)	S_1 T_1 T_2 T_3 T_4 T_5 T_6 T_7 S_2 T_8 T_9 T_{10} T_{11} T_{12} T_{13} T_{14} . .	T_{20} T_{21} S_4 T_{22}

From another perspective, however, tests or examinations are known to provoke anxiety/nervousness. Test anxiety (TA) has been a universal problem for all age groups, for example, from elementary schools (e.g., Hill & Wilgfield, 1984; Luckie & Smethurst, 1998) to colleges/universities, graduate/medical schools (e.g., Frierson & Hoban, 1992; Schonwetter, 1997); and crosses geographical and racial boundaries throughout the world from the West, America and Europe (e.g., Ferrando, Varea, & Lorenzo, 1999; Hodapp & Benson, 1997; Meijer, Elshout, & Van Hout, 2001) to the East, Japan (e.g., Shioya, 1995), China (e.g., Wang, 2001; Wang & Liu, 2000), and Korea (e.g., Kim & Park, 1995). TA is present throughout wide variations of mental abilities and disabilities (e.g., Bryan, Sonnefeld, & Grabowski, 1983; Kutchukian-Boxley, 1993).

To help combat the problem of TA, scales/inventories have been developed (e.g., Test Anxiety Scale, Sarason, Kestenbaum, & Smith, 1972; Friedben Test Anxiety Scale, Friedman & Bendas-Jacob, 1997; Test Anxiety Scale for Children, Ludlow & Guida, 1991; Revised Children Manifest Anxiety Scale, Reynolds & Richmond, 1978; State-Trait Anxiety Inventory, Marteau & Bekker, 1992; etc. and their modifications world-wide). The bulk of research in this field is described in Hagtvet and Johnsen (1992). For example, TA seems primarily defined by participants' subjective responses to self-evaluation, whose criteria may differ from individual to individual.

We preferred to address TA related issues more objectively via observable physical responses such as perspiration and heart rate (HR), which reflect test takers' discomfort. Questions: How common is discomfort? Are those symptoms limited to T trials? How about S trials? Are the levels of these symptoms different for S and T trials? If so, how? To examine such psychophysiological responses, we used two objective measures, heart rate (HR) and the galvanic-skin-response (GSR). We also asked: Do these psychophysiological phenomena enhance or interfere with optimal learning?

The above questions have long been neglected in extant studies. We see only provocative fragments of information; some suggest that high HR and GSR make learning more efficient (Andreassi, 1966; Blatt, 1961; Berlyne, 1964; Berry & Davis, 1958; Geisleman, Woodward, & Beatty, 1982; Germana, 1968; Kahneman, 1969; Kleinsmith & Kaplan, 1964; Lindholm & Cheatham, 1983; Thompson & Obrist, 1964).

To make matters worse, no study seems to have compared HR or GSR for S and T trials. With that hiatus in mind, note Kintsch's (1965) study: While GSR in PAL increased initially, and remained at peak levels during intermediate trials, it habituated after learning occurred. HR has been shown to follow a similar pattern (e.g., Andreassi & Whalen, 1967).

In general, an arousal response is strongly correlated with enhanced learning (e.g., Andreassi, 1966; Berlyne, 1964; Blatt, 1961). It has been argued that in certain instances such arousal has been accompanied by memory consolidation that effectively strengthens the ability to recall items learned earlier (Kleinsmith & Kaplan, 1964). However, the same study found that associates learned under low arousal conditions showed high immediate recall, but weak long-term memory (LTM). This is in contrast to a high arousal condition in which learners demonstrated low immediate recall but stronger LTM.

Their results reinforce our position that there exists links between physiological arousal and learning/memory. Furthermore, GSR data were reported to be similar to those for HR (Andreassi & Whalen, 1967) and EEG (Electroencephalogram; e.g., Thompson & Obrist, 1963). Andreassi (1966) showed that subject's arousal varied in relation to the association value of nonsense syllables. A list of high association value nonsense syllables (easy learning) generated heightened GSR and HR. While a majority of studies indicate high arousal leads to better learning, others suggest lower arousal would result in superior serial learning (e.g., Taylor & Spence, 1952).

From an extensive review of the literature, we conclude that there is seldom a distinction in acquisition between S and T trials in respect to the extant HR and/or GSR. These studies concentrated on physiological reactions during the response phase (T trial), while neglecting the encoding phase or S trials. Thus, both the physiological data and interpretations thereof are often inconsistent.

The present experiment was designed, in part, to rectify some of these difficulties and to embark on a new voyage to examine S and T trials separately, wherever feasible, for all current response measures via systematic variations of experimental conditions (Table 2.1) in standard PAL/cued-recall learning under the traditional list design.

THEORETICAL PREDICTIONS

Another major purpose of the present study (Table 2.1) was to conduct crucial tests of the larger families of influential learning theories in order to explore optimal learning.

General Learning Theories/Models (Equation 2.1)

Most general learning theories that have influenced ideas about human learning assumed that learning is a positive function of S (study) trials in the sense that the greater N (number of S trials), the greater the acquisition (Eq. 2.1). If correct, said theories predict equal performance among the current five S-T presentation programs/conditions when N is constant, (Table 2.1).

The Total Time Hypotheses (TTHs)

According to another influential theoretical approach, the total time hypothesis (TTH. e.g., Bugelski, 1962; see Izawa & Hayden, 1993 for an extensive review of studies for and against it), total S time expended controls learning irrespective of how that total S time is utilized (strictly speaking therefore, the original TTH = TSTH, the total study time hypothesis). If so, this TTH/TSTH also anticipates the same performance levels among all S-T presentation programs as a function of total S time.

However, there has been confusion about what total time means. Thus, Izawa and Hayden (1993) and Izawa (1993a, 1993b) examined modified versions of it. One such modification is the total study and test time hypothesis, TS+TTH. It predicts similar performance among all S-T presentation programs, given a constant total S+T time. The same principle also holds true for other modified TTHs.

The Test Trial Learning or S=T Hypothesis

For nearly a century, the function of T trials received little attention under the traditional list design (LTM), but starting about 1960, fragments of T trial information became available via, for example, short-term memory (STM) item design concerning T effects after a very small number of S trials (very early in acquisition) as reviewed by Izawa (1992). Some such studies found positive T effects: Birnbaum and Eichner (1971) obtained similar results for S and T trials (Conds. STST = STTT). Lachman and Laughery (1968), Rosner (1970), and Tulving (1967) also observed data consistent with these positive T effects. The most extreme example came from Whitten and Bjork (1977) and Whitten and Leonard (1980) who assumed the virtual identity of S and T trials. Let's call such a position the S=T or the T trial learning hypothesis.

If T trials are de facto the same as S trials, this hypothesis perforce predicts that performance in all programs must produce the same level of responses, because all S-T presentation programs (Table 2.1) could be construed as being identical. The T trial learning hypothesis also predicts increasingly better performances over successive T trials in each replication of Programs STTT and STTTTTTT, due to its equivalent learning assumption on both T and S trials (S=T).

The Hypermnesia View

Similar improvements over the successive T trials could be interpreted as hypermnesia, as has been done by a group of investigators in this field. (cf. Izawa et al., Chap. 5 of this volume). If so, improved performances over a T trial block may be expected. Unlike the T trial learning hypothesis (acquisition processes), this group attributes improvements to retrieval processes/search strategies at the retention phase, however.

Since Ballard (1913), a sizable family of hypermnesia studies has occasionally obtained enhanced retrievals over consecutive Ts. These typically occurred after a small number of S trials, often after only one S trial, S_1 (very early in acquisition). Hypermnesia, however, is not universal; it can occur sometimes when one or more factors are manipulated to induce it (e.g., pictorial or high-imagery items, longer recall times (e.g., Roediger & Challis, 1989), categorized lists, encoding strategies/manipulations after a S trial and before Ts, free recall T mode, instructions that encourage better performances before each repetitive T, a large number of items within the list, a "think" period to enhance performance on subsequent Ts, recalling old knowledge such as names of high school classmates, states in the US, US presidents, or the like).

Even for most favorable conditions of verbal learning (such as free or forced recall), Payne (1987, Tables 3 & 4) reported 46.2% of the cases (56/121) produced hypermnesia, and 53.8% did not. When response time was less than 4.9 s, a mere 18.6% (8/43) obtained hypermnesia, and a whopping 81.4% did not. There are much less favorable response measures, such as recognition or PAL/cued recall (notwithstanding rare exceptions such as Otani & Hodge, 1991; Otani, Widner, Whiteman, & St. Louis, 1999).

Although the hypermnesia position typically applies following the first S (S_1) or a very small number of S trials and as a retention process phenomenon in a sense, it is worthwhile to examine whether it applies to later acquisition stages utilizing repetitive replication patterns for optimal learning.

The Study-Test-Rest (S-T-R) Presentation Program Hypothesis

According to Izawa's S-T-R presentation program hypothesis (e.g., 1999, 2000; derivative of the original/modified test trial potentiating model, 1971/1992; retention interval model, 1981; identity model, 1985, 1989), learning is a function of how the learner's time is programmed for S, T, and R trials/cycles/times.

Therefore, contrasting sharply with the first three theories above which predict identity among all five S-T presentation programs/ conditions is the S-T-R presentation program hypothesis which predicts substantial differences among the S-T presentation programs, either as a function of the number of S trials (*N*), or the total time, or both.

Also, unlike one-factor theories (the first three above), the S-T-R hypothesis rests on an interactive multi-factor foundation, and thus is better equipped to handle deviant cases, for example, one factor overriding another under special circumstances (e.g., Izawa, 1999, 2000).

Physiological Response Measures

The above theoretical analyses apply also to HR and GSR responses, although none specifically addressed these physiological measures at the outset.

The T trial learning/S=T hypothesis raises the expectation that all S-T programs will produce equivalent results in HR and GSR, respectively. Furthermore, because S=T for this hypothesis, psychophysiological outcome measures are expected to be uniform (**a**) across performances for each of the five S-T presentation programs, and (**b**) over both successive S and successive T trials.

In contrast, each of the general learning, TSTH, and S-T-R presentation program positions predict differential performances between S and T trials, for different theoretical reasons. For both general learning theories and TSTH, some differences are expected between S and T trials, because each postulates that S trials control learning, but that T trials do not. Indeed, the latter have no place in any of these theories and it is implicit that such predicted differences may be inferred from the theoretical presence of S effects in the absence of T trial effects.

In sharp contrast, Izawa's S-T-R presentation program hypothesis explicitly differentiates S and T trials, which are formally postulated to participate in the production of learning and retention phenomena (via her test trial potentiating, retention interval, and identity models, respectively 1971/1992, 1981, 1985). Thus, S and T performances are generally expected to be different for all response measures including HR and GSR, except for a few special cases touched on above.

In comparing HR and GSR measures, heart rate values are constrained by the boundaries of normal physiology. There is no possibility for a HR of 0, or for values too high for the human organism. Whereas, near 0 perspiration is possible as the GSR's lowest value, while even extreme sweating would not be disabling.

METHOD

Fifty college freshmen volunteers, 10 per program, learned a list of 20 CVC-two digit number paired-associates at a 6 sec presentation rate (to facilitate hypermnesia) with 6 sec inter-cycle intervals. Both HR and GSR were recorded for both S and T trials using the accepted standard procedures in the field (cf. Acknowledgements).

For heart rate (HR), a series of six heartbeats were taken as a sample for each condition per participant. The shortest distance between each of the two adjacent heartbeats was measured in millimeters, and the smallest one was taken as the HR for that sample. Thus, widely spaced beats represent a slower HR and vice versa. The GSR was measured in ohms of electrical resistance per centimeter, and subsequently converted to micromhos of electrical conductance to facilitate easier understanding of GSR data.

RESULTS AND DISCUSSION

Learning Performances

Large Families of General Learning Theories/Models (Equation 2.1) and Total Time Hypotheses (TTHs)

Fig. 2.1 shows performances as a function of S trials, *N*, or total S time. Both general learning theories and TSTH (total study time hypothesis) predict that all five curves here will be about the same. Contradicting such an expectation is that a larger number of T trials enhance performance, thus replicating Izawa (1966, 1992, 1999, 2000, cf. Slamecka & Katsaiti, 1988). For example, on the S_3 trial, the applicable programs differed significantly: $F(3, 36) = 4.616$, $p = 0.03$: Program STTTTTTT with the highest T density performed reliably better than those in Programs SSST, ST, and STTT with less Ts. Similarly large trends $F(2, 27) = 2.355$, $p = 0.13$ were detected on S_7 trial, again S-T programs lined up according to the T density.

To examine the TS+TTH (total study and test time hypothesis), see Fig. 2.2 drawn as a function of the total S+T time/cycles. Here, the prediction of identity from TS+TTH among the five S-T programs was decisively rejected by the highly significant differences obtained throughout: $F(4, 45) = 7.35, p < 0.001$; as well as by the ultimate levels of performance: $F(4, 45) = 2.98, p = 0.028$. The same held true for another modified TTH, the total test time hypothesis (TTTH, e.g., Izawa, 1993b). Data clearly rejected both modified or unmodified TTH.

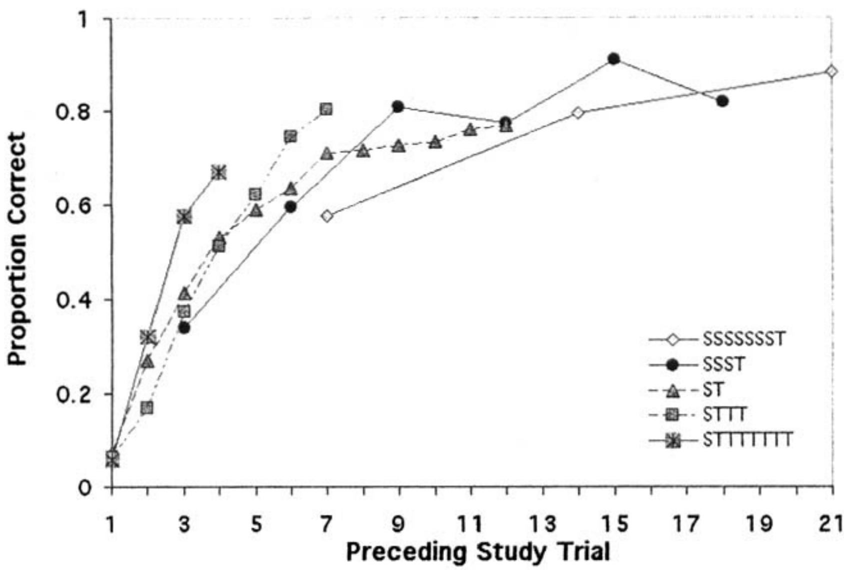


FIG. 2.1. Learning performances of the five S-T (study-test) presentation programs as a function of preceding S (study) trials.

Most intriguingly, Program STTTTTTT that had the most T trials and the fewest S trials showed the greatest potentiating effect on its S trials (Fig. 2.1). For example, on S_3 trials, the acquisition curve for Program STTTTTTT was about 75% better than Program SSST! Similarly, Program STTT was much better than Program SSSSSSST on the S_7 trial. Also, on Cycle 24 in Fig. 2.2, Programs SSST and STTT differed little: The encoding effects of 18 S trials in Program SSST in total were amazingly equivalent to only 6 S trials in Program STTT! 18 = 6? What kind of math is that?

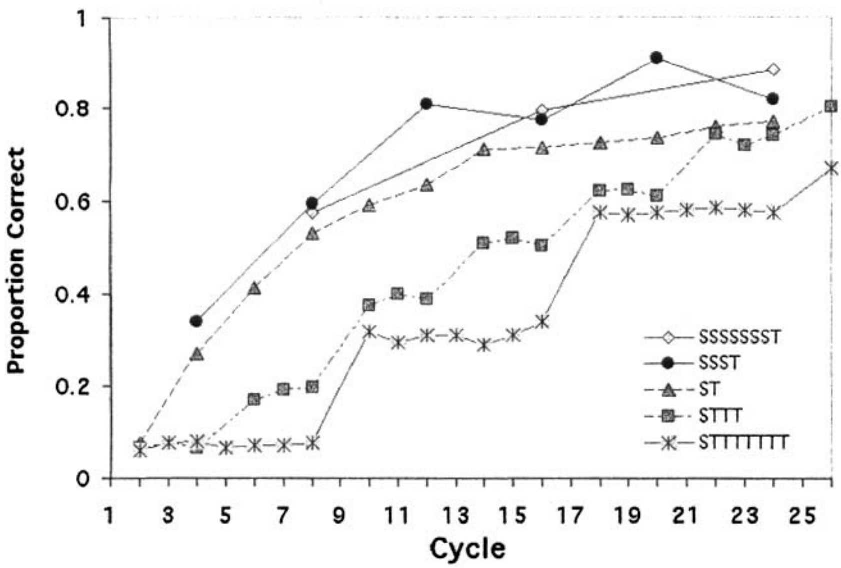


FIG. 2.2. Learning performances of the five study-test (S-T) presentation programs as a function of cycle/total S+T time.

TABLE 2.2
Mean Probability of Encoding per Study (S) Trial for Each Study-Test (S-T) Presentation Program

Presentation Program (Condition No.)	Encoding Probability
SSSSSSST (1)	0.042
SSST (2)	0.051
ST (3)	0.051
STTT (4)	0.134
STTTTTTT (5)	0.192

The magnitude of each S trial effect generally rises as the density of T trials increases as seen in Table 2.2. This is powerful evidence for rejecting general learning theories that postulate the more S trials (*N*), the greater learning (Eq. 2.1) independent of presentation programs. On the contrary, data clearly indicate control by the presentation program, viz, fewer S trials lead to greater acquisition! *The S effects are neither constant nor static across all learning methods or S-T-R presentation programs. They are highly dynamic and variable, contingent on presentation programs/methods used (cf. Izawa, 1999, 2000), as expected from the S-T-R presentation program hypothesis and observed data.*

The Test-Trial-Learning/S=T Hypothesis and the Hypermnesia View

The unambiguous differences seen in both Figs. 2.2 and 2.1 decisively reject non-significance as predicted by the S=T hypothesis.

Further damaging evidence against the T trial learning hypothesis emerged from performances during successive T trials in Programs STTT and STTTTTTT. If learning had occurred on T trials per se over the successive Ts, proportions of correct responses would have increased from the first to the last T. For Programs STTT and STTTTTTT learning curves never rose significantly from T_1 to T_3 or T_7, (Fig. 2.2). Instead, they stayed dramatically flat.

These unchanging performances over all consecutive Ts in Programs STTT and STTTTTTT also negated the hypermnesia view of ever-increasing improvements over the T block. These negative findings did not occur only after S_1 (frequently used in this field), but also held true for every S trial during all acquisition stages (Fig. 2.2).

Additional negative evidence, contrary to both T trial learning and hypermnesia hypotheses, emerged from current encoding/one-T-to-the-next probability analyses that decisively rejected the S=T hypothesis (for similar earlier results, see, e.g., Izawa, 1988, 1992, 1999, 2000).

The Study-Test-Rest (S-T-R) Presentation Program Hypothesis

Evidence challenging the general learning theories, TTHs, and the T trial learning (S=T) hypothesis is found in the large differences among the five S-T presentation programs in Figs. 2.1 and 2.2, strengthening

Izawa's S-T-R presentation program hypothesis from two different perspectives (cf. 1993a, 1993b, 1999, 2000). Indeed, learning is a function of how a constant time X is programmed (as replicated here).

Still more support for the S-T-R program hypothesis can be marshaled from other aspects of current data. For example, the stationary learning curves over successive Ts in Programs STTT and STTTTTTT provide dramatic support for the S-T-R program hypothesis: See (**a**) forgetting-prevention-effects of T trials and (**b**) no new learning on T trials per se (cf. Appendix 2.3). Our current encoding probability analyses reinforced the S-T-R program hypothesis' validity as well.

Thus, our acquisition data provided strong, internally consistent support for the S-T-R presentation program hypothesis, as well as convincing evidence against the four other theories examined.

Heart Rate (HR) Data

Overall HR as a Function of Cycles

In dealing with HR, note that a large value indicates a greater recorded distance in real time between two adjacent heartbeats. Therefore, the higher value on the vertical axis in HR graphs, the slower the HR; lower values indicate a faster HR. Despite a natural restriction on the HR's response range, important variations were nevertheless revealed.

See Fig. 2.3 for HR data of each S-T program as a function of total time/cycle. Tested concurrently, S-T program curves rose significantly: $F(23, 45) = 2.177, p = 0.001$ over the 24 cycles. The HR was generally faster early in the experiment, indicating substantial alertness or tension, but it gradually slowed toward the end. Calmness and a slower HR prevailed in all S-T presentation programs as participants settled into their tasks. Similar results were observed by others, such as Andreassi and Whalen (1967) and Blatt (1961).

HR Differences between S (Study) and T (Test) Trials

We discovered overall HR differed significantly between S and T trials, $F(1, 49) = 13.238, p = 0.007$ (suggestive in Fig. 2.3; verified in Fig. 2.4). Most intriguingly, counterintuitive results were obtained: Within the current experimental situation, the overall S trial HR was higher

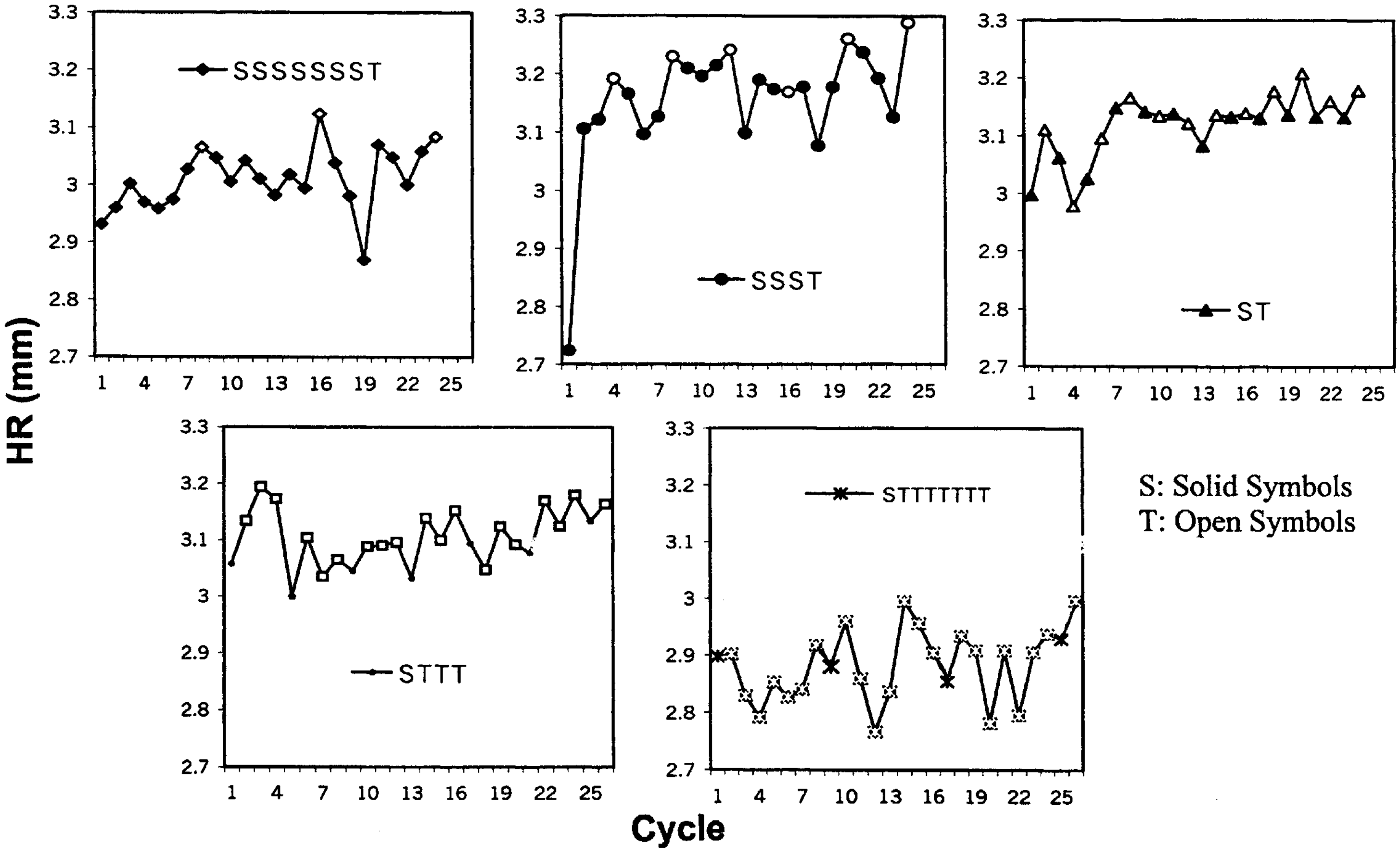


FIG 2.3. Mean HR (Heart Rate, distance between adjacent heartbeats in mm) for each of the 5 S-T (study-test) presentation programs as a function of cycle.

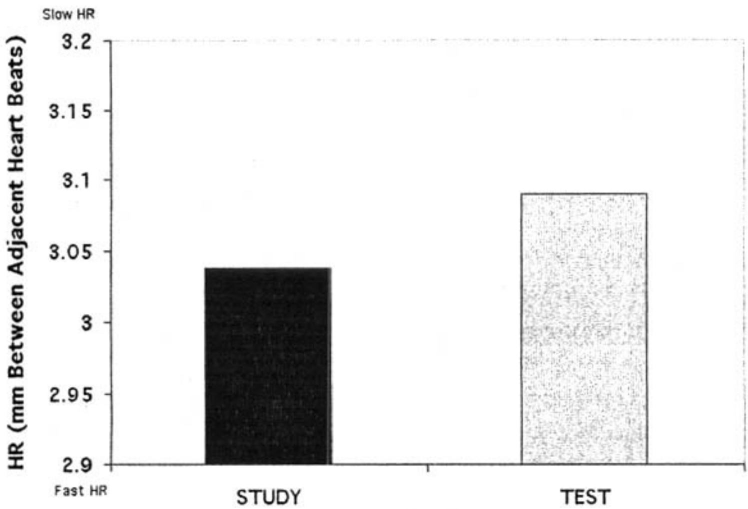


FIG. 2.4. Overall mean HR (heart rate) for S (study) and T (test) trials in the present experiment (Table 2.1) with the equal number of S and T trials in total.

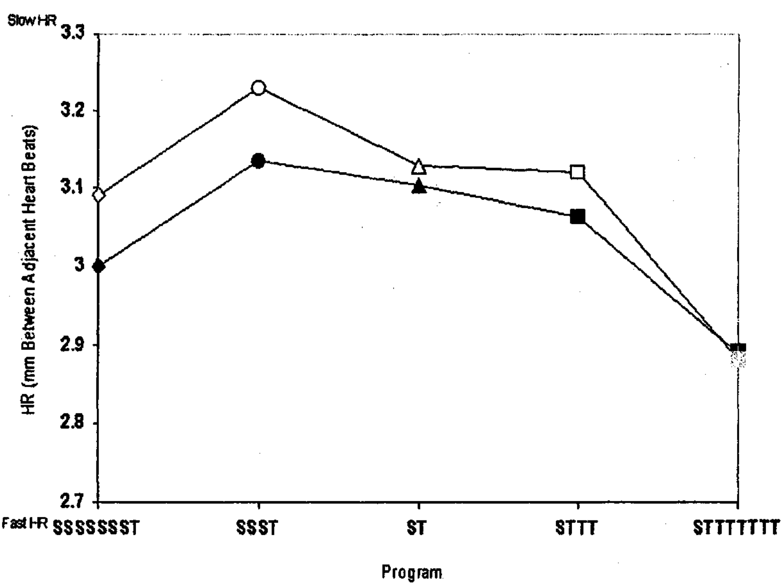


FIG. 2.5. Mean HR (heart rate) for S (study) and T (test) trials by S-T (study-test) presentation program.

unexpectedly and significantly (smaller/lower on the Y axis) than the overall T trial HR! This brand new finding, however, makes sense, in that the learner's activation level and overall alertness must rise to acquire, as quickly as possible, new and often difficult materials over a very brief span of time.

These newly discovered large HR differences between S and T trials are definitively contrary to the parity of results predicted by the S=T (T trial learning) hypothesis. Similarly, this hypothesis' expectation that an orderly change would be associated with learning on T trials from the first T to the last over the block of successive Ts in Programs STTT and STTTTTTT did not materialize for HR (Fig. 2.3) just as was the case for performance in learning (Fig. 2.2).

These non-significant curves in HR (Fig. 2.3), in addition to those in learning, are also at odds with the hypermnesia view of expected improvements over successive T trials. The negative evidence prevailed not only in early, but also in all subsequent stages of acquisition.

Instead, data support differential results for the general learning theories, TSTH, and S-T-R presentation program hypothesis. To determine which of the three surviving theories best account for our HR data, we looked into the roles played by the S-T presentation program when its interactions with S and T trial differentials influenced the HR.

The vertical distances between open (T) and solid (S) dots in Fig. 2.5 illustrate the differentials between the two trials. The quicker heartbeats on S trials (more alert/attentive) as compared with T trials were contingent on S-T programs, suggesting that these differences represent large overall trends: $F(4, 45) = 2.124$, at $p = 0.09$.

Finer grained analyses confirmed a significantly lower HR on T trials vis-à-vis S trials mostly for higher S density Programs SSSSSSST, SSST, and STTT: $F(1, 9)$s being 6.436, 6.993, and 6.175, with respective $ps < 0.05$. However, in higher T density programs such as ST or STTTTTTT, T and S trial differentials were very minute. Therefore, S and T trial differentials in HR differed substantially depending on S-T presentation programs.

Generally, stronger attentiveness/concentration (faster HR, smaller mm differences between beats on the Y-axis) in S trials as compared with T trials was greatest for high S density programs, but diminished as S density decreased and T density rose. Furthermore, HR activations in Program STTTTTTT generated higher alertness/effort levels for both S and T trials than any other program, while also producing the largest of all acquisition gains per S trial (Fig. 2.1, Table 2.2). This is an additional, newly discovered, unique, and positive effect of T trials (the 9th, Izawa, 1993b; Appendix 2.3).

These newly uncovered results of HR and S-T program interactions can only be successfully explained by the S-T-R presentation program hypothesis. However, the general learning theories or TSTH cannot do so, because they lack the theoretical mechanisms to deal with T effects.

From the standpoint of a heightened (faster) HR in both S and T trials in Program STTTTTTT, our data share some common features with fragmented information from extant studies reviewed earlier in this chapter. In that context also, consider the fact that we did not deal with conditions of extreme pressure, tension, or anxiety here. Therefore, it may also be conjectured that the present HR level may have been optimal, neither too fast nor too relaxed for circumstances of serious concentration (e.g., Dienstbier, 1989; Johansson & Frankenhaeuser, 1973; Levine, 1960; Meaney, Aitkens, Berkel, Bhatnagar, Sarrieau, & Sapolsky, 1987; Ursin, 1978; Weiss, Glazer, Pohorecky, Brick, & Miller, 1975), a situation which apparently leads to greater potentiating effects of T trials ($T \leq 7$, Izawa, 1970) and greater acquisition per S trial!

Galvanic Skin Response (GSR) Performances

Overall GSR Adaptation Processes as a Function of Cycles

Fig. 2.6 shows GSR as a function of total time/cycles. Overall, downward trends emerged: Participants' GSR, sweat levels, indicating an anxious/tense psychological state, started out high at the onset of the experiment, but gradually declined toward asymptotes, reflecting highly significant GSR adaptation processes: $F(23, 45) = 6.269$, $p = 0.0001$. Similar phenomena have also been observed by Brown (1937), Eisenstein, Eisenstein, & Bonheim (1991), Kintsch (1965), and Pendery & Maltzman (1979). The GSR adaptation processes are somewhat similar to those of HR, but are as per our a priori hunch, more variable for the GSR (Fig. 2.3 vs. 2.6/2.8).

It is particularly important to note that Cycle x S-T program interactions were also significant: $F(92, 1035) = 1.313$, $p = 0.03$. Thus, the powerful effects of S-T presentation programs were also crucial for controlling the differential GSR adaptation processes.

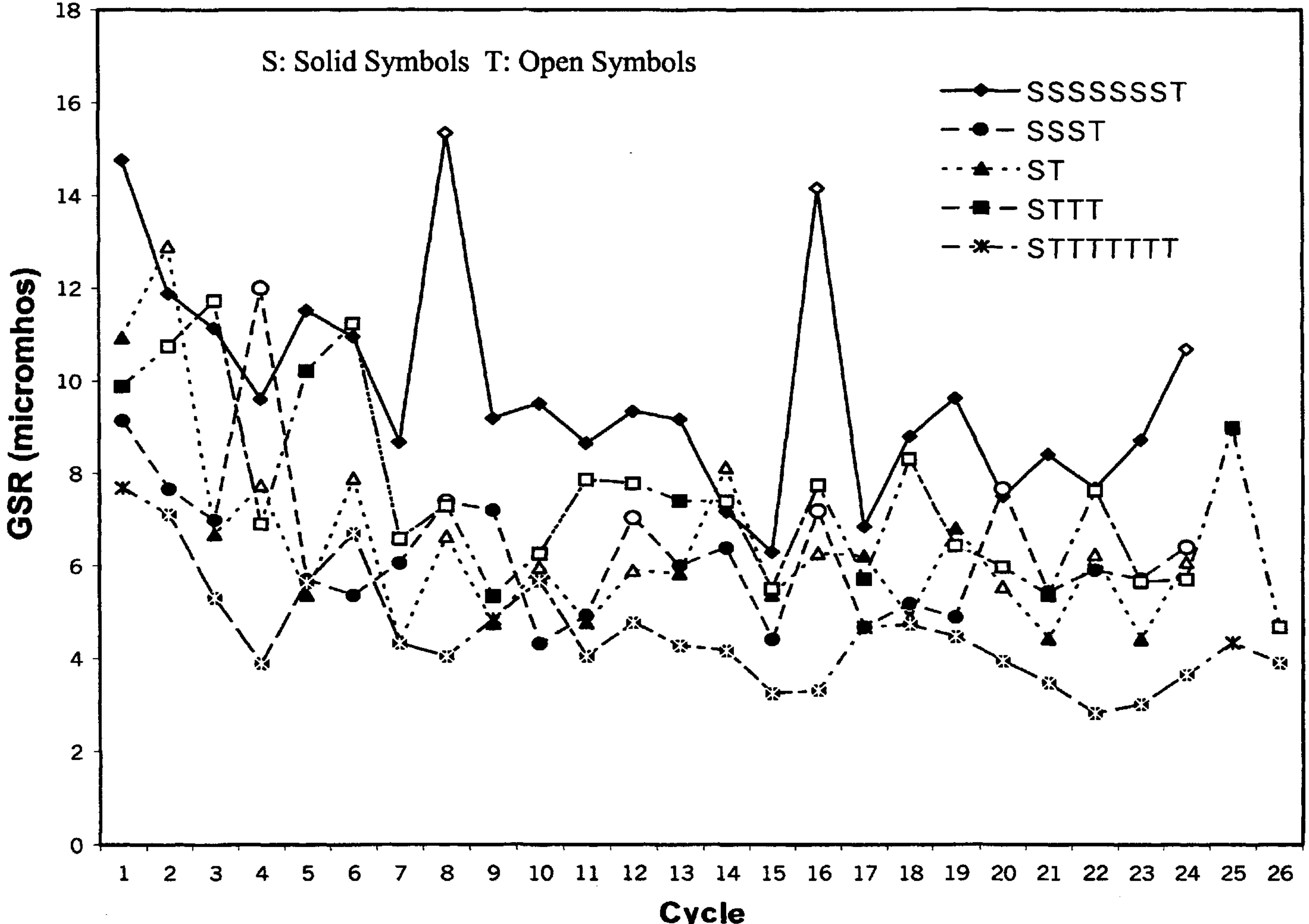


FIG. 2.6. Mean GSR (galvanic skin response) for each of the 5 S-T (study-test) presentation programs as a function of cycle.

Overall GSR Differences among the S-T Presentation Programs

Fig. 2.6 also depicts large S-T presentation program effects as a main factor: Overall GSR decisively differed among the five S-T programs: $F(4, 45) = 3.365$, $p = 0.017$. Said significance provides clear support for the S-T-R presentation program hypothesis via GSR measures, and again, reinforces serious doubts concerning all alternatives: General learning, TTHs, and S=T (T trial learning) hypotheses, all of which predicted non-significance.

While conducting the experiment, we unearthed another brand-new function of T trials (the 10^{th}, Izawa, 1993b; Appendix 2.3). Generally, the greater the T density, the smaller is GSR activation (less sweat), because learners are calmer and more comfortable.

Differences in GSR between S (study) and T (test) Trials

Most interestingly, we found that the overall GSR results for S and T trials differed highly significantly as seen in Fig. 2.7. Although counterintuitive, most learners sweated significantly more overall on S trials than T trials! $F(1, 45) = 13.27$, $p = 0.0007$.

These remarkable findings did not support the concept of equality between S and T trials predicted by the S=T hypothesis, by confirming differential outcomes, which may be inferred from both general learning theories and TSTH, as was explicitly predicted by the S-T-R presentation program hypothesis.

A solid rationale for distinguishing between these theories relies on the levels of interactions between trials (S vs. T) and S-T presentation programs, which turned out to be highly significant: $F(4, 45) = 6.898$, $p = 0.0002$. The results decisively boosted the S-T-R program hypothesis, while negating the two alternatives.

In order to illustrate the intriguing and orderly interactions, see Fig, 2.8 and zoom into each S-T presentation program individually. Compare differences between S trials in solid dots and T trials in open dots. The learners in Program SSSSSSST (highest S density, top left, a common everyday situation/occurrence with more Ss and less Ts) had much larger GSR responses on T trials than S trials: The mean difference was 4.08 micromhos, resulting in $F(1, 9) = 12.426$, $p = .0065$. The learners were significantly more nervous/sweaty on T than S trials.

When the density of S trials decreased and that of T trials increased to Program SSST (top center), the mean conductivity differentials between T and S trials decreased significantly to 2.06 micromhos, although T trials still generated reliably higher GSR: $F(1, 9) = 8.908$, $p = .015$.

T and S trial GSR-differentials continued to diminish as the density of S trials decreased and that of T trials were raised in Program ST (top right). The mean difference became 1.185 micromhos, demonstrating only a trend: $F(1, 9) = 3.991$, $p = .080$, although the T trials trend continued to produce more perspiration than S trials. However, when the density of T trials exceeds that of S trials in Program STTT (lower left), the differences between S and T trials became a minuscule 0.217 micromhos, with no differences between S and T trials at all!

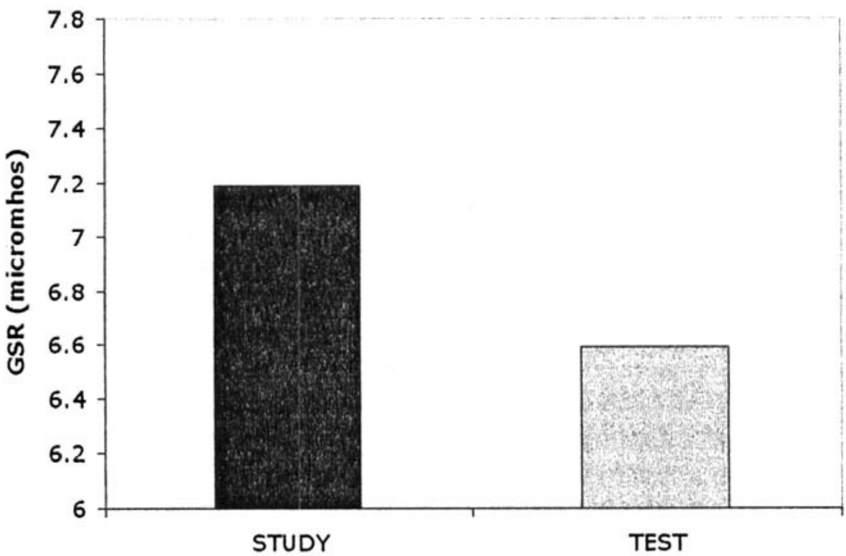


FIG. 2.7. Overall mean GSR (galvanic skin response) for S (study) and T (test) trials in the present experiment (Table 2.1) with the equal number of S and T trials in total.

Behold the right bottom panel (Fig. 2.8) for Program STTTTTTT, the highest T and lowest S density program. A major reversal of GSR differentials resulted between S and T trials and to a significant degree! $F(1, 9) = 5.933$, $p = 0.038$! T trials here generated much less sweating than did S trials, by as much as 1.013 micromhos! The dramatic interactions between the trials (S vs. T) and S-T programs factors are summarized in Fig. 2.9. Thus, the S-T presentation programs employed determined the direction and quantity of S vs. T trial differentials in GSR.

GSR Performances over Successive S (Study) and T (Test) trials

Another novel phenomenon was uncovered via GSR measurements over successive S trials (S blocks) in Programs SSSSSSST and SSST (Fig. 2.8, two top left). Overall, learners' perspiration significantly decreased from the first S to the last S trials in an S block: GSR decrements over S blocks were both significant: $F(6, 9) = 4.929$, $p = .0196$ from S_1 to S_7 in Program SSSSSSST, and $F(2, 9) = 3.939$, $p = .0381$ from S_1 to S_3 in Program SSST. Apparently, after getting used to the respective S-T programs, participants also became more comfortable as consecutive S trials progressed. However, such decrements were barely perceptible toward the end of the experiment.

By contrast, Presentation Programs STTT and STTTTTTT featured successive T trials (T blocks). See the bottom two panels of Fig. 2.8 that display another significant GSR adaptation pattern. Notwithstanding some noise, overall GSR conductivity substantially diminished over successive T trials: GSR declines from T_1 to T_3 in T blocks of Program STTT resulted in $F(2, 9) = 3.684$, $p < 0.05$, and from T_1 to T_7 in Program STTTTTTT, $F(6, 9) = 3.018$, $p = 0.013$. Toward the end, habituation occurred to a lesser degree over successive T trials. Further analyses indicate that decreasing GSR levels for both S and T trial blocks were significantly different depending on acquisition phases: $F(2, 216) = 4.416$, $p = 0.019$ for Programs SSSSSSST/STTTTTTT, and $F(5, 180) = 2.773$, $p = 0.022$ for Programs SSST/STTT.

These significant diminutions of nervous sweat levels on successive T trials contrast sharply to (**a**) the perfectly flat learning curves in Fig. 2.2, indicating neither new learning nor forgetting, and (**b**) the absence of significant HR change in Fig. 2.3, indicating *constant learner alertness over consecutive T trials.*

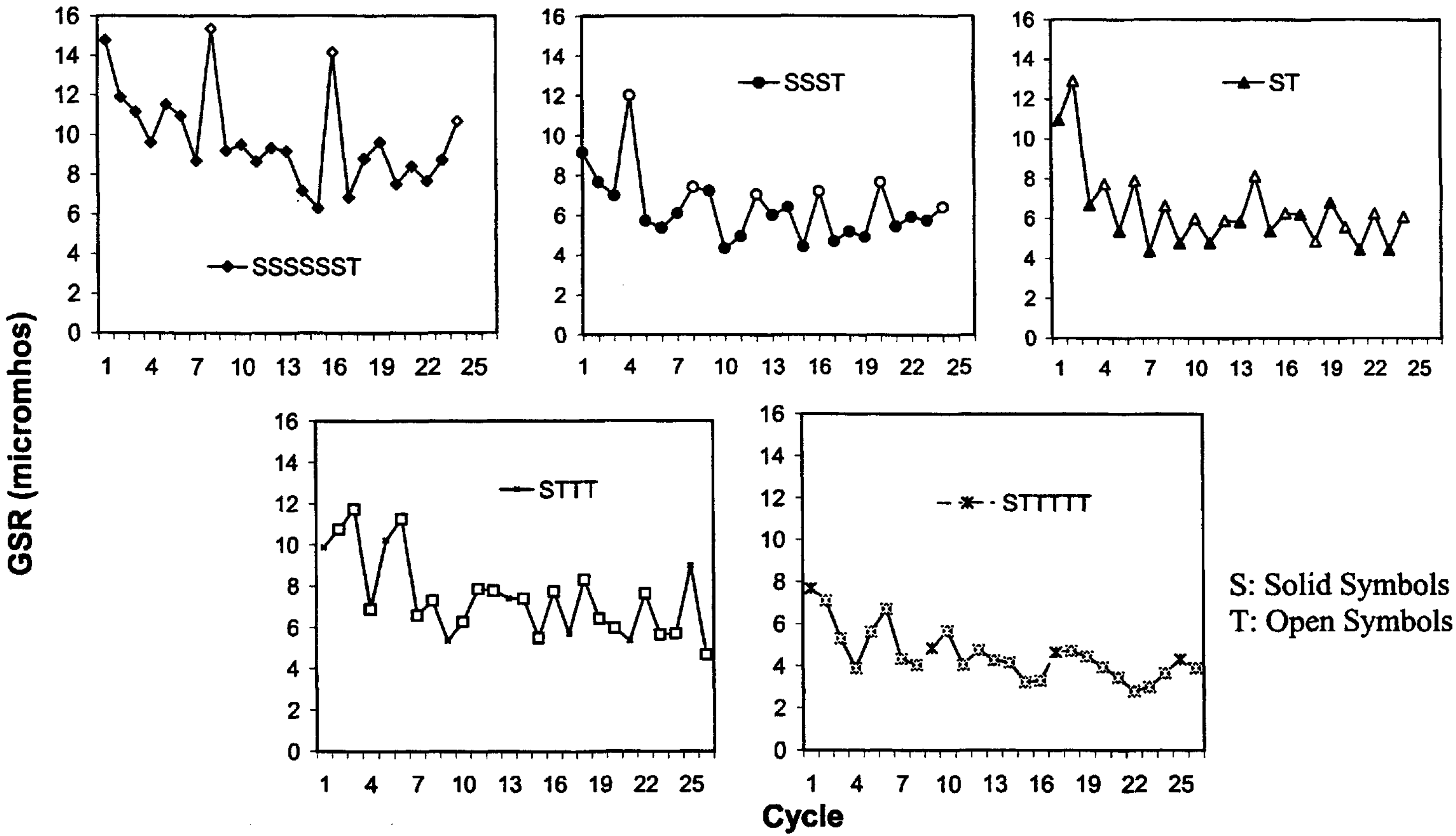


FIG. 2.8. Mean GSR for each study-test (S-T) presentation program as a function of cycle, presented separately to zoom in interactions between S and T trial differentials and presentation programs.

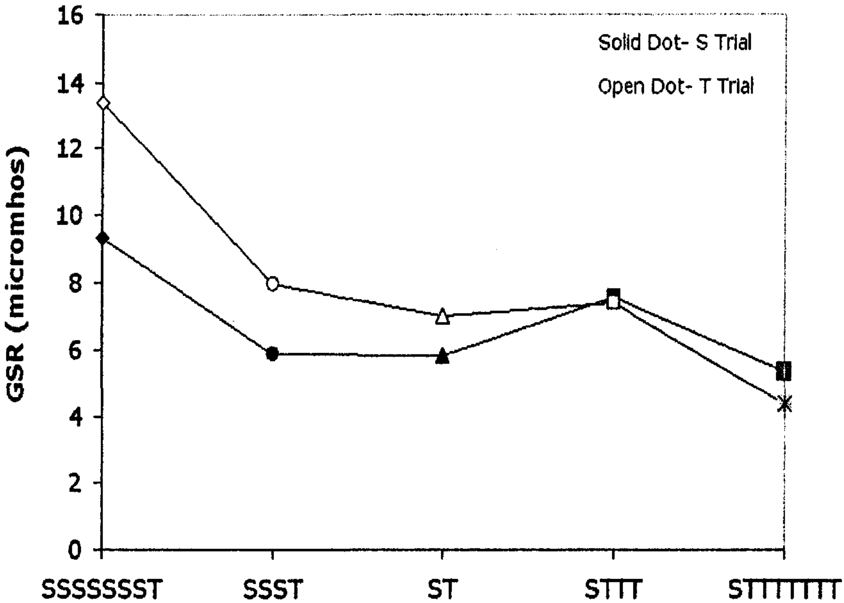


FIG. 2.9. Mean GSR (galvanic skin response) for S (study) and T (test) trials for S and T trials by S-T presentation program: Overall interactions were highly significant.

Here, again, is another challenge to the S=T hypothesis: If indeed S and T trials are virtually identical, GSR phenomena over successive S and T trials should reflect the same state of affairs when Programs SSSSSSST and STTTTTTT are analyzed. Contrary to the S=T hypothesis, habituation processes were quite reliably different for S-block and T-block groups: $F(1, 36) = 7.315, p = 0.015$. With far smaller sizes of S and T blocks, e.g., as in Programs SSST vs. STTT, the absolute differences between S and T blocks decreased, but continued to be in the same direction. Although these results create additional difficulties for the S=T hypothesis, they continue to add to the credibility of the S-T-R presentation program hypothesis, which clearly postulates qualitative and quantitative differentials for S and T trials.

SUMMARY

In search of optimized learning, we examined five theories: General learning theories (Eq. 2.1), both original and modified total time (TTHs), T (test) trial learning/S=T hypothesis, the hypermnesia view, and study-test-rest (S-T-R) presentation program (Eq. 2.2) hypotheses, via hitherto unexamined psychophysiological reactions to both S and T trials (measured separately), for five S-T presentation programs: SSSSSSST, SSST, ST, STTT, and STTTTTTT repetitive patterns. Fifty college freshmen learned a 20-pair list, while GSR (Galvanic Skin Responses) and HR (heart rates) were recorded. Learning curve analyses affirmed large differences among presentation program effects. Each response differed significantly as a function of cycles/total time. Both HR and GSR revealed overall habituation from early to late acquisition stages.

New psychophysiological findings include: (**a**) Overall S and T trials differed significantly in HR and GSR, respectively, but (**b**) the amount and directions of S and T differentials were controlled by S-T presentation programs. (**c**) Over both successive S and T trials (blocks), HR remained stable from the first to the last trial of each block, while (**d**) GSR declined significantly within the S or T block early in acquisition, although to different degrees. Toward the end, however, the GSR decline within the block became asymptotic. (**e**) The main S-T program effects were very large for GSR, but smaller for HR. (**f**) However, the S-T-R Program effects (R variable held constant) strongly interacted with S and T trial differentials with respect to GSR, and to a lesser degree with HR. (**g**) Overall, the mean HR on S trials was lower, but it was greater with T trials. (**h**) Heartbeat differentials between S and T trials were large for the higher S density programs but decreased as T density rose. (**i**) Alertness (HR) in both S and T trials was highest in Program STTTTTTT. (**j**) The same interactions between trials and programs were more dramatic for GSR effects. (**k**) The greater the T trial density, the more efficient learning per S trial, the higher the attention/alertness activation, and the greater was participant comfort (as shown by GSR)!

Overall, Izawa's S-T-R (study-test-rest) presentation program hypothesis is well supported. None of its four alternatives fare as well.

ACKNOWLEDGEMENTS

Support was provided, in part, by Grant MH 12979-02 of the National Institutes of Mental Health and by Flowerree Foundation Grant 624178 to Chizuko Izawa. Edwin Katkin's laboratory provided both psychophysiological equipment and operator for the current experiment.

We are most appreciative of assistance provided by scores of devoted students. The following students not only participated in the experiment, but some also assisted with verifying and proofing data entries: Joe Bevenzi, Stephano Biancho, Brian Brooks, M. Brottner, Barry Brown, Richard Chiarillo, Kathy Cohan, Richard Cohen, Joann Dolce, William Draper, J. Freidenberg, Gary Frys, Barbara Glovier, Gorden Grant, Pati Griffin, Martha Janish, Louis Maiino, M. Pasternak, Anthony Presley, Ralph Spiegel, John Stacey, R. Stark, Karen Szafranski, Sue Terrill, Donna Vincent, Barb Wiedemann, Harney Wineke, and G. Wisniewski.

The conduct of the experiment was placed in the competent hands of Amy Wisnieweki. The devotion to and interest in this study by several research assistants and especially that of uncompensated volunteer, Charles Kluge, made it possible to manually transcribe, decode, transform GSR electrical resistance ohms to conductance mhos scores, and enter massive amounts of data into the computer for analyses. Involved were 12,400 learning performance observations, 178,560 heart rate (HR) measurements (sampling 24,800 data points, see Method), plus 24,800 GSR measurements. In total, 240,560 data points were dealt with. That entire process took nearly a decade. When data were nearly ready to be written up, one box holding analyzed data mysteriously disappeared from our laboratory! Unnecessary forced moves of principal investigator's laboratories and office immediately after major surgery, in an apparent attempt to relocate her (a person of color) to the flood-prone dank basement (long abandoned from faculty office/labs; hazardous to your health as well), in a city below sea level, created additional blows to this already disabled study. It was our determination to combat all adversities, odds, interferences, and malice, no matter what! Notwithstanding several more years that were required, we redid all that was lost (from raw data), and restored what was water-damaged. This chapter is the splendid fruit of a long, treacherous odyssey that surmounted severe adversities in the form of flooding, interference from entities beyond our control, and a remarkable spontaneous outpouring of countervailing support from student collaborators.

Parts of the preliminary data analyses were presented at meetings of the 2001 Psychonomic Society (Orlando, FL) and at the 2003 Southeastern Psychological Association (New Orleans, LA). Salvador Iloreta, and Leith Edgar competently assisted with the proofing and indexing of this chapter. Althea J.E.K. Izawa-Hayden diligently lent hands as an editorial assistant, and Sal Iloreta further demonstrated his talents in graphic illustrations of psychophysiological data. Without the assistance of dedicated students and research interns, this investigation would never have been completed. We thank each and everyone for their steadfast assistance.

Direct correspondence to: Chizuko Izawa, 411 Pine Street, New Orleans, LA 70118-3715, USA. E-mail: Izawa@tulane.edu.

REFERENCES

Andreassi, J. (1966). Some physiological correlates of verbal learning task difficulty. *Psychonomic Science*, *6*(2), 69-70.

Andreassi, J. L., & Whalen, P. (1967). Some physiological correlates of learning and overlearning. *Psychophysiology*, *3*(4), 406-413.

Ballard, P. (1913). Obliviscence and Reminiscence. *British Journal of Psychology Monograph Supplements*, *1*, 1-82.

Berlyne, D. E. (1964). Emotional aspects of learning. *Annual Review of Psychology*, *15*, 115-142.

Berry, R. N., & Davis, R. C. (1958). Muscle responses and their relation to rote learning. *Journal of Experimental Psychology*, *55*(2), 188-194.

Birmbaum, I. M., & Eichner, J. T. (1971). Study versus test trials and long-term retention in free-recall learning. *Journul of Verbal Learning and Verbal Behavior*, *10*, 516-521.

Blatt, S. (1961). Patterns of cardiac arousal during complex mental activity. *Journal of Abnormal and Social Psychology*, *63*(2), 272-282.

Brown, C. H. (1937). The relation of magnitude of galvanic skin responses and resistance levels to the rate of learning. *Journal of Experimental Psychology*, *20*, 262-278.

Bryan, J., Sonnefeld, L., & Grabowski, B. (1983). The relationship between fear of failure and learning disabilities. *Learning Disability Quarterly*, *6*(2), 217-222.

Bugelski, B. (1962). Presentation time, total time, and mediation in paired associate learning. *Journal of Experimental Psychology*, *63*(4), 409-412.

Dienstbier, R. (1989). Arousal and physiological toughness: Implications for mental and physical health. *Psychological Review*, *96*(1), 84-100.

Ebbinghaus, H. (1885). *Über das Gedächtnis: Untersuchungen zur experimentellen Psychologie*. Leipzig, Germany: Duncker & Humbolt.

Eisenstein, E. M., Eisenstein, D., & Bonheim, P. (1991). Initial habituation or sensitization of the GSR depends on magnitude of first response. *Physiology and Behavior*, *49*, 211-215.

Estes, W. K. (1950). Toward a statistical theory of learning. *Psychological Review*, *57*, 94-107.

Ferrando, P., Varea, M., & Lorenzo, U. (1999). A psychometric study of the Test Anxiety Scale for Children in a Spanish sample. *Personality and Individual Differences, 27*(1), 37-44.

Friedman, I., & Bendas-Jacob, O. (1997). Measuring perceived test anxiety in adolescents: A self-report scale. *Educational and Psychological Measurement, 57*(6), 1035-1046.

Frierson, H., & Hoban, J. (1992). The effects of acute test anxiety on NBME Part I performance. *Journal of the National Medical Association, 84*(8), 686-689.

Gates, A. (1917). Recitation as a factor in memorizing. *Archives of Psychology, 40*, 1-105.

Geisleman, R. E., Woodward, A., & Beatty, J. (1982). Individual Differences in Verbal Memory Performance: A Test of Alternative Information-Processing Models. *Journal of Experimental Psychology: General, 111*(1), 109-134.

Germana, J. (1968). Psychophysiological correlates of conditioned response formation. *Psychological Bulletin, 70*(2), 105-114.

Hagtvet, K., & Johnsen, T. (1992). *Advances in test anxiety research* (Vol. 7): Bristol, PA, US: Swets and Zeitlinger Publishers.

Hill, K., & Wigfield, A. (1984). Test anxiety: A major educational problem and what can be done about it. *Elementary School Journal, 85*(1), 105-126.

Hodapp, V., & Benson, J. (1997). The multidimensionality of test anxiety: A test of different models. *Anxiety, Stress, and Coping: An International Journal, 10*(3), 219-244.

Hull, C. (1943). *Principles of behavior*. New York: Appleton-Century-Crofts.

Izawa, C. (1966). Reinforcement-test sequences in paired-associate learning. *Psychological Reports, Monograph Supplement, 3-V18*, 879-919.

Izawa, C. (1970). Optimal potentiating effects and forgetting-prevention effects of tests in paired-associate learning. *Journal of Experimental Psychology, 83*(2, Pt. 1), 340-344.

Izawa, C. (1971). The test trial potentiating model. *Journal of Mathematical Psychology, 8*, 200-224.

Izawa, C. (1981). Toward a quantitative theory for performance differences between anticipation and study-test procedures: The retention interval model. *Scandinavian Journal of Psychology, 22*, 79-91.

Izawa, C. (1985a). A test of the differences between anticipation and study-test methods of paired-associate learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 165-184.

Izawa, C. (1985b). The identity model and factors controlling the superiority of the study-test method over the anticipation method. *Journal of General Psychology, 112*(1), 65-78.

Izawa, C. (1988). A search for the control factors of the repetition effect. *Japanese Psychological Review, 31*(3), 367-403.

Izawa, C. (1989). *Current issues in cognitive processes: The Tulane Flowerree Symposium on Cognition*. Hillsdale, NJ: Lawrence Erlbaum Assoc., Inc.

Izawa, C. (1992). Test trials contributions to optimization of learning processes: Study/test trial interactions. In A. Healy, S. Kosslyn, & R. M. Shiffrin (Eds.), *Essays in honor of William K. Estes: From learning processes to cognitive processes* (Vol. 2, pp. 1-33). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Izawa, C. (1993a). Efficient learning: The total time, exposure duration, frequency, and programming of the study phase. In C. Izawa (Ed.), *Cognitive psychology applied* (pp. 43-78). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Izawa, C. (1993b). Power behind the scenes: Hidden effects of test trials unveiled. In C. Izawa (Ed.), *Cognitive psychology applied* (pp. 79-110). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Izawa, C. (1999). Efficiency in short-term memory: Study-test-rest presentation program and learning difficulty. In C. Izawa (Ed.), *On human memory: Evolution, progress, and reflections on the 30th Anniversary of the Atkinson-Shiffrin model* (pp. 165-213). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Izawa, C. (2000). Total time and efficient time management: In search of optimal learning and retention via study-test-rest presentation programs. *American Journal of Psychology, 113*(2), 221-248.

Izawa, C. (2003, March). *Theoretical and empirical examinations of optimal learning: Study-test-rest presentation program effects and neuroscience contributions.* Invited Symposium Presentation at the 49th Annual Meeting of the Southeastern Psychological Association, New Orleans, LA.

Izawa, C., & Hayden, R. G. (1993). *Race against time: Toward the principle of optimization in learning and retention.* In C. Izawa (Ed.), *Cognitive psychology applied* (pp. 15-42). Hillsdale, NJ: L. Erlbaum Associates, Inc.

Johansson, G., & Frankenhaeuser, J. (1973). Temporal factors in sympatho-adrenomedullary activity following acute behavioral activation. *Biological Psychology, 1*, 63-73.

Kahneman, D. (1969). Pupillary, heart rate, and skin resistance changes during a mental task. *Journal of Experimental Psychology, 79*(1), 164-167.

Kim, S., & Park, K. (1995). Kindergarten children's entry behaviors into peer groups. *Korean Journal of Child Studies, 16*(1), 147-165.

Kintsch, W. (1965). Habituation of the GSR component of the orienting reflex during paired-associate learning before and after learning has taken place. *Journal of Mathematical Psychology, 2*(2), 330-341.

Kleinsmith, L., & Kaplan, S. (1964). Interaction of arousal and recall interval in nonsense syllable paired-associate learning. *Journal of Experimental Psychology, 67*(2), 124-126.

Kutchukian-Boxley, C. (1993). The effect of anxiety on incidental learning and examination performance in college students. *Dissertation Abstracts International, 53*(12-A), 4256.

Lachman, R., & Laughery, K. (1968). Is a test trial a training trial in free recall learning? *Journal of Experimental Psychology, 76*(1), 40-50.

Levine, S. (1960). Stimulation in infancy. *Scientific American, 202*, 80-86.

Lindholm, E., & Cheatham, C. (1983). Autonomic activity and workload during learning of a simulated aircraft carrier landing task. *Aerospace Medical Association, 54*(5), 435-439.

Luckie, W., & Smethurst, W. (1998) . *Study power: Study skills to improve your learning and your grades* (Vol. viii): Cambridge, MA, US: Brookline Books.

Ludlow, L., & Guida, F. (1991). The Test Anxiety Scale for Children as a generalized measure of academic anxiety. *Educational and Psychological Measurement, 51*(4), 1013-1021.

Marteau, T., & Bekker, H. (1992). The development of a six-item short-form of the state scale of the Spielberger State-Trait Anxiety Inventory (STAI). *British Journal of Clinical Psychology, 31*(3), 301-306.

Meaney, M., Aitkens, D., Berkel, C., Bhatnagar, S., Sarrieau, A., & Sapolsky, R. (1987). *Post-natal handling attenuates age-related changes in the adrenocortical stress response and spatial memory deficits in the rat.* Paper presented at the 17th Annual Meeting of the Society of Neuroscience, New Orleans, LA.

Meijer, J., Elshout, M., & Van Hout, B. (2001). An instrument for the assessment of cross-curricular skills. *Educational Research and Evaluation, 7*(1), 79-107.

Otani, H., & Hodge, M. (1991). Does hypermnesia occur in recognition and cued recall? *American Journal of Psychology, 104*(1), 101-116.

Otani, H., Widner, R., Whiteman, H., & St. Louis, J. (1999). Hypermnesia: The role of multiple retrieval cues. *Memory & Cognition, 27*(5), 928-934.

Payne, D. G. (1987). Hypermnesia and reminiscence in recall: A historical and empirical review. *Psychological Bulletin, 101*(1), 5-27.

Pendery, M., & Maltzman, I. (1979). Verbal conditioning and extinction of the GSR index of the orienting reflex. *Physiological Psychology, 7*, 185-192.

Reynolds, C., & Richmond, B. (1978). What I think and feel: A revised measure of children's manifest anxiety. *Journal of Abnormal Child Psychology, 6*(2), 271-280.

Roediger, H. L., III, & Challis, B. (1989). *Hypermnesia: Improvements in recall with repeated testing.* In C. Izawa (Ed.). *Current issues in cognitive processes: The Tulane Flowerree Symposium on Cognition.* (pp. 175-200). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Rosner, S. (1970). Serial mediation: Effects of associative strength and structure on serial learning and transfer. *Journal of Experimental Psychology, 84*(2), 370-372.

Sarason, I., Kestenbaum, J., & Smith, D. (1972). Test anxiety and the effects of being interviewed. *Journal of Personality, 40*(2), 242-249.

Schonwetter, D. (1997). Effective instruction and student differences in the college classroom. *Dissertation Abstracts International Section A:*

Humanities and Social Sciences, 58(4-A), 1200.

Shioya, S. (1995). Relationship of test anxiety and learning behavior to cognitive appraisals in high school students. *Japanese Journal of Educational Psychology, 43*(2), 13-21.

Slamecka, N., & Katsaiti, L. (1988). Normal forgetting of verbal lists as a function of prior testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(4), 716-727.

Taylor, J., & Spence, K. (1952). The relationship of anxiety level to performance in serial learning. *Journal of Experimental Psychology, 44*, 61-64.

Thompson, L. W., & Obrist, W. D. (1964). EEG correlates of verbal learning and overlearning. *Electroencephalography and Clinical Neurophysiology, 16*, 332-342.

Tulving, E. (1967). The effects of presentation and recall of material in free-recall learning. *Journal of Verbal Learning and Verbal Behavior, 6*(2), 175-184.

Ursin, H. (1978). Activation, coping, and psychosomatics. In H. Ursin, E. Baade, & S. Levine (Eds.), *Psychobiology of stress: A study of coping men.* New York: Academic Press.

Wang, C. (2001). Reliability and validity of the Test Anxiety Scale-Chinese Version. *Chinese Mental Health Journal, 15*(2), 95-97.

Wang, C., & Liu, Y. (2000). Correlation among general self-efficacy, trait anxiety, state anxiety, and test anxiety. *Chinese Journal of Clinical Psychology, 8*(4), 229-230.

Weiss, J., Glazer, H., Pohorecky, L., Brick, J., & Miller, N. (1975). Effects of chronic exposure to stressors on avoidance-escape behavior and on brain norepinephrine. *Psychosomatic Medicine, 37*, 522-534.

Whitten, W., & Bjork, R. (1977). Learning from tests: Effects of spacing. *Journal of Verbal Learning and Verbal Behavior, 16*(4), 465-478.

Whitten, W., & Leonard, J. (1980). Learning from tests: Facilitation of delayed recall by initial recognition. *Journal of Experimental Psychology: Human Learning and Memory, 6*(2), 127-134.

APPENDIXES

Appendixes 2.1 and 2.2 on the following two pages illustrate the traditional standard anticipation and study-test (S-T) methods of item presentation methods for the standard cued-recall/paired-associate learning (PAL) situations under the list design. Illustrations are made for a 12 pair-list (n = 12), and Item A_i and B_i, respectively, indicate the stimulus/cue and response/target terms of Pair i, where $1 \leq i \leq 12$. Appendix 2.3 summarized positive effects of unreinforced T (test) trials by Izawa (1993). The current study adds several more of the positive T effects newly discovered herein.

APPENDIX 2.1

The Standard Traditional Anticipation Method
Illustrated for a 12-Pair List:
T (test) and S (study) events come alternately for each item within anticipation cycle.

Anticipation Cycle 1	Anticipation Cycle 2
A_1 - ?	A_8 - ?
A_1 - B_1	A_8 - B_8
A_2 - ?	A_5 - ?
A_2 - B_2	A_5 - B_5
A_3 - ?	A_2 - ?
A_3 - B_3	A_2 - B_2
A_4 - ?	A_{12} - ?
A_4 - B_4	A_{12} - B_{12}
.	.
.	.
.	.
A_{11} - ?	A_4 - ?
A_{11} - B_{11}	A_4 - B_4
A_{12} - ?	A_7 - ?
A_{12} - B_{12}	A_7 - B_7

A_i = Cue or Stimulus Term of Pair *i*

B_i = Target or Response Term of Pair *i*

APPENDIX 2.2

The Standard Study-Test Method:
S-T Presentation Program ST
Illustrated for a 12-pair List
S (study) and T (test) cycles come alternately

S_1 - Cycle	T_1 - Cycle	S_2 - Cycle	T_2 - Cycle
A_1 - B_1	A_8 - ?	A_2 - B_2	A_{10} - ?
A_2 - B_2	A_5 - ?	A_7 - B_7	A_{12} - ?
A_3 - B_3	A_2 - ?	A_5 - B_5	A_8 - ?
A_4 - B_4	A_{12} - ?	A_3 - B_3	A_6 - ?
A_5 - B_5	A_{10} - ?	A_{10} - B_{10}	A_5 - ?
A_6 - B_6	A_3 - ?	A_8 - B_8	A_3 - ?
A_7 - B_7	A_9 - ?	A_6 - B_6	A_7 - ?
A_8 - B_8	A_{11} - ?	A_1 - B_1	A_1 - ?
A_9 - B_9	A_1 - ?	A_4 - B_4	A_2 - ?
A_{10} - B_{10}	A_6 - ?	A_{12} - B_{12}	A_9 - ?
A_{11} - B_{11}	A_7 - ?	A_{11} - B_{11}	A_{11} - ?
A_{12} - B_{12}	A_4 - ?	A_9 - B_9	A_4 - ?

A_i = Cue or Stimulus Term of Pair *i*

B_i = Target or Response Term of Pair *i*

APPENDIX 2.3

Positive Effects of Unreinforced T (test) Trials From Izawa (1993, p.80)

1. **The forgetting-prevention effect** (preserving information in both short-term memory, STM, and long-term memory, LTM); for one of the most striking instances, refer to Izawa (1970; see also McDaniel & Fischer, 1991; Runquist, 1986).
2. **The potentiating effect** (making the subsequent S trials more effective; Izawa, 1966-1992).
3. **The LTM retrieval facilitation effect** (e.g., Hagman, 1983; Izawa & Patterson, 1989).
4. **The T generation effect** (the subject's active participation in the learning task by expressing the response overtly, e.g., Izawa, 1976).
5. **The positive S/T trials interactions** to the benefit of both S and T trials (e.g., Izawa, 1966, 1971a, 1971b, 1988, 1992; LaPorte & Voss, 1974).
6. **The T trial spacing (spaced/distributed practice) effect** (e.g., Izawa, 1992; Landauer & Bjork, 1978).
7. **The immediate test, $T_{(1)}$, effect** (the T, immediately following the S, is especially effective in preserving the just acquired information; e.g., Izawa, 1992; McDaniel, Kowitz, & Dunay, 1989; McDaniel & Masson, 1985).
8. **The retrieval practice effect** (e.g., Gross & Bjork, 1991; Izawa, 1992; King, Zechmeister, & Shaughnessy, 1980; Longstreth, 1971).

This page intentionally left blank

Chapter 3

Recollection and Familiarity: Redundancy at the Item Level

Michael S. Humphreys
Angela M. Maguire
The Key Centre for Human Factors and Applied Cognitive Psychology
University of Queensland, Australia.

When identifying words that occurred in a study list, participants report remembering how they processed the word (e.g., whether the word was rated or memorized); details about the word's presentation (e.g., the modality of presentation); and individual reactions to the word (e.g., what they thought of when they encountered the word). The retrieval of such information is referred to as recollection, which is assumed to be analogous to recall (as observed in free and cued recall paradigms), but the exact relationship to recall is not always specified (Gardiner, 1988; Jacoby, 1991; Qin, Raye, Johnson, & Mitchell, 2001; Rotello, 2001; Tulving, 1985; Yonelinas, 2002).

Recollection is generally contrasted with familiarity (Jacoby, 1991). There are instances in which familiarity is an unambiguously context-sensitive process (Gillund & Shiffrin, 1984); instances in which it is an unambiguously context-insensitive process (Jacoby & Dallas, 1981; Mandler, 1980); and instances in which the role of context is ambiguous (Jacoby, 1991; Yonelinas, 2002). It seems likely that many authors are not concerned with the role that context might or might not play in familiarity. Instead they contrast a recall-like process (possibly a thresholded process) that produces course-grained or symbolic information (a word or concept), with a fine-grained or subsymbolic process. The former process is generally considered relatively easy to disrupt using activities at either study or test, while the latter is not so easy to disrupt (Humphreys & Bain, 1983; Jacoby 1991).

Just as authors tend not to concern themselves with the role of context in familiarity, the question of whether context and/or test instructions play a role in guiding recollection is also generally ignored. That is, given that a participant is trying to remember information that would indicate an item has been read, is he/she more likely to access (or become aware of) information indicating that the item had been read, and less likely to access (or become aware of) information indicating that an item had, for example, been heard? Jacoby's (1991) process dissociation procedure assumes that context and/or test instruction does not affect recollection in this manner. Consider, for example, an instantiation of the process-dissociation procedure in which participants study a list of both visually and auditorially presented words. In the inclusion condition participants are asked to respond *yes* to old words and *no* to new words. In the exclusion condition they are asked to respond *yes* to words presented in one modality and *no* to words presented in the other modality and to new words. If recollection changes with the change in instructions, Jacoby (1991) could not obtain meaningful estimates of recollection and familiarity by combining the results from the inclusion and exclusion conditions. This conclusion also holds for the variant on the process-dissociation procedure introduced by Yonelinas (1994).

Although the possible role of context in familiarity, and in guiding retrieval in recollection, has largely been ignored, these issues are bound to two other issues that have received widespread attention. That is, do recollection and familiarity refer to fundamentally different kinds of memory (Jacoby & Dallas, 1981; Mandler, 1980; Tulving, 1985); or do they refer to different ways of accessing what is fundamentally the same memory (Gillund & Shiffrin, 1984; Humphreys & Bain, 1983; 1991)? If context is involved in familiarity, and is used to guide memory access in recollection, it is unlikely that recollection and familiarity refer to fundamentally different forms of memory. If familiarity and recollection do refer to fundamentally different forms of memory, then it may be reasonable to assume that, at the item level, recollection and familiarity are independent.

In a major review of the literature on recollection and familiarity, Yonelinas (2002) concluded that there was substantial evidence that recollection and familiarity are independent. It is unclear just what Yonelinas (2002) means by independence. Nevertheless, he does talk about independence in the process-dissociation and remember-know procedures and this requires independence at the item level. In support of independence, Yonelinas cited two findings. First, findings that demonstrated the pattern of results with task-dissociation procedures (e.g., those requiring highly speeded responses in the hope of eliminating recollection) were very similar to the pattern obtained with process-estimation procedures. "The estimation methods all assume that recollection and familiarity are independent, whereas the task-dissociation methods make no explicit assumptions about how the two processes are related. To the extent that the task-dissociation methods verify the results of the estimation methods, the results indicate that the assumptions underlying the estimation methods were not violated because such violations would have biased the parameter estimates" (Yonelinas, 2002, p. 478).

Second, he cited findings that demonstrate the existence of dissociations and double dissociations: "In order to determine whether recollection and familiarity are operating independently, it is necessary to assess whether they can be functionally dissociated. That is, if one process operates independently of the other, it should be possible to find variables that influence one process without influencing the other" (Yonelinas, 2002, p. 477).

Before discussing why similar results from the task-dissociation and process-estimation methods cannot support independence over redundancy at the item level, we need to provide details about these methods. We must also detail the relationship between the observable statistics and the underlying constructs of recollection and familiarity. In the remember-know paradigm, participants are asked to respond *remember* to a word if they recollect some aspect of its presentation and to respond *know* if they are confident that it is old but cannot recollect the features of its presentation. We assume that participants follow instructions such that the probability of a *remember* response equals the probability of recollection, P(R). If we also assume that participants respond *yes* to familiar words (more precisely, words that exceed a familiarity criterion), then the probability of a *know* response divided by one minus the probability of a *remember* response, P(K)/[1 – P(R)], is the probability that an old word is familiar conditional on it not supporting recollection, $P(F_{old}|\bar{R})$. If recollection and familiarity are independent, then the conditional probability equals the unconditional probability that a word is familiar, $P(F_{old})$.

In the process-dissociation procedure, participants study two groups of words. At test, participants are asked to make one of two decisions: either respond *yes* to all old words and *no* to new words (the inclusion condition); or respond *yes* to one group of old words and *no* to the other group and to new words (the exclusion condition). Like Jacoby (1991) and Yonelinas (2002) we assume that under inclusion instructions participants respond *yes* if they can recollect (R) or if the word is familiar (F), and under exclusion instructions they will incorrectly respond *yes* if recollection fails but there is familiarity. These assumptions yield Equations 3.1 and 3.2 where P(I) is the probability of responding *yes* to an old word under inclusion instructions, and P(E) is the probability of incorrectly responding *yes* to an old word under exclusion instructions.

$$P(I) = P(R) + [1 - P(R)]\, P(F_{old} | \bar{R}) \qquad (3.1)$$
$$P(E) = [1 - P(R)]\, P(F_{old} | \bar{R}) \qquad (3.2)$$

From Equations 3.1 and 3.2, it is easy to show that $P(R) = P(I) - P(E)$ and that $P(E)/[1 - P(R)] = P(F_{old} | \bar{R})$. If we assume that familiarity and recollection are independent, then $P(F_{old}) = P(F_{old} | \bar{R})$.

In a confidence-rating recognition task, participants are asked to rate their confidence in the "oldness" or "newness" of a word. Assume that higher ratings indicate more confidence in the oldness of the word, or less confidence in the newness of the word, and that k confidence ratings are used. It follows then that for each level of confidence, the probability that the participant is that confident, or more confident, in the oldness of the word is calculated as $P(C \geq i)$. In keeping with Yonelinas (2002), we assume that a participant sets k-1 criterion levels such that their response is confidence level i if the observed familiarity is greater than confidence level $i - 1$ and less than confidence level i. We also assume that the participant uses the highest confidence rating if the word supports recollection. Given these assumptions and an old word, the probability that a participant is as confident, or more confident, than level i is given by Equation 3.3.

$$P(C \geq i) = P(R)\, P(F_{old} \geq i \mid \bar{R}) \qquad (3.3)$$

Yonelinas' estimation procedure then estimates P(R) and the distance between the mean of the familiarity distribution for new items, mean(F_{new}), and the mean of the familiarity distribution for old items conditional on recollection failure, mean($F_{old} | \bar{R}$). Again, if recollection and familiarity are independent then the distribution of $F_{old}|\bar{R}$ is the same as the distribution of F_{old}.

When there is a correspondence between the estimates of recollection obtained with the three estimation procedures, it is powerful evidence for the psychological reality of the subjective experience of recollection. That is, it establishes that the same subjective state is associated with list discrimination, remember responses, and high-confidence recognition judgments. Likewise, when independent variables affect the estimates of recollection in the same way, it further establishes the psychological reality of the subjective state. However, when independent variables affect the estimates of familiarity in the same way, or even produce the same estimates of familiarity, it conveys nothing about the relationship between recollection and familiarity. This is because all three methods produce estimates of the $P(F_{old}|\bar{R})$, or the distance between the mean of the familiarity distribution for new items, mean(F_{new}), and the mean of the familiarity distribution for old items conditional on recollection failure, mean($F_{old}|\bar{R}$). If we make the assumption of independence, we can also estimate $P(F_{old})$, or the distance between the means of the familiarity distributions for old and new items. There is nothing in the procedure, however, that allows a test of the independence assumption.

Task-dissociation methods (e.g., requiring speeded responding at test) seek to exploit the greater vulnerability of recollection to disruption. They produce estimates of the probability of familiarity, $P(F_{old})$, not estimates of familiarity conditional on recollection failure, $(F_{old}|\bar{R})$. If the estimates of familiarity obtained from task-dissociation methods were the same as the estimates obtained from process-dissociation methods it would be strong evidence for independence. Humphreys, Dennis, Chalmers, & Finnigan (2000) noted this outcome would be unlikely because task-dissociation procedures are rather blunt instruments. Familiarity may be resistant to disruption, but it is not totally insensitive to disruption. For example, if participants do not see a visual test item, there will be no recollection and no familiarity.

Likewise, there is no guarantee that any specific form of disruption will totally eliminate recollection. Thus, if a form of disruption produces estimates of familiarity that are the same as those produced by the process-dissociation methods, a slightly greater or lesser amount of disruption may produce different estimates. There is no *a priori* way to determine the amount of disruption sufficient to eliminate recollection, but leave familiarity unaffected. Yonelinas (2002), however, does not make the claim that these different methods produce the same estimates. Instead, he claims that independent variables affect the estimates obtained from the process-dissociation and task-dissociation procedures in the same way. However, finding that independent variables affect estimates of familiarity and estimates of familiarity conditional on recollection in a similar way is no more surprising than finding that independent variables affect the highest confidence old response and all old responses in a similar way.

Dissociations, including double dissociations, establish a relationship between estimates of recollection and familiarity at one level of aggregation: the list, experimental condition, or experiment-wide level. Using a relationship established at one level of aggregation to draw a conclusion about the relationship at another level is simply not logically valid. In sociology this is referred to as the ecological fallacy, and when applied to discrete data it is known as Simpson's paradox. In fact, if one could use a relationship established at the list-wide or condition level to infer the item-level relationship, we could reject several theories about recollection and familiarity.

For example, Bain and Humphreys (1988) provided a test of Mandler's (1980) theory by having participants learn a list of word pairs. At study, one group of 40 participants was given joint-rehearsal instructions (repeat both words in a left to right order for 10 seconds), and a second group of 40 was given separation-rehearsal instructions (repeat the first word for 5 seconds then the second word for 5 seconds). As expected, subsequent tests indicated that joint-rehearsal participants were better at cued recall and recognition-in-context. The critical test, however, was single-item recognition (recognizing one member of a Rearranged test pair, or the old word in a Mixed test pair). Mandler's theory asserted that recall (recollection) and familiarity made independent contributions to single-item recognition. Bain and Humphreys therefore argued that because recall (recollection) was reduced in the separation-rehearsal condition, recognition should also be reduced. Instead, single-item recognition did not differ between the two rehearsal conditions (see Table 3.1).

This example demonstrates that, at the level of the conditions within the experiment, there was a *negative* relationship between recollection (cued recall) and familiarity. As Bain and Humphreys (1988) pointed out: Mandler's theory, which addressed the item level, could be supported if it was assumed that separation rehearsal produced higher levels of familiarity. Such an assumption would produce the negative relationship that was observed across the conditions of the Bain and Humphreys experiment. Nevertheless, they argued that redundancy between familiarity and recollection was a more parsimonious explanation for the invariance in single-item recognition than the precise trade-off between recollection and familiarity the results required. One might also argue that separation rehearsal reduced cued recall while increasing other forms of recollection. However, this assertion still requires a precise trade-off between the different forms of recollection. Therefore, redundancy between this form of recollection and familiarity still appears to be the most parsimonious explanation. Thus, the Bain and Humphreys' result, coupled with Humphreys' (1976) demonstration that single-item recognition does not depend on whether participants study word pairs or single items, provides the first evidence in support of redundancy.

TABLE 3.1
Mean Hit, False Alarm, Recall, and Familiarity Rates for Intact (Two Old Words Studied Together), Rearranged (Two Old Words Studied in Different Pairs), Mixed (One Old and One New Word), and New (Two New Words) Test Pairs (from Bain & Humphreys, 1988).

Measure	Pair Type	Rehearsal Instruction Separate	Rehearsal Instruction Joint
Hit Rate			
	Intact	.67	.72
	Rearranged	.62	.60
	Mixed	.61	.60
False Alarm Rate			
	Mixed	.12	.10
	New	.10	.08
Cued Recall			
	Rearranged + Mixed	.14	.30
Familiarity			
	Rearranged + Mixed	.56	.42

Yonelinas and Jacoby's (1996) experiment offers a second example where the relationship observed at the level of the list or experiment may differ from the relationship at the item level. They had participants study a list of words. Equal numbers of words were presented on the left and right of the screen, in large and small fonts. Participants were instructed to learn the screen location in which a word appeared. At test one group of participants was instructed to respond *yes* if the word had appeared in one screen location, and *no* if it had appeared in the other location or if it was new. A second group was asked to respond *yes* if the word had appeared in one font size, and *no* if it had appeared in the other font size or if it was new. In addition, both groups were encouraged to respond *yes* if they thought that the word was old but could not remember the details relevant to the decision (i.e., screen location or font size).

Estimates of recollection were higher and those of familiarity were lower for participants instructed to discriminate on the basis of screen location relative to those instructed to discriminate on the basis of font size. In an attempt to explain this trade-off, Yonelinas and Jacoby (1996) proposed that recollection criterial for one purpose (e.g., screen location information when making screen location judgments) would be non-criterial for another purpose (e.g., screen location information when making font size judgments). Furthermore, they assumed that the recollection of non-criterial information would lead to a *yes* response and inflate the estimate of familiarity. In making this assumption, they could preserve their supposition that recollection and familiarity were independent at the item level, despite the negative relationship they observed at the group level. However, nothing in their results establishes that participants actually recollect screen location information when they attempt to discriminate on the basis of font size. It is possible that recollection became familiarity as the basis of the discrimination changed from screen location to font size, which would imply a redundancy between recollection and familiarity.

Some evidence in the Yonelinas and Jacoby (1996) study supports this conjecture about words supporting recollection under one set of retrieval instructions and appearing familiar under another set. In both the screen location and font size discrimination conditions, their speeded memory test reduced estimates of recollection, but left the estimates of familiarity (which were presumably inflated by non-criterial recollection) unchanged. To explain why non-criterial recollection was not affected by the requirement to respond quickly, Yonelinas and Jacoby assumed that the process was automatic, like familiarity. However, the transformation of recollection to familiarity is a more parsimonious explanation for the unchanging estimates of familiarity.

Perhaps Yonelinas (2002) is not committing an ecological fallacy when he asserts that the presence of dissociations and double dissociations in estimates of recollection and familiarity support the independence of these two constructs. Instead, he may be misinterpreting the redundancy hypothesis. The redundancy hypothesis asserts that recollection cannot occur in the absence of familiarity, such that the probability of recollection is bounded above by the probability of familiarity. The independence hypothesis asserts that the probability of recollection is bounded above by the probability of recognition. However, because the redundancy hypothesis also assumes that the probability of familiarity equals the probability of recognition, familiarity and recollection are free to dissociate under both hypotheses.

Although both the redundancy and independence hypotheses are compatible with dissociations, they predict different patterns when coupled with the hypothesis that familiarity is more resistant to disruption than is recollection (Humphreys, Dennis, Maguire, Reynolds, Bolland, & Hughes, in press). That is, independence plus the resistance-to-disruption hypothesis predicts that disruptions will tend to leave the probability of familiarity conditional on recollection failure, $P(F_{old}| \bar{R})$, constant, while estimates of recollection vary. In contrast, redundancy plus the resistance-to-disruption hypothesis predicts that disruption will leave the probability of recognition constant, while estimates of recollection vary. The independence hypothesis also predicts that there will be relatively little variation in the probability of recognition (see the discussion later in this paper of Macken, 2002). Similarly, the redundancy hypothesis predicts that there will be relatively little variation in the probability of familiarity conditional on recognition failure, $P(F_{old}| \bar{R})$. Neither hypothesis is completely supported in the literature. One possibility is that the disruptions used also have some impact on familiarity. Another possibility is that the true status of the relationship lies somewhere between independence and redundancy.

Thus, it appears extremely difficult to discriminate between independence and redundancy by examining patterns of invariance. However, it is possible to examine the conjecture that criterial recollection becomes familiarity, rather than non-criterial recollection. This was investigated by Reynolds (2001), the results of which are described next.

Support for Redundancy: Instructions Direct the Memory Access Process

In Reynolds (2001), three groups of participants studied a list containing both visually and auditorially presented words. Two groups were instructed to simply learn the words for a subsequent test. The third group, presented with the same study list, was asked to rate the visually presented words for pleasantness. Following study, all three groups received the first of two tests. Test 1 consisted of old visual words, old auditory words, and new words. Participants were asked to make a *remember-know-new* judgment for each word. That is, did they *remember* some aspect of how they studied the word, just *know* that it was old, or believe it to be *new*. Following the first test, all participants received a second test. For the two groups of participants that had simply read and heard the items, one group was asked to respond *yes* if they read the word, and *no* if they heard the word or if it had not been studied (visual test condition); the other group was asked to respond *yes* if they heard the word, and *no* if they read the word or if it had not been studied (auditory test condition). The third group (who in addition to reading and hearing words had also rated the visually presented words for pleasantness) were asked to respond *yes* to words they had read and rated, and *no* to words they had heard and to words that had not been studied (visual + pleasantness test condition). For all participants, the second test contained the same words as the first test, presented in the same order.

The data of principle interest was the probability of responding *yes* on the source monitoring task conditional on whether the participant had responded *remember*, *know*, or *new* on the prior recognition test. These results are presented in Table 3.2. First, compare the probability of a *yes* response conditional on a *know* response to the same probability conditional on a *new* response. The former is consistently higher than the latter. This finding supports two conclusions. First, words which elicit a *know* response on Test 1 are more likely to appear familiar on Test 2 than words which elicit a *new* response on Test 1. Second, on Test 2 (the source monitoring task) participants have a tendency to respond *yes* to familiar words. This tendency to respond *yes* to familiar words occurs both when a *yes* response is correct and when it is incorrect.

TABLE 3.2
Mean Proportion of *Yes* Responses During Source Monitoring, Conditional on a Remember (R), Know (K), or New (N) Judgment to Read Words, Heard Words, and New Words in the Recognition Task, for Participants in the Visual, Auditory, and Visual + Pleasantness Conditions (from Reynolds, 2001).

Test Condition		N	Recognition Response		
And Item Type			R	K	N
Visual					
	Read	19	.71(.05)	.52(.06)	.30(.06)
	Heard	20	.38(.05)	.38(.07)	.20(.04)
	New	16	.56(.09)	.48(.07)	.27(.05)
Auditory					
	Read	20	.40(.07)	.28(.04)	.16(.02)
	Heard	20	.67(.05)	.55(.05)	.21(.04)
	New	13	.31(.08)	.34(.08)	.14(.02)
Visual + Pleasantness					
	Read	21	.90(.02)	.57(.06)	.31(.07)
	Heard	24	.49(.07)	.35(.05)	.24(.05)
	New	17	.69(.10)	.36(.08)	.19(.06)

Note: Bracketed figures denote standard errors.

Next, for those conditions where the study status of the test item is *congruent* with the test instructions, the probability of a *yes* response on Test 2, conditional on a *remember* response on Test 1, was consistently higher than the same probability conditional on a *know* response. Again, this supports two conclusions. First, the probability of recollection on Test 2 is higher following a *remember* response than it is following a *know* response. Second, participants are recollecting information about the modality of presentation because this information is assisting with the source monitoring decision on Test 2.

Given the previous conclusions, it is possible to make a prediction about the probability of responding *yes* conditional on *remember* and *know* responses in those conditions where the test item is *incongruent* with the test instructions (i.e., heard words in the two visual test conditions and read words in the auditory test condition). In the incongruent conditions, participants should be more likely to recollect modality specific information on Test 2 if their response on Test 1 was *remember* rather than *know*. However, in contrast to the congruent conditions, the recollection of modality specific information should lead to a *no* response. This prediction is not confirmed – the probability of incorrectly responding *yes* in the incongruent conditions, following a *remember* response on Test 1, is as large as, or larger than, that following a *know* response.

The most obvious conclusion is that participants are less likely to retrieve auditory information if they are looking for visual information, and less likely to retrieve visual information if they are looking for auditory information. Not only are participants responding to the incongruent words which previously elicited a *remember* response as if they are *highly* familiar, they also exhibit a tendency to respond *yes* to both congruent and incongruent words following a *know* response.

Thus, Reynolds' findings are consistent with our conjecture about Yonelinas and Jacoby's (1996) results. However, in their design, when participants were faced with a font size decision, one could not differentiate those who responded *yes* on the basis of recollected screen location information, from those who responded *yes* on the basis of familiarity. In contrast, in Reynolds' experiment, when participants are making a heard decision about a visual word, we can determine whether participants are responding on the basis of a feeling of familiarity or on the basis of recollected visual information. Since participants were responding on the basis of familiarity, not recollection, it appears that words that support recollection under one set of retrieval instructions, appear familiar under another set. This supports a degree of redundancy between recollection and familiarity.

Further Support for Dependency? Matching and Mismatching Contexts

Humphreys, Bain, and Pike (1989) presented an extensive argument for the involvement of context in recognition and recall. As part of that argument they addressed the problem caused by the apparent insensitivity of recognition to the reinstatement of environmental and internal contexts. They proposed two partial solutions to this problem. First, they argued that test instruction must allow participants to reinstate most, if not all, of the relevant context. Second, they noted that some models (cf. Humphreys, Pike, Bain, & Tehan, 1989) permit a distinction between what they referred to as the additive and interactive use of context. Specifically, they proposed that recall would be improved by the additive use of context, whereas in recognition, the additive use of context would increase both the hit and false alarm rate, leaving sensitivity unchanged.

This conjecture received partial support from Murnane and Phelps (1993, 1994, 1995), who used a combination of font color, screen background color, and screen location as their context manipulation. Participants studied words in one contextual combination and were either tested in the same or in a different combination. In accordance with the assumption that an additive combination rule was used with some forms of context: Murnane and Phelps found that testing in the same context produced higher hit and false alarm rates than testing in a different context. However, over multiple experiments there was no trend for sensitivity to differ between the two conditions.

Macken (2002) extended the Murnane and Phelps design by asking participants to make *remember* and *know* judgments under same- and different-context conditions. His assumption was that recollection would increase from the different- to same-context condition, but that there would be no change in familiarity. In addition, he asserted that this pattern of results could explain the insensitivity of recognition to contextual reinstatement. However, this assertion is only partially correct. Macken's hypothesis, that context affects recollection but not familiarity, predicts that there will be a larger effect of context in measures of recollection than in overall measures of recognition sensitivity, but that there will still be a small effect on overall sensitivity.

The assertion that Murnane and Phelps' contextual manipulation will produce a small effect on sensitivity seems unlikely given Murnane and Phelps' (1993, 1994, 1995) extensive null results. Nevertheless, we decided to examine Macken's (2002) data to see if there was a trend towards a small effect of context. In Table 3.3 we present results from his three experiments[1]. Note that in Macken's experiments there are two Different Context conditions. In the Different Context (new) condition, words are tested in a new unstudied combination of font color, screen background color, and screen location. In the Different Context (old) condition, participants study words in one of two contexts, A and B. At test, half the words studied in the A context are tested in the A context, and the other half are tested in the B context. Similarly, half the words studied in the B context are tested in the B context, and the other half are tested in the A context. The false alarm rate (FAR) is calculated by

[1] We would like to thank William Macken for making his data available to us.

averaging the FAR to new words tested in the A context and new words tested in the B context. That is, there is no same and different context for new words, so the FAR is identical for the Same and Different Context (old) conditions. This manner of testing does not allow the participant to adopt a different criterion for the Same and Different Context (old) conditions, so the difference in the hit rate (HR) is the appropriate measure of sensitivity.

TABLE 3.3
Results from Macken (2002) Experiments 1, 2 and 3: the Probability of Responding Remember (R), Remember and Know Responses to Old Words (Hit Rate: HR), Remember and Know Responses to New Words (False Alarm Rate: FAR), and the Signal Detection Measures of Sensitivity (d' and A').

Experiment Number and Condition	R	HR	FAR	d'	A'
Experiment 1					
Same Context	.59	.72	.09	2.06	.88
Different Context (new)	.49	.66	.07	2.00	.88
Experiment 2					
Same Context	.32	.69	.26	1.23	.80
Different Context (new)	.23	.66	.22	1.30	.81
Experiment 3					
Same Context	.59	.81	.21	1.68	.88
Different Context (new)	.49	.76	.14	1.79	.88
Different Context (old)	.51	.78	.21	-	-

Note: The d' and A' measures of sensitivity are not included for the Different Context (old) Condition of Experiment 3 because in this design the FAR is constrained to be identical in the Same and Different Context (old) conditions. Therefore, the HR is the appropriate measure of sensitivity.

Table 3.3 provides the probability of a *remember* response, the probability of either a *remember* or *know* response to an old word (HR), the probability of either a *remember* or *know* response to a new word (FAR), and the signal detection measures of sensitivity, d' (parametric) and A' (non-parametric). As Macken (2002) predicted, the probability of a *remember* response was consistently higher for the Same Context conditions relative to the Different Context conditions. In Macken's data (not included in Table 3.3) there was also a small trend for the probability of a *remember* response to be higher for new words in the

Same Context conditions. In his comparisons, Macken used a high-threshold guessing correction to ensure that the increase for old words was not the result of a bias process. However, given the low level of responding *remember* to new words, the high-threshold guessing correction makes little difference. As we have noted, the correction is also unnecessary when comparing the Same Context and Different Context (old) conditions of Experiment 3.

Table 3.3 also demonstrates that, in keeping with the assumption that context is being used additively, the Same Context conditions have a higher HR and FAR than do the Different Context (new) conditions. Furthermore, the signal detection measures of sensitivity show no sign of a trend towards increased sensitivity in the Same Context conditions. This conclusion depends on the adequacy of the d' and A' measures. However, there is an alternative. As the Same Context and Different Context (old) conditions share a common FAR, performance can be evaluated using the HR. In Macken's Experiment 3 data (the only experiment including a Different Context (old) condition), there is a small difference in favor of the Same Context condition (.81 to .78). If an average difference of this magnitude were maintained across a number of studies it would provide evidence to support the small difference in sensitivity that Macken's hypothesis requires. Alternatively, if this small difference is not maintained, as would be suggested by Murnane and Phelps' (1993, 1994, 1995) null results, it would be evidence that most of the improvement in recollection is due to words which would have been given *know* responses in the Different Context (new) condition. This changing of a *know* response to a *remember* response would also provide evidence for a degree of redundancy between recollection and familiarity. Consequently, we decided to closely replicate Macken's (2002) Same Context and Different Context (old) conditions.

EXPERIMENT

Design and Participants

Twenty-two first-year psychology students from the University of Queensland participated to receive credit for an introductory psychology course. A 2x2 factorial design was employed, manipulating Study-Test

Context Condition (Same or Different) and Test Item (Old or New) within participants. Four contexts were implemented across two lists. Within a list, items could appear in one of two contexts at study (e.g., A or B) and either context at test. At test, half the studied items were re-presented in their original context (Same Context Condition: e.g., A-A & B-B), while the remaining items were presented in the other context that appeared during the study episode, but not for those items (Different Context Condition: e.g., A-B & B-A). Half the new items were presented in one context (e.g., A), the remaining in the other context (e.g., B). Each context was a unique combination of font color, screen background color, and screen location.

Materials

The stimuli were 240 5-letter words of intermediate frequency (20-50 occurrences per million) derived from the 1994 issues of The Sydney Morning Herald (TSMH Word Database – Dennis, 1995). The experiment was administered using Intel PentiumIII® computers and Diamond View® 17-inch monitors. All words were presented in lower case (24-point MS Sans Serif). Four different contexts were employed, two of which were used in Macken (2002): Context A: a word in red font, on a green background, in the bottom right corner of the screen; Context B: a word in yellow font, on a brown background, in the top left corner of the screen; Context C: a word in blue font, on a pink background, in the top right corner of the screen; Context D: a word in white font, on an orange background, in the bottom left corner of the screen.

Procedure

Each participant studied two lists of words (Lists 1 and 2). The 240 words were randomly assigned to be presented at study in either context A or B (List 1), or C or D (List 2), or to serve as new items presented at test in either context A or B (List 1), or C or D (List 2). Each study list contained 80 words: 40 presented in each of the two list contexts. Each test list contained 120 words: 40 targets re-presented in their original context (e.g., 20 A-A & 20 B-B), 40 targets presented in the alternate context (e.g., 20 A-B & 20 B-A), and 40 distractors – 20 new words appearing in each of the list contexts (e.g., 20 new A & 20 new B). The order of the lists was counterbalanced such that half the participants received List 1 (Contexts A & B) followed by List 2 (Contexts C & D), while the remaining participants received List 2 followed by List 1. The assignment of words to the list conditions, and the order of presentation of items (in their various contexts), was randomized for each participant.

Participants were instructed that they would receive a list of words to study; that the words would appear in different colors, on different backgrounds, and in different screen locations, but that they were simply required to learn the words for a memory test later in the experiment. During study, each word remained on screen for 3 seconds, followed by a 1-second gray screen, after which the next list item appeared. The 80 words took just over 5 minutes to present, after which participants were informed of the nature of the recognition test. The first page of instructions indicated that they would receive a list of words containing both old and new items and that they would be required to discriminate the words they did study from those they did not. Additionally, they were informed that they would be required to characterize their memory for "old" words into a separate "remember" or "know" judgment. A brief description of the memorial experience that supports a *remember* response and that which supports a *know* response followed. The second page of instructions informed participants that test words would appear in different colors, on different backgrounds, and in different screen locations, but that their recognition judgment should be made purely on the basis of whether they recognized the word or not – whether they believed they studied it earlier. At test, items remained on screen for 2 seconds, after which a screen appeared with two buttons marked "old" and "new". Any *old* response made by the participant automatically generated a screen with two buttons marked "remember" and "know". Participants simply clicked the buttons using a mouse in order to make their responses. The buttons stayed on screen until a response was recorded, at which time the next test word appeared. A page indicating that the participant would receive a second version of the learning task they had just completed separated the two lists of words comprising the experiment. The procedure in the second list was identical to that of the first list. Both accuracy and reaction time data were recorded.

RESULTS

The mean probability of responding *remember* and *old* (*remember* plus *know*) for the Same and Different Context (old) targets and distractors are presented in Table 3.4. The standard error of the mean for each type of test item is given in parentheses as a measure of variability. The results are presented across both lists, and separately for the first and second list that participants received. Recall that the experiment was counterbalanced such that half the participants received a list with contexts A and B first, followed by a list with contexts C and D, and vice versa. Consequently, the breakdown of the data reflects the order in which participants performed the two tasks rather than the contrast between the use of contexts A and B and contexts C and D. The data were examined to check whether there was an advantage for one pair of contexts over the other, and no mean differences greater than .03 were observed for *remember*, *know* or *new* responses across Same and Different Context (old) targets and distractors.

TABLE 3.4
Mean Probability of Responding Remember and Old (Remember or Know) for Same and Different Context (old) Targets and Distractors for Both Lists (A&B; C&D), the First List (either A&B or C&D), and the Second List (either A&B or C&D) received by participants.

Test Condition	Item Type	Recognition Response	
		Remember	Remember or Know
Both Lists			
	Old (Same)	.424 (.046)	.678 (.035)
	Old (Different)	.425 (.046)	.678 (.033)
	New	.073 (.021)	.203 (.022)
First List			
	Old (Same)	.380 (.048)	.652 (.033)
	Old (Different)	.391 (.047)	.659 (.032)
	New	.072 (.018)	.205 (.021)
Second List			
	Old (Same)	.468 (.054)	.705 (.041)
	Old (Different)	.459 (.054)	.698 (.040)
	New	.075 (.024)	.201 (.026)

Note: Bracketed figures denote standard errors.

While there was no advantage for one pair of contexts over the other, performance on the second list appears superior to the first, particularly for *remember* responses to Same and Different Context (old) targets. It is important to note that the key difference between our experiment and Macken's (2002) is the addition of the second list. Interestingly, the increase in performance for the second list (greater sensitivity) is characterized by a higher proportion of *remember* responses, a lower proportion of *know* responses, and an unchanged proportion of *new* responses, relative to the first list.

The most significant result of our experiment was the failure to find any effect of context for both *remember* and *old* (*remember* plus *know*) responses in the Same and Different Context (old) conditions. The common FAR precludes the need for signal detection measures of sensitivity, allowing a direct comparison of the HR for the two conditions.

DISCUSSION

As expected, we found no effect of context on the probability of an *old* response when we compared a Same Context condition with a Different Context (old) condition. If we had also replicated Macken's (2002) finding that the Same Context condition had higher levels of *remember* responses relative to the Different Context (old) condition, we would have had strong support for a degree of redundancy between recollection and familiarity. However, we found no evidence for any change in the probability of a *remember* response, despite obtaining more than twice Macken's number of observations.

The most ready explanation for our failure to replicate Macken is some difference in our procedures. The most significant change was the use of two study-test sessions. However, an examination of the separate sessions suggests that this is not the problem. We also presented our materials at a somewhat faster rate than Macken, both at study and test. In addition, Macken included a Different Context (new) condition in his Experiment 3, while we only employed the Same and Different Context (old) conditions. All other details of the respective experiments (including instructions, list lengths, screen locations, and colors) were very similar.

Future research will endeavor to determine whether any of these differences are responsible for the disparity in results. If we can establish a way to reproduce a Same versus Different Context (old) effect on *remember* responses, we will be able to test the hypothesis that recollection and familiarity have a degree of redundancy.

Macken had three comparisons in which a *remember* response was more likely in Same Context than in Different Context (new) conditions, and only one comparison between a Same Context and a Different Context (old) condition. Although he found a significant effect, our null result must cast some doubts on his finding. It is possible that the effect of context on *remember* responses, just like the effect of context on *old* responses, disappears when the Different Context (old) condition is compared to the Same Context condition. If this conjecture proves correct, it would be evidence against the prevailing assumption that *remember* responses index something like recall. That is, they would simply look like high-confidence recognition responses (Donaldson, 1996; Hirshman & Master, 1997).

CONCLUSIONS

To obtain estimates of familiarity from the process dissociation procedure, the remember-know paradigm, and recognition confidence judgments, it is necessary to assume that recollection and familiarity are independent at the item level. As we have demonstrated, the arguments supporting independence do not apply to the item level.

The first evidence against independence has been in the literature for several years. Bain and Humphreys (1988) demonstrated that after studying pairs of words, item recognition was invariant, even when rehearsal instructions produced large differences in recall. Although this finding can be reconciled with assumptions about independence, such reconciliation requires an implausibly exact trade-off.

Additional evidence derives from a two-test procedure, where participants are presented with auditory and visual words at study, make *remember-know* judgments on the first test, and on the second test, either respond *yes* to auditory words and *no* to visual words or respond *yes* to visual words and *no* to auditory words. In those conditions where the study presentation was congruent with the test instructions, the probability of correctly responding *yes* on Test 2, following a *remember* response on Test 1, was higher than the same probability following a *know* response. Thus, *remember* responses did predict the availability of source specific information on a subsequent test. However, in those conditions where the study presentation was incongruent with the test instructions, the probability of responding *yes* given a prior *remember* response was also as large as, or larger than, the same probability given a prior *know* response. We concluded that words that support recollection under one set of test instructions simply appear familiar under other sets.

Finally, we examined whether the reported increase in *remember* responses, when study-test context matched, occurred because items which only support a *know* response under mismatching conditions support a *remember* response under matching conditions. As predicted, there was no evidence that *old* responses increased in the Same Context condition relative to the Different Context (old) condition. However, we did not replicate the contextual-matching advantage with *remember* responses. More work is required to turn this task into a rigorous test of the independence hypothesis.

SUMMARY

We considered evidence from a variety of sources to test the assumption that recollection and familiarity are independent at the item level. A review of previous arguments revealed that they do not address independence at this level. Previous research also suggests that an implausible trade-off is required if the independence assumption is to be maintained. An examination of conditional probabilities in a two-test procedure revealed that words that support recollection under one set of test instructions, support familiarity under other sets. Contextual reinstatement effects were also examined. However, a failure to replicate previous findings, despite having more than twice as many observations, prevented us from a definitive test of the independence assumption.

REFERENCES

Bain, J. D., & Humphreys, M. S. (1988). The relational context effect: Cues, meanings or configurations. In G. Davies & D. Thomson (Eds.), *Memory in context: Context in memory* (pp. 97-138). London: Wiley.

Dennis, S. (1995). The Sydney Morning Herald Word Database. *Noetica: Open Forum, 1(4),* [Online] Available: http://www.psy.uq.edu.au/

CogPsych/Noetica/OpenForumIssue4/SMH.html.
Donaldson, W. (1996). The role of decision processes in remembering and knowing. *Memory & Cognition, 24,* 523-533.
Gardiner, J. M. (1988). Functional aspects of recollective experience. *Memory & Cognition, 16,* 309-313.
Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91,* 1-67.
Hirshman, E., & Master, S. (1997). Modeling the conscious correlates of recognition memory: Reflections on the remember-know paradigm. *Memory & Cognition, 25,* 345-351.
Humphreys, M. S. (1976). Relational information and the context effect in recognition memory. *Memory & Cognition, 4,* 221-232.
Humphreys, M. S., & Bain, J. D. (1983). Recognition memory: A cue and information analysis. *Memory & Cognition, 11,* 583-600.
Humphreys, M. S., & Bain, J. D. (1991). Episodic and strength components in recognition memory. In W. E. Hockley & S. Lewandowsky (Eds.), *Relating theory and data: Essays on human memory in honor of Bennet B. Murdock* (pp. 309-330). Hillsdale, NJ: Erlbaum.
Humphreys, M. S., Bain, J. D., & Pike, R. (1989). Different ways to cue a coherent memory system: A theory for episodic, semantic and procedural tasks. *Psychological Review, 96,* 208-233.
Humphreys, M. S., Dennis, S., Chalmers, K. A., & Finnigan, S. (2000). Dual processes in recognition: Does a focus on measurement operations provide a sufficient foundation? *Psychonomic Bulletin & Review, 7,* 593-603.
Humphreys, M. S., Dennis, S., Maguire, A. M., Reynolds, K., Bolland, S. W., & Hughes, J. D. (in press). What you get out of memory depends on the question you ask. *Journal of Experimental Psychology: Learning, Memory, and Cognition.*
Humphreys, M. S., Pike, R., Bain, J. D., & Tehan, G. (1989). Global matching: A comparison of the SAM, Minerva II, Matrix and TODAM models. *Journal of Mathematical Psychology, 33,* 36-67.
Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language, 30,* 513-541.
Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General, 110,* 306-340.
Macken, W. J. (2002). Environmental context and recognition: The role of recollection and familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 153-161.
Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review, 87,* 252-271
Murnane, K., & Phelps M. P. (1993). A global activation approach to the effect

of changes in environmental context on recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 882-894.

Murnane, K., & Phelps M. P. (1994). When does a different environmental context make a difference in recognition? A global activation model. *Memory & Cognition, 22,* 584-590.

Murnane, K., & Phelps M. P. (1995). Effects of changes in relative cue strength on context-dependent recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 158-172.

Qin, J., Raye, C. L., Johnson, M. K., & Mitchell, K. J. (2001). Source ROC's are (typically) curvilinear: Comment on Yonelinas (1999). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 1110-1115.

Reynolds, K. (2001). *An Examination of the Information Used to Guide Recognition and Source Monitoring Judgments.* Unpublished honors thesis, University of Queensland, Brisbane, Australia.

Rotello, C. M. (2001). Recall processes in recognition memory. In D. L. Medin (Ed.), *The Psychology of Learning and Motivation: Vol. 40. Advances in Research and Theory* (pp. 183-221). San Diego, CA: Academic Press.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychologist, 26,* 1-12.

Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: Evidence for a dual-process model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 1341-1354.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language, 46,* 441-517.

Yonelinas, A. P., & Jacoby, L. L. (1996). Noncriterial recollection: Familiarity as automatic, irrelevant recollection. *Consciousness and Cognition, 5,* 131-141.

This page intentionally left blank

Chapter 4

Modeling Implicit and Explicit Memory

Jeroen G. W. Raaijmakers
University of Amsterdam, The Netherlands

Quantitative modeling approaches to human memory are currently more prominent than ever. Widely used textbooks such as Haberlandt (1999) and Neath and Surprenant (2003) as well as the recently published new edition of Stevens' Handbook of Experimental Psychology (Raaijmakers & Shiffrin, 2002) all contain chapters on memory models. This is encouraging and contrasts with the still quite common attitude among experimental psychologists who regard mathematical modeling at best as "too complicated" but more often as a bit suspect.

One reason why many experimental researchers are a bit suspicious about mathematical models is the identification of mathematical modeling with data fitting. There was, indeed, a time when just showing that a particular equation could fit the data from an experiment was considered a major accomplishment. Even nowadays we sometimes see such exercises, for example, in the controversy over the exact nature of the learning curve: is it a negatively accelerated exponential or does a power law describe the data better? Such an approach may be important under some circumstances, for example, in practical applications if one wants to predict the amount of learning that is to be expected after a given number of study periods.

However, there are drawbacks to such approaches. First, there is often no underlying theory about memory processes that leads to the specific equation. If so, we are still in the dark about what it tells us about the memory system itself. Ideally, we would like to know whether a particular set of assumptions regarding the memory system generates a curve such as observed. Second, it does not generalize to anything else. In order to evaluate the proposal, we can only look at other data from the same type, but we cannot devise new experiments that test the underlying assumptions in a different task, because there are no such underlying assumptions.

These two criteria summarize what I believe to be the major ones by which the success of mathematical models of memory should be evaluated. Much more important are models that are based on general frameworks for a large variety of memory tasks rather than models for just a single task, even if these general models are somewhat less detailed for predicting specific experimental data. Over the past 25 years or so, we have seen a number of such models that have been quite successful as general models for episodic memory. Examples are: the ACT model, SAM/REM, MINERVA2, and TODAM. These models differ in a number of respects, but they all focus mainly on episodic memory paradigms. Raaijmakers and Shiffrin (2002) have given an extensive review of this work. What is important is that these models have not just "fitted the data" but give detailed explanations of several puzzling phenomena and have led to the discovery of a number of important new facts about human memory. In that respect, global memory models distinguished themselves above simple curve-fitting models.

Some examples from the SAM/REM theory that Richard Shiffrin and I have encountered over the past 25 years illustrate these points. One of the initial accomplishments of the SAM model was its explanation for the part-list cuing effect. This effect refers to the phenomenon in free recall when one cues the subject with a random sample of the list items. The effect is not an increase in the number of items recalled, as one would expect, based on the notion that performance in such a task depends heavily on the formation of interitem associations. The interitem associations do not seem to help and may even seem to hurt, despite that other aspects of the data do show a positive effect of such associations on recall performance. Application of the SAM model (Raaijmakers & Shiffrin, 1981) showed that this counterintuitive finding follows quite naturally (without making any special assumptions). The explanation provided by SAM was based on two considerations. First, since both groups will be using cues to generate additional items during the recall process (as assumed in the SAM model), large differences should not be expected. Second, the slight negative effect was due to a rather subtle difference between the cues that were given to the cued group (the experimenter-provided cues) and the cues that were (mainly) used by the control group, i.e. the subject-generated cues.

We also predicted under which circumstances the effect would reverse. For example, the model predicts that when the list is tested after a delay, the effect becomes positive. Such an effect was indeed obtained:

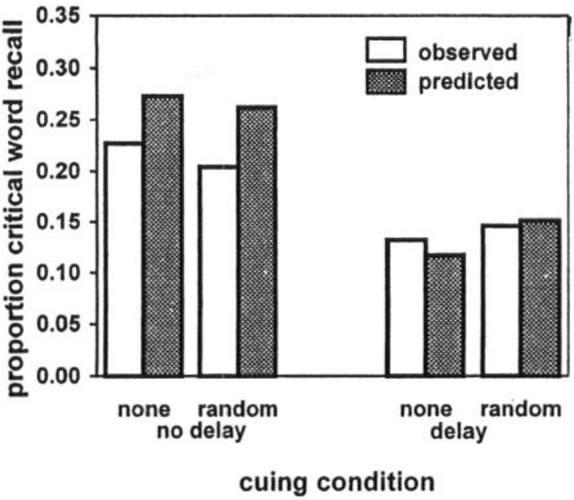


FIG. 4.1. Observed and predicted proportions of critical word recall for the cued (random cues) and noncued conditions in Experiment 2 of Raaijmakers and Phaf (1999).

Raaijmakers and Phaf (1999) gave the part-list cues (i.e., a randomly selected half of the list items) either immediately after study or after a delay filled with the learning of an unrelated list. When the cues were given immediately after study, the usual negative effect was observed (Fig. 4.1). However, when the cues were given after a delay, a reversal occurred and the cued group now recalled slightly more critical items than the noncued group.

The second example comes from Mensink and Raaijmakers (1988). We applied the SAM model to interference and forgetting. The model resolved a number of inconsistencies and controversies that had plagued the traditional interference theory for years. Interference theory had been quite successful in the 1950s as is evident from this quote from Postman:

> "Interference theory occupies an unchallenged position as the major significant analysis of the process of forgetting" (1961, p. 152).

However, by the early 1970s the situation had changed and the problems for the theory had become so large that the same author had to conclude:

> "Interference theory today is in a state of ferment if not disarray ... There is no lack of new data ... but so far they have failed to resolve the basic theoretical issues" (Postman, 1975, p. 327).

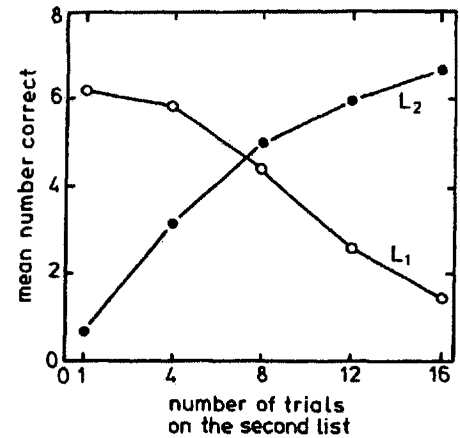


FIG. 4.2. Predictions of the SAM model for first-list and second-list recall in the Barnes and Underwood study showing retroactive interference in a MMFR design. (Fig. 2 from Mensink & Raaijmakers, 1988).

This remarkable shift of opinion was caused by the problems with the concept of "unlearning," one of the cornerstones of the then standard Two-Factor theory for interference. The major problem was that it assumed that a particular type of test, the MMFR (Modified Modified Free Recall) method, was immune to the effects of competition and that any form of interference that was observed using such a test method was due to "unlearning." However, proactive interference effects were also observed using MMFR tests and these could not be due to unlearning (since the critical learning took place after the learning of the proactive interfering list). Hence, the theory failed to provide a satisfactory explanation for proactive interference. However, using the SAM model, Mensink and Raaijmakers (1988) showed that the conflicting results could be resolved, basically because that model did not assume that MMFR testing eliminated response competition. As shown in Figs. 4.2 and 4.3, the model successfully predicted both interference effects in MMFR testing as well as proactive interference.

The third example comes from the global familiarity models for recognition memory (the SAM model for recognition being an example). This research led to the discovery of the list-strength effect, or rather the absence of it, in recognition. Although performance on both recognition and recall tests is affected by the number of other items on the list (the list-length effect), only recall is affected by the strength of those other items; recognition is not so affected (Ratcliff, Clark & Shiffrin, 1990).

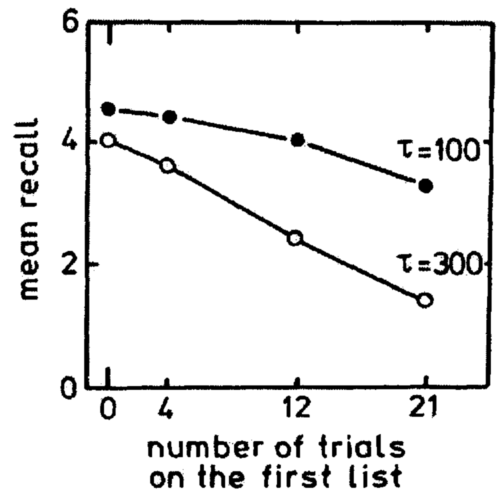


FIG. 4.3. Predictions of the SAM model showing proactive interference effects as a function of the retention interval (τ) and the number of trials on the first list.
(Fig. 6 from Mensink & Raaijmakers, 1988).

Shiffrin, Ratcliff and Clark (1990) showed that such a difference between the number of other items and their strength presents serious problems for many models for recognition. They also proposed a solution based on the differentiation hypothesis, assuming that the interfering effect of an item decreases as the item becomes more different, more differentiated, from the target item. This idea subsequently became very important in the development of the REM model for recognition memory (Shiffrin & Steyvers, 1997).

Finally, recent work by Malmberg and Shiffrin (in press) led to the "one shot of context" hypothesis, the idea that roughly only the first second of study is important for the storage of contextual information in a trace; additional study time will increase the amount of semantic and associative information stored in the trace, but will not have an effect on the amount of context information stored in the trace. This hypothesis appears to explain a large number of findings but would probably not have arisen outside the context of mathematical modeling.

All in all, these examples show that these more complex models do much more than simple data-fitting, providing new insights and leading to the discovery of important new phenomena, and imposing strong restrictions on the form of models for human memory.

Modeling Implicit Memory

As illustrated above, the SAM/REM theory is quite successful, and represents one of the most fully developed models for episodic memory, accounting in detail for the data from a variety of episodic memory tasks. Further, we have recently extended the scope of the theory to deal with semantic memory and, especially, implicit memory. First, let me briefly discuss the basic phenomenon and the major explanations that have been provided in the literature.

Generally, implicit memory stems from paradigms where some initial study (either intentional or incidental) takes place followed by a semantic memory task involving both old (presented during the initial study) and new items. Semantic memory tasks include word identification, category decisions, naming, word or fragment completion, etc. Note that in all of these cases, it is possible to do the task even if there was no initial study trial. In this respect, implicit memory paradigms are very different from episodic memory tasks (quite confusing for a subject without a prior study list).

The fact that such implicit memory tasks can be performed even without prior study implies, in my opinion, that any reasonably complete model for implicit memory must at the same time be a model for semantic memory, if only to account for the performance on the new items. Many explanations for implicit memory or repetition priming do not provide a model for performance on the new items and hence are not precise enough to enable quantitative predictions.

There are a number of phenomena for which any theory for implicit memory should provide an explanation. First, implicit memory is sensitive to variables that do not affect explicit memory. Examples of this include the finding that implicit memory is sensitive to the perceptual format of the items (auditory/visual) and this usually does not affect explicit memory performance. (This is not a universal law, however: some more conceptual implicit memory tasks are also not affected by the perceptual format.)

Second, subject populations that show a deficit in performance on standard memory tasks often show a relatively normal performance on implicit tasks. The most obvious example is the finding that amnesic patients usually show a relatively normal repetition priming effect. (Again, this should not be exaggerated: there often is a slight difference in the size of the priming effect; however, the difference is much less dramatic than that seen on explicit memory tasks.)

Third, there is usually no correlation between the scores on explicit and implicit memory tasks. This finding has been used in the past as evidence for the claim that such priming effects are not dependent on episodic memory. However, these findings should be interpreted quite cautiously since episodic tasks such as recall and recognition may also show a large amount of independence. Moreover, contrary to explicit memory scores, scores on implicit memory tasks are usually based on difference scores. The priming effect is defined as the difference between performance on the repeated versus the nonrepeated items. Such difference scores are known to be quite unreliable and hence would not be expected to correlate high with other measures (cf. Buchner & Wippich, 2000; Meier & Perrig, 2000; Buchner & Brandt, 2003).

Three types of theoretical accounts in the literature for implicit memory seem to be most popular. The first is that priming or implicit memory effects are due to *temporary strengthening of semantic or lexical traces*. This is perhaps the oldest explanation for priming effects, dating back to Morton's Logogen model (Morton, 1969). A problem for such an explanation was that priming effects are dependent on the perceptual similarity between the initial study and the implicit memory test. If priming was due to the strengthening of lexical traces, it appeared to require not just one mental lexicon but several. However, since priming effects are also influenced by such factors as whether it is a male or a female voice, this would then seem to require one mental lexicon for male voices and one for female voices. It is clear that such a proposal quickly becomes quite ridiculous and thus, this type of explanation lost its appeal.

The second type of explanation attributes implicit memory effects to the contribution of episodic memory traces. This explanation seems to be especially popular among language researchers who are mainly interested in semantic and lexical memory and have no special interest in memory itself. A reason why this might be appealing to these researchers is that it relegates such effects to another memory system, one that they have no interest in and, hence, they do not have to bother about such priming effects. They are merely an experimental nuisance. An obvious problem with such an explanation is that it seems to predict a correlation between episodic memory and implicit memory performance since both are based on the same memory traces. Another problem, not often mentioned but also important, is that it is difficult to come up with a reasonable model for, say, lexical decision in which those episodic traces would be activated so quickly as to affect the processing of the item in semantic or lexical memory. That is, it is one thing to claim that such effects are due to episodic memory but quite another to show that such an idea would also work in practice. After all, the episodic trace would often be quite weak, and retrieval is nearly always assumed to be faster for stronger traces. Hence, one would have to predict that the semantic trace would be retrieved much faster than the episodic trace. Finally, such an explanation does not seem able to account for the finding that in many priming tasks, the priming effect is not affected by the nature of the encoding task (elaborative vs. superficial encoding) while this has, of course, a huge effect on explicit memory performance.

The third and perhaps most popular account is that implicit memory performance is due to a separate memory system or memory systems, distinct from semantic and episodic memory. Thus, researchers such as Schacter (1990) propose that priming effects are due to perceptual representation systems. For example, Schacter argues that "visual priming may make it easier ... to extract visual information from the test cue" (1990, p. 237). Here, the priming effect appears to be due to low level perceptual learning. Such an account seems to provide a simple explanation for various dissociations between implicit and explicit memory tasks, because these are simply due to different memory systems. For example, to account for the finding that amnesics show normal priming, this account simply assumes that amnesics have a normal implicit memory system and only a deficit in the episodic memory system. However, what is often overlooked is that such an account has its own problems when one would try to formulate it in terms of a quantitative model.

One problem: Just as the explanation based on episodic memory, this explanation would have to show that it also accounts for performance on the control items, and this would require a model for semantic or lexical memory. It is not at all certain that the dynamics of such a multiple systems approach would generate an adequate account of reaction times in naming or lexical decision.

Bias Effects in Implicit Memory

In addition, recent new findings have posed a problem for both the episodic and multiple systems explanations of implicit memory effects. Ratcliff and McKoon (1997) showed that priming effects seem to be due to a bias in the system in favor of recently presented items rather than

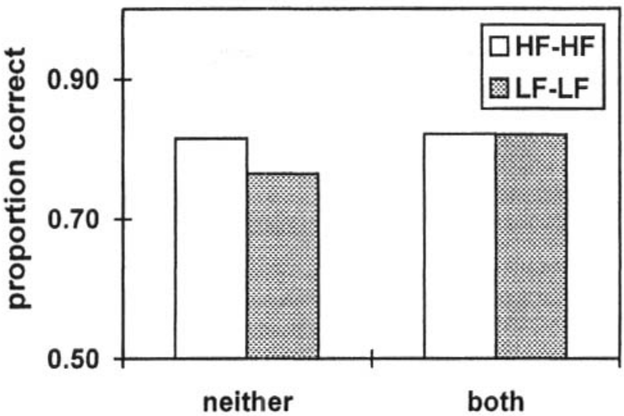


FIG. 4.4. Probabilities of correct identification for high and low frequency items in the 2AFC task as a function of whether both or neither of the two alternatives had been previously studied. Data from Wagenmakers et al. (2000).

better or more efficient processing of repeated items. The rationale for this hypothesis is best illustrated by their example in a perceptual identification experiment. The subject is presented with briefly flashed words for their subsequent identification. After the brief, tachistoscopic presentation of the word, two alternatives are presented for the subject to identify the presented word. A critical aspect of their experiments is that the alternatives may be either perceptually similar or dissimilar. The general result: There was a priming effect in the sense that the previously studied items are more likely to be chosen, independent of the word presented. That is, when the word LIED is studied and the alternatives are LIED and DIED, the subject is more likely to choose LIED, irrespective of whether LIED or DIED was flashed. This, by itself, might perhaps be reconciled with a multiple systems approach. However, they also showed that this effect is only obtained in the case of similar test alternatives (such as LIED and DIED). With dissimilar alternatives (such as LIED and SOFA), no effect of prior study is obtained. A multiple systems approach would be hard pressed to come up with an explanation why such an effect would only be obtained for similar alternatives if the effect is indeed based on more efficient processing of a previously studied item.

Ratcliff and McKoon (1997) provided an elegant model to explain

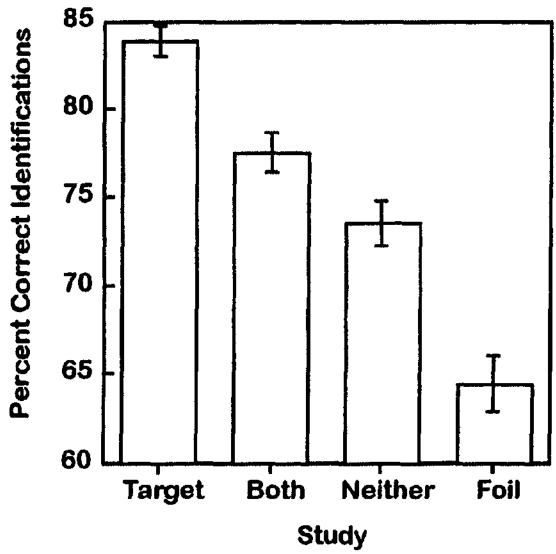


FIG. 4.5. Percentages of correct identification in 2 AFC auditory word identification as a function of whether the target, the foil, both or neither of the alternatives had been studied. (Fig. 2 from Zeelenberg et al., 2002).

these results based on the assumption that the system accumulates evidence in favor of each of the alternatives and that there is a bias to assign ambiguous evidence to recently presented items. When the alternatives are dissimilar, there is no competition between the alternatives, and hence no effect of prior study. Although this explanation gives a good account of the major effects, we have shown in a series of experiments that this is not the whole story. When both test alternatives are studied, the bias account of Ratcliff and McKoon predicts no difference compared with the case where neither alternative has been studied (the two biases cancel each other). However, in a number of experiments we observed better performance when both alternatives were studied.

Wagenmakers, Zeelenberg and Raaijmakers (2000) found, in perceptual identification, an advantage for low frequency items when test alternatives had both been studied (Fig. 4.4). Furthermore, Zeelenberg, Wagenmakers and Raaijmakers (2002) report a similar benefit in a series of experiments with both alternatives studied in a variety of priming tasks such as: auditory word identification, word fragment completion and picture identification (Fig. 4.5). In addition, we have demonstrated an advantage for the both-studied case with multiple study trials.

Such results appear to be at variance with the counter model by Ratcliff and McKoon (1997). In reaction to these results, they (McKoon & Ratcliff, 2001; Ratcliff & McKoon, 2000) modified the counter model by assuming that studied low-frequency words have a higher rate of feature extraction compared to nonstudied low-frequency words. However, it is unclear whether such a revision, running counter to the basic bias explanation, is necessary. Wagenmakers, Zeelenberg, Schooler, and Raaijmakers (2000) have shown that an alternative version of the Counter model can handle the both-primed benefit without altering the rate of feature extraction for studied words.

However, despite these slight deviations from the bias explanation, the overall picture still seems to provide strong evidence against any account that explains priming effects in terms of better or more efficient perceptual processing of the studied items as maintained by a multiple systems approach or the pure episodic account. We are thus left with the conclusion that neither the episodic nor the multiple systems account provides a satisfactory explanation for priming effects. Because the explanation based on strengthening of semantic or lexical traces was also discredited, the question becomes how priming effects should be explained.

An alternative account was developed by Schooler, Shiffrin and Raaijmakers (2001), based on the REM theory (Shiffrin & Steyvers, 1997). This account is a modification of the explanation that assigns the effect to changes in semantic or lexical memory. However, instead of assuming that semantic memory is a system that encodes only abstract information, we make the assumption that semantic memory is a dynamic system that is sensitive to all kinds of contextual factors.

The model by Schooler, Shiffrin and Raaijmakers assumes that the semantic or lexical memory is the result of the accumulation of many episodic memory traces: When a new semantic unit is first encountered, it is stored just as any normal episodic experience. However, upon a second presentation, the old trace may be retrieved. If so, new information will be added to that trace. With repeated presentations, the trace will accumulate many different semantic as well as context features. The end result is a trace that has all the properties normally associated with semantic memory: The semantic traces are relatively complete and hence easily accessible, and they are associated to so many different contexts that they are for all practical purposes context-independent. The idea that information may be accumulated in a trace is an assumption that we have

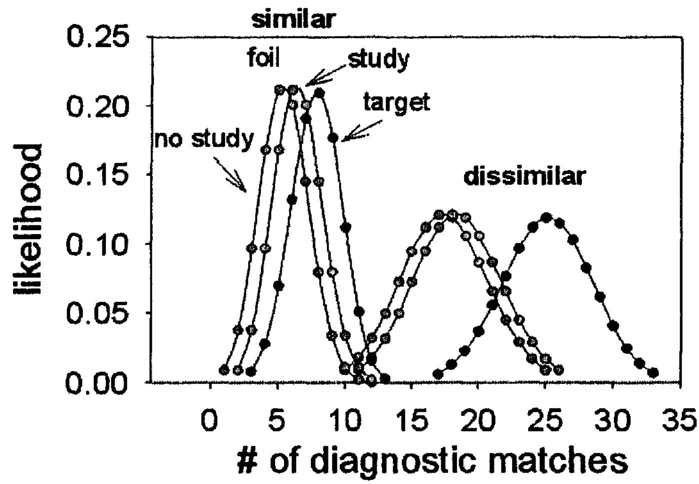


FIG. 4.6. Predicted distributions of the number of diagnostic matches in a 2AFC perceptual identification task for similar and dissimilar alternatives. Adding one extra match due to prior study of the foil item has more effect on the overlap for the similar case than for the dissimilar case due to the higher variance in the latter case.

been using in SAM all along. Moreover, it corresponds to the common idea that semantic memory grows out of episodic memory (although this is not universally accepted, e.g. Tulving (1983) has always maintained that it is the other way around: Semantic memory comes first and episodic memory builds upon semantic memory).

How could such a theory account for priming effects in the 2AFC (two-alternatives forced-choice) paradigm? The basic idea is that noisy perceptual information extracted from the flash is compared with the lexical traces of the two test-alternatives. The system simply determines which alternative better matches the perceptual information. However, contrary to other views of semantic memory, we assume that the current context does take part in this comparison: The set of features from the flashed item that is compared to the lexical traces includes the current context. Due to the previous study episode, the lexical trace of a studied item contains a small amount of context information that matches the features from the flash and this provides an advantage for the studied alternative.

Although this is not immediately obvious, this account also provides an explanation for the result that a priming effect is only obtained for the similar alternatives and not for the dissimilar alternatives. It is because for similar alternatives such as LIED and DIED, many of the features from the lexical representation will give the same result for the comparison with the features of the flashed stimulus. In this example, only the first letter as well as the stored context features are relevant for the comparison between LIED and DIED. When the alternatives are dissimilar, however, many more features are relevant for the comparison. Such a model predicts that the effect of the additional contextual matches for a recently studied alternative will be more or less washed out, if the number of relevant feature comparisons is high (Fig. 4.6). The result is that an effect is predicted for similar alternatives but not for dissimilar alternatives.

REM (Retrieving Effectively from Memory) as a Model for Implicit Memory

REM was developed initially to explain performance in standard explicit memory tasks. Hence, this model is the first, as far as I am aware, that provides a theoretical account for both episodic and semantic memory, and for both explicit and implicit memory tasks. As such, it sets a new standard for future mathematical modeling attempts.

First, let us examine the dissociations between explicit and implicit memory that have been given so much attention recently. The model accounts for the finding that priming effects are affected by the perceptual format of the stimuli since the features that are compared to the lexical traces are the perceptual and contextual features. Any semantic features that might have been activated or strengthened as a result of the prior study are irrelevant when a perceptually based task is used. This explains why levels-of-processing effects will have little or no effect on performance when a task is used where the cues are mainly perceptual in nature. It also explains why perceptual modality has a major effect on the priming effect that is observed since this directly affects the match between the perceptual features of the test item and the perceptual information stored in the semantic or lexical trace.

Second, the present model accounts for the finding that explicit and implicit memory performance are largely independent. Implicit memory is based on the semantic/lexical traces, whereas episodic performance is based on the episodic traces. Finally, the model explains the finding that amnesics have a normal implicit memory performance by pointing to the fact that such performance is assumed to be based on the semantic/lexical memory system, a system that we may assume to be relatively spared in amnesics. That is, we know that amnesics are able to use their semantic/lexical memory systems (otherwise they would have great difficulties with simple conversations), hence, access to these systems may be assumed to be unimpaired. If this is the case (perhaps not for all amnesics, e.g. Alzheimer patients do seem to have problems with semantic memory), then there is no reason to assume that they should not also show priming effects, if such effects are based on modifications of the semantic/lexical system. Thus, the present theory holds that the implicit memory performance of patients will be normal provided that they also show normal performance on standard semantic memory tasks.

How can we extend this theory to other implicit and semantic memory paradigms? First, let's examine related priming phenomena such as associative priming. This effect refers to the finding that performance in, say, lexical decision is affected if just before the target item, another word is presented that is either semantically or associatively related or unrelated to the target item. Such a result is explained by the assumption that in such a task, features from the prime are still in STS or still being processed when the target is presented, and hence these features combine with the perceptual features of the target item and this combined set of features is then compared to the lexical traces. If the features of the prime are related to the stored semantic features of the target item, the match will be better compared to the case where the prime is unrelated to the target item (see Ratcliff & McKoon, 1988, for a similar approach based on SAM). Note that even when the prime is related to the target item, the features from the prime will also increase the level of noise in the comparison but this will also be the case for unrelated primes. In fact, in such experiments it is indeed observed that performance is often better without a prime than with a related prime. However, the usual comparison is between the conditions with related and unrelated primes, resulting in a reliable advantage for related primes.

Finally, in some paradigms, the target has to be classified in terms of its semantic features such as animate versus inanimate. Such a judgment cannot be made on the basis of the perceptual features, but has to be made on the basis of the semantic features, recovered from the lexical/semantic system. If so, the present model *would* predict an effect of the nature of the study tasks, i.e., the priming effect should be affected by the level-of-processing of the study task. Although no studies have

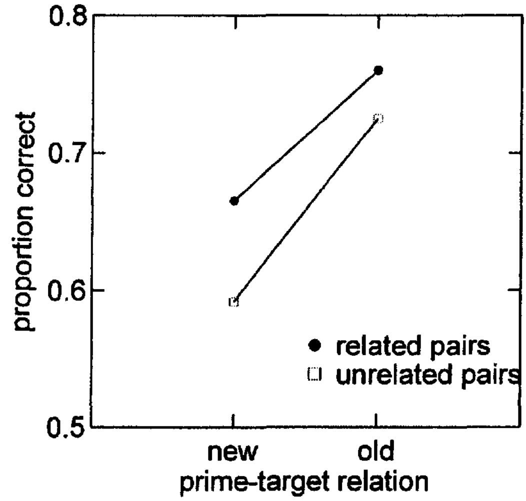


FIG. 4.7. Proportion correct in primed perceptual identification after four study trials for intact and rearranged prime-target pairs of pre-experimentally related or unrelated items. Data from Schrijnemakers (1994).

been done that directly examined level-of-processing effects in semantic classification tasks, indirect support is provided by the finding that many other conceptual implicit memory tests do show reliable effects of level-of-processing (see Challis & Sidhu, 1993; Hamann, 1990; Srinivas & Roediger, 1990).

What is new and different in the approach that I advocate is that we conceive semantic memory not as a relatively static system, but as a system that is quite dynamic in nature. In particular, we assume that the semantic traces include contextual information and, hence, are sensitive to recent episodes in which that particular word was encountered.

Associative Repetition Priming

Thus far, most of the implicit literature has focused on pure repetition priming effects. However, some researchers have also examined repetition priming for associations rather than merely single items (e.g., Dagenbach, Horst, & Carr, 1990; Goshen-Gottstein & Moscovitch, 1995a, 1995b; Graf & Schacter, 1987). We have performed a number of experiments to demonstrate such associative repetition priming effects for novel associations and to determine whether such effects are larger for novel than for existing relations. Our focus was on those types of associative priming tasks where the priming effects were most likely due to automatic activation of associative information rather than on strategic

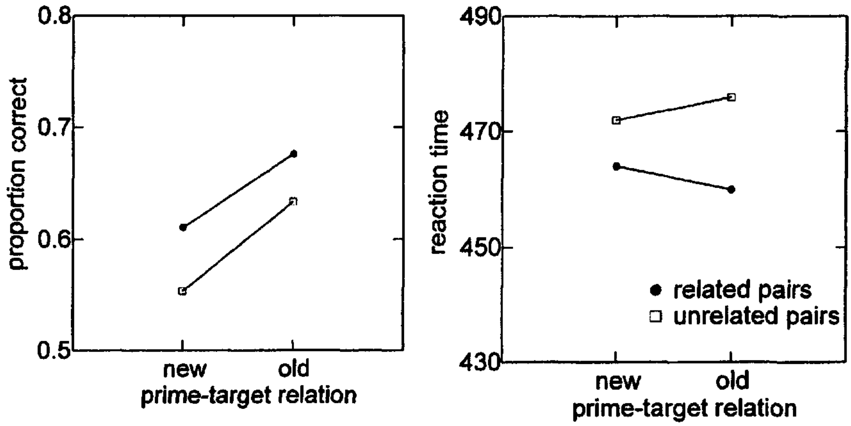


FIG. 4.8. Performance on the final test for intact and rearranged prime-target pairs of pre-experimentally related and unrelated items after four training trials using perceptual identification. Left: Proportion correct on the primed perceptual identification test. Right: Reaction times on the primed lexical decision test. Data from Schrijnemakers (1994).

factors. We have used lexical decision tasks with relatively short SOA's and an associative priming variant of perceptual identification, in which the prime item is briefly flashed prior to the tachistoscopic presentation of the target item (Schrijnemakers & Raaijmakers, 1997; Pecher & Raaijmakers, 1999; Pecher, Zeelenberg & Raaijmakers, 2002; Zeelenberg, Pecher & Raaijmakers, 2003). Although such tasks previously produced mixed results, we obtain clear and consistent effects of prior study for novel associations, provided that (a) a relatively large number of study trials is given, and (b) the same task is used during both the initial study and the final test. The first requirement indicates that the effect is relatively small, suggesting a reason why such effects have been difficult to obtain in the past.

For example, Schrijnemakers (1994) ran an experiment with four study trials under the following design. On the initial study trials, both a perceptual identification test and a paired-associate study was given, the pairs being either related or unrelated. After four such trials, the pairs were either rearranged or kept the same. The final pairing was either pre-experimentally related or unrelated. The results (Fig. 4.7) show that there was a clear effect of prior study, but the effect was the same for pre-experimentally related and unrelated pairs. In a following experiment, we gave subjects the same type of initial study, but on the final test, the task was either the same as the one in the prior study (primed perceptual identification) or different (primed lexical decision). The results (Fig. 4.8) showed that an associative repetition priming effect was only obtained when the task at test was the same as at initial study. This is not due to the fact that such associative repetition priming effects cannot be obtained with lexical decision, because in other experiments where lexical decision was used both at study and at the final test, we did obtain reliable priming effects. Hence, whatever was learned was restricted to that particular task and did not generalize to other associative priming tasks. We interpret these findings as showing that performance is affected by the prior episodic study, but what is stored is not some abstract semantic information but associative information that is specific to a particular task.

Pecher and Raaijmakers (1999) replicated this result, again finding effects when the same task was used as during study. In all of these experiments, we have found that the effects of pre-experimental relatedness and episodic study are additive: The effect is just as large for previously related as for previously unrelated pairs. Pecher and Raaijmakers (in press) have replicated this finding using yet another priming task, in which subjects have to classify words into 'animate' or 'inanimate' categories.

All of these findings are difficult to reconcile with theories that assume that standard associative priming effects are based on a semantic or lexical memory system that is abstract and relatively static. Rather, the effects seem to be much more compatible with a view that the semantic system is highly flexible and dynamic and is sensitive to all kinds of contextual variables.

In the coming years, the research groups at Indiana and Amsterdam hope to extend the REM model to deal with such semantic or lexical tasks along the lines described above. We have already begun to develop a model for lexical decision that appears to be at least capable of explaining the major findings in that area (e.g., Wagenmakers, Steyvers, Raaijmakers, Shiffrin, Van Rijn, & Zeelenberg, in press). Hence, at the present time, the prospects for such a unified theory for both implicit and explicit and for both episodic and semantic memory seem to be quite promising. I expect that others will develop competing models with similar aims so that within a few years we will be able to do comparative evaluations of different models for both explicit and implicit memory.

SUMMARY

Mathematical models of memory are useful for describing basic processes of memory in a way that enables generalization across a number of experimental paradigms. Models that have these characteristics do not just engage in empirical curve-fitting, but may also provide explanations for puzzling phenomena and may lead to new discoveries. We provided a number of examples, taken from previous research with the SAM model. Although previous research has focused exclusively on the explanation of episodic memory, recent research within the SAM/REM approach has extended this model to implicit and semantic memory phenomena. This review provided some speculations on how this approach may be extended to deal with a number of basic data in implicit memory. It was emphasized that constructing a model for implicit memory necessitates the development of detailed models of lexical-semantic processing.

AUTHOR NOTES

Address correspondence to Jeroen G. W. Raaijmakers, Department of Psychology, University of Amsterdam, Roetersstraat 15, 1018 WB Amsterdam, The Netherlands. E-mail: J.G.W.Raaijmakers@uva.nl.

REFERENCES

Buchner, A., & Brandt, W. (2003). Further evidence for systematic reliability differences between explicit and implicit memory tests. *Quarterly Journal of Experimental Psychology, 56A*, 193-209.

Buchner, A., & Wippich, W. (2000). On the reliability of implicit and explicit memory measures. *Cognitive Psychology, 40*, 227-259.

Challis, B. H., & Sidhu, R. (1993). The dissociative effect of massed repetition on implicit and explicit measures of memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 115–127.

Dagenbach, D., Horst, S., & Carr, T. H. (1990). Adding new information to semantic memory: How much learning is enough to produce automatic priming? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 581-591.

Goshen-Gottstein, Y., & Moscovitch, M. (1995a). Repetition priming for newly formed and preexisting associations: Perceptual and conceptual influences. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* *21*, 1229-1248.

Goshen-Gottstein, Y., & Moscovitch, M. (1995b). Repetition priming effects of newly formed associations are perceptually based: evidence from shallow encoding and format specificity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 1249-1262.

Graf, P., & Schacter, D. L. (1987). Selective effects of interference on implicit and explicit memory for new associations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*, 45-53.

Haberlandt, K. (1999). *Human memory: Exploration and application.* (2nd ed.) London: Allyn & Bacon.

Hamann, S. B. (1990). Level-of-processing effects in conceptually driven implicit tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 970–977.

Malmberg, K. J., & Shiffrin, R. M. (in press). The "one-shot" hypothesis and the list- strength effect for free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition.*

McKoon, G., & Ratcliff, R. (2001). Counter model for word identification: Reply to Bowers (1999). *Psychological Review*, *108*, 674-681.

Meier, B., & Perrig, W. J. (2000). Low reliability of perceptual priming: Consequences for the interpretation of functional dissociations between explicit and implicit memory. *Quarterly Journal of Experimental Psychology*, *53A*, 211-233.

Mensink, G. J., & Raaijmakers, J. G. W. (1988). A model for interference and forgetting. *Psychological Review*, *95*, 434-455.

Morton, J. (1969). Interaction of information in word recognition. *Psychological Review*, *76*, 165-178.

Neath, I., & Surprenant, A. (2003). *Human memory.* (2nd ed.) Belmont, CA: Wadsworth.

Pecher, D., & Raaijmakers, J. G. W. (1999). Automatic priming effects for new associations in lexical decision and perceptual identification. *Quarterly Journal of Experimental Psychology*, *52A*, 593-614.

Pecher, D., & Raaijmakers, J. G. W. (in press). Priming for new associations in animacy decision: Evidence for context-dependency. *Quarterly Journal of Experimental Psychology.*

Pecher, D., Zeelenberg, R., & Raaijmakers, J. G. W. (2002). Associative priming in a masked perceptual identification task: Evidence for automatic processes. *Quarterly Journal of Experimental Psychology*, *55A*, 1157-1173.

Postman, L. (1961). The present status of interference theory. In C. N. Cofer (Ed.), *Verbal learning and verbal behavior.* (pp. 152-196). New York: McGraw-Hill.

Postman, L. (1975). Verbal learning and memory. *Annual Review of Psychology*, *26*, 291-335.

Raaijmakers, J. G. W., & Phaf, R. H. (1999). Part-list cuing revisited: A test of the SAM explanation. In C. Izawa (Ed.), *On memory: Evolution, progress* *and reflection on the 30th anniversary of the Atkinson-Shiffrin model.* (pp 87-104). Mahwah, N.J.: Erlbaum.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88,* 93-134.

Raaijmakers, J. G. W., & Shiffrin, R. M. (2002). Models of memory. In H. Pashler & D. Medin (Eds.), *Stevens' Handbook of Experimental Psychology, Third Edition, Volume 2: Memory and Cognitive Processes.* (pp. 43-76). New York: Wiley.

Ratcliff, R., Clark, S., & Shiffrin, R. M. (1990). The list-strength effect: I. Data and discussion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 163-178.

Ratcliff, R., & McKoon, G. (1988). A retrieval theory of priming. *Psychological Review, 95,* 385-408.

Ratcliff, R., & McKoon, G. (1997). A counter model for implicit priming in perceptual word identification. *Psychological Review, 104,* 319-343.

Ratcliff, R., & McKoon, G. (2000). Modeling the effects of repetition and word frequency in perceptual identification. *Psychonomic Bulletin & Review, 7,* 713-717.

Schacter, D. L. (1990). Priming and multiple memory systems: Perceptual mechanisms of implicit memory. In D. L. Schacter & E. Tulving (Eds.), *Memory systems 1994* (pp. 233-268). Cambridge, MA: MIT Press.

Schooler, L. J., Shiffrin, R. M., & Raaijmakers, J. G. W. (2001). A Bayesian model for implicit effects in perceptual identification. *Psychological Review, 108,* 257-272.

Schrijnemakers, J. M. C (1994). *The storage of newly learned information in semantic memory.* Unpublished Doctoral Dissertation, Catholic University of Nijmegen, Nijmegen, The Netherlands.

Schrijnemakers, J. M. C., & Raaijmakers, J. G. W. (1997). Adding new word associations to semantic memory: Evidence for two interactive learning components. *Acta Psychologica, 96,* 103-132.

Shiffrin, R. M., Ratcliff, R., & Clark, S. (1990). The list-strength effect: II. Theoretical mechanisms. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 179-195.

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM: Retrieving effectively from memory. *Psychonomic Bulletin and Review, 4,* 145-166.

Srinivas, K., & Roediger, H. L. (1990). Classifying implicit memory tests: category association and anagram solution. *Journal of Memory and Language, 29,* 389–412.

Tulving, E. (1983). *Elements of episodic memory.* Oxford: Oxford University Press.

Wagenmakers, E. J. M., Steyvers, M., Raaijmakers, J. G. W., Shiffrin, R. M., Van Rijn, H., & Zeelenberg, R. (in press). A Bayesian model for lexical decision. *Cognitive Psychology.*

Wagenmakers, E. J. M., Zeelenberg, R., & Raaijmakers, J. G. W. (2000). Testing the counter model for perceptual identification: Effects of repetition priming and word frequency. *Psychonomic Bulletin & Review, 7*, 662-667.

Wagenmakers, E. J. M., Zeelenberg, R., Schooler, L. J., & Raaijmakers, J. G. W. (2000). A criterion-shift model for enhanced discriminability in perceptual identification: A note on the counter model. *Psychonomic Bulletin & Review, 7*, 718-726.

Zeelenberg, R. Pecher, D., & Raaijmakers, J. G. W. (2003). Associative repetition priming: A selective review and theoretical implications. In J. S. Bowers & C. S. Marsolek (Eds.), *Rethinking implicit memory*. (pp. 261-283). Oxford: Oxford University Press.

Zeelenberg, R., Wagenmakers, E. J. M., & Raaijmakers, J. G. W. (2002). Priming in implicit memory tasks: Prior study causes enhanced discriminability, not only bias. *Journal of Experimental Psychology: General, 131*, 38-47.

This page intentionally left blank

In Honor of Robert L. Solso, an Outstanding Psychologist and Esteemed Colleague:

Chapter 5

Optimal Foreign Language Learning and Retention: Theoretical and Applied Investigations on the Effects of Presentation Repetition Programs

Chizuko Izawa, Sally Maxwell, Robert G. Hayden, Marc Matrana
Tulane University, USA; and
Althea J.E.K. Izawa-Hayden
American Bar Association, USA

"Only the educated are free." (Epictetus, Discourses, Book II, Chap 1, c A.D. 100)

In the current treacherous state of unsettled wars/military occupations, and continued acts of violence, the world is in need of more effective education in its schools and elsewhere to upgrade international understanding and mutual respect. This chapter, therefore, addresses applied aspects of learning theories to facilitate world peace and freedom from violence.

EFFICIENCY IN LEARNING AND EDUCATION: APPLICABLE LEARNING THEORIES TESTED

Sound application requires logical derivation from sound theory. Learning efficiency is no exception. Suppose that we have Time X for people to learn Y amount. Efficiency can be achieved in two ways: (**a**) Given fixed Time X, efficient learning occurs if Time X produces more than learning Y (i.e., Y+). Or, (**b**) given a fixed amount of learning Y, efficient learning also results, if Y is achieved in a time smaller than X (i.e., X-).

While these two ways may appear identical (learning takes time, and they describe the relationship between the amount learned and the time consumed), there are distinct differences in terms of adaptability to scientific experiments (in pre-experimental manipulations).

It is difficult, by nature, to hold the amount of learning Y constant in (**b**), because it is a dependent variable and cannot be easily and precisely manipulated by the experimenter. In contrast, total time, Time X, is an independent variable, which can be easily and precisely manipulated by the experimenter pre-experimentally. Thus, we can ascertain the degree of efficient learning readily and accurately under the (**a**) situation above.

What, then, is the value of extant learning theories? What can they suggest for actual educational settings? According to the most influential general learning theories such as Ebbinghaus (1885), Hull (1943), and modern mathematical learning models pioneered by Estes (e.g., 1950), discussed by Izawa, Hayden, and Franklin, the amount of learning is primarily a function of the number of S (study) trials, N (Chap. 2, Eq. 2.1, in this volume). This implies that in order to increase learning (Y+), the number of S trials/sessions must increase. If the administration of N trials consumes Time X, an additional S trial, i.e., a total of $(N+1)$ S trials requires Time X+ (i.e., Time X+x, if an S trial requires Time x). Thus, learning Y+ cannot be achieved within Time X. Efficient learning seems, therefore, impossible within the prescribed time here.

In the case of the total time hypothesis (TTH, e.g., Bugelski, 1962), quite explicitly: Time X is needed for learning Y, regardless of how Time X is scheduled or programmed. Thus, this hypothesis denies any possibility for efficient learning. For TTH (similar to general learning theories), the only way to increase the amount of learning is, therefore, to increase the total Time X of S trials/sessions, i.e., X+ to produce Y+. Hence, efficiency cannot be modified.

In contrast, Izawa's study-test-rest (S-T-R) presentation program hypothesis (a derivative of the test trial potentiating, retention interval, and identity models, Izawa, 1971b, 1981, 1985a-1985b/1989a-1989b, respectively) expects that optimization of learning is possible depending on how total Time X is programmed for study (S), test (T), and rest (R) trials and on how they are presented and repeated to the learner (for more details, see, e.g., Izawa, 1999, 2000; Izawa, Hayden, & Franklin, Chap. 2, this volume).

It is the *quantity* of S time/trials that is crucial for both the TTH and general learning theories. For the S-T-R presentation program hypothesis, however, it is the *quality* of S, T, and R presentation programs within Time X that determines the amount of learning. Thus, for the S-T-R presentation program hypothesis, Time X may produce either Y-, Y, or Y+, depending on presentation programming and related factors that control a given learning situation.

In four investigations utilizing 10 experiments (1993a, 1993b, 1999, 2000), Izawa discovered that learning efficiency may (**a**) vary as a function of S-T-R programs, (e.g., Izawa, 1966, 1967, 1968, 1970), repetition modes/programs, learning and testing difficulty levels, and participant attributes, (**b**) that it is a function of non-monotonic (U-shaped) multi-factors, and (**c**) the best performances are likely to be found at some intermediate points among several factors, including T effects, S and T trial interactions, and spacing of S and T trials, respectively (which produce positive effects on both short-term memory, STM and long-term memory, LTM).

In most previous studies, however, the experimental learning materials were manipulated by strictly measured nonsense syllables to control learning difficulty levels objectively and pre-experimentally (cf. Ebbinghaus, 1885. Nelson & McEvoy, Chap. 8 of this volume). However, there are concerns/preoccupations in certain quarters of cognitive psychology (especially applied psychology), which favor the view that nonsense syllables are of little relevance in applied settings.

For example, some anonymous reviewers of certain journals suggest in negative reviews that the mere fact that a study utilized nonsense syllables as experimental learning materials is sufficient to reject the manuscript (MS). (cf. Appendix PS.4: Shiffrin has intriguing comments about article reviews as an author of the famous chapter on the Atkinson-Shiffrin model and numerous other quality publications, as well as the former editor of the *Journal of Experimental Psychology: Learning, Memory, and Cognition*, and coeditor of several volumes).

It is our view that such views are ill founded, if not misguided. After all, independent of learning materials, students must master new items/concepts without prior knowledge. Thus, learning nonsense syllables may be thought of as an artificial language. For instance, for students who know nothing about the French language, the word "École" is equivalent to a nonsense syllable (e.g., XOJ) until it is learned to mean

"School" in English, "Escuela" in Spanish, "Schule" in German, or "学校" in Japanese.

If one should insist on some common characteristics among learning materials, e.g., ease of learning when comparing extant languages (French, English, Spanish, German, and Japanese, as above) with nonsense syllables (artificial language), one could ask, "What are the similarities among natural languages that are distinctively different from artificial-languages, for example, nonsense-syllables?"

Nevertheless, there may be some qualitative differences between artificial-language/nonsense-syllables in comparison with real languages, which evolved naturally. The truth will only be known after experimenting. Our current experiment was conducted with the German language, involving a native German speaker (Hayden), and those who acquired it as a second and a fifth language (Maxwell & Izawa), respectively.

OTHER EFFICIENCY RELATED HYPOTHESES

Most learners need more than one trial for acquiring knowledge or mastering skills, necessitating repeated presentations/practices. What are the optimally efficient ways to repeat/present S and T items? To delve into the efficiencies of natural language learning, the current experiment utilized the following innovative presentation methods, introduced by Izawa (1993a, 1993b, 1999, & 2000) to empirically examine two additional efficiency-related hypotheses below.

Study plus Test (S+T, S'+T', S/T) Presentation Duration Hypothesis

Given a total Time X, we can program that time by presenting each item of the learning list for a long period with a lower frequency, or a shorter duration for each item, with a higher frequency. Thus, item presentation duration and frequency are inversely related.

Each item/pair of a cued-recall/paired-associate learning (PAL) situation, utilized in the current experiment, is presented to the participants as an S (study) event where both cue/stimulus (A) and target/response (B) terms are jointly presented: As "A_j-B_j" for Pair j, or a T (test) event where only the former is presented as "A_j-?" for the participants to supply the B_j term (cf. Izawa, Hayden, & Franklin, Chap.

2, Appendix 2.2 of this volume).

When experimental variations were introduced for the S events alone (e.g., Izawa 1993a), the data provided allowed us to investigate both S presentation duration and frequency hypotheses, under the constant S-Time X, but if T events alone were varied (e.g., Izawa, 1993b), it became possible to examine both T presentation duration and frequency effects under the constant T-Time X conditions.

Similarly, co-variations of S and T events created experimental situations that shed light on the gross total effects of both S plus T (S+T, S'+T', or S/T) presentation duration and frequency hypotheses (e.g., Izawa, 1999, 2000). Co-varied S and T trial situations may be most relevant to applied settings such as foreign language learning classes. Here, both presentation rates and frequencies were the same for S and T trials. This also allowed us to include another important repetition program (S/T alternation) to be examined below, where the same logic holds.

We note here a well-established fact that *ceteris paribus*, longer presentations facilitate greater total learning than shorter ones. If so, the S+T, S'+T', or S/T presentation duration hypothesis would expect that the longer the presentation duration, the better the learning (see, e.g., Izawa, 1999, 2000 for more details). The " 's " will be explained shortly.

Study Plus Test (S+T, S'+T', S/T) Presentation Frequency Hypothesis

Another well-known factor that influences the outcome of learning is frequency of S (and T) trials. Other things being equal, the more often the S trial is repeated, the greater the learning. When considering this particular factor in isolation, the study plus test presentation frequency hypothesis does not seem different from general learning theories in Izawa, Hayden, and Franklin's Equation 2.1 (Chap. 2, this volume). However, there are fundamental differences: The current presentation frequency hypothesis **(a)** is confined to the fixed total Time X, and **(b)** provides a legitimate theoretical function for T trials in the S+T, S'+T', or S/T frequency hypothesis, while general learning theories do not recognize the effects of T trials at all, and each S trial is set for exactly the same duration, independent of the frequency of S trials.

Yet, the predictions from both theoretical positions are similar, in that the S+T, S'+T', or S/T frequency hypothesis expects that the greater frequency lead to the greater acquisition and retention. However, due to the constraint stemming from a constant S+T, S'+T', or S/T time (Time X), the greater frequency programs had shorter durations. Thus, the S+T, S'+T', or S/T frequency hypothesis has an inverse relationship with the corresponding duration hypothesis above. (cf. Izawa, 1999, 2000).

REPETITION PROGRAMS

All parents and schoolteachers know repetition is mandatory for children's learning. The same applies to adults. It is very rare, indeed, that a large quantity of very difficult new items can be mastered in a single S presentation. (We are sure that this applied to even Lady Murasaki's legendary feats of memory, cf. Izawa, 1988; Chap. 1, this volume). The question remains: Is there a better repetition program/ method under the restriction of the constant S+T (or S'+T' or S/T alternation) Time X? In search of the correct answer, we examined the following three repetition programs:

The S+T (Study plus Test) List-Repetition Presentation Program

The traditional standard PAL (paired-associate/cued-recall learning) proceeds under the list-design, suitable to the applied setting such as students' daily/weekly homework assignment of learning a list of *n* (=10 here) chemical symbols, historical facts, or German words. For example, in the multiple trial situations, trials/cycles are repeated with respect to the list, in which items within a list are presented in a random order from cycle to cycle, where a cycle is a run-through of the entire list, presenting each item only once, one at a time. On an S cycle, only S items (e.g., "A_j-B_j" for Pair *j*) are presented for the participants to learn, while on a T cycle, only T items (e.g., "A_j-?") are shown for them to provide corresponding B_j term responses (cf. Chap 2, Appendix 2.2 of this volume).

This procedure is similar to students in a language class trying to learn foreign words, one by one, in a different order, and in each S or T cycle (cf. a flash card technique, shuffling a deck of flash cards, each card representing an S or a T event) so that they learn those words without the help from the presentation order (such as the first one is "Mutter," instead of the corresponding English word, "Mother").

The S'+T' (Study plus Test) Item Repetition Presentation Program

The above traditional/standard list-repetition method/program is not the only way to learn thousands of foreign words, however. While learning several languages, almost simultaneously, Izawa found, as a participant in applied natural learning situations, that list-repetition is not always most efficient, especially when a large number of unlearned foreign words have to be mastered in a short time (i.e., a very difficult learning situation).

Instead, she discovered that the more effective method in high learning difficulty situations, a situation where she found herself perpetually, was to repeat an unlearned item several times successively (e.g., A_j-B_j, A_j-B_j, A_j-B_j, . . .) before proceeding to the next unlearned item within the list. She utilized these personal experiences in the laboratory and in relevant applied experiments.

Izawa named this new method the *item-repetition program* (1993a, 1993b, 1999, 2000). In the current experiment, both S and T cycles were presented with the item-repetition program that was identified by primes (" 's, " e.g., S' and T'). For example, in Condition 3, each S and T item is presented 3 times in succession: A_j-B_j, A_j-B_j, A_j-B_j, for Pair j prior to proceeding to the next S item on S cycle/trial, and the same was done for three consecutive T events, A_j-?, A_j-?, A_j-? on the T cycle. For an illustration of the S'+T' item repetitions, see Izawa (e.g., 2000, Fig. 4).

The S/T (Study/Test) Alternation Program

Within the same Time X (S plus T total time held constant), there is a third way to program S and T trials: The standard study-test method, or Program ST, a special case of Izawa's general S-T-R presentation program (Chap. 2 of this volume, Eq. 2.2, Table 2.1). In this program, S and T cycles alternate until the end of the experiment. The S/T alternation program is also a special case of the list-repetition program.

The advantages of the S/T alternation program, absent in the above two programs, include **(a)** the maximum positive S and T trial interactions and **(b)** positive S and T trial spaced practice effects, respectively. It would be interesting to examine whether the earlier results (e.g., Izawa, 1993a, 1993b, 1999, 2000) obtained with nonsense syllables could also be obtained in a moderately challenging German language-learning situation in a medium difficulty dimension.

TABLE 5.1
Experimental Design: Three Study (S) plus Test (T) Presentation Repetition Programs in Acquisition with Total S, T, S+T and R (Rest) Times Held Constant

S and T Presentation Programs Condition (Name, Rate)	Study (S) and Test (T) Phases $S_{(1)}$	$T_{(1)}$	. . .	$S_{(10)}$	$T_{(10)}$
S+T List-Repetition Program					
1 (ST, 3 sec)	S_1	T_1*	. . .	S_{10}	T_{10}*
2 (SSSTTT, 1 sec)	S_1 S_2 S_3	T_1 T_2 T_3*	. . .	S_{28} S_{29} S_{30}	T_{28} T_{29} T_{30}*
S'+T' Item-Repetition Program					
1 (S'T', 3 sec)	S'_1	T'_1*	. . .	S'_{10}	T'_{10}*
3 (S'S'S'T'T'T', 1 sec)	S'_1 S'_2 S'_3	T'_1 T'_2 T'_3*	. . .	S'_{28} S'_{29} S'_{30}	T'_{28} T'_{29} T'_{30}*
S/T Alternation Program					
1 (ST, 3 sec)	S_1	T_1*	. . .	S_{10}	T_{10}*
4 (STSTST, 1 sec)	S_1 T_1 S_2	T_2 S_3 T_3*	. . .	S_{28} T_{28} S_{29}	T_{29} S_{30} T_{30}*

T*= Target T

S = Study trial presented by the S list-repetition program

T = Test trial presented by the T list-repetition program

S' = Study trial presented by the S item-repetition program

T' = Test trial presented by the T item-repetition program

EXPERIMENTAL DESIGN

A list of 10 English-German paired-associates was programmed in four different ways as shown in Table 5.1 (derived from Izawa 2000, Table 2). Here a phase is defined as the cycle for the control (Cond. 1), 30 sec. in duration, in which 10 S or T items/events in the learning list were presented at a 3 sec. rate on an S or T cycle, respectively. On each S or T cycle, each S item (A_j-B_j for Pair j, where $1 \leq j \leq 10$ throughout the present investigation) or T item (A_j-? for Pair j) was presented once, one at a time, respectively. The presentation order of the pairs was randomized from cycle to cycle. Each phase is identified by a subscript in parenthesis, for example, $S_{(1)}$ and $T_{(1)}$ indicate the first S and T phases, respectively; whereas corresponding cycle S_1 or T_2 without parenthesis, for example, identify the first S cycle or the second T cycle in each condition within the S phase or the T phase, respectively.

Condition 1 (where the cycle uniquely coincides with the phase) is special in another sense as well: Throughout all three presentation programs in the current experiment, this condition can be construed as S+T list-repetition, S'+T' item-repetition, and S/T alternation programs simultaneously, and therefore serves as the control for all three experimental presentation programs (Table 5.1).

During acquisition, S and T phases were administered alternately until the 10th S and T phases were reached. Then, a 10 min. interpolated task followed, after which learners were given two unanticipated delayed retention tests, summarized in Presentation Sequence 5.1 below:

$$S_{(1)}\, T_{(1)}\, S_{(2)}\, T_{(2)} \ldots S_{(10)}\, T_{(10)}\text{--<Intpl. Task>--}T_{d1} T_{d2} \qquad (5.1)$$

S+T List-Repetition Programs

In Condition 2, each item from the same list was presented at a 1 sec. rate, three times faster than Condition 1. Because the list items were given at a faster presentation rate under the same 30 sec. total for each S (or T) phase as in Condition 1, the items were presented three times more frequently to the participants in Condition 2: Three S (or T) cycles per S (or T) phase as in the top of Table 5.1. Conversely, the control (Cond. 1) enjoyed a presentation three times longer per item, but a lower cycle frequency (only a 1/3) than the experimental condition (Cond. 2).

S'+T' Item-Repetition Programs

The primes (S' & T') are employed to distinguish between list- and item-repetitions. Conditions 1 and 3 were administered under the innovative item-repetition program (Izawa, 1999, 2000). Each item in the control and experimental conditions underwent longer vs. shorter presentation durations and fewer vs. greater presentation frequencies, respectively at 1 or 3 ratios (Table 5.1, center).

Participants were either presented with the same item once in Condition 1 (A_j-B_j on S cycles and A_j-? on T cycles, respectively, for Pair *j*), or 3 successive times in Condition 3 (A_j-B_j, A_j-B_j, A_j-B_j on S cycles and A_j-? A_j-? A_j-? on T cycles, respectively, for Pair *j*). Here, the two item-repetition program conditions (1 & 3) had only one S or one T cycle, each during one S phase or one T phase, respectively.

S/T Alternation Programs

The third type of presentation program under current investigation is the S/T alternation program, a special case of Izawa's S-T-R presentation programs or that of S+T list-repetition programs, i.e., the traditional study-test method. With this presentation program, given the conditions' constant total time, S and T cycles alternated throughout, and were utilized in Conditions 1 and 4: The same list was presented with S and T cycles alternating at 3 and 1 sec. presentation rates, respectively (Table 5.1, bottom). Presentation frequencies: One S and T cycle per 30-sec. phase for Condition 1, but three S and T cycles per respective phase for Condition 4.

With the S/T alternation program at the end of the target T^*s (Table 5.1), all four conditions had *identical total* S, T, and R times. (For all four conditions, the total inter-cycle R (rest) time in each S or T phase was 3 sec.). This fact facilitated rigorous tests of several learning theories targeted in the current investigation.

In summary, the present experiment utilized 2 presentation speeds (3 vs. 1 sec.), and 2 different frequencies (10 vs. 30). The control, Condition 1, afforded 10 S and 10 T trials at a 3 sec. rate, but experimental conditions, 30 S and 30 T trials at a 1 sec. rate in Conditions 2, 3, and 4, under the S+T list-repetition, S'+T' item-repletion, and S/T alternation programs, respectively.

THEORETICAL PREDICTIONS

Care was taken to test as many major learning theories in a natural language learning situation as possible (experimental design, Table 5.1). All conditions utilized a strictly *constant total time* (30 sec. per S or T phase, with 3 sec. inter-phase R interval). The conditions differed only in the *program types* of the S and T events (subdivisions of the S and T cycles/phases), allowing for very strong tests of Bugelski's (1962) TTH, (both original and modified, Izawa and Hayden, 1993), Izawa's (1999, 2000) S-T-R presentation program hypothesis, and the presentation duration and frequency hypotheses. The current design also provided opportunities for examining the general learning theories, the T trial learning/S=T hypothesis, and the hypermnesia view.

The Total Time Hypotheses (TTHs, Original and Modified)

From Bugelski's TSTH, (original, total study time hypothesis) as well as, the modified hypotheses, such as TTTH (total test time hypothesis) and TS+TTH (total study plus test time hypothesis, Izawa, 1993a, 1993b; Izawa and Hayden, 1993), all our conditions were expected to produce the same levels of performance.

The item-repetition program conditions maximally favor TTHs, because these two conditions (Table 5.1, center) were nearly identical procedurally. Each item was given 3 sec. for studying or testing before proceeding to the next item on the list: One long 3 sec. exposure vs. 3 successive 1 sec. exposures differ little in reality for learners. Non-significance between item-repetition programs, predicted by TTHs, was one of the very few situations that was borne out in data with nonsense syllables (e.g., Izawa, 1993a, 1993b, 1999, 2000). Thus, similar non-significance is also likely for unlearned foreign words, which could be regarded as equivalent to nonsense syllables.

The Study-Test-Rest (S-T-R) Presentation Program Hypothesis

In sharp contrast with expected TTHs identity predictions everywhere, Izawa's S-T-R presentation program hypothesis generally predicts substantial differences among and between program conditions depending on the presentation programs and how total Time X is scheduled.

However, the S-T-R presentation program hypothesis is also equipped to accommodate occasional identity in performances with certain combinations of presentation programs, repetition methods/programs (especially the item repetition program), and the operation of multiple factors including the learning difficulty dimension controlled by the quality and quantity of learning materials, participants' attributes, and the amount of Time X.

The Presentation Duration and Frequency Hypotheses

Under all three types of repetition programs, *the presentation duration hypothesis* anticipates better performances (fewer errors) with the control (Cond. 1). Its S+T, S'+T', and S/T alternation presentation duration was three times longer than that of the experimental conditions (Conds. 2, 3, & 4, respectively) with much shorter ones: 3 vs. 1 sec.

In contrast, *the presentation frequency hypothesis* expects the complete opposite. This follows from the fact that the experimental conditions afforded three times greater frequency in S+T, S'+T', and S/T presentation programs than the control: 30 vs. 10 exposures in total.

The General Learning Theories

The general learning theories (T effects are not recognized) predict identical performances among the conditions with an identical number of S trials (*N*). Such identity is expected **(a)** when all four learning curves are plotted as a function of preceding S trials, and **(b)** when three experimental conditions (Conds. 2, 3, & 4) are compared (identical number of S trials experienced at an identical S presentation rate) with each other.

The T Trial Learning/S=T Hypothesis and the Hypermnesia View

Notice in Table 5.1 that the current experiment has two situations (Conds. 2 & 3) with successive T trials. Hence, the experiment also provides a testing ground for Whitten's T trial learning/S=T hypothesis, (Whitten & Bjork, 1977; Whitten and Leonard, 1980) as well as the hypermnesia view (e.g., Ballard, 1913; Payne, 1987; Roediger & Challis, 1989). Both positions predict significant growth in correct responses (decrement in incorrect responses) from the first to the last T/T' per T/T' block, based on the respective assumptions concerning acquisition and retrieval processes by the proponents of these two positions.

Furthermore, it is of great interest to examine whether consecutive Ts/T's given with respect to the list-repetition program (Cond. 2) and to the item-repetition program (Cond. 3) produce similar or different results in a natural language learning situation.

METHOD

Forty undergraduate men and women with no knowledge of the German language volunteered to participate. They were assigned randomly to one of the four conditions (one control, three experimental conditions for three S-T-R presentation programs), 10 participants per condition.

A list of 10 English-German paired-associates was presented on a Stowe Memory Drum. German words were relatively low in meaningfulness, without contemporary relevance, impersonal, and neutral. The list was judged by authors familiar with German to be easy. Note, however, foreign language learning is generally considered to be difficult, thus the current situation can be regarded as intermediate on the learning difficulty scale for those unfamiliar with German, requiring neither too extreme effort, nor too trivial: It was a moderately challenging situation for learners, just as it is for many subjects in school.

Prior to the main English-German paired-associate learning task, a practice list that reflected all three program-types was administered to every participant. Practice task data were reassuring in that they revealed neither group differences, nor indication that uncontrolled participant variables probably influenced main task data only to a minimal degree.

RESULTS AND DISCUSSION

Overall Learning Performance Comparisons of All Four Conditions

Fig. 5.1 shows all four conditions in the three presentation programs as a function of the number of preceding S trials, *N*, in terms of proportion of incorrect responses. For Condition 2 or 3 that had successive T trials under either S+T list-repetition or S'+T' item-repetition programs, there were no significant differences over the three successive T trials in any T blocks; thus, the first T of each T block was entered in Fig. 5.1.

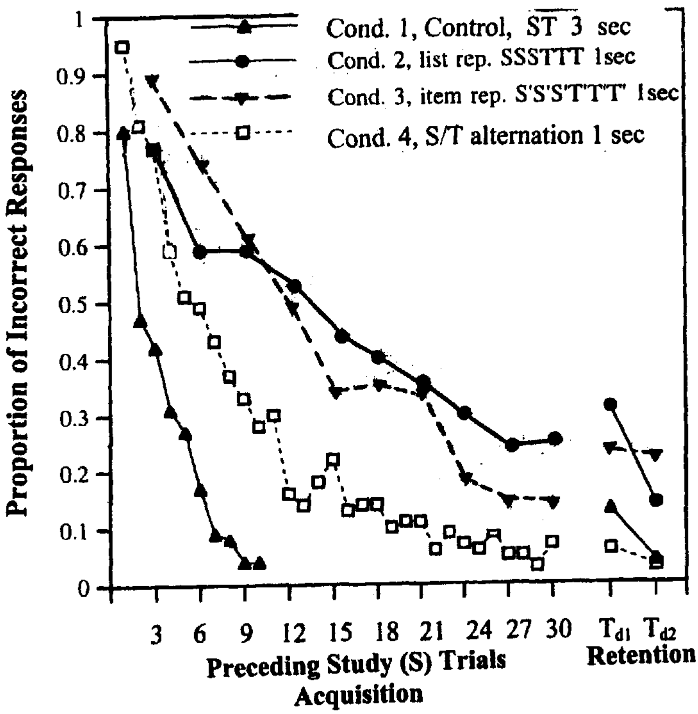


FIG. 5.1. Learning performances under all conditions in terms of proportion of incorrect responses as a function of preceding S (study) trials.

From both general learning theories (the first 10 S trials) and TTHs (total time hypothesis, both original TSTH and modified TTTH, TS+TTH), the four conditions were expected to be identical for both short-term acquisition (STM) and long-term retention (LTM) performances, because all conditions were characterized by an identical total number of S trials (=*N* for the first 10 S trials) and S, T, S+T, S'+T', and R times, respectively for the four curves in Fig. 5.1.

Results clearly contradicted the anticipated identity: Highly significant differences among the four conditions resulted in both acquisition tests (target T^*s, Table 5.1) and in retention T_ds (combined);

respective $F(3, 36)$s were 19.17 and 9.11, $ps < 0.01$.

For the STM acquisition period, Condition 1 with a longer S+T list repetition program was advantageous. However, for the LTM (on delayed retention Ts, which experienced identical S, T, and R times, as well as retrieval response time), significant differences were obtained among the four conditions, the S/T alternation program being the best.

In contrast, the data provides substantial support for Izawa's S-T-R presentation program hypothesis, given its general prediction of varied results among/between different S and T presentation programs. These results with natural language materials differ little from the earlier findings with nonsense syllables. Our *a priori* hunch proved correct. Previously unlearned natural languages are the functional equivalents of synthetic languages from the laboratory (since 1885).

Study plus Test (S+T) List-Repetition Program Performances

To identify the determinants of the strong positive results among the S-T-R programs, as well as to examine both presentation duration and frequency hypotheses, more fine-grained analyses were conducted both within and between presentation programs.

Subtle differences between conditions may be hard to detect because: (**a**) The variations utilized here are limited to only two frequencies at 3 and 1 sec. rates (Izawa, 2000 had four, ranging from 6 to 1 sec. with difficult materials), and (**b**) the current learning situations were only moderately challenging (mimicking applied settings), as apparent from some figures, affected by the floor effects (in incorrect responses).

Figure 5.2 exhibits both S+T list-repetition programs (Conds. 1 & 2). This and all remaining figures show T performances in terms of *incorrect responses* on the vertical (Y) axis, as a function of the *preceding S phases (total S time/trials)* on the horizontal (X) axis. The horizontal bars accompanying the control (Cond. 1) data points indicate a response time/duration of T events/trials three times longer than those in the experimental conditions.

Differences between the S+T list-repetition programs were not very large early in acquisition, but later became more prominent and significant in both acquisition (best data per 3 sec. T block) and delayed retention Ts: $F(1, 18)$s being 9.11, and 4.66, $ps < 0.01$ and .05, respectively. These strong overall differences sharply contradict TTH's prediction of nonsignificance, but provide solid support for the S-T-R presentation program hypothesis.

Likewise, there was no support for the ascending performance order (the greater the frequency, the better the performances) predicted by the S+T presentation frequency hypotheses. In fact, the rank order was the exact opposite: Condition 2, with a 3 times greater frequency, although predicted to be better, resulted in higher errors! Instead, Condition 1 with only 1/3 of the frequency, which was predicted to be worse, turned out to be significantly better. The same negative verdict applies to general learning theories as well: Condition 2 with three times more S trials (N = 3) demonstrated significantly poorer performance than Condition 1 with only $N = 1$ per replication!

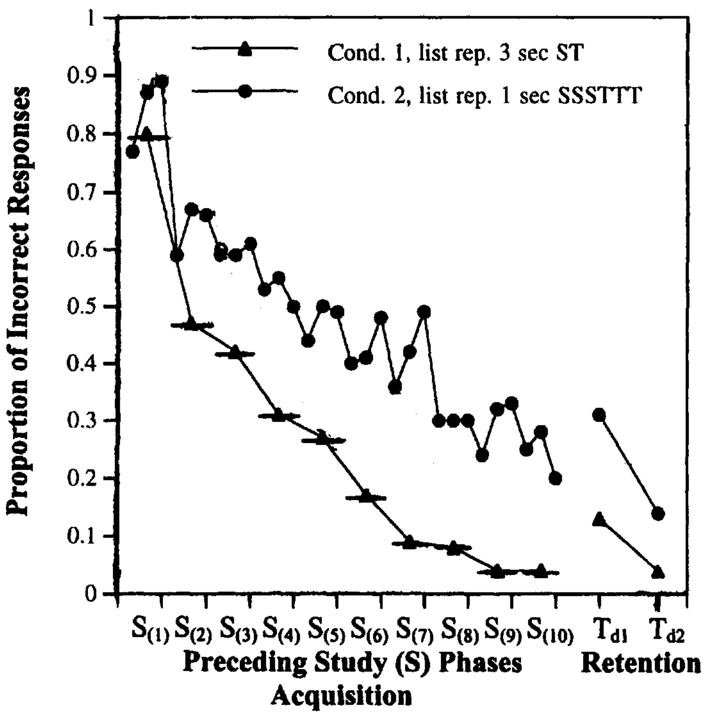


FIG. 5.2. S+T list-repetition program performances in Conditions 1 and 2, with the total S (study), T (test), and R (rest) times held constant, respectively. (This and all subsequent figures show performances on T phases in terms of proportion of incorrect responses as a function of preceding S phases).

The fact that the longer S+T list-repetition condition (Cond. 1) was significantly superior to the shorter one (Cond. 2) is consistent with the S+T presentation duration hypothesis' prediction of a descending performance order (longer duration boosts performance).

Following each S phase (3 S cycles), Condition 2 (SSSTTT, 1 sec) had three successive T cycles under the S+T list-repetition program. Over these three successive T trials in each T block, there were no significant differences among them in any T block. That is, lack of significant error reduction indicated no new learning on T per se, or improved response search strategies. These results provided evidence against both the T trial learning/S=T hypothesis and the hypermnesia view for the German language learning data. These negative results buttress similar outcomes in Izawa and colleagues' Chapter 2 and her other earlier nonsense syllables studies. T trials are indeed distinctively different from S trials (S $\neq$ T, e.g., Izawa, 1969, and all related studies).

Study plus Test (S'+T') Item-Repetition Program Performances

As far as the learners' experiences of the two S'+T' item-repetition conditions (1 & 3) are concerned, they differed little procedurally, because each list item was exposed for a total of 3 sec. in a block during both S and T cycles/phases, regardless of various presentation frequencies and durations. This S'+T' item-repetition program was designed to (**a**) maximally favor TTH and (**b**) explore if and how S and T presentation program variations may produce meaningful differences when applied to natural language as compared to "artificial" nonsense syllables. As expected, no significant differences resulted overall, between the two variations within the S'+T' item-repetition program, as in Fig. 5.3. The findings for a natural language were similar to those for nonsense syllables (e.g., Izawa, 1999, 2000).

Nonsignificance between the two S'+T' item-repetition programs does support TTH's basic assumptions. It is also consistent with the S-T-R presentation program hypothesis postulate that, among other things, learning and retention are a function of repetition programs. One such program that produces nonsignificance is the item-repetition program, as is the case here.

In contrast to that lack of significance, which clearly contradicted such single factor theoretical positions as the S'+T' presentation duration and frequency hypotheses. These positions call for either descending or ascending performance increments, respectively, when moving from Condition 1 (3 times longer duration with a 1/3 frequency) to Condition 3 (a 1/3 shorter duration, but 3 times greater frequency).

The same data also contradicted general learning theories that assert learning to be a function of the number of S trials: The more S trials, the greater acquisition. Condition 3 with 3 times more S trials differed little from Condition 1 (only a 1/3 of S trials) that showed nonsignificant trends of superior performances (fewer errors).

Note also: Over three successive T' trials in every T' block with the

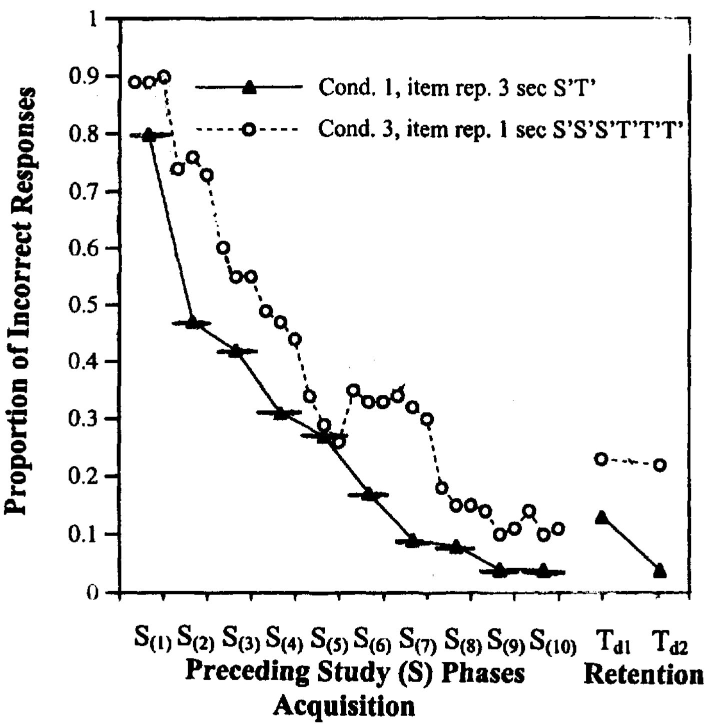


FIG. 5.3. S'+T' item-repetition program performances in Conditions 1 and 3, with the total S' (study), T' (test), and R (rest) times held constant, respectively.

item repetition program, there was no significant decrements of incorrect responses (Fig. 5.3). These findings negate both test trial learning/S=T hypothesis and the hypermnesia view. Both positions predicted orderly improvements over successive T' events/trials.

Comparisons of Study-plus-Test S+T List- and S'+T' Item-Repetition Programs with the Same Presentation Duration and Frequency

The S+T list- and S'+T' item-repetition programs are compared at a 1 sec. rate and at the same presentation speeds/frequencies in Fig. 5.4. There were some general trends during acquisition periods for superior performances for the item-repetition program, especially the mid to late acquisition periods. However, overall, superiority was not large enough to reach an accepted significance level with current German language learning (intermediate in the general learning difficulty dimension; see the Method section above). According to Izawa (1999, Fig. 9.14; 2000), there is a U-shape relationships between list- vs. item-repetition programs, the latter excels the former when learning is extremely difficult, and the opposite holds when extremely easy, whereas when learning is intermediately difficult/easy, as is the case here, the two programs differ little. Thus, current data do support such predictions by Izawa.

Nonsignificance here is consistent with TTH expectations and those of the presentation duration and frequency hypotheses because the two conditions in Fig. 5.4 had constant S/S', T/T', S+T/S'+T', and R total times, as well as constant presentation durations and frequencies. However, the extent of the support is weak.

Also, owing to the fact that the two conditions in Fig. 5.4 had the same number of S/S' trials, general learning theories are also supported, though weakly, from this aspect of data.

However, neither the test trial learning/S=T hypothesis nor the hypermnesia predictions were supported by Fig. 5.4 data. None of the conditions, either under the list- or item-repetition program, produced significant improvement over 3 successive T/T' trials in T/T' blocks. Thus, no successive tests in the list- or item-repetition program produced any decline of incorrect responses. The results do support the S-T-R presentation program hypothesis, which predicts neither new learning nor forgetting on unreinforced T/T' trials.

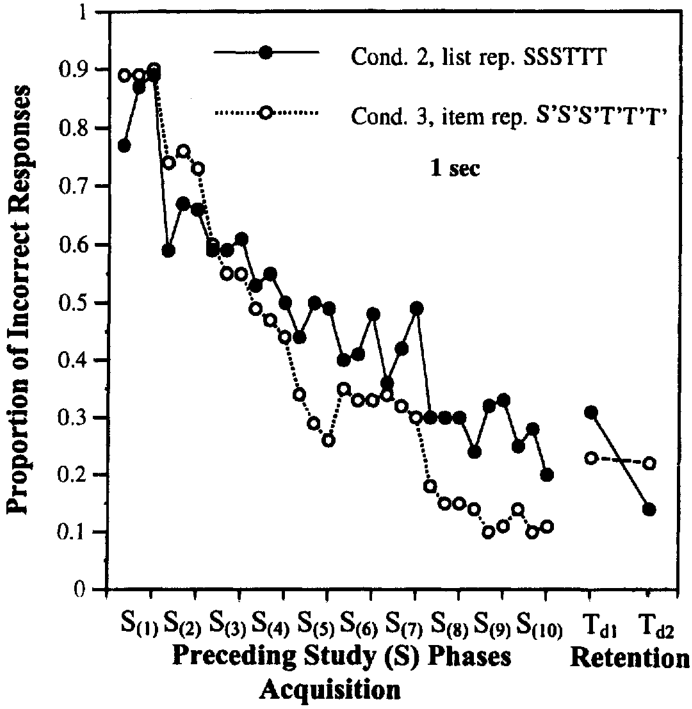


FIG. 5.4. A comparison of the S+T list- and S'+T' item-repetition programs in learning performances between Conditions 2 and 3, with the total S/S' (study), T/T' (test), and R (rest) times as well as presentation durations and frequencies held constant, respectively.

Study/Test (S/T) Alternation Program

See Fig. 5.5. Unlike very difficult nonsense syllable learning tasks, this shows an intermediately difficult/easy German language learning situation with some floor effects, the two S/T alternation conditions could not be differentiated, and did not reach significance.

These findings are consistent with both TTH and S-T-R presentation program hypotheses (in an intermediately easy/hard learning situation), but inconsistent with theoretical predictions from the presentation duration hypothesis that predicted significantly better performances for Condition 1, the control. Current nonsignificance also contradicts the presentation frequency hypothesis and general learning theories: Both of these expected that performance on experimental Condition 4 would be better given that the frequency per S trials were 3 times greater than that of the control, Condition 1.

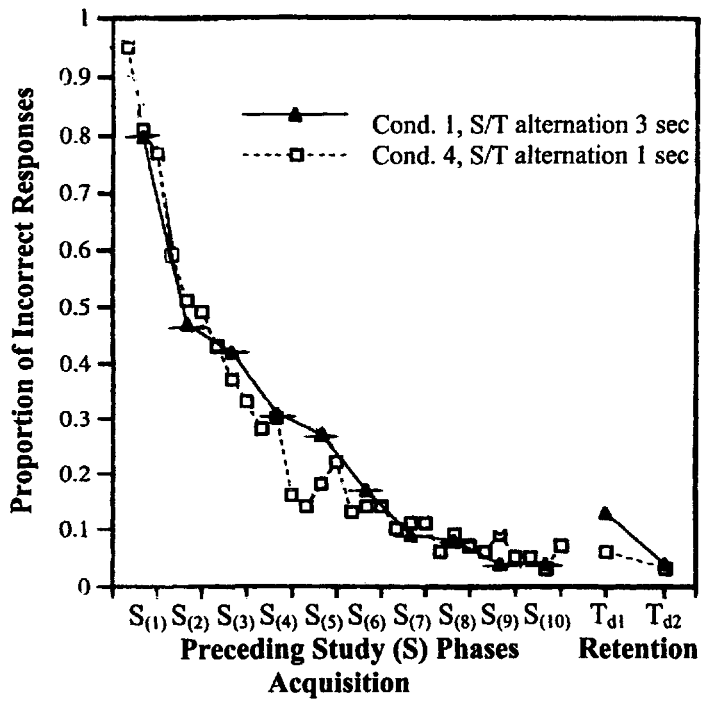


FIG. 5.5. S/T alternation program performances in Conditions 1 and 4, with the total S (study), T (test), and R (rest) times held constant, respectively.

Simultaneous Comparison of All Three Presentation Programs

To identify optimal learning and retention strategies, we compared all experimental conditions: S+T list- and S'+T' item-repetition and S/T alternation programs simultaneously in Fig. 5.6. A strong test can be made for each of the TTHs, presentation duration, and frequency hypotheses, together with the general learning theories. Each predicted identical outcomes because the three conditions all had identical total S+T+R times, identical presentation duration and frequency, along with an identical number of S trials (*N*), respectively.

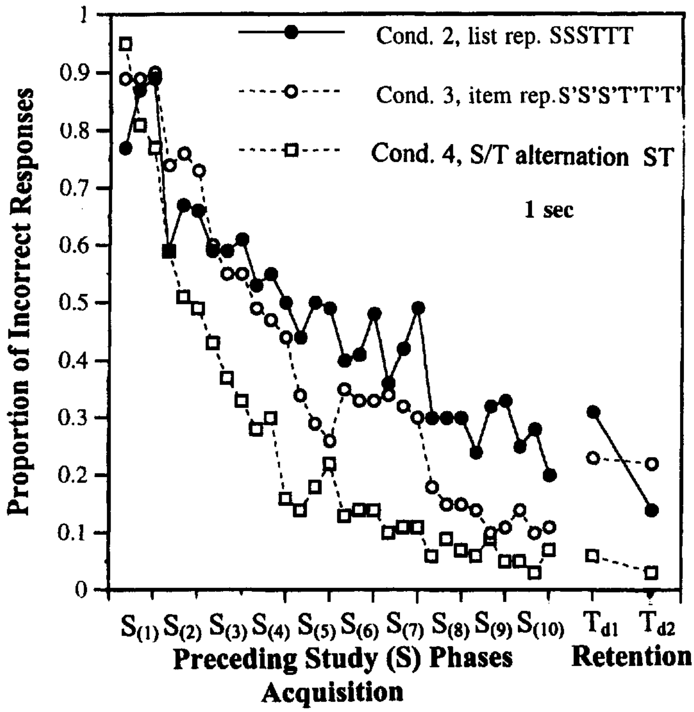


FIG. 5.6. Simultaneous comparisons of all three repetition programs as to learning performances in the experimental conditions: The S+T list-repetition program (Cond. 2), the S'+T' item-repetition program (Cond. 3), and the S/T alternation program (Cond. 4, target T*, Table 5.1) with identical presentation durations (1 sec.) and frequencies (30 Ss and 30 Ts).

The predicted identity of outcomes from each of these three conditions was not fully realized, and substantial superiority in favor of the S/T alternation program condition was obtained, followed by the S'+T' item-repetition, and S+T list-repetition programs. Not only approximating significance during acquisition alone, among the 3 conditions: $F(2, 27) = 2.78$, $p = 0.08$, $MS_e = 298.52$, in spite of dominating floor effects; the S/T alternation program unambiguously surpassed the S+T list-repetition program: $F(1, 18) = 4.43$, $p < 0.05$. Furthermore, the S/T alternation condition was significantly better than the S+T list- and S'+T' item-repetition conditions in retentio: $F(1, 28) = 10.81$, $p < 0.005$, $MS_e = 5.99$. Also, in more difficult learning situations without floor effects, all three item-presentation programs produced highly significant differences (e.g., Izawa, 2000).

The large advantages of the S/T alternation condition over the other item presentation programs observed here and elsewhere can only be explained by Izawa's S-T-R presentation program hypothesis. Substantial differences (Fig. 5.6) among the conditions with an identical total S, T, S+T, and R times/trials, as well as an identical presentation duration and frequency, are decidedly contrary to the TTH, presentation duration and frequency hypotheses, as well as general learning theories.

GENERAL DISCUSSION AND CONCLUSIONS

The current investigation clearly dispelled concerns that the unlearned natural language materials (real foreign language learning situations at school) may be fundamentally different from the unlearned artificial words, such as nonsense syllables. Overall, allowing for minor deviations, the two did not differ in the quality of their influence on the learning process. The present data on German language learning (moderately challenging) replicated, to a major extent, earlier results on the learning of nonsense syllables (e.g. Izawa, 1999, 2000).

Efficient learning is indeed possible under the identical number/ frequency of S trials or identical durations of S, T, S+T, and R trials, largely rejecting the utility of general learning theories, TTHs (original and modified), and S/S', T/T', S+T/S'+T' and S/T alternation presentation duration and frequency hypotheses. Instead, current data fits best with Izawa's S-T-R presentation program hypothesis that predicts efficient learning is largely the function of how Time X is programmed: It is the programming/scheduling of Time X and repetition programs/methods that produces Y-, Y, or Y+ amount of learning, under the constant time X. Consistent affirmation of the S-T-R presentation program hypothesis, in spite of a very small range of variations in the present investigations, strongly reinforces the value of the S, T, and R presentation programs in their control of efficient learning and retention.

The present findings give hope that learning in schools may be managed to enhance optimal learning by challenging educators and students to program their time efficiently.

The current investigations of moderately challenging natural language learning (as well as moderately difficult/easy nonsense-syllable learning situations in many of Izawa's earlier studies), concluded that the S/T alternation programming (both Conds. 1 & 4, Fig. 5.5; see also Figs. 5.1 & 5.6) was optimal. An advantage of a slightly greater S density program such as Program SST seems limited to extremely difficult learning situations (e.g., Izawa, 1968, 1970), which may not generally apply in the current moderately easy/difficult learning situation.

The S/T alternation (STSTST...) programming here had advantages in the following ways: (a) It enjoys the greatest S/T interactions (including faster feedback on the participants' correctness of the response on the S cycle immediately following each T cycle), (b) all S trials are spaced (positive S distributed practice, e.g., Estes, 1955; Izawa, 1971a), and (c) all T trials are also spaced (positive T distributed practice, e.g., Izawa, 1992), while some S and T trials are given in massed practice in list- and item-repetition programs. All of these positive effects contribute to better acquisition and retention (e.g., Izawa, 1992, 1993b).

That, therefore, appears to be sufficient reason to recommend that moderately challenging routine school situations such as natural language acquisition and other learning be programmed in a way that the students schedule time for studying and testing alternately as STSTST... sessions, instead of giving blocks of studying/learning as SSSSSS... sessions alone, as is in fact often done in school settings around the world (cf. Izawa, 1968 for ineffective learning and retention for Program SSSSS...).

Encouraged by the current positive findings with German language learning situation, Maxwell (1998) wished to expand these investigations into Japanese language learning using English-Japanese pairs with the current program design. Similarly, Griffin (2000) who learned Japanese as his second language is ready to launch an experiment to investigate Izawa's S-T-R presentation hypothesis, which he calls "Izawa's efficiency hypothesis" using a Japanese-English learning situation. He will use Izawa's (1993a) design, with 175 participants. We look forward to the data from these experiments. At this juncture, we surmise that replications of our current results will be likely. However, there may be surprises, since the Japanese language is considered to be linguistically unique and hard for English speaking students to master.

SUMMARY

Enhancing efficiency of acquisition and retention is our cardinal concern for applied school settings and elsewhere, and is best guided by sound theories. The seven theories evaluated include a large family of general learning theories, Bugelski's TTH, and Izawa's S-T-R presentation program, presentation duration and frequency hypotheses, plus Whitten's test trial learning/S=T hypothesis, as well as the hypermnesia view.

Data were collected while holding total S, T, S+T, and R times constant, in a moderately challenging (medium difficulty) natural German language learning situation in Conditions 1 (ST, 3 sec.) and 2 (SSSTTT, 1 sec.) under the S+T list-repetition program, Condition 3 (S'S'S'T'T'T', 1 sec.) under the S'+T' item-repetition program, and Condition 4 (S/T, 1 sec.) under the S/T alternation program. Condition 1 at 3 sec. presentation rate is a special case that represents all three repetition programs simultaneously, and thus served as the control for all other (experimental) conditions at a 1 sec. presentation rate.

Forty undergraduates, unfamiliar with German, 10 per condition, volunteered to learn a list of 10 English-German paired-associates. Current natural language learning data best supported Izawa's S-T-R (study-test-rest) presentation program hypothesis, but six alternative theories did not fare as well. Overall large differences among conditions were found in both short-term acquisition and long-term retention data in favor of the S/T alternation program. Thus, it is recommended that students' study time be interspersed with tests as: STST... schedules, or via an S/T alternation program in moderately challenging school learning situations.

NOTES AND ACKNOWLEDGMENTS

This chapter was written in honor of Bob Solso, a prominent psychologist and Professor at the University of Nevada, Reno, USA, a former president of the Western Psychological Association (WPA), and a friend of many Tic4 participants. Bob was really looking forward to his trip to Tic4 in Tsukuba, but he became seriously ill shortly before his flight to his favorite country, Japan. In this way, he joined all of our invited speakers in spirit.

Supported by the Japan-Tulane Friendship Award grant, underwritten by Peter Aron and the Aron Foundation, preliminary data analyses of the current investigation were reported at the November 1999 Psychonomic Meeting in Los Angeles. The authors are indebted to Leith Edgar in whose competent hands lay the final proofing, indexing, and illustration modifications. Salvador Iloreta made extensive contributions to the general proofing process and K. J. Yuan provided expert assistance in the computer related phases of the project. Jeannie Stier McGuire made the general clerical work for the chapter seem deceptively easy. Their devotions and conscientious assistance made completion of this chapter in a timely and efficient manner possible.

Direct all correspondence to: Chizuko Izawa, 411 Pine Street, New Orleans, LA 70118-3715, USA. E-mail: Izawa@tulane.edu.

RERERENCES

Ballard, P. (1913). Obliviscence and reminiscence. *British Journal of Psychology Monograph Supplements, 1*, 1-82.

Bugelski, B. R. (1962). Presentation time, total time, and mediation in paired associate learning. *Journal of Experimental Psychology, 63*(4), 409-412.

Ebbinghaus, H. (1885). *Über das Gedächtnis: Untersuchungen zur experimentellen Psychologie.* Leipzig, Germany: Duncker & Humbolt.

Estes, W. K. (1950). Toward a statistical theory of learning. *Psychological Review, 57*, 94-107.

Estes, W. K. (1955). Statistical theory of distributional phenomena in learning. *Psychological Review, 62*, 369-377.

Griffin, L. J. (2000). *Generalizability of the efficiency (S-T-R) hypothesis to Japanese language learning.* Unpublished thesis, Tulane University.

Hull, C. (1943). *Principles of behavior.* New York: Appleton-Century-Crofts.

Izawa, C. (1966). Reinforcement-test sequences in paired-associate learning. *Psychological Reports, Monograph Supplement, 3-V18*, 879-919.

Izawa, C. (1967). Function of test trials in paired-associate learning. *Journal of Experimental Psychology, 75*, 194-209.

Izawa, C. (1968). Effects of reinforcement, neutral and test trials upon paired-associate acquisition and retention. *Psychological Reports, 23*, 947-959.

Izawa, C. (1969). Comparison of reinforcement and test trials in paired-associate learning. *Journal of Experimental Psychology, 81*, 600-603.

Izawa, C. (1970). Optimal potentiating effects and forgetting-prevention effects of tests in paired-associate learning. *Journal of Experimental Psychology, 83*(2, Pt. 1), 340-344.

Izawa, C. (1971a). Massed and spaced practice in paired-associate learning: List versus item distributions. *Journal of Experimental Psychology, 89*, 10-21.

Izawa, C. (1971b). The test trial potentiating model. *Journal of Mathematical Psychology, 8*, 200-224.

Izawa, C. (1981). Toward a quantitative theory for performance differences between anticipation and study-test procedures: The retention interval model. *Scandinavian Journal of Psychology, 22*, 79-91.

Izawa, C. (1985a). A test of the differences between anticipation and study-test methods of paired-associate learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 165-184.

Izawa, C. (1985b). The identity model and factors controlling the superiority of the study-test method over the anticipation method. *Journal of General Psychology, 112*(1), 65-78.

Izawa, C. (1988). A search for the control factors of the repetition effect. *Japanese Psychological Review, 31*(3), 367-403.

Izawa, C. (1989a). A test of the identity model: Encoding processes differ little between anticipation and study-test methods. In C. Izawa (Ed.). *Current issues in cognitive processes: The Tulane Flowerree symposium on cognition.* (pp. 210-245). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Izawa, C. (1989b). Similarities and differences between anticipation and study-test item information presentation methods. In C. Izawa (Ed.). *Current issues in cognitive processes: The Tulane Flowerree symposium on cognition.* (pp. 201-209). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Izawa, C. (1992). Test trials contributions to optimization of learning processes: Study/test trial interactions. In A. Healy, S. Kosslyn, & R. M. Shiffrin (Eds.), *Essays in honor of William K. Estes: From learning processes to cognitive processes* (Vol. 2, pp. 1-33). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Izawa, C. (1993a). Efficient learning: The total time, exposure duration, frequency, and programming of the study phase. In C. Izawa (Ed.), *Cognitive psychology applied* (pp. 43-78). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Izawa, C. (1993b). Power behind the scenes: Hidden effects of test trials unveiled. In C. Izawa (Ed.), *Cognitive psychology applied* (pp. 79-110). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Izawa, C. (1999). Efficiency in short-term memory: Study-test-rest presentation program and learning difficulty. In C. Izawa (Ed.), *On human memory: Evolution, progress, and reflections on the 30th Anniversary of the Atkinson-Shiffrin model* (pp. 165-213). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Izawa, C. (2000). Total time and efficient time management: In search of optimal learning and retention via study-test-rest presentation programs. *American Journal of Psychology, 113*(2), 221-248.

Izawa, C., & Hayden, R. G. (1993). *Race against time: Toward the principle of* *optimization in learning and retention*. In C. Izawa (Ed.), *Cognitive psychology applied* (pp. 15-42). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Maxwell, S. (1998). *Presentation programs and meaningfulness elements in foreign language acquisition*. Unpublished honors thesis, Tulane University.

Payne, D. G. (1987). Hypermnesia and reminiscence in recall: A historical and empirical review. *Psychological Bulletin, 101*(1), 5-27.

Roediger, H., & Challis, B. (1989). Hypermnesia: Improvements in recall with repeated testing. In C. Izawa (Ed.). *Current issues in cognitive processes: The Tulane Flowerree Symposium on Cognition*. (pp. 175-200). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Whitten, W., & Bjork, R. (1977). Learning from tests: Effects of spacing. *Journal of Verbal Learning and Verbal Behavior, 16*(4), 465-478.

Whitten, W., & Leonard, J. (1980). Learning from tests: Facilitation of delayed recall by initial recognition. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(2), 127-134.

Chapter 6

Optimizing the Speed, Durability, and Transferability of Training

Alice F. Healy, James A. Kole, Erica L. Wohldmann, Carolyn J. Buck-Gengler, James T. Parker, and Lyle E. Bourne, Jr.
University of Colorado at Boulder, U. S. A.

Many educational, military, and industrial organizations spend much time and many resources in personnel training. Because training is costly, it is important to ensure that it be accomplished as quickly and as efficiently as possible. However, increasing training speed should not be the only consideration. If individuals have successfully learned how to perform a task during training but then forget how to perform it at the time that they need to do so, the training clearly is inadequate. Training also needs to be durable. But even durable training cannot guarantee that learned knowledge and skills will be applied successfully to situations different from those encountered during training. It is, therefore, essential that training also be transferable. It is the aim of our research program to develop principles that optimize separately and in combination the three major aspects of training: (a) its speed, (b) its durability, and (c) its transferability to new situations.

In some of our earlier studies, we discovered that training that minimizes the time to acquire knowledge or skills may be detrimental to long-term retention (e.g., Schneider, Healy, & Bourne, 1998, 2002; Schneider, Healy, Ericsson, & Bourne, 1995). In other words, what is learned quickly often is not what is remembered best. Likewise, in other studies, we found that training that maximizes long-term retention of material may severely limit the transferability of that material (e.g., Healy & Bourne, 1995). For example, being able to retain a fact does not guarantee that its relevance will be recognized in new situations. Thus, in designing an effective instructional program, the goal should be to optimize *simultaneously* all three characteristics of eventual performance

– speed, durability, and transferability of training – taking into account the tradeoffs among them. Simultaneous optimization would not necessarily optimize any one characteristic individually, but would require instead a balanced consideration of them all.

The balance of the three aspects of training is not necessarily fixed across tasks or even within a given task but rather may depend on a variety of external factors, such as stress, frustration, fatigue, speed of information presentation, and information load, that can change from time to time. Variations in any factor can affect the interaction of these aspects of training.

The studies summarized in this chapter illustrate our recent work on two topics. The first topic is managing factual overload, rapidly presented information, stress, frustration, and fatigue, with an emphasis on tasks involving perceptual and motoric processing. The experiments reported on this topic focus on the specific issue of initiating and executing response components under fatigue produced by prolonged work. The second topic is optimizing the balance of the three major aspects of training. The study reported on this topic focuses on the specific issue of ways to promote transfer of training.

Initiating and Executing Response Components Under Fatigue Produced by Prolonged Work

We have been exploring the underlying causes of durability and specificity of skill using the data entry task, in which participants see numbers and type them into a computer. Fendrich, Healy, and Bourne (1991) showed that individuals type numbers that have been typed previously during training significantly faster than numbers that, in the context of the laboratory experiment, are "new." This advantage for typing "old" numbers, known as the repetition priming effect, is based largely on implicit or nondeliberate memory. We found that both perceptual and motoric processes contribute to the repetition effect. Buck-Gengler and Healy (2001) demonstrated that the abstract numerical concept, rather than the surface percept, contributes to repetition priming. To manipulate the surface percept, the presentation format of the numbers was varied. Half of the numbers were presented symbolically as words (e.g., "two four one seven"), whereas the other half were presented as numerals corresponding to the labels on the data entry keys (e.g., "2 4 1 7").

Thirty-two college students were trained in the data entry task over a single session and then tested in a second session after a 1-week delay. In both sessions, four-digit numbers were presented one at a time on a computer display screen until the participant responded by typing the number onto the computer keyboard followed by the "enter" key. The presentation and subsequent entry of a four-digit number constituted one trial. Total response times for entering all four digits in a number and the final enter key as well as the accuracy of the entries were recorded on every trial. During the training session, a set of 64 four-digit numbers was used as the learning set. This set was repeated for entry by the participant five times across five blocks of 64 trials during the training session, with a different order of numbers for each block.

The procedure in the testing session was the same as in the training session except that there were only two blocks of trials. Sixty-four "new" four-digit numbers not previously entered by the participants were intermixed with the 64 "old" numbers of the learning set. For half of the old numbers, the presentation format was the same in the testing session as in the training session, and for the remaining half the presentation format in training and testing were different.

One group of participants was trained with the digit keypad to the right of the letter keys on the standard keyboard and was tested with the digit row above the letters on the keyboard. A second group of participants was trained with the digit row and was tested with the keypad. Thus, for both groups, the key configuration used in training did not match the key configuration used in testing to ensure that any repetition effect could not be attributed to the retention of the motoric component of the task.

Numbers presented as numerals were typed more quickly than those presented as words. Also, old numbers were typed significantly faster than new numbers, reflecting repetition priming. There was no overall difference for old numbers between those in the same format and those in different formats at test and at training. However, for both test presentation formats, numbers presented as words in training were faster at test than those presented as numerals, suggesting that having to encode numbers from words in training led to more processing than encoding them from numerals.

In Healy, Kole, Buck-Gengler, and Bourne (2003; also reported by Healy, Buck-Gengler, Kole, & Bourne, 2001), we used a variation of this method to assess the effects of fatigue on data entry performance, to see whether any negative effects of fatigue could be reduced, and to evaluate whether fatigue affects the magnitude of repetition priming.

When individuals work at a continuous task, such as data entry, two opposing processes might affect accuracy and response time. First, the individuals might improve, becoming more accurate, faster, or both in their responses as they master the required skills. Second, one or both aspects of performance might deteriorate as the individuals suffer the effects of fatigue and boredom over long trial periods.

To examine these effects, in Experiment 1 of Healy et al. (2003), there were two sessions in which 32 participants were given strings of four-digit numbers always viewed as numerals displayed on a screen, and they always entered the numbers by typing them on the keypad at the right of the keyboard. In the first session, which was twice the length of that used in the Buck-Gengler and Healy (2001) study, participants saw 64 numbers repeated in different random orders five times over five blocks in the first half of the session, and they saw five blocks of 64 different numbers in the second half. In the second session 1 week later, there were four blocks of 64 numbers; in each block half of the numbers were repeated from each half of the first session, and half were new. Only right-handed individuals participated, and to promote fatigue, they were required to use their left hand. Also to promote fatigue, the 1/2-s intertrial delay between response and presentation of the next stimulus was eliminated. A short break, up to 5 min in length, halfway through the training session (after the fifth block) allowed participants to have some level of recovery from fatigue.

We found that accuracy significantly decreased from the first to the second half of the session and across blocks within each session half (Table 6.1). This finding documents the fact that we successfully induced fatigue during training. Nevertheless, we also found that total correct response time significantly decreased (i.e., improved) from the first to the second half of the session and across blocks within each session half, suggesting that fatigue and practice combined to lead to an increase in the speed-accuracy tradeoff during training.

This experiment yielded significant repetition priming at test 1 week later (i.e., an advantage for old numbers previously entered relative to new numbers) for the measure of total correct response time. There was also a significant interaction of training half and test half on accuracy at test for the old numbers previously entered during training. Accuracy was lowest for those old numbers occurring in both the second half of training and the second half of the test. This finding is consistent with the simple hypothesis that accuracy deteriorates with fatigue, so accuracy should be worst when the participants' state of fatigue is high during both training and testing.

TABLE 6.1
Accuracy (Proportion Correct) and Response Time (RT) in ms (Total RT, Initiation Time, Execution Time, and Conclusion Time). During Training in Experiments 1 and 2 of Healy, Kole, Buck-Gengler, and Bourne (2003) by Session Half and Block

	First Half					Second Half				
Measure	1	2	3	4	5	1	2	3	4	5
Experiment 1										
Accuracy	.920	.895	.893	.880	.878	.881	.872	.865	.857	.857
Total RT	2718	2699	2688	2657	2635	2644	2607	2543	2573	2537
Experiment 2										
Accuracy	.889	.896	.875	.889	.873	.880	.875	.870	.876	.863
Total RT	2498	2500	2527	2454	2445	2372	2340	2378	2354	2384
Init time	1108	1079	1098	1069	1085	1058	1068	1095	1079	1124
Exec time	326	336	333	326	324	306	301	302	303	297
Conc time	282	282	284	280	275	265	257	254	253	249

Note. Init = Initiation; Exec = Execution; Conc = Conclusion. Execution time is the average time per keystroke for the second, third, and fourth digits. Means of medians are provided; thus Total RT is not necessarily equal to the sum of the component RTs.

At the beginning of this study, two opposing ways in which performance might be affected by deliberate practice were proposed. One possibility was that performance should improve, becoming more accurate, faster, or both as participants learned the task. Alternatively, performance might deteriorate, becoming less accurate, slower, or both as participants suffered from the effects of fatigue. We found that practice showed opposite effects on speed and accuracy, reflecting an increasing speed-accuracy tradeoff. With practice, responses became increasingly faster, but also increasingly less accurate.

Experiment 2 of Healy et al. (2003) focused on these findings concerning the increasing speed-accuracy tradeoff with training. The improvement in response times across blocks during training in Experiment 1 could be due to either general or specific training factors.

In terms of general factors, the improvement could be due simply to practice with the task. In terms of specific factors, the improvement could be due to within-session repetition priming because each number was repeated five times, once in each of the five blocks within a given half of the training session. To isolate the effects of practice and fatigue and to eliminate the effects of repetition priming within training, in Experiment 2, each number occurred only once, with no numbers repeated during training. Also, to determine whether the speed-up in response time and the decline in accuracy across session halves were due in part to peripheral motoric factors involving the specific hands, half of the participants (all of whom were right-handed) switched the hand they used to type from the first to the second half of the session. This *switch* condition was compared to a *no switch* condition in which the same hand was used for typing in both halves. We counterbalanced the hands employed, so that half of the participants in both switch and no switch conditions used their right hand during the first half of the session, and the remaining half of the participants used their left hand. To insure that participants used the hand assigned to them for a given session half, they wore socks on the unassigned hand. Participants received a 5-minute break between session halves. Thirty-two participants were tested individually in a single session.

We found that, as in Experiment 1, accuracy declined across session halves overall (Table 6.1), even though the right hand was used on half of the trials and half of the participants switched from one hand to another halfway through the session. This finding implies that the effect of fatigue is not limited to peripheral motoric processes involving the specific hands but has some central, cognitive component.

In contrast, as in Experiment 1, we found that total correct response times significantly decreased (i.e., improved) across session halves, even though any effects of within-session repetition priming were eliminated in this experiment. We also found a significant interaction of session half and block: There was a general decline in total response time across the five blocks in the first half of the session but no consistent change across the five blocks in the second half of the session.

The speed-up across session halves depended on both switch condition and the hand used in the first half because participants were usually faster with their right hand. The decrease in total response time across session halves was greatest for the participants who switched from using their left hand in the first half of the session to using their right hand in the second half, and there was a small increase in total response time across session halves for those who switched from right hand to left. Importantly, finding the average speed-up across session halves comparable overall for the switch condition and for the no-switch condition implies that, like fatigue, any learning responsible for the speed-up is not limited to peripheral motoric processes involving the specific hands, but has a central, cognitive component.

We also examined component response times, and we found different effects of practice on the different components (Table 6.1). Most interesting is the fact that initiation time (the time to enter the first digit, including time to encode the number) showed a significant interaction of session half and block; initiation times generally got faster across the first five blocks, but got slower across the second five blocks. This pattern suggests that for the relatively difficult and time-consuming process of encoding, in the second half of the session the effect of fatigue overcame the effect of practice found in the first half. In contrast, there was no consistent change across blocks in either session half for execution time (the average time to enter the second, third, and fourth digits) although there was a speed-up across session halves. Finally, there was a strong speed-up in conclusion time (pressing the final "enter" key) not only across session halves but also across blocks in both session halves. This pattern leads to the unintuitive suggestion that for purely motoric processes, there may be no effect of fatigue and only an improvement due to practice.

Experiment 2, therefore, supports findings from Experiment 1 of a changing speed-accuracy tradeoff as a function of practice and indicates that the speed-up with practice cannot be attributed to within-session repetition priming. Further, Experiment 2 illustrates that fatigue affects the component processes of the data entry task differentially and at different points in time, with fatigue having its largest effect on initiation time, which is the most cognitively demanding component.

In the basic version of the data entry experiment, participants are allowed to use any means they wish to remember the number, including subvocal or vocal phonological rehearsal, thus activating the phonological loop of working memory. In Kole, Healy, and Buck-Gengler (in press), we determined whether articulatory suppression, which would disrupt this means of rehearsal, would thus alter performance on this task. We conducted another variant of the initial Buck-Gengler and Healy (2001) data entry experiment to assess this issue, by repeating it with 32 new participants and one important change: Half the participants were in an articulatory suppression group, in which they repeated the word "the" continuously while they typed the digits in both sessions, and the remaining participants were in a silent group in which they entered the digits silently, with no secondary articulatory suppression task.

We found that during training, accuracy decreased for the articulatory suppression group relative to the silent group, but only for numbers presented as words. The same interaction was also significant for total correct response time. In this case, however, response times actually *improved* (got faster) under articulatory suppression, but now only for numbers presented as numerals. An explanation for this unexpected finding is that under articulatory suppression, participants go directly from the printed numeral to the typing response, without subvocalizing the number. Such a strategy could reduce response time but not affect response accuracy. Thus, considering each presentation format (words or numerals) separately, there was no speed-accuracy tradeoff. However, speed and accuracy did show complementary patterns: Articulatory suppression had negative effects on accuracy for words and had positive effects on speed for numerals.

There was a change in the response time pattern as a function of training block. As in Healy et al. (2003), all participants showed improvement across training blocks; however responses to words presented under articulatory suppression improved at a faster rate than did the other combinations of conditions (Table 6.2). This finding is presumably due to the fact that across blocks, participants under articulatory suppression learned to use a more efficient, non-phonological strategy for the word format.

Total response times for old numbers were significantly faster than those for new numbers at test 1 week after training, demonstrating repetition priming. This effect was evident only for numbers presented as words at test, not for those presented as numerals at test. The same pattern was found for initiation time. Further, for initiation times, there was an advantage at test for numbers presented as words during training, as found by Buck-Gengler and Healy (2001), resulting in an interaction of presentation format by format continuity for the old items at test. Thus, whenever participants practiced during training typing a given four-digit number presented as a sequence of words, they were faster at test 1 week later than when they practiced during training typing the same number presented as a sequence of numerals, independent of the test presentation format. Importantly, the advantage for numbers presented as words at training was significant for the silent group, but not for the articulatory suppression group. This finding implies that at least some of the advantage here might be due to subvocalization or use of the articulatory loop.

TABLE 6.2
Total Response Time in ms in Experiment by Kole, Healy, and Buck-Gengler (in press) During Training by Group, Presentation Format, and Block.

	Block				
Presentation Format	1	2	3	4	5
Articulatory Suppression					
Numeral	2497	2449	2361	2283	2264
Word	3478	3305	3205	3104	2958
Silent					
Numeral	2699	2622	2564	2542	2492
Word	3330	3233	3145	3105	3082

Initiation times also yielded a surprising main effect of group. Participants who completed the task in the articulatory suppression group initiated trials significantly *faster* at test than did participants who completed the task in the silent group. In contrast, when we examined execution time, we found faster times for the silent group than for the articulatory suppression group, with the disadvantage for articulatory suppression greater for words than for numerals. Hence, once again, we found a different pattern of results for the different response time measures. In this case, the pattern seems to imply that the response for the initial digit of the four-digit number can be based on visual input alone, without any phonological code, but that the responses for the subsequent digits do seem to rely on phonological coding, presumably in order to maintain those digits in working memory. These results underline our findings from and our interpretation of the fatigue study that the various response time components reflect different underlying processes. Further, these results suggest that the effects of adding a secondary task cannot simply be described as lowering performance. The effects are more complex, not always negative, and can even be in some respects performance enhancing.

Ways to Promote Transfer of Training

The second issue to be addressed also involves the effects on performance of adding a secondary task, in this case in an effort to understand ways to promote transfer of training. We have explored many different laboratory and natural tasks to investigate the long-term retention and transfer of knowledge/skills.

To summarize many studies, we formulated the principle of procedural reinstatement (see, e.g., Clawson, Healy, Ericsson, & Bourne, 2001; Healy et al., 1992, 1993; Jensen & Healy, 1998), which is related to the earlier concepts of transfer appropriate processing (e.g., Kolers & Roediger, 1984; Morris, Bransford, & Franks, 1977) and encoding specificity (Tulving & Thomson, 1973). According to this principle, to optimize long-term retention, the procedures (i.e., motoric, perceptual, and cognitive operations) required of participants during training must duplicate the procedures used at the retention test. We found that when tasks met this duplication criterion, performance was highly durable over long delay intervals between training and testing. On the other hand, we also found that for such durable tasks, performance was highly specific to the training procedures and did not generalize well even when only minor changes were made in those procedures (e.g., Clawson, King, Healy, & Ericsson, 1995; Healy, Wohldmann, & Bourne, 2002; Rickard, Healy, & Bourne, 1994).

Recently, we addressed this issue with a new task, the estimation of short temporal intervals (Parker, Healy, & Bourne, 2000). This task is a component in many everyday situations, such as when a speaker has to estimate how much time is remaining in a talk. Although time estimation was widely studied in the past, the influences of prior training on this skill have not been fully addressed. Because of participants' prior experience in estimating seconds or minutes, we chose a slightly different fundamental unit of time to obtain a purer assessment of the effects of training the skill of duration estimation; 1 unit equals 783 ms.

In our prospective production estimation task, participants practice estimating six specified intervals of time expressed in these arbitrary units. Participants are not told how long a given unit is; instead they learn this information through training with feedback, although they naturally know that a larger number of units corresponds to a longer duration than a smaller number of units. For example, participants might see "After the beep, estimate 32 units." They would wait until they thought that 32 units had passed and then press the space bar. Then, they would receive feedback like the following: "Your estimate was 29 units. The difference is -3 units." During training, the intervals were presented in six blocks of six trials.

In Experiment 1 of Parker et al. (2000), after training, participants were immediately given a transfer test with no feedback on three types of intervals. Some of the transfer intervals were repetitions of the actual intervals used during training; some were outside of the practiced range, and others were new within the practiced range. Based on procedural reinstatement, we expected that participants would be most accurate during the transfer test on the actual practiced intervals and least accurate on intervals outside the practiced range. Surprisingly, however, there was no consistent effect of interval type.

These findings concerning transfer interval can be understood in terms of the strategy participants used to estimate intervals. According to retrospective self-reports, almost all participants used some method of counting in order to estimate the durations. Thus, it appeared that participants were using a counting strategy that was highly generalizable to other intervals and independent of the actual estimated intervals, although naturally dependent on the fixed size of the fundamental time unit.

We assumed that counting would rely on some method of articulation in order to maintain an accurate representation of the number of units and the pacing between those units. To test this hypothesis, Parker et al. (2000) conducted a second experiment that included three conditions that varied in the articulation required. The no-secondary-task condition was like that used in our first experiment with no concomitant articulation required. The letter condition required participants simply to repeat continuously a random letter given to them at the beginning of each trial, and the alphabet condition required them to recite aloud the alphabet backward by every third letter, also beginning from a random letter given to them at the start of each trial. For example, participants in the alphabet condition might receive the cue "J" and then while estimating the specified interval of 32 units, would say "j, g, d, a, x, u, ..." We expected the counting strategy to be disrupted by this relatively difficult secondary task. No secondary tasks were used during the transfer phase in any of these training conditions. The alphabet task used here requires keeping track of where one is in a sequence of events, and this skill is a component of many everyday situations, such as when a speaker needs to keep track of what he or she has already said and what has to be said next.

We found a large effect of training condition on accuracy during training, with much worse performance for alphabet training relative to training with no secondary task. This finding was consistent with our prediction that this condition would disrupt the usual counting strategy. We also found that performance during transfer was worse in the alphabet condition than in the condition with no secondary task. This finding suggests that the strategies acquired during training were specific to the secondary task, which was not present in any condition during the transfer test. In support of this suggestion, we found that performance on the transfer test was worse than that on the last block of training for both the alphabet and letter conditions, but not for the condition with no secondary task. Removing the secondary task actually hurt performance for those participants trained with the secondary task. This is a unique finding that is surprising in many respects but consistent with our suggestion concerning the specificity of the strategies learned.

As in Experiment 1, Experiment 2 gave no support for the prediction that accuracy would be better overall for actual practiced intervals than for new intervals and better for new intervals within the practiced range than for those outside the practiced range. Thus, these two experiments demonstrated that improvement in estimating intervals during training was not specific to the training intervals. On the other hand, we also found that improvement in estimating intervals during training was specific to the secondary task employed. Thus, participants were worse during transfer than at the end of training when the secondary task used during training was not employed during transfer. These findings can be understood within the procedural reinstatement framework on the assumption that the same procedures, based on the fixed fundamental time unit, are used for the different target intervals, but the procedures differ as a function of the secondary task. It appears that participants used a generalizable counting strategy even with the secondary alphabet task.

The aim of Healy, Wohldmann, Parker, and Bourne (2002) was to provide a confirmation and extension of our unexpected findings concerning secondary task effects and to investigate the long-term retention of the duration estimation skill. We examined training of duration estimation skill as well as retraining after a 1-week delay. In both sessions, there were six blocks of trials with feedback, and each block contained six different intervals to estimate. The intervals we used included three that are relatively low (below 30 s) and three that are relatively high (above 30 s but less than 1 min).

We tested 48 participants in this experiment. There were two different conditions – no switch and switch, depending on whether participants did or did not perform the same task during training and retraining. During both sessions, half of the participants in each condition performed the alphabet task, and the remaining half performed no secondary task. In the previous study, testing always occurred without a secondary task, but retraining occurred with the alphabet task for half of the participants in the present experiment.

We examined two measures of error in this experiment, the proportional absolute error and the proportional relative error. The proportional absolute error is defined as the absolute (unsigned) difference between the estimated interval and the specified interval divided by the specified interval. This measure gives us a normalized assessment of error magnitude. The proportional relative error is defined as the signed difference between the estimated interval and the specified interval divided by the specified interval. This measure is just like the other one but uses signed values rather than absolute values. It provides us an index of response bias. When the estimated interval is longer than the specified interval, there is a positive bias, by this index, and when the estimated interval is shorter than the specified interval, there is a negative bias.

For proportional *absolute* error, we found significant improvement across blocks, especially across the first two blocks, reflecting the effects of practice, and the improvement was greater in Week 1 than in Week 2. For proportional *relative* error, we found that improvement in estimating intervals reflected a decreasing bias to make estimates toward the central tendency of the practiced range. Thus, there was a negative bias for the intervals of high magnitude and, to a lesser degree, a positive bias for the intervals of low magnitude, especially at the beginning of practice.

Of great interest are the changes that occur between the end of training and the beginning of retraining, 1 week later when half of the participants switch tasks. We conducted an analysis comparing performance on Week 1 Block 6 to that on Week 2 Block 1. In terms of proportional absolute error, we found that participants who did not switch tasks showed no change in error across blocks despite the 1-week delay. This finding is consistent with our procedural reinstatement framework and demonstrates remarkable skill durability. Also consistent with the procedural reinstatement framework is the dramatic increase in error across blocks for participants with a switch in tasks, demonstrating remarkable skill specificity.

TABLE 6.3
Mean Proportional Absolute Error in Experiment by Healy, Wohldmann, Parker, and Bourne (2002) During Critical Blocks

Condition	Week and Block Wk 1 Bl 1	Wk 1 Bl 6	Wk 2 Bl 1	Wk 2 Bl 6
Switch A-N	.306	.188	.216	.138
Switch N-A	.218	.100	.273	.190
No Switch A-A	.303	.185	.185	.172
No Switch N-N	.214	.147	.148	.123

Note. Wk = Week, Bl = Block, A = Alphabet, N = No Alphabet

The decline across blocks as a result of the switch in tasks occurred for both switching directions, although the increase was greater when participants switched from the no alphabet to the alphabet task. When participants switched from the alphabet to the no alphabet task, the level of performance during Week 2 Block 1 was equivalent to that during Week 1 Block 1 for the participants who switched from the no alphabet task to the alphabet task (Table 6.3). This result indicates that despite the great improvement from Block 1 to Block 6 in Week 1 for the group who switched from the alphabet to the no alphabet task, that group did not benefit at all from that improvement when it switched from the more difficult alphabet task in Week 1 to the simpler no alphabet task in Week 2. A similar set of results was found by comparing performance on Week 2 Block 1 for the group who switched from the no alphabet task to the alphabet task with performance on Week 1 Block 1 for those who switched from the alphabet task to the no alphabet task.

Nevertheless, by comparing the last block of training during Week 2 with the first block of training during Week 2, it is clear that improvement occurred during retraining for all four groups, especially the groups who switched to a different task during retraining. These groups ultimately reached the same level of performance on their tasks as the two groups who did not switch tasks (Table 6.3).

A similar analysis for the proportional relative error compared Week 1 Block 6 to Week 2 Block 1 to examine performance surrounding the switch in tasks when it occurred. This analysis revealed that the decline in performance for the switch condition relative to the no switch condition at the start of Week 2 was due to an increase in positive bias for low intervals and an even greater increase in negative bias for high intervals.

In summary, Healy, Wohldmann, Parker, and Bourne (2002) demonstrated clearly that improvement of the time estimation skill during training was completely specific to the task used during training. Participants who were trained with the alphabet task as well as those who were not given that task showed considerable improvement during training. After a 1-week delay, participants from those two groups showed perfect retention for the estimation skill that they had acquired if they were in the no switch condition. However, when participants from the switch condition returned 1 week later, they showed an increase in errors, to a large extent when they trained on the easier no alphabet task and retrained on the harder alphabet task but even to some extent when they trained on the harder alphabet task and retrained on the easier no alphabet task. In fact, for both switch groups, performance at the start of retraining was equivalent to that for the same task at the start of training. Thus, we found that when the tasks were the same there was no forgetting across the 1-week delay, and when the tasks were changed there was no transfer across the same delay. In our current research, we are exploring ideas concerning methods that may overcome the lack of transfer and promote generalizability of training for duration estimation.

In theoretical terms, the present results are consistent with the procedural reinstatement principle (e.g., Healy et al., 1992, 1993), assuming that the procedures used are different in the two tasks because duration estimation is integrated with the differing demands in the two tasks. Participants may use a counting strategy in both tasks, but that strategy might involve number counting in the no alphabet task and letter counting in the alphabet task. That is, in the alphabet task, the requirement to say the alphabet backwards by three's becomes integral with the requirement to estimate durations.

In practical terms, these findings highlight the importance of training individuals on the same operations that they will use subsequently. If the operations are changed, then they will start at square one and training will have been totally useless. Thus, for example, in military or certain industrial applications, if a simulator is used for training, the training may not be worthwhile at all, even if the simulator seems generally realistic, if the full set of operations required in the simulator are different from those required in the field. For example, if the simulator lacks specific operations required in the field – such as keeping track of where one is in certain sequences – then training in the simulator may not transfer to the field situation. Therefore, it is important to take into consideration the complete set of field task requirements when developing a training simulator that will effectively prepare individuals for performance in the field.

SUMMARY

Our research program aims to develop principles that optimize simultaneously all three characteristics of training – speed, durability, and transferability of learned knowledge and skills. The studies summarized illustrate our current work on two topics. The experiments reported on the first topic involved a data entry task. They focused on the specific issue of initiating and executing response components under fatigue produced by prolonged work. We found an increasing speed-accuracy tradeoff as a function of practice in the data entry task. As training progressed, participants became less and less accurate but generally faster and faster. Further, we found that fatigue affected the component processes of the data entry task differentially and at different points in time during training, with the relatively difficult encoding component (reflected in initiation times) slowing down earlier during training than the other components and the purely motoric concluding component showing no evidence at all of slowing down. In another data entry experiment, we stressed the participants by requiring simultaneous performance of a secondary task. The secondary task was articulatory suppression, which eliminates phonological coding of the numbers. Again, we found both harmful and beneficial effects of the stressor. Articulatory suppression reduced typing accuracy for numbers presented as words but enhanced overall typing speed for numbers presented as numerals. Further, like fatigue, articulatory suppression had different effects on the various response time components. It led to a speed-up in initiation time but a slow-down in execution time, suggesting that encoding can be based on visual input alone, without any phonological code, but that response execution relies on phonological coding, presumably in order to store the digits after the first in working memory. The implications of these findings for training are that adding stressors to a training regime could be harmful or useful depending on what aspects of the task are most crucial.

The experiments reported on the second topic involved a duration estimation task, which was in some cases coupled with a secondary articulatory suppression task. They focused on the specific issue of ways to promote transfer of training. We found that learning was highly specific to the conditions of training, so that participants learned how to estimate durations in a way that critically depended on whether or not they were required to perform a secondary task. The implication for training is that to be effective, it must incorporate the complete set of transfer task requirements, including any secondary task requirements imposed.

ACKNOWLEDGMENTS

This research was supported primarily by Army Research Institute Contracts DASW01-96-K-0010, DASW01-99-K-0002, and DASW01-03-K-0002 (Alice F. Healy, Principal Investigator; Lyle E. Bourne, Jr., Co-Principal Investigator) to the University of Colorado. Supplemental support was also received from Army Research Office Grant DAAG55-98-1-0214 (Alice F. Healy, Principal Investigator).

We wish to thank Hama Watanabe, who served as Japanese translator of this report during the conference, for her helpful comments and suggestions concerning the oral version of this chapter.

Correspondence concerning this chapter should be addressed to Alice F. Healy, Department of Psychology, 345 UCB, University of Colorado, Boulder, Colorado, 80309-0345. Electronic mail may be sent to *ahealy@psych.colorado.edu*.

REFERENCES

Buck-Gengler, C. J., & Healy, A. F. (2001). Processes underlying long-term repetition priming in digit data entry. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 879-888.

Clawson, D. M., Healy, A. F., Ericsson, K. A., & Bourne, L. E., Jr. (2001). Retention and transfer of Morse code reception skill by novices: Part-whole training. *Journal of Experimental Psychology: Applied, 7,* 129-142.

Clawson, D. M., King, C. L., Healy, A. F., & Ericsson, K. A. (1995). Training and retention of the classic Stroop task: Specificity of practice effects. In A. F. Healy & L. E. Bourne, Jr. (Eds.), *Learning and memory of knowledge and skills: Durability and specificity* (pp. 234-254). Thousand Oaks, CA: Sage.

Fendrich, D. W., Healy, A. F., & Bourne, L. E., Jr. (1991). Long-term repetition effects for motoric and perceptual procedures. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 137-151.

Healy, A. F., & Bourne, L. E., Jr. (Eds.). (1995). *Learning and memory of knowledge and skills: Durability and specificity.* Thousand Oaks, CA: Sage.

Healy, A. F., Buck-Gengler, C. J., Kole, J. A., & Bourne, L. E., Jr. (2001, November). *Effects of prolonged work on data entry speed and accuracy.* Paper presented at the 42nd annual meeting of the Psychonomic Society, Orlando, FL.

Healy, A. F., Clawson, D. M., McNamara, D. S., Marmie, W. R., Schneider, V. I., Rickard, T. C., Crutcher, R. J., King, C., Ericsson, K. A., & Bourne, L. E., Jr. (1993). The long-term retention of knowledge and skills. In D. Medin (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 30, pp. 135-164). New York: Academic Press.

Healy, A. F., Fendrich, D. W., Crutcher, R. J., Wittman, W. T., Gesi, A. T., Ericsson, K. A., & Bourne, L. E., Jr. (1992). The long-term retention of skills. In A. F. Healy, S. M. Kosslyn, & R. M. Shiffrin (Eds.), *From learning processes to cognitive processes: Essays in honor of William K. Estes, Volume 2* (pp. 87-118). Hillsdale, NJ: Erlbaum.

Healy, A. F., Kole, J. A., Buck-Gengler, C. J., & Bourne, L. E., Jr. (2003). *Effects of prolonged work on data entry speed and accuracy.* Manuscript in preparation.

Healy, A. F., Wohldmann, E. L., & Bourne, L. E., Jr. (2002, November). *Specificity of hand-eye coordination: Training and transfer effects.* Paper presented at the 43rd Annual Meeting of the Psychonomic Society, Kansas City, MO.

Healy, A. F., Wohldmann, E. L., Parker, J. T., & Bourne, L. E., Jr. (2002). *Training, retention, and transfer in duration estimation: The procedural reinstatement framework.* Manuscript submitted for publication.

Jensen, M. B., & Healy, A. F. (1998). Retention of procedural and declarative information from the *Colorado Drivers' Manual.* In M. J. Intons-Peterson & D. Best (Eds.), *Memory distortions and their prevention* (pp. 113-124). Mahwah, NJ: Erlbaum.

Kole, J. A., Healy, A. F., & Buck-Gengler, C. J. (in press). Does number data entry rely on the phonological loop? *Memory.*

Kolers, P. A., & Roediger, H. L. (1984). Procedures of mind. *Journal of Verbal Learning and Verbal Behavior, 23,* 425-449.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior, 16,* 519-533.

Parker, J. T., Healy, A. F., & Bourne, L. E., Jr. (2000, November). *Training and transfer of duration estimation: The procedural reinstatement framework.* Paper presented at the 41st annual meeting of the Psychonomic Society, New Orleans, LA.

Rickard, T. C., Healy, A. F., & Bourne, L. E., Jr. (1994). On the cognitive structure of basic arithmetic skills: Operation, order, and symbol transfer effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 1139-1153.

Schneider, V. I., Healy, A. F., & Bourne, L. E., Jr. (1998). Contextual interference effects in foreign language vocabulary acquisition and retention. In A. F. Healy & L. E. Bourne, Jr. (Eds.), *Foreign language learning: Psycholinguistic studies on training and retention* (pp. 77-90). Mahwah, NJ: Erlbaum.

Schneider, V. I., Healy, A. F., & Bourne, L. E., Jr. (2002). What is learned under difficult conditions is hard to forget: Contextual interference effects in foreign vocabulary acquisition, retention, and transfer. *Journal of Memory and Language, 46,* 419-440.

Schneider, V. I., Healy, A. F., Ericsson, K. A., & Bourne, L. E., Jr. (1995). The effects of contextual interference on the acquisition and retention of logical rules. In A. F. Healy & L. E. Bourne, Jr. (Eds.), *Learning and memory of knowledge and skills: Durability and specificity* (pp. 95-131). Thousand Oaks, CA: Sage.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80,* 352-373.

This page intentionally left blank

Chapter 7

Working-Memory Capacity Limits in a Theoretical Context

Nelson Cowan
University of Missouri, U.S.A.

After a brief definition and description of working memory (WM), several theoretical views will be discussed. The view most emphasized is based on the idea that attention is critically important for WM. However, there are several ways in which attention can be used. The present suggestion is that all of the functions of attention are relevant to WM.

A number of specific experimental tasks developed from different theoretical points of view will be examined. This will be done to make the case that the types of tasks that have been used most often to look at individual differences in WM are not as special as has typically been assumed. These tasks correlate well with scholastic abilities but so do other tasks that are suggested as possible alternatives. Correlations from various data sets between different WM measures and various measures of scholastic ability exemplify this point. Finally, conclusions will be reached about the value of various measures of WM.

DEFINITION AND DESCRIPTION OF WM (Working Memory)

Definitions of WM differ (Miyake & Shah, 1999). However, one useful, broad definition is the retention of information in a temporarily accessible form, through all available mental processing mechanisms (Cowan, 1999). That definition has some boundaries. For some theorists, but not others, WM includes the ability to manipulate the information being held in mind.

The importance of WM is that it must be used to complete many intellectual tasks. When we perceive the outside world, WM is needed to extend the perception. For example, imagine that you are walking along the street and reading a sign off to your right. While you are doing so, your WM is retaining a conception of what the world is like to the left, where you are not looking. It retains information about where the street is, how busy the traffic is, and what buildings are found along the opposite side of the street. To take another example, in comprehending speech, it is necessary to remember some of the words that are spoken, or at least the ideas behind those words, long enough for the entire idea to be constructed. WM is similarly important, to hold in mind the assumptions and partial results in calculations, in problem-solving generally, and in various sorts of logical reasoning.

THEORETICAL VIEWS OF WM

Several different views of WM have served as the basis of different kinds of experimentation and tests: a psychometric or modal view, a multicomponent view, a storage-plus-processing view, and an attentional view. According to the traditional, psychometric point of view, it is useful to measure peoples' ability to repeat simple lists of digits or words in the presented order. For example, digit span is a regular part of standardized intelligence tests. It is useful because serial recall of lists of items has much in common with the use of memory in other intellectual tasks. It reflects the storage of needed data in a temporary form and correlates with aptitude.

Baddeley (1986) challenged this approach by adopting what he saw as a multicomponent view. This view holds that there are different types of storage media in the brain, such as phonological storage of speech information on one hand and visual or spatial information on the other hand, and that one must also consider the quality of executive processes that maintain and transform the information. Daneman and Carpenter (1980) and others did not exactly challenge this view, but they concentrated on what kinds of memory tasks might be best if the goal is to measure WM in a way that reveals its quality in an individual. In order to measure the capacity of WM in a meaningful way, they suggested to tie up both its ability to store information and to carry out a processing task.

Although many researchers who accept the multicomponent point of view acknowledge that the type of task that Daneman and Carpenter (1980) pioneered works well in predicting scholastic abilities, not everyone agrees with the explanation involving tying up storage and processing together. For example, Engle, Kane, and Tuholski (1999, and various other articles by Engle and colleagues) suggest the alternative possibility that what is most important about WM tasks developed by Daneman and Carpenter is that they require the intensive use of controlled attention to carry out the task. It may be the control of attention that distinguishes between people with better or poorer WM spans.

In brief, here is what has happened to the different views of WM. The traditional psychometric or modal view is considered disconfirmed because Baddeley (1986 and elsewhere) showed that a multicomponent view is needed. Different types of brain damage, for example, produce different types of WM deficit that can be interpreted as resulting from damage to different parts of his multicomponent system. The multicomponent view is still viable. However, the methods developed from that point of view are generally not designed to study individual differences in the ability to carry out complex cognitive tasks. The storage - plus - processing view (Daneman & Carpenter, 1980) has produced the testing procedures that have become dominant in an examination of individual differences, used in hundreds of studies. Finally, though, the attentional view (Engle et al., 1999) has provided what is now probably the most generally accepted interpretation of how these storage-plus-processing tasks actually work. According to that reinterpretation, their success does not depend on storage plus processing per se, after all. It depends on controlled attention.

This chapter focuses on the attentional view, but within that view, notes several subtypes. Hasher and Zacks (1988) proposed that the ability to inhibit irrelevant information is what is critical and distinguishes between people with good versus poor WM and intellectual abilities. For example, when you recall a list of names you may think of a word that you already said and you must inhibit any inclination to say the same word again. Engle et al. (1999) and others have suggested that this view is too restricted and that any kind of executive function requiring attention is critical to WM. The relevant executive functions can include inhibition but also other functions (Miyake, Friedman, Emerson, Witzki, & Howerter, 2000), such as updating (e.g., during a baseball game if one is trying to keep track of how many balls and strikes there are, which is difficult because it keeps changing) and coordination (e.g., at a restaurant when a waiter tries to keep in mind which diner gets which meal while writing them down). Even keeping in mind the goal of a task that one is doing is a WM function (e.g., Kane & Engle, 2003).

The present approach is not very different from these other investigators. However, it suggests that attention is always critical for WM. In particular, attention can zoom in to processes as little as a single item or zoom out to apprehend up to about 4 independent items (chunks), though not much more than that (Cowan, 2001). However, whereas Engle and his associates have preferred to measure the capability of attention when it is zoomed in, the present view is that it is useful to try to measure performance when it is zoomed out because that may provide a meaningful measure of WM capacity. The emphasis on measuring how many chunks can be held in the focus of attention at one time can be associated with several investigators (Broadbent, 1975; Cowan, 2001; Miller, 1956). Also, the recent concept of the episodic buffer (Baddeley, 2000, 2001) may serve the same purpose as holding items in the focus of attention.

It is worthwhile to illustrate what a wide variety of functions the focus of attention is assumed to carry out. WM measures may be based on any of these functions, or at least may relate to them. The first function is filtering; for example, when one listens intently to one speaker at a party, while tuning out others. Fig. 7.1 shows that within the memory system, some items are in an activated state (represented by the jagged line), but that only some of the activated information is sufficiently active to be the focus of attention (represented by the large circle). The information in the focus of attention receives much more complete information processing. Each small dot represents an activated element of the memory system. Incoming sensory information activates representations in the memory system automatically, and any number of sensory stimuli can do this simultaneously. However, most of this information cannot make it into the focus of attention, which has limited capacity. This figure shows the ability of attention to focus on one stream of stimulation (horizontal, solid line) and to effectively filter out other streams (horizontal, dotted lines) on the basis of their differences in physical characteristics such as voice quality, or color if they are visual stimuli. As shown here (rising dashed lines), pertinent information in the memory system is used rather automatically to help interpret the incoming information in the focus of attention.

Another point that should be added is that this view is not as different as it seems from Baddeley's (1986) WM model. Any function that Baddeley would attribute to the phonological buffer or the visuo-spatial buffer could be attributed here to properties of the activated memory outside of the focus of attention. The differences between models have to do with assumptions about how specialized or general the storage devices are (Cowan, 1999).

Fig. 7.2 illustrates how the focus of attention may be involved in the process of comparing incoming stimulation to representations already in memory. This may occur, for example, when one is checking to make sure that a series of numbers is written down correctly. Fig. 7.3 shows the retention function of attention. For example, in watching a busy, urban street in Tokyo you can examine what a few people are doing at any one time but you cannot observe what, say, 10 people are doing all at the same time. There is a limit to how much independent information can be held at one time. Finally, Fig. 7.4 shows the chunking function of attention, in the spirit of Miller (1956). Information that resides in the focus of attention at the same time tends to be linked together or associated to form a larger chunk. For example, if you attend to the telephone number 7 5 7 8 0 3 2 you may soon focus on 7-5-7 as one chunk, 8-0 as a second chunk, and 3-2 as a third chunk. Then you are able to focus attention on the three chunks together; 757-80-32 and, before long, you have memorized the telephone number. A great deal of learning takes place through chunking.

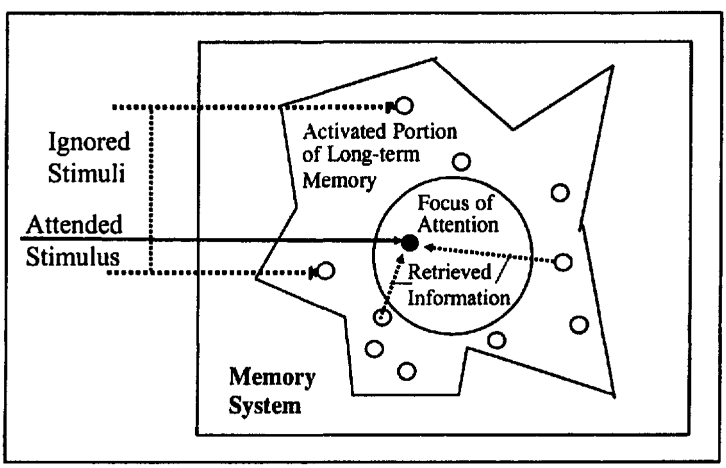


FIG. 7.1. Filtering and interpretation functions of attention (see text for details).

One can use these principles to distinguish between *compound* and *pure* estimates of WM capacity (Cowan, 2001). A compound estimate, exemplified by Miller's estimate of 7 ± 2 items in memory, comes from situations in which one item is presented at a time. The 7 items recalled presumably do not result from 7 separate chunks in memory. As discussed above, 7 items can be combined to form a smaller number of chunks. In contrast, a pure estimate presumably comes from situations in which familiar items are used but they cannot be grouped to form larger, fewer chunks of information. This can occur when information is presented too quickly or with too many items all at once, making chunking difficult. Under these circumstances, it turns out that adult participants recall about 4 items. The convergence of results from many procedures tends to lend support to the theoretical analysis of the tasks. Presumably, each item is a separate chunk in WM in these tasks, which is why they yield similar estimates of capacity.

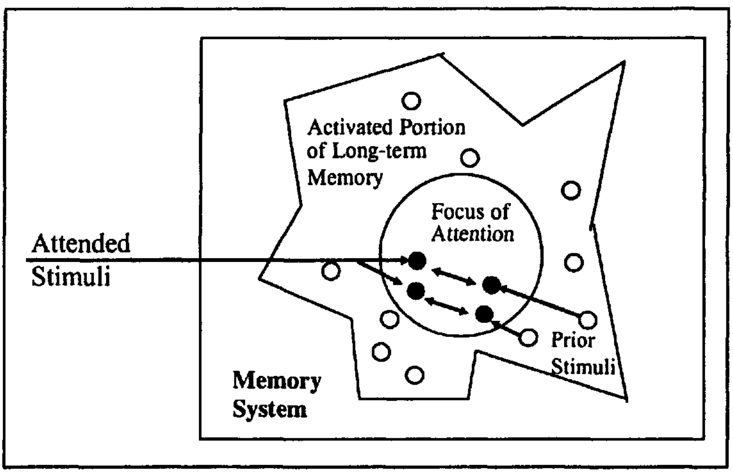


FIG. 7.2. Comparison and updating functions of attention (see text for details).

Cowan (2001) suggested that the focus of attention may be the *only* basis of true capacity limits, conceived as limits in how many chunks of information can be held in WM. Other faculties of WM, such as Baddeley's (1986) "slave systems" or Cowan's (1988, 1995) activated memory, would be limited by other factors such as decay, interference, and temporal distinctiveness but would not have chunk limits per se. Baddeley's (2001) newer, episodic buffer component may be an alternative conception of the chunk limit; it holds information that does not neatly fit into phonological or spatial stores. Yet, an episodic buffer might encounter attention limits at least in information acquisition, if not in its maintenance.

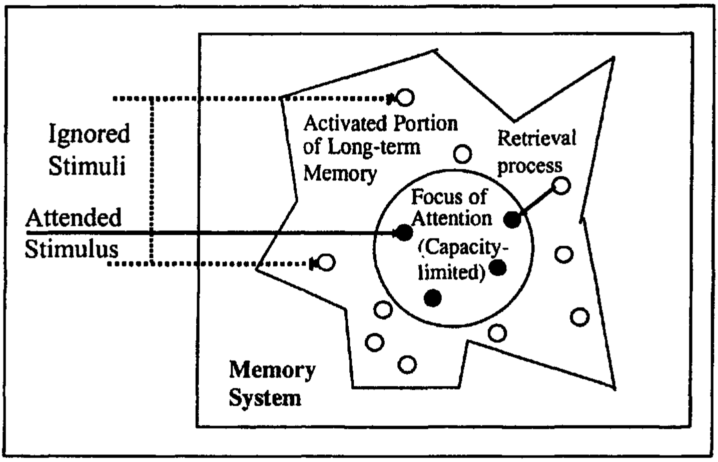


FIG. 7.3. Encoding and maintenance functions of attention (see text for details).

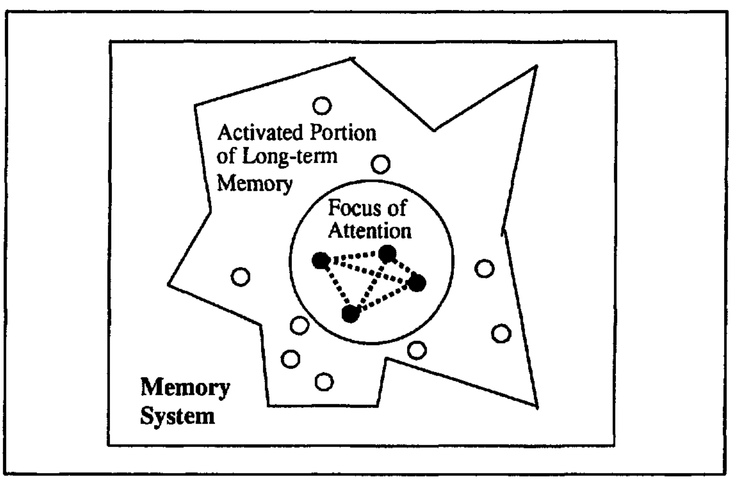


FIG. 7.4. Chunking and learning functions of attention (see text for details).

What is advocated here is that we now need to do more research in which the WM task itself is based on how much information the focus of attention can hold. The actual capacity or retention function of the focus of attention is important to study, for at least three reasons. The first is the validity of the WM concept as a memory concept as opposed to just an attention concept. Only items that are held in attention at the same time can be combined into one new chunk of information, and chunking is the main mechanism of new learning. A second, related reason is psychometric. Only the retention function of the focus of attention may lead to a meaningful numerical estimate of the capacity of WM. The third reason is philosophical. Limits on the contents of consciousness (as discussed by William James, 1890) can be estimated by the number of chunks in the focus of attention.

MEASURES OF THE CAPACITY OF WM

This section will review briefly the common methods of measuring WM span and suggest some alternative measures, which are taken as measures of the capacity of the focus of attention.

Measures Often Used

The types of measures that are used to examine WM from the traditional, psychometric approach and from the storage-plus-processing approach are well known (for a review see Daneman & Merikle, 1996). In the psychometric approach, on each trial, a list of items is presented and must be repeated back in the presented order. The list length grows until a point at which the participant can no longer repeat the lists correctly and span is defined in various ways with reference to the list length (e.g., as the list length at which 50% of the lists can be recalled correctly).

According to the storage-plus-processing approach, there are processing episodes interweaved with items to be recalled, in order, after the last processing episode. For example, in a counting-span task, the participant must count the number of dots on each screen and then recall the sums. In a sentence span task, spoken sentences are presented, whereas in a reading span task, written sentences are presented. Each sentence is evaluated in some way and then the final word of each sentence is recalled or, in another version, a separate word following each sentence is recalled.

In an arithmetic operation span task, an arithmetic operation is carried out and the result is retained in memory for later recall or a separate word is retained for recall. It is the number of presentations of the processing task (displays of dots, sentences, or arithmetic problems) that can be carried out along with correct recall of the task-final memoranda, that defines the WM span. This WM span correlates with complex cognitive task performance better than simple span (Daneman & Merikle, 1996).

Some Possible Measures of the Focus of Attention

In contrast, it is unclear how to measure WM from the theoretical standpoint in which it reflects attentional capacity. One might examine the effects of attention when zoomed in to focus on a goal (e.g., see Engle et al., 1999; Kane & Engle, 2003). However, this does not yield a task-independent estimate of some theoretical quantity. Perhaps, in the future, it will be possible to do so (e.g., to obtain an estimate of the number of seconds for which an individual can keep in mind a goal in the face of, say, a constantly-present competing task). What is proposed here, though, is that we can measure the capacity of WM in terms of the number of chunks that can be held in mind when attention is zoomed out to apprehend as many unconnected items as possible in a currently-relevant array. Methods by which one can do so were suggested by Cowan (2001).

According to the logic of Cowan (2001), there is a reason why it is difficult to obtain a theoretically-pure estimate of WM capacity. When an experimental participant is presented with a stimulus set, one typically does not know how the participant groups the stimulus set into a smaller number of chunks. Therefore, it is not possible to estimate the number of chunks held in WM in a manner that can be compared across stimulus situations. To overcome this problem, Cowan suggested examining situations in which there is good reason to suspect that grouping processes cannot be carried out (e.g., when the participant is engaged in a rehearsal-suppression task during the encoding of a verbal stimulus set). One also must restrict the examination to sets of stimuli that are familiar so that each item is represented in memory initially as an integrated chunk (e.g., studies using words in the participant's native language, but not foreign or nonsense words, would qualify). A wide variety of situations taken to fit these constraints provide estimates of WM capacity of about 4 chunks, with young-adult means in various experimental conditions ranging from about 3 to 5 chunks and individual scores ranging from about 2 to 6 chunks. By implication, these same limits might apply to all WM situations although that assumption cannot be verified in situations in which the chunking processes are unclear.

The striking convergence in the capacity observed in many different procedures was taken by Cowan (2001) to suggest that the analysis of these procedures was correct; that, in these procedures, to a close approximation, each item is retained as a separate chunk in WM, allowing an estimate of the number of chunks in WM. When other mechanisms are allowed to operate (rehearsal, chunking, sensory memory, and automatic forms of storage) the result is presumably a larger number of items retained, as in the ordinary memory span of about 7 items (Miller, 1956).

Cowan (2001) proposed four types of situations that lead to estimates of WM capacity in chunks: (1) when information overload limits chunks to individual stimulus items, (2) when other steps are taken specifically to block the recoding of stimulus items into larger chunks, (3) in performance discontinuities caused by the capacity limit, and (4) in various indirect effects of the capacity limit. Here, however, we focus on measures for which evidence exists, relating them to cognitive aptitude. These include memory for visual arrays, multi-object tracking, running memory span, memory for ignored speech, and conceptual span.

Before describing these measures it is important to note that the capacity limit of about 4 items cannot easily be attributed to the rate of sensory forgetting or the rate of transfer of information from sensory memory once attention is focused on it. Similar limits are obtained no matter whether the items come from briefly presented visual arrays, as in the seminal research with character arrays by Sperling (1960) and the more recent work with color arrays by Luck and Vogel (1997), or auditory arrays, as in the research of Darwin, Turvey, and Crowder (1972). That is true even though sensory memory seems to be useful for a much longer period in the auditory arrays. The present explanation is that the common result reflects a limit on how many independent pieces of information can be held in the focus of attention (or perhaps in an episodic buffer).

Visual-array Measure

The first measure is memory for visual arrays (Luck & Vogel, 1997). An array of randomly arranged colored squares is presented for a half-second or less and a second array is presented shortly afterward, at the same location that the first array was presented. The second array is identical to the first or differs in the color of one square. A cued (encircled) square is the one that may have changed and the required response is to indicate whether it has changed or not. Young adults can carry out the task very well with up to 4 squares per array, then performance levels begin to decline markedly across set sizes. Even at the larger array sizes, a simple formula to correct for guessing shows that people can retain about 4 colors in mind from the first array, to be compared to the second array. A formula that works well is $k = N * [h + c - 1]$, where k is the capacity of WM, N is the set size in the array, and h and c are the probabilities of hits and correct rejections (Cowan, 2001). This formula was calculated by assuming that k items are apprehended from the first array and that, if the cued item is one of those k items, the participant will know whether it has changed color or not; if the cued item is not among the k items, the participant will guess "different" with some fixed rate g (drops out of the final equation). The formula works well in that the calculated k remains relatively constant across set sizes higher than 4, more so than a slightly different formula (Pashler, 1988).

The present interpretation of the visual array task is that items from the first array cannot be retained in an automatically-held form of memory activation that is not limited in capacity per se, such as visual sensory memory. The reason is that the second array presumably overwrites the visual memory of the first array. Luck and Vogel (1997) also showed that a memory load to suppress articulation during the trial has no effect. It is apparently necessary to hold items from the first array in an interference-resistant form, at least momentarily when that array is seen (presumably, in the focus of attention). It is possible that the k items that are apprehended in the focus of attention can be transferred to a form that does not require attention for maintenance; perhaps a form of activated memory such as Baddeley's (1986) visuo-spatial sketchpad.

Multi-Object Tracking

In this procedure (Pylyshyn & Storm, 1988), several dots flash and it is those dots that are to be tracked. When they stop flashing, all of the dots move around randomly but the participant must keep track of which dots had been flashing initially. People can track up to 3 or 4 dots simultaneously. The limit in tracking dots is presumably a limit in how many can be held in the focus of attention at once.

Running-Memory Span

In this task (Pollack, Johnson, & Knaff, 1959), typically using digit stimuli, spoken digits are presented rather quickly and continue until an unpredictable point. At that point, the list ends and the participant must recall a certain number of items from the end of the list. Under these circumstances, it is difficult or impossible to rehearse and group the items. Participants may adopt a passive attitude. Then, when the list ends, they presumably use auditory sensory memory or phonological memory to retrieve some items from the end of the list. This retrieval is limited by the amount that the focus of attention can apprehend from sensory memory. Young adults can remember about 4 items in the correct serial position relative to the end of the list. If the usable phonological memory is assumed to last about 2 s (Baddeley, 1986), it is clear that only about half that amount can be transferred to the focus of attention in the running span task.

Memory for Ignored Speech

In this procedure (Cowan, Lichty, & Grove, 1990; Cowan, Nugent, Elliott, Ponomarev, & Saults, 1999), more direct means are used to prevent rehearsal and grouping of items. Lists of spoken items are presented through headphones, one after another. Meanwhile, participants engage in a task designed to distract attention from the spoken items. Cowan et al. (1999) used spoken digits and a primary visual task in which rhymes are to be formed among the names of pictures that are presented, but without speaking. This task strongly discourages both rehearsal and attention to the spoken digits. Just occasionally, the rhyming game is interrupted and a display is presented on the computer screen, indicating that it is time to try to recall a spoken list that just ended. This can only be accomplished by suddenly shifting attention to a sensory memory trace of the spoken list and transferring as many items as possible into the focus of attention. Young adults recall an average of about 4 digits in their correct serial positions in this ignored-speech task, whereas children recalled fewer. Presumably, there is plenty of sensory memory but the attentional focus is limited. Therefore, regardless of the list length, which ranged from a maximum equal to the longest list the participant recalled in an ordinary span task to a minimum of 3 less than that, the number of digits recalled correctly stayed constant.

It is clear from the results of the ignored-speech task, including various safeguards that were taken, that it powerfully manipulates attention. First, no tradeoffs are found between visual and auditory task performance levels. For example, the rhyming game is carried out no more quickly when it is carried out alone than when there are digits to be ignored, suggesting that attention is not deflected to the digits. Second, the digits that are more frequent in the language, the low digits 1 - 3, are not recalled any better than the high digits 7 - 9, by children or adults. Third, the patterns of performance in memory for attended versus ignored lists of digits look very different from one another. Whereas the number of ignored digits recalled stays roughly constant across list lengths, the number of attended digits recalled climbs steadily across the same lists lengths. Fourth, the age and individual differences in memory for ignored lists of digits cannot be attributed to less forgetting of sensory memory over time in more capable subjects; those at different developmental levels forget the list at roughly the same rate, except for the final digit (Cowan, Nugent, Elliott, & Saults, 2000). It seems to occur instead because more information is transferred from sensory memory to the focus of attention in more capable subjects. These points are explained in more depth by Cowan, Elliott, and Saults (2002).

Some people comment that it is odd to use an ignored-speech task to measure the capacity of the focus of attention. An explanation might help. The point is to restrict the use of attention so that, hopefully, it is only used after the list is presented, to extract information from auditory sensory memory. The logic is similar to that of Sperling's (1960) study of visual sensory memory. It is presumably not possible in these procedures to use attention to group items when they are presented, only to extract information from sensory memory afterward. Sperling obtained a whole-report limit of about 4 items, similar to the tasks highlighted here.

It is worth pointing out that, in the attention-related WM tasks that have been described, what is meant by an "item" in WM is actually a binding between features. In the visual array task, it is a binding between the location of an object and the color, given that a particular color can occur more than once in an array. In the multi-object tracking task, similarly, it is a binding between an object and a present location. In the running span and ignored-speech procedures, it is the binding between a digit and a serial position in the list (in running span, at least, counted from the end of the list). It is assumed here that there is no limit on how many objects, colors, or digits can be in an activated state at one time (Cowan, 1988, 1995, 1999), but that there is a limit in how many feature bindings can be retained. (For a direct demonstration of the latter see Wheeler & Treisman, 2002.)

Conceptual Span

One more potential measure of the capacity of the focus of attention is a conceptual span task developed by Haarmann, Davelaar, and Usher (2003). A list of 9 words from 3 semantic categories is presented randomly and is followed by a cue to recall all of the words from one category (e.g., "*lamp, pear, tiger, apple, grape, elephant, horse, fax, phone,* FRUIT*?* Correct answer*: apple, pear, grape*"). Words are drawn repeatedly from a limited pool and are presented at a rapid rate of one word per second, minimizing the contribution of long-term memorization of the lists. People recalled an average of about 2 to 3 items in the cued category in this task. It is not known exactly how this task is carried out but the results differ in various ways from an ordinary word span; it seems likely that a conceptual structure is held in mind. If participants do not rehearse the phonological sequence as they often appear to do when serial recall is required (Baddeley, 1986), the alternative would be to retain concepts in an active form and to maximize the amount of activation of these concepts by recycling them through the focus of attention.

CORRELATIONS BETWEEN WM MEASURES AND COMPLEX COGNITIVE TASKS

Although it is clear that storage-plus-processing measures correlate with intellectual and scholastic types of aptitude better than do simple spans (Daneman & Merikle, 1996), it is not clear why this difference occurs. Daneman and Carpenter (1980) proposed that it is because only the storage-plus-processing tasks explicitly tie up both storage and processing components of WM. However, another possibility is that the storage-plus-processing tasks represent just one situation in which items cannot be retained through an uninterrupted phonological rehearsal of the memoranda. With rehearsal processes out of the picture, the WM tests may reflect individual differences in how much can be held in the focus of attention or how well the attention processes can function when there is a need to shift from one task to another. If this is the case, it seems worth checking whether the tasks that have been reviewed above, as possible indicators of the capacity of the focus of attention, also will correlate well with aptitudes.

One set of correlations comes from an often-ignored meta-analysis of past research conducted by Mukunda and Hall (1992). The analysis included studies with children and adults, using within-age correlations between spans and various achievement and aptitude tests. The measures that they looked at included measures of conventional span, measures requiring both storage and processing, and one measure that may simply index the capacity of the focus of attention, running memory span. Whereas digit span (based on 53 independent tests) correlated with aptitude tests at a combined R = .22, word span did much better, (9 tests, R = .43). An often-used type of storage-plus-processing span, reading span, produced the expected high correlation (11 tests, R = .43). However, running memory span produced an almost equally good outcome (11 tests, R = .40). In contrast, correlations were lower for two other storage-plus-processing measures, operation span (6 tests, R = .23) and counting span (3 tests, R = .28). Inasmuch as running memory span does not require the verbal or mathematical ability that various storage-plus-processing spans require, but still produces hefty correlations with aptitude, it may be a purer measure of WM capacity.

In two unpublished studies of our own with elementary-school children and college students (with collaborators J. Scott Saults, Emily M. Elliott, Candice C. Morey, & Anna Hismjatullina), we found comparably high correlations between aptitude and achievement measures, on one hand, and measures of the capacity of the focus of attention, on the other hand. The aptitude and achievement measures include the American College Test and high school grades in college students; the Cognitive Abilities Test in children; and in all of the age groups, Stanford-Binet vocabulary and pattern recognition scores, Ravens Progressive Matrices, and the Peabody Picture Vocabulary Test. The latter include in all of the age groups, memory for ignored speech, running memory span, the visual array task, and an auditory analogue of that task. Although storage-plus-processing tasks that require verbal proficiency (specifically, listening and reading span, but not counting span) seem to contribute something extra that is not present within the focus-of-attention measures, it is likely that this part of the correlation is inappropriate and does not truly tap WM processes.

Oberauer, Süß, Schulze, Wilhelm, and Wittmann (2000) carried out a large-scale study of different types of WM tasks and included one task that was suggested as a possible measure of the capacity of the focus of attention, the multi-object tracking task. That task is rather special in measuring the capacity of the focus of attention directly, rather than measuring its mnemonic aftermath. Oberauer et al. did not show the correlations between this task and the scholastic tasks separately, but it was a valid predictor that was combined with other tasks in latent variable analyses. Its correlation with other WM tasks was highest for the spatial WM tasks, ranging from $r = .30$ to $r = .42$. This is a promising task for future correlative work on WM.

Haarmann et al. (2003) compared conceptual span to word span and reading span in terms of their correlations with text comprehension and spoken sentence comprehension. In both cases, conceptual span did at least as well as those types of other WM tasks. Once more, this result questions how essential the storage-plus-processing view really is in accounting for individual differences in complex cognitive activity.

There is one more relevant theoretical question. The storage-plus-processing tasks were developed originally as a type of span task that would improve the correlation between span and scholastic abilities measures, as compared to the simple digit span test that is used within tests of intelligence. The storage-plus-processing tests are successful in that regard, at least in adults; but why? From the point of view in which it is the quality of attentional processes that is important for WM, the critical difference between simple span and WM tasks may be the benefit of rehearsal in simple span tasks only. In storage-plus-processing tasks, the processing component may prevent covert rehearsal, perhaps just inadvertently. In the tasks that have been suggested to reflect the capacity of the attentional focus, the blocking of rehearsal is completely deliberate.

If this task analysis is correct, there is an interesting developmental prediction. It is well established that young children do not use rehearsal well. For them, we should find little difference in the way that simple span tasks and WM tasks correlate with measures of scholastic ability. When we look at developmental results, is that the case? Tentatively, the answer appears to be "yes." At least in some studies, digit span does just as well as storage-plus-processing tasks in predicting scholastic success in elementary-school children (Cowan et al., 2003; Hutton & Towse, 2001).

In conclusion, there is no reason to remain obsessed with storage-plus-processing tasks as a means to measure WM capacity. Measures designed to estimate how much information can be brought into the focus of attention at once are conceptually simpler. They may be less likely to confound WM with special knowledge such as linguistic knowledge, as the reading and listening span tests are likely to do, for example.

As the commentaries following Cowan (2001) attest, there are still open controversies regarding WM capacity and the focus of attention. For example, there could be separate capacity limits for various types of features, outside of the focus of attention (Wheeler & Treisman, 2002), or just a limit that emerges when attention is turned to any one feature field. A metric of complexity that takes into account the number of dimensions that must be considered to identify a stimulus correctly (Halford, Phillips, & Wilson, 2001; Phillips & Niki, 2002) might be unrelated to a chunk storage limit or might have to be combined with it to form a single, comprehensive theory.

SUMMARY

Although the concept of WM is at a forefront of research in cognitive psychology and cognitive neuroscience, there is little agreement on the definition of WM (Miyake & Shah, 1999) or how it should be measured. An apparent truism in the field is that, in order to measure WM capacity, one must tie up both storage and processing mechanisms within WM (Daneman & Carpenter, 1980). However, an alternative conception of WM holds that it relies on the ability to use the focus of attention in processing (Engle et al., 1999). The present chapter is consonant with that view. Yet, numerical estimates of capacity, in terms of chunks held in the focus of attention, can be obtained in situations where the focus is zoomed out to apprehend multiple items in a set (that cannot be combined into a fewer number of chunks; Cowan, 2001), rather than zoomed in to keep a goal in mind notwithstanding competing interference. Several WM measures were described as potential measures of the capacity of the focus of attention. They correlate with scholastic and intellectual aptitude measures just about as well as storage-plus-processing types of WM measures. In children too young to use sophisticated means of verbal rehearsal and grouping, simple digit span also serves as a good correlate of aptitude. To investigate individual and developmental differences, this research advocates vigorous attempts to find WM measures that are as simple as possible and are designed to index the capacity of the focus of attention without relying on verbal knowledge that inadvertently contributes to more complicated WM performance. Note that this approach is compatible with efforts to determine individual and developmental differences in faculties other than the focus of attention, such as the persistence of activation in memory (Cowan et al., 2000; Towse, Hitch, & Hutton, 2000), the speed of processing of information (Kail & Salthouse, 1994), or changes in the use of strategies (Cowan et al., 2003; Hitch, Towse, & Hutton, 2001).

ACKNOWLEDGMENTS

Preparation of this chapter was supported by NIH grant R01 HD21338.
Author address: Nelson Cowan, Dept. of Psychological Sciences, University of Missouri, 207 McAlester Hall, Columbia, MO 65211 USA. Electronic mail: CowanN@missouri.edu.

REFERENCES

Baddeley, A. D. (1986). *Working memory.* Oxford Psychology Series #11. Oxford: Clarendon Press.

Baddeley, A. D. (2000). The episodic buffer: a new component of working memory? *Trends in cognitive sciences, 4,* 417-423.

Baddeley, A. D. (2001). The magic number and the episodic buffer. *Behavioral and Brain Sciences, 24,* 117-118.

Broadbent, D. E. (1975). The magic number seven after fifteen years. In A. Kennedy & A. Wilkes (Eds.), *Studies in long-term memory.* Wiley. (pp. 3-18)

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information processing system. *Psychological Bulletin, 104*, 163-191.

Cowan, N. (1995). *Attention and memory: An integrated framework.* Oxford Psychology Series, No. 26. New York: Oxford University Press.

Cowan, N. (1999). An embedded-processes model of working memory. In A. Miyake & P. Shah (Eds.), *Models of Working Memory: Mechanisms of active maintenance and executive control* (pp. 62-101). Cambridge, U.K.: Cambridge University Press.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*, 87-185.

Cowan, N., Elliott, E. M., & Saults, J. S. (2002). The search for what is fundamental in the development of working memory. In R. Kail & H. Reese (Eds.), *Advances in Child Development and Behavior, 29*, 1-49.

Cowan, N., Lichty, W., & Grove, T. R. (1990). Properties of memory for unattended spoken syllables. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 258-269.

Cowan, N., Nugent, L. D., Elliott, E. M., Ponomarev, I., & Saults, J. S. (1999). The role of attention in the development of short-term memory: Age differences in the verbal span of apprehension. *Child Development, 70*, 1082-1097.

Cowan, N., Nugent, L. D., Elliott, E. M., & Saults, J. S. (2000). Persistence of memory for ignored lists of digits: Areas of developmental constancy and change. *Journal of Experimental Child Psychology, 76*, 151-172

Cowan, N., Towse, J. N., Hamilton, Z., Saults, J. S., Elliott, E. M., Lacey, J. F., Moreno, M. V., & Hitch, G. J. (2003). Children's working-memory processes: A response-timing analysis. *Journal of Experimental Psychology: General, 132*, 113-132.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning & Verbal Behavior, 19*, 450-466.

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A Meta-Analysis. *Psychonomic Bulletin & Review, 3*, 422-433.

Darwin, C. J., Turvey, M. T., & Crowder, R. G. (1972). An auditory analogue of the Sperling partial report procedure: Evidence for brief auditory storage. *Cognitive Psychology, 3*, 255-267.

Engle, R. W., Kane, M. J., & Tuholski, S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence and functions of the prefrontal cortex. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of* *active maintenance and executive control* (pp. 102-134). Cambridge, UK: Cambridge University Press.

Haarman, H. J., Davelaar, E. J., & Usher, M. (2003). Individual differences in semantic short-term memory capacity and reading comprehension. *Journal of Memory and Language, 48*, 320-345.

Halford, G. S., Phillips, S., & Wilson, W. H. (2001). Processing capacity limits are not explained by storage limits. *Behavioral and Brain Sciences, 24*, 123-124.

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In G. H. Bower (Ed.), *The psychology of learning and motivation*, Vol. 22 (pp. 193-225). Orlando, FL: Academic Press.

Hitch, G. J., Towse, J. N., & Hutton, U. (2001). What limits children's working memory span? Theoretical accounts and applications for scholastic development. *Journal of Experimental Psychology: General, 130*, 184-198.

Hutton, U. M. Z., & Towse, J. N. (2001). Short-term memory and working memory as indices of children's cognitive skills. *Memory, 9*, 383-394.

James, W. (1890). *The principles of psychology*. NY: Henry Holt.

Kail, R., & Salthouse, T. A. (1994). Processing speed as a mental capacity. *Acta Psychologica, 86*, 199-255.

Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General, 132*, 47-70.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390*, 279-281.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81-97.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., & Howerter, A. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41*, 49-100.

Miyake, A., & Shah, P. (Eds.), (1999). *Models of Working Memory: Mechanisms of active maintenance and executive control.* Cambridge, U.K.: Cambridge University Press.

Mukunda, K. V., & Hall, V. C. (1992). Does performance on memory for order correlate with performance on standardized measures of ability? A meta-analysis. *Intelligence, 16*, 81-97.

Oberauer, K., Süß, H. M., Schulze, R., Wilhelm, O., & Wittmann, W. W. (2000). Working memory capacity - facets of a cognitive ability construct. *Personality and Individual Differences, 29*, 1017-1045.

Pashler, H. (1988). Familiarity and visual change detection. *Perception &*

Psychophysics, 44, 369-378.

Phillips, S., & Niki, K. (2002). Separating relational from item load effects in paired recognition: Tempoparietal and middle frontal gyral activity with increased associates, but not items during encoding and retention. *Neuroimage, 17*, 1031-1055.

Pollack, I., Johnson, I. B., & Knaff, P. R. (1959). Running memory span. *Journal of Experimental Psychology, 57*, 137-146.

Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision, 3*, 179-197.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs, 74* (Whole No. 498).

Towse, J. N., Hitch, G. J., & Hutton, U. (2000). On the interpretation of working memory span in adults. *Memory & Cognition, 28*, 341-348.

Wheeler, M. E., & Treisman, A. M. (2002). Binding in short-term visual memory. *Journal of Experimental Psychology: General, 131*, 48-64.

This page intentionally left blank

Chapter 8

Implicitly Activated Memories: The Missing Links of Remembering

Douglas L. Nelson and Cathy L. McEvoy
University of South Florida, U. S. A.

Cosmologists map the universe, geologists the land, biologists the genome, and psychologists map word knowledge. They map their domains to gain theoretical and practical insights that would not be forthcoming without an atlas to guide their inquiry. The common assumption is that maps are essential. In our research, we assume that experiencing a familiar word implicitly activates related words (e.g., seeing PLANET activates *earth*, *mars*, and so on). Understanding how implicit activation affects memory requires a map of links among known words. Implicitly activated memories represent the missing links of remembering because we tend to be unaware of their activation and the effects they have on our ability to remember recent episodes. To date, formal theories of memory do not incorporate such memories into the modeling. This chapter shows how implicitly activated memories affect the recall and recognition of recent episodes and presents a model for explaining such effects.

MAPPING WORD KNOWLEDGE

The first task in trying to understand the effects of word knowledge on remembering is to construct an associative map. Such a map tells us what words are likely to be activated, how many there are likely to be, and how they are organized. Co-occurrence of words in text represents one means for constructing a map of word connectedness (Landauer & Dumais, 1997), but is a poor predictor of cued recall (Steyvers, Shiffrin, & Nelson, in press). We use free association to construct associative maps, and this task requires people to produce the first word to come to mind along some dimension (e.g., rhyme or meaning). Following Deese (1965), we assume the associative links are acquired in pair-wise learning that occurs during language acquisition. We started collecting free association norms in the 1970s so we could compare the effectiveness of rhyme and meaning cues on the same scale (e.g., Nelson, Wheeler, Borden, & Brooks, 1974). We were interested in controlling the probability that each type of cue would produce a word in the absence of study, and free association provided a solution for indexing pre-existing strength between test cues and to-be-remembered targets.

This work evolved when we noticed that some words produced smaller sets of rhyme- and meaning-related associates than others. Although there were no theories then about the potential effects of associative set size on memory, we launched a series of exploratory studies to see what would happen when it was varied. In the 1980s, we realized that the strength and set size indices were limited because they ignored connections *from* and *among* a word's associates, and over 5,000 words were normed (Nelson, McEvoy, & Schreiber, 1999).

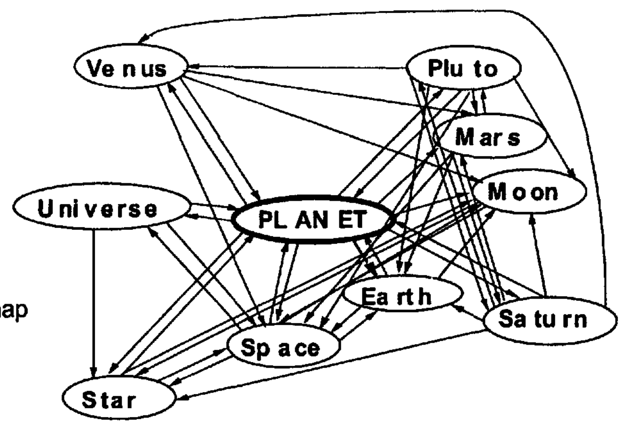


FIG. 8.1. Associative map for the word PLANET

Fig. 8.1 shows the associative map for the word PLANET. This word has 9 associates, defined as those given by two or more people in a sample of 150. The arrows indicate the direction of pre-existing links between words indexed by free association. Arrows pointing *to* the words appearing around the perimeter indicate its associates, e.g., from PLANET to *earth*, and so on. Note that some associates have arrows that point back to PLANET. These represent links from a word's associates, and are referred to as its *resonance*. Also, note that arrows link some of the associates together. These links reflect connections among the target's associates, and index its *connectivity*.

Our norms indicate that different words have different numbers of associates at varying strengths. Words also differ in terms of resonance and connectivity. Such differences allow us to explore the effects of a word's associative structure on its recall and recognition. In our standard cued recall task, a list of 24 words is presented one at a time for 3 seconds. Participants are asked to remember as many words as possible without being told how they will be tested. Immediately after the last word, the test instructions are read to them, which indicate that meaningfully related words will be presented as cues in order to help them recall each word just seen. This test is self-paced. This is an extralist cuing task because the test cues are unavailable during study. Participants must rely on pre-existing knowledge to recall the intended target, and the procedure models hundreds of everyday tasks where cues are used to recall associated knowledge.

There are many variations on this task, e.g., incidental instead of intentional study instructions. Similarly, implicit instead of explicit test instructions can be given. In the primed free association variation, participants can be asked to rate the study words for pleasantness, and then at test the cues used in the standard task can be shown with a request to produce the first word to come to mind to each cue (e.g., Humphreys, Tehan, O'Shea, & Bolland, 2000; Nelson & Goodmon, 2002; Zeelenberg, Shiffrin, & Raaijmakers, 1999). In the recognition test variation, a longer list of words is studied, and the study words serve as test cues. Participants are asked to recognize the study words by discriminating them from equal numbers of new words.

The rationale underlying this research is that we can come to understand how pre-existing knowledge affects memory for a recent episode by using associative word maps. Words for the episodic tasks are selected to differ systematically in resonance, connectivity, or set size in order to determine how these manipulations affect memory performance. This approach has been used successfully in studying word concreteness and frequency, but this research was more dedicated to understanding these variables than to understanding the interaction between known information and newly acquired episodic information. Known information refers to information acquired from worldly experience (Tulving, 1983), and new information involves learning recent facts. In our approach, the relative effects of variables related to word knowledge provide indicators of how important such knowledge is in episodic tasks. In what follows, we provide a description of the basic findings and suggest how they can be explained.

THE BASIC FINDINGS

Associative Resonance and Connectivity

Experiments show that variations in resonance and connectivity affect both extralist cued recall (e.g., Nelson, McKinney, Gee, & Janczura, 1998; Nelson, McEvoy, & Pointer, 2003) and recognition (Nelson, Zhang, & McKinney, 2001). In these experiments, study words were selected so that resonance and connectivity were varied in a factorial design at high or low levels. At high levels of resonance, about 75% of the target's associates had resonant links to the target. At low levels, less than 25% had resonant links. At high levels of connectivity, each associate in the target's set was connected to an average of nearly three other associates, whereas at low levels each associate was linked to an average of less than one. Other variables were either varied or held constant.

Fig. 8.2 shows the results of a portion of a standard extralist cuing experiment for medium strength pairs that produced their targets in free association with an average probability of .17 (Nelson et al., 2003). This probability estimates the likelihood that a cue will produce its target in the absence of a study trial and serves as a lower boundary on expected recall. As can be seen, recall was well above this boundary for all conditions. In addition, recall was more likely when the target words had higher levels of resonance or connectivity. Importantly, the interaction between the two variables was unreliable.

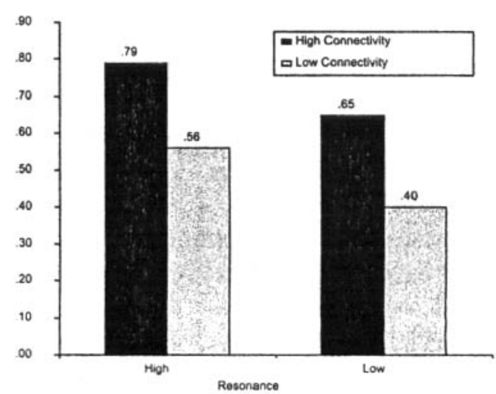


FIG. 8.2. Probability of cued recall as a function of resonance and connectivity (from Nelson et al., 2003)

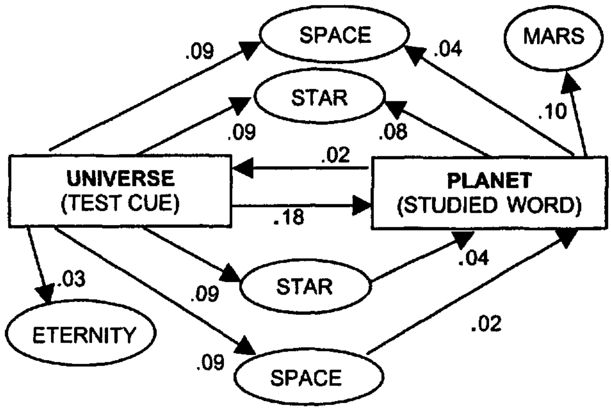
Such results are interesting for two reasons. First, they indicate that targets having more links *from* their associates and *among* their associates are easier to recall. Memory for a recent experience, for what is new, is influenced by what is known. Second, the effects of connectivity among the associates did not depend on the presence of resonant links. This finding is important because it is inconsistent with spreading activation theory. In spreading activation, connectivity among the associates of a word can affect recall only when activation can return to the target through its resonant links. Activation supposedly spreads from the target to its associates, among its associates, and back to the target. With fewer resonant links the activation has a reduced chance of returning, and spreading activation theory predicts that connectivity will have reduced effects, but this did not occur. Connectivity had essentially the same effects on recall at both high and low levels of resonance. Similar results are obtained in recognition (Nelson et al., 2001). The additive effects of these variables suggest that the target activates the links that bind an associative set in parallel.

Other findings indicate that connectivity effects are uninfluenced by variations in word frequency, concreteness, and set size (Gee, Nelson, & Krawczyk, 1999; Nelson, Bennett, Gee, Schreiber, & McKinney, 1993; Nelson & Goodmon, 2002). High connectivity among the associates of a studied word facilitates recall for young and old participants, for strong and weak cues, and for short and long study times (Nelson et al., 2003; Nelson, Bennett et al., 1993). It also facilitates recall when participants incidentally name the vowels of the study words or rate them for concreteness (Nelson, Bennett et al., 1993). Collectively, these findings indicate that connections among the associates of a studied word have similar effects for different types of words and for different study conditions. Connectivity is a robust effect that influences target recovery in recognition, cued recall and primed free association (Nelson & Goodmon, 2002).

Strength of the Cue-Target Relationship

Resonance and connectivity refer to pre-existing links within the associative set of a single word. In contrast, linking connections that join a test cue with its target define the strength of a cue-target relationship. Any two words can be linked because of prior language learning, and such links vary in strength, direction, and directness (Nelson et al., 1998). Strength varies from weak to strong, indexed by the probability that one word is given as a response to another in free association. Higher response probabilities are used to infer stronger pre-existing links, and in the model described later, they are used as estimates of activation levels.

FIG. 8.3. Pre-existing links between the test cue UNIVERSE and the studied target word PLANET (from Nelson et al., 1999)



As shown in Fig. 8.3, pre-existing connections are directional. Forward links run from the test cue to the target, as from UNIVERSE to PLANET. Backward links run from the target to the test cue, as from PLANET to UNIVERSE. Theoretically, the backward link refers to the probability that the studied target activates the test cue during learning, and the forward link refers to the probability that the test cue will produce the target during recall. In experiments, test cues can have forward, backward or both links with their targets. Finally, connections can be indirect as well as direct. Shared associate connections occur when the cue and target produce a common associate, e.g., the cue UNIVERSE → *star* and the target PLANET → *star* (Fig. 8.3). Mediated connections involve associated words that intervene between the cue and target, e.g., UNIVERSE → *space* → PLANET. A given associate can be a mediator, a shared associate, or can serve both roles.

Fig. 8.4 compares the effects of forward and backward strength in the standard extralist cued recall task (Nelson et al., 1998). All participants studied the same list of targets, and different subgroups were provided with test cues that held forward-and-backward links with their targets, only forward links, only backward links, or neither. Regardless of direction, strength averaged .13 when a link was present and .00 when it was not according to free association norms. As shown in Fig. 8.4, the presence of a backward link facilitated recall regardless of whether a forward link was present. Link direction had additive effects on probability of recall.

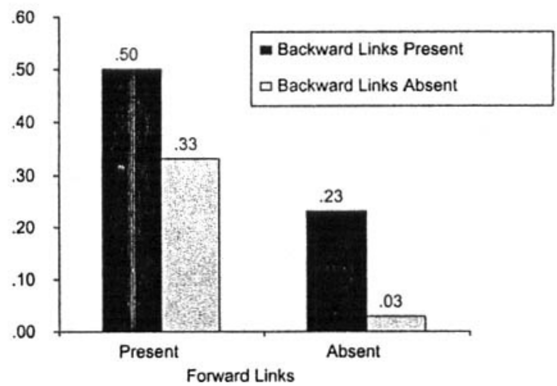


FIG. 8.4 Probability of cued recall as a function of forward and backward links (from Nelson et al., 1998)

Similar effects of forward and backward links appear in implicit primed free association (Nelson & Goodmon, 2002). The study procedures and items were the same as those used in cued recall, but participants produced the first word to come to mind. The effects of prior study are stronger in the cued recall task because of the greater likelihood of recovering explicit information. Studying the target words a second time before testing has a large effect on the probability of cued recall and no effect on primed free association (Humphreys et al., 2000; Nelson & Goodmon, 2002).

Finally, connections can be direct or indirect. Fig. 8.5 shows the effects of indirect shared associate and mediated links in extralist cuing (Nelson et al., 1998). All of the test cues had relatively weak direct forward links to their targets, and both shared associates and mediators add their effects to the benefits produced by these links. The effects of indirect connections are generally not as robust as the effects of direct connections (Nelson & Zhang, 2000). Three-step indirect links (cue → mediator 1 → mediator 2 → target) appear to have no effects on recall (Nelson, Bennett, & Leibert, 1997) because they generate too much noise to be effective. An average of only three steps is needed to get from one word to any other in our norms (Steyvers & Tenenbaum, submitted).

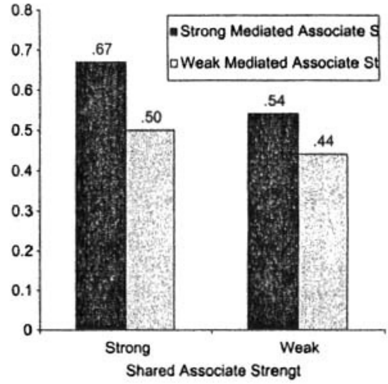


FIG 8.5 Probability of cued recall as a function of shared associate strength and mediated strength (from Nelson et al., 1998)

These findings indicate that the likelihood of recalling a recent event is not only a function of the strength of the forward link (Bahrick, 1970). Recall is also not only a function of the backward link, or what has been called encoding specificity (Tulving & Thomson, 1973). Each link is important but does not tell the whole story. A cue can be entangled with its target in multiple ways, and as the findings in Figs. 8.4 and 8.5 show, each link adds something to the cue's success.

Set Size

Note that in Fig. 8.3, some of the associates of the test cue and of the target fail to join these items. UNIVERSE is linked to words that are not connected within two-steps to the target (e.g. Eternity), and PLANET is linked to words that are not connected within two steps to the cue (e.g. Mars). Such associates hinder recall. Words differ in how many associates they produce in free association, or what we call set size. Given the preceding section on linking connections, it would be tempting to infer that net cue-target strength would be greater when the test cue and its target have larger associative sets. There should be more opportunities for shared associates and mediators. However, this temptation needs to be resisted. Using the normative database, we added cue set size to target set size for 48,572 pairs of words to determine the total number of associates for each pair. This total was correlated with shared associate and mediated strength and, with respective correlations of -.29 and -.18, the results indicated that the reverse was true. Pairs with larger sets had weaker shared associate strengths and weaker mediated strengths. Cue and target set size are not strongly correlated with any index of strength, or with resonance or connectivity (Nelson & Zhang, 2000).

In contrast to associates that link the cue with the target, associates that fail to link them compete during retrieval and decrease recall (e.g., Nelson, Schreiber, & McEvoy, 1992). This competition effect is illustrated in Fig. 8.6, which shows the probability of cued recall in the extralist cuing task as a function of cue and target set size. Small and large set size words had averages of 6.8 and 21.2 associates, respectively. Words with more associates were less likely to be recalled than those with small sets, regardless of whether the manipulation of set size was among the cues or the targets. Accuracy declines and response latencies increase as set size increases because more competitors are likely to be present when set size is larger (Schreiber, 1998; Schreiber & Nelson, 1998).

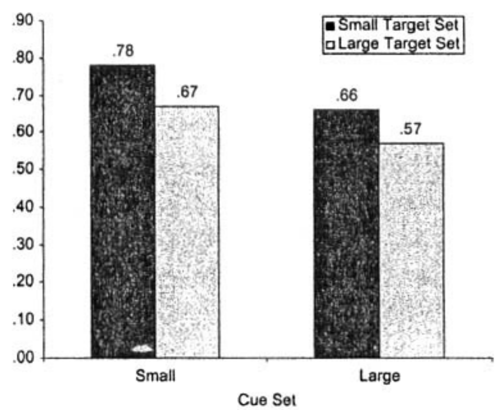


FIG. 8.6. Probability of cued recall as a function of cue and target set size (from Nelson, Schreiber, & McEvoy, 1992)

Target set size affects recall regardless of word ambiguity (Gee, 1997), concreteness (Nelson & Schreiber, 1992), and frequency of occurrence (Nelson & Xu, 1995). The magnitude of set size effects is uninfluenced by variables that affect how well the target has been encoded, such as study time and level of processing (Nelson, Schreiber, & McEvoy, 1992). Furthermore, such effects are found on implicit tests as well as explicit tests of memory (e.g., Nelson, Schreiber, & Holley, 1992). The size of a word's associative network has robust effects on word retrieval for many different types of words under many different types of conditions. Experiencing a familiar stimulus appears to access much related knowledge and this knowledge either helps or hinders recall. It helps recall when it strengthens the link between what has been encoded and the available cue but it hinders recall when it competes with what needs to be recalled (Nelson & McEvoy, 2002).

THE MODEL: PROCESSING IMPLICIT AND EXPLICIT REPRESENTATIONS (PIER2)

PIER2 was designed to explain how resonance, connectivity, linking connections and competing associates affect extralist cued recall and recognition (Nelson et al., 1998; Nelson & Zhang, 2000). Its general assumptions are that experiencing a word produces an explicit representation plus an implicit representation. The explicit representation is produced as a result of explicit processing operations applied to a list word, e.g., rehearsing it, rating its attributes, and naming its vowels. These operations bind the word to its context and make it more accessible to retrieval cues that normally produce it as a response. The implicit representation is a byproduct of a comprehension process that automatically activates related words in memory. In terms of its current development, PIER2 has focused on the contribution of the implicit representation. This representation was formalized to explain how the features of pre-existing associative structures affect memory for recent episodes. The model assumes that retrieving the implicit representation involves two processes, a target activation process and a cue-target intersection process.

The activation process explains the effects of target resonance and connectivity on cued recall, recognition and priming. In this process, experiencing a word produces a parallel activation of its associates and the links that bind them into a coherent associative structure. It is an integrating process that brings related knowledge together, with the integration achieved by summing the strengths of the links between the pairs of words, which comprise the network. The model assumes that free association strengths estimate the level of link activation between any two words.

The top portion of Table 8.1 displays an *nxn* association matrix for a hypothetical target and its three associates. The initial self-strength of the target is assigned a nominal value of 1.00, and remaining strengths come from free associations norms. The probabilities that the target produces

TABLE 8.1.
A Computational Example of PIER2's Equations.

The Target Activation Process

Equation 8.1: $S(T_i)=\left[S(T,T)+\sum_{i=1}^{n}S(A_i,T)\right]+\sum_{j=1}^{n}\left[S(T,A_j)+\sum_{i=1}^{n}S(A_i,A_j)\right]=$

[(1.00+.25+.10)]+[(.40+.35)+(.30+.10)+(.30)]=2.80

Equation 8.1a: $S(A_i)=\sum_{j=1}^{n}\left[S(T,A_j)+\sum_{i=1}^{n}S(A_i,A_j)\right]$; So, $S(A_1)=[(.40)+(.35)]=.75$

Association Matrix

	Target	Associate 1	Associate 2	Associate 3
Target	1.00	.40	.30	.30
Associate 1	.25	--	.10	
Associate 2		.35	--	
Associate 3	.10			--
Activation Strength	1.35	.75	.40	.30

Resonance **Associative Connectivity**

Cue-Target Intersection Process

Eq. 8.2: $S(Q_j,T_i)=\sum_{k}^{n}S_{jk}S_{ik}+\sum_{k}^{n}S_{jk}S_{ki}$ =(.25x2.80)+(1.00x.75)+(.10x.40)+(.30x.15)=1.54

Retrieval Matrix

	Target	Test Cue Associate 1	Shared Associate 2	Mediator	Target Competitor Associate 3	Cue Competitor
Target	2.80	.75	.40	.15	.30	---
Test Cue (Associate 1)	.25	1.00	.10	.30	---	.22

Eq. 8.3: $P(T_i/Q_j)=\dfrac{S(Q_j,T_i)}{S(Q_j,T_i)+\sum_{q}^{n}S(Q,A_q)+\sum_{t}^{n}S(T,A_t)}$ =1.54 ÷ 1.54+.22+.30=.75

Associates 1-3 are, respectively, .40, .30, and .30. The probabilities that Associate 1 produces the target and the target's Associates 2 and 3 are .25, .10 and .00, and so forth for Associates 2 and 3. The strength values for any pair in the structure can be determined by reading the matrix row-by-row from left to right. From Row 1, we see how strongly the target activates each of its associates. From the remaining rows, we see how strongly an associate of the target activates the target and its associates. Column 1 indicates how strongly each of the target's associates are linked to it. The target's resonance is determined by adding the strengths of the individual resonant links. The remaining columns index the target's connectivity, associate by associate.

PIER2 assumes that target activation strength increases as an additive function of the strength of all pre-existing links in the network. Each link adds to activation regardless of its direction or source. The strength of the target (T_i) in this example is 2.80. Similarly, the strengths of each associate are determined by adding the input from the target to the inputs from the other associates. Some associates will be activated to higher levels because of stronger input from the target, from other associates, or both.

PIER2 is not a spreading activation model in the sense that activation has to spread back to the target in order to heighten its activation. Each link contributes separately to target activation regardless of link direction, so resonant links are no more important than connective links among the associates. The model is an activation-at-a-distance model that allows links among the target's associates to affect its activation level even in the absence of resonant links. The important principle is the simultaneity of activation, not its spread (Nelson et al., 2003). Targets with more and stronger links to, from and among their associates are activated to higher levels than those with fewer and weaker links. PIER2 uses the results of the activation process directly, to predict the effects of resonance and connectivity on recognition. Cued recall and primed free association, however, rely on both activation and intersection processes.

Cue-target intersection is conceptualized as a *separating* process that selects the target from the associates activated by the cue and target. This aspect of the model was designed to explain the effects that direct links, indirect links, and set size have on recall. Table 8.1 uses a numerical example to illustrate how the model's equations are computed. The results of the activation process are incorporated into the computation of the intersection process, which ensures that resonance and connectivity will influence cued recall as a matter of degree.

The effects of linking connections and set size are computed in separate steps. The initial step computes the net strength of a cue-target relationship, and the second step incorporates the noise generated by competing associates of the cue and target. Table 8.1 shows that the first step enters the primed activation levels of the target and its associates into the top row of a retrieval matrix that has to be constructed for each word pair. The first value in this row is the activated self-strength of the target (Eq. 8.1), and the second is the activation level of the test cue, if it is a member of the target's set (Eq. 8.1a). In successive columns, the primed activation levels of the target's shared associates are shown (also Eq. 8.1a). Finally, the top row shows the mediated links, if any. The values for mediators that do not appear in the target's association matrix are not primed, and free association values are used to estimate the strengths of such links.

The bottom row in the retrieval matrix enters the strengths of the forward links to the target, its shared associates, and any cue-to-mediator links. All of these values are taken from free association norms. As a simplifying assumption, the activation process computed in Equation 8.1 is not applied to the test cue. Consequently, this version of the model predicts that cue resonance and cue connectivity will have no effect. The cue merely activates the target and its associates in accordance with pre-existing strengths. As the example indicates (Eq. 8.2), cross-multiplying and adding the values in the retrieval matrix is what determines the net strength of the intersection.

As target resonance and connectivity increase and as forward, backward, shared and mediated strength increase, net cue-target strength increases. Each link in the retrieval matrix adds to net strength. However, as shown in the second step of the intersection process, competing associates curtail net strength. Each associate is "friend" or "foe," depending on whether it links or competes with target recovery. Normative free association values are used to estimate the strengths of all competitors. We assume that target competitors are no stronger or weaker than those activated by the test cue. Priming during the study phase affects net strength only when it is directly supported by links coming from the test cue. The cue and the information it brings to bear on the targeted information is a key feature of the model. In the final computation, the results of net cue-target strength are made relative to noise produced by competing associates (Eq. 8.3). A good retrieval cue can be overwhelmed by noise from competing associates.

PIER2 incorporates the effects of cue and target set size, which have negative effects on extralist cued recall, with features that contribute positive effects. The model correctly predicts that forward and backward strength will have additive effects, and that the effects of resonance-connectivity will be more apparent for stronger cues (Nelson et al., 2003). The model sums the strengths of various types of cue-target connections, and it multiplies target activation strength by forward strength in its computations. The model was evaluated using a cued recall database that includes the results of 29 studies and the probabilities of correct recall for 2,062 pairs (Nelson & Zhang, 2000; http://luna.cas.usf.edu/~nelson/). When Equation 8.3 was entered in a regression analysis as the only predictor of cued recall, it explained 26% of the variance, which increased to 40% when variance explained was adjusted for test reliability. When 148 pairs were eliminated as extreme outliers ($\pm$1.80 SDs), 65% was explained. For comparison purposes, a simultaneous multiple regression was computed on the same purged data set using the 8 features discussed in this paper as predictors (e.g., connectivity). The resulting equation, also adjusted for reliability, explained 64% of the variance. The single predictor based on PIER2's calculations predicted cued recall as effectively as 8 separate estimates. For the majority of the pairs in these experiments, PIER2 was a good predictor of cued recall even though no parameters were estimated from the cued recall data.

A contrast between PIER2 and the standard models of what make a cue effective is instructive. The forward strength only model is the modal model for manipulating and controlling strength between pairs of related words in many tasks, including, among others, paired-associate learning, lexical decision, category decision, and naming. When entered as a single predictor in the purged database and adjusted for reliability, forward strength explains 30% of the variance. Similarly, encoding specificity, or backward strength in the extralist cuing task, explains some of the variance, but not as much as would be expected on the basis of a strong encoding specificity principle, 10%. These findings indicate that the effectiveness of a retrieval cue is determined by a multiplicity of links that both help and hinder recall, and that models incorporating this complexity will fare better than single feature models in predicting performance.

INTERACTION EFFECTS

Three interaction effects played significant roles in developing PIER2.

The first involved differences in the effects of set size versus resonance-connectivity on various retention tests. Because of these differences, the effects of these variables were attributed to different processes in the model. Target set size has robust positive effects on extralist cued recall, but this advantage is not normally found in recognition (e.g., Nelson, Canas, & Bajo, 1987). The null effect is not the result of insensitivity, as 18 attempts to find such effects under a variety of conditions failed.

We attribute this interaction to differences in the processes involved in the two tasks. Extralist recall requires using a novel cue to recover the target from competing associates. This task requires both the activation and cue-target intersection processes. The recognition task, in contrast, normally requires only the activation process because the target serves as the test cue. This cue directly re-activates its implicit representation, and can be recognized on the basis of apparent familiarity (e.g., Shiffrin & Steyvers, 1997). Support for process differences in cued recall and recognition comes from data showing that set size effects can be produced in a recognition test when the testing procedure requires computing the intersection for some list words (Nelson et al., 1987).

In contrast to the set size pattern, target resonance and connectivity influence cued recall and recognition in the same way. Stronger connections of either type increase cued recall and recognition. PIER2 predicts this result because these variables affect the target activation process common to both tasks. These variables are thought to reflect the integration process, not the separation process. We also assume that low frequency target words produce a higher level of activation than high frequency words, and that this difference explains the advantage that low frequency words have on recognition (e.g., Glanzer & Bowles, 1976). This assumption requires PIER2 to predict that low frequency words will be better recalled in cued recall because the activation process is common to both retention tests. As expected, low frequency words are more likely to be recalled in extralist cuing (Nelson & Xu, 1995), and they also show stronger priming effects in primed free association (Nelson & Goodmon, 2002).

The second interaction concerns the effects of disrupting a memory task prior to the retrieval test. Experiments on disruptions suggest that retention failures of implicitly activated memories are produced by failures to retrieve context features rather than by interference or decay. Despite test instructions that explicitly refer to the memory task, access to the contextual cues that define this experience falters, with the degree of failure related to the nature of the disruption.

In our disruption research, participants study a list of words in an extralist cuing task, and instead of being tested immediately, they solve multiplication problems or study additional lists for 10-20 minutes prior to testing (e.g., Nelson, McEvoy, Janczura, & Xu, 1993). As shown in Fig. 8.7, recall declines after these disruptions, compared to immediate testing. More importantly, the studies show that target set size effects are reduced after the math task. Words with smaller sets no longer have the same recall advantage shown on immediate tests. The reduction of set size effects after the math task is not produced by interference generated by response competition (McGeoch, 1942). Compared to the multiplication task, studying additional word lists after studying the original list produces declines in recall but has no effect on the magnitude of set size effects (Nelson, Bennett et al., 1993). Whether the interfering lists are associatively related or unrelated to the target makes no difference, nor does the number of interfering lists up to three. A recent experiment, yet unpublished, shows similar results for connectivity. If interference was the key to understanding the effects of disruption, then, as with math, studying interference lists should have reduced the effects of set size and connectivity, especially when comprised of related words.

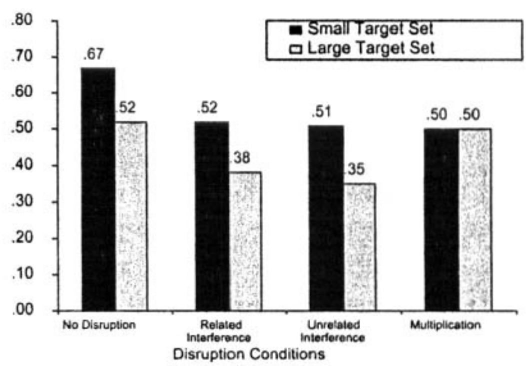


FIG. 8.7. Probability of cued recall as a function of disrupting attention and target set size (from Nelson, Bennett et al., 1993)

A decay explanation (e.g., Cowan, 1999) for these findings also fails, but for additional reasons. Relatively weak cues were used in these experiments. For stronger cue-target relationships, the math task lowers recall but has no apparent influence on the magnitude of set size and connectivity effects (Nelson et al., 1998). Such effects are as large after math as they were on the immediate task. A stronger cue-target relationship overcomes the effects of disruption produced by the math task. Decay theory can explain why set size effects disappear after a delay but not why such effects fail to disappear for stronger cues.

Initially, these findings were puzzling, but seem to have nothing to do with either interference or decay and much to do with attention switching, disrupting access to context. While studying interference lists, attention is maintained on the context of the memory task. In contrast, doing math disrupts this context because it requires participants to switch attention away from memory to a conceptually different task. When only weak cues are available, this switch reduces access to context of the original encoding. When stronger cues are available, access to the original context is more likely and the effects of implicitly activated memories are apparent.

This interpretation is also supported by results showing that studying a list of words after the math task reinstates set size effects (Nelson, McEvoy et al., 1993). Studying words after doing math helps weak cues reinstate memory for the encoding context. Furthermore, recent work shows that both direct and indirect linking connections are susceptible to the effects of switching attention (Nelson & Goodmon, 2003). After 20 minutes of math, cues, having backward but no forward links with their target, produce target recall that is just above chance. Shared associate cues produce chance recall. In addition, changing the test room substantially reduced the benefits of forward and backward links. These findings indicate that the retention of implicitly activated memories depends on accessing memory for the original episode. When this fails, implicitly activated memories, like explicitly encoded memories, fail to have an influence.

In constructing PIER2, we assumed that context cues were bound only to the explicit representation. The findings, however, indicate that this supposition must be changed. Context and the activation of known information may be processed simultaneously in a working memory designed to encode current experience and prior knowledge interactively. This interpretation links attention to context retrieval failure and, in so doing, links PIER2 to theories of working memory (e.g., Cowan, 1999; Engle, Tuholski, Laughlin, & Conway, 1999). The model attributes the reductions in set size and connectivity effects, caused by math disruptions, to context retrieval failures induced by attention switching. Primed information is available, but it cannot be accessed by a test cue unless it instantiates the study context. This explanation assumes that context cues are associated with both the implicit and the explicit representations of the target. This assumption is consistent with recent attempts to extend REM to include implicit as well as episodic memory. In the extended model, the two forms of memory are separated, but context information is represented in each form (Raaijmakers, in press; Shiffrin, in press).

The third interaction shows that dramatic effects occur when targets are studied in the presence of meaningfully related words, such as STAR, PLANET (e.g., Nelson, Bennett, et al., 1993; Nelson, Schreiber, & McEvoy, 1992). Target set size and connectivity effects are substantially reduced, regardless of whether recall is cued by the study word (STAR), an extralist associate (UNIVERSE), or a rhyming stem (ANET). Retrieving the specific episodic context of an encoding is a more important cause of these reductions than the nature of the test cue. Associates of the studied word and target are activated, then inhibited, and this inhibition is specific to the episodic context. Implicitly activated associates that have been inhibited will have reduced effects as long as the test cue recovers this inhibiting context.

Two lines of evidence are consistent with the context driven inhibition interpretation. First, such effects are sensitive to context timing. When the context word appears slightly before or after the target, there is a corresponding reduction in set size effects (Nelson, Gee, & Schreiber, 1992). Even when pairs are presented simultaneously, they must be displayed for relatively long intervals to observe the reduction in set size effects. With extremely rapid pair presentation, normal set size effects are observed (Schreiber & Carter, 1997). Second, when related word pairs are studied under optimal conditions, the multiplication disruption has the opposite effect of what occurs in extralist cuing. When related pairs are studied, set size effects are found after the math task but not on the immediate test (Nelson, McEvoy et al., 1993). The math disruption reduces context-driven inhibition effects and the effects of previously inhibited associates return. Context-driven inhibition effects, unlike activation processes, require more processing time, but switching attention disrupts inhibitory processes just as it disrupts activation processes. The activation and inhibition of associative information is largely determined by the retrieval of context, suggesting that context is linked to both implicit as well as explicit representations of the target.

PIER2 AND FEATURE THEORIES

PIER2 predicts many effects linked to implicitly activated associates in recognition and recall. Formal models of memory based on features have been successful in predicting episodic phenomena in these tasks (e.g., Shiffrin & Steyvers, 1997). The feature approach differs substantially from PIER2. Among other differences, the features are undefined and treated as random variables, whereas PIER2 uses measured connection strengths to make its predictions. The two approaches seem fundamentally incompatible because the units of analysis are fundamentally different. They may be incompatible, but another interpretation assumes that PIER2 offers algorithms for indexing an associative feature that links a word to its associates (Eq. 8.1) and an associative feature that links any two words together (Eq. 8.3). Theoretically, these equations can be understood as indexing a word's associative similarity to its own associates or to another related word. Alternatively, methods other than those used in PIER2, such as single value decomposition (Steyvers et al., in press), may prove useful for capturing associative relationships. In either case, association simply becomes one or more of the features in a hypothetical feature array, along with orthographic, phonological, categorical and semantic features. With this orientation, what are needed are different algorithms for computing different types of features. The apparent advantage of combining association and feature approaches is that predictions would be constrained by real measurements taken on specific words and specific word pairs. We stress "apparent" because, at this point, the success of such a venture is uncertain, but it could offer a reasonable solution to the problem of how prior knowledge interacts with episodic knowledge to determine memory for an experience.

SUMMARY

Just as scientists in many disciplines are mapping information in their domains, we are mapping word knowledge that reveals the associative structure of specific words. We can build such a map because familiar words remind us of related words, and by using free association procedures, we can learn what these words are and how they are linked. The associative structures of known words vary systematically in terms of three different features: Resonance, connectivity, and set size.

Resonance refers to the probability that a word's associates produce it as an associate. Connectivity refers to links among a word's associates, and set size refers to how many relatively strong associates there are in its set. Our interest lies not in constructing word maps, but in determining how the pre-existing associative structure of a word affects its recognition and cued recall. The broader issue lies in understanding how pre-existing knowledge influences recent episodic memory. Recognition and cued recall processes are best understood as the result of an interaction between known and new information. Our model assumes that processing a familiar word activates related words in memory, and that disrupting attention causes forgetting not by interference, but by reducing access to the context of the original encoding.

ACKNOWLEDGEMENTS

This research was supported by grants MH16360 from the National Institute of Mental Health to D. L. Nelson and AG13973 from the National Institute on Aging to C. L. McEvoy. The authors would also like to thank Amie Wilbanks for her help in preparing the materials.

REFERENCES

Bahrick, H. P. (1970). Two-phase model for prompted recall. *Psychological Review, 77,* 215-222.

Cowan, N. (1999). An embedded-process model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 62-101). Cambridge: Cambridge University Press.

Deese, J. (1965). *The structure of associations in language and thought.* Baltimore, MD: The Johns Hopkins Press.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General, 128,* 309-331.

Gee, N. R. (1997). Implicit memory and word ambiguity. *Journal of Memory and Language, 36,* 253-275.

Gee, N. R., Nelson, D. L., & Krawczyk, D. (1999). Is the concreteness effect a result of underlying network interconnectivity? *Journal of Memory and Language, 40,* 479-497.

Glanzer, M., & Bowles, N. (1976). Analysis of the word frequency effect in recognition memory. *Journal of Experimental Psychology: Human Learning* *and Memory, 2*, 21-31.

Humphreys, M. S., Tehan, G., O'Shea, A., & Bolland, S. W. (2000). Target similarity effects: Support for the parallel distributed processing assumptions. *Memory & Cognition, 28*, 798-811.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and the representation of knowledge. *Psychological Review, 104*, 211-240.

McGeoch, J. A. (1942). *The psychology of human learning.* New York: David MaKay.

Nelson, D. L., Bennett, D. J., Gee, N. R., Schreiber, T. A., & McKinney, M. (1993). Implicit memory: Effects of network size and interconnectivity on cued recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 747-764.

Nelson, D. L., Bennett, D. J., & Leibert, T. (1997). One step is not enough: Making better use of association norms to predict cued recall. *Memory & Cognition, 25*, 785-796.

Nelson, D. L., Canas, J. J., & Bajo, M. T. (1987). The effects of natural category size on memory for episodic encodings. *Memory & Cognition, 15*, 133-140.

Nelson, D. L., Gee, N. R., & Schreiber, T. A. (1992). Sentence encoding and implicitly activated memories. *Memory & Cognition, 20*, 643-654.

Nelson, D. L., & Goodmon, L. B. (2002). Experiencing a word can prime its accessibility and its associative connections to related words. *Memory & Cognition, 30*, 380-398.

Nelson, D. L., & Goodmon, L. B. (2003). Disrupting attention: The need for retrieval cues in working memory theories. *Memory & Cognition*, 31, 65-76.

Nelson, D. L., & McEvoy, C. L. (2002). How can the same type of prior knowledge both help and hinder recall? *Journal of Memory and Language, 46*, 652-663.

Nelson, D. L., McEvoy, C. L., Janczura, G. A., & Xu, J. (1993). Implicit memory and inhibition. *Journal of Memory and Language, 32*, 667-691.

Nelson, D. L., McEvoy, C. L., & Pointer, L. (2003). Spreading activation or spooky action at a distance? *Journal of Experimental Psychology: Learning, Memory and Cognition, 29*, 42-52.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1999). *The University of South Florida Word Association, Rhyme and Fragment Norms.* http://luna.cas.usf.edu/~nelson/.

Nelson, D. L., McKinney, V. M., Gee, N. R., & Janczura, G. A. (1998). Interpreting the influence of implicitly activated memories on recall and recognition. *Psychological Review, 105*, 299-324.

Nelson, D. L., & Schreiber, T. A. (1992). Word concreteness and word structure as independent determinants of recall. *Journal of Memory and Language, 31*, 237-260.

Nelson, D. L., Schreiber, T. A., & Holley, P. E. (1992). The retrieval of controlled and automatic aspects of meaning on direct and indirect tests. *Memory &*

Cognition, 20, 671-684.

Nelson, D. L., Schreiber, T. A., & McEvoy, C. L. (1992). Processing implicit and explicit representations. *Psychological Review, 99*, 322-348.

Nelson, D. L., Wheeler, J. W., Borden, R. C., & Brooks, D. H. (1974). Levels of processing and cuing: Sensory vs. meaning features. *Journal of Experimental Psychology, 103*, 971-977.

Nelson, D. L., & Xu, J. (1995). Effects of implicit memory on explicit recall: Set size and word frequency effects. *Psychological Research, 57*, 203-214.

Nelson, D. L., & Zhang, N. (2000). The ties that bind what is known to the recall of what is new. *Psychonomic Bulletin & Review, 7*, 604-617.

Nelson, D. L., Zhang, N., & McKinney, V. M. (2001). The ties that bind what is known to the recognition of what is new. *Journal of Experimental Psychology: Learning, Memory and Cognition, 27*, 1147-1159.

Raaijmakers, G. W. (in press). Modeling implicit and explicit memory. In C. Izawa & N. Ohta (Eds.) *Learning and Memory: Advances in Theory and Applications*. New Jersey: Erlbaum.

Schreiber, T. A. (1998). Effects of target set size on feelings of knowing and cued recall: Implications for the cue effectiveness and partial-retrieval hypotheses. *Memory & Cognition, 26*, 553-571.

Schreiber, T. A., & Carter, K. (1997, November). The activation and suppression of pre-existing fan. Presented at the 28th annual meeting of the Psychonomic Society, Philadelphia, PA.

Schreiber, T. A., & Nelson, D. L. (1998). The relation between feelings of knowing and the number of neighboring concepts linked to the test cue. *Memory & Cognition, 26*, 869-883.

Shiffrin, R. M. (in press). Bayesian modeling of memory and perception. In C. Izawa & N. Ohta (Eds.) *Learning and Memory: Advances in Theory and Applications*. New Jersey: Erlbaum.

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM--retrieving effectively from memory. *Psychonomic Bulletin & Review, 4*, 145-166.

Steyvers, M., Shiffrin, R. M., & Nelson, D. L. (in press). Semantic spaces based on free association that predict memory performance. In A. Healy (Ed.), *Experimental Cognitive Psychology and its Applications.*

Steyvers, M., & Tenenbaum, J. (submitted). The large scale structure of semantic networks: Statistical analysis and a model of semantic growth. *Cognitive Science.*

Tulving, E. (1983). *Elements of episodic memory*. New York: Oxford University Press.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80*, 352-373.

Zeelenberg, R., Shiffrin, R. M., & Raaijmakers, J. G. (1999). Priming in a free association task as a function of association directionality. *Memory & Cognition, 27*, 956-961.

Chapter 9

It's About Time: Circadian Rhythms, Memory, and Aging

Lynn Hasher, David Goldstein,
University of Toronto, Canada, and
Cynthia P. May
College of Charleston, USA

Psychologists have long been interested in variations in behavior over the course of the day. The first article in the first issue of *The American Journal of Psychology* (the first psychology journal published in the US) reported data on the impact of time of day on the magnitude of the knee jerk response (Lombard, 1887). Much of this work has assumed that there is one general pattern or one optimal time of day that characterizes everyone's performance, regardless of variables such as age (e.g., Baddeley, Hatter, Scott, & Snashall, 1970; Blake, 1967; but see Bergstrum, 1894, for an early, albeit rare, exception). This neglect of individual and developmental differences is mirrored in longstanding western cultural beliefs, as indexed by aphorisms such as "early to bed, early to rise, makes a man healthy, wealthy, and wise" or "the early bird gets the worm." The message is clear: it is best for all of us to be morning people.

However, even casual observation provides many examples of deviations from this general pattern. For example, parents of high school students frequently complain about how difficult it is to get teens out of bed early in the morning, and their teachers complain about how difficult it is to get them to pay attention (or even stay awake) in early morning classes. Indeed, these casual observations have been confirmed in surveys of time of day preferences among adolescents and young adults (e.g., Kim, Dueker, Hasher, & Goldstein, 2002; May & Hasher, 1998).

Despite these observations and the evidence of a relative absence of morning type preferences in adolescents and young adults, the school day tends to reflect the cultural belief in the wisdom of learning in the morning. In fact, the school day typically begins at an earlier hour as students get older, potentially exacerbating any problems created by a mismatch between circadian preferences and the timing of learning opportunities (Carskadon, Wolfson, Acebo, Tzischinsky, & Seifer, 1998). Indeed, there is a widespread belief that the critical subjects of English and Mathematics should be taught early in the school day while other, possibly less important topics such as Physical Education, Art and Music, should be taught later in the day (Dunn, Dunn, Primavera, Sinatra, & Virostko, 1987). The tacit assumption is that students are - or possibly should be - most alert in the morning and the most important and intellectually demanding subjects should naturally be taught during this time. Oddly, there is very little empirical evidence to support this position. One of the few citations buttressing this assertion is to a paper by Gates published in 1916.

Another example of the neglect of individual and developmental differences is actually seen in the event that initiated the line of research discussed in this paper. We started a project that involved testing a large number of older and younger adults and to use our lab space efficiently, we planned to test people from 8:30AM to 6PM. However, we quickly found that older adults did not wish to be tested in the afternoon and certainly not in the late afternoon. Undergraduate students, by contrast, did not want to be tested early in the day; they really only wanted to come in after 11AM, preferably after noon.

How might this puzzling pattern of time preferences be explained? For the young adults, we considered that it was a social phenomenon because US campuses have lots of visiting and partying (and some studying) late in the evening. More puzzling to us was the behavior of older adults; they were mostly retired and so were free to set their daily schedule as they liked. Why did so few of them want to be tested in the afternoon? The answer is tied to individual as well as developmental differences in circadian rhythms.

CIRCADIAN RHYTHMS

Daily fluctuations in behavior are ubiquitous features of virtually all organisms from plants, to unicellular algae, to humans. These fluctuations, which occur on a roughly 24-hour cycle, are known as circadian rhythms (from the Latin *circa*, "approximately" and *dies*, "day"). Circadian rhythms – a term that entered the literature in 1959 (see Moore-Ede, Sulzman, & Fuller, 1982) - persist even in the absence of environmental cues and, as a result, are considered to be endogenous. However, this internal "clock" is entrained, or set, to a precise 24-hour period by exogenous cues (known as zeitgebers). The locus of this biological clock is thought to be in the suprachiasmatic nucleus of the hypothalamus and recent evidence suggests a role for the Opn4 gene which codes for the protein Melanopsin (Panda, Sato, Castrucci, Rollag, DeGrip, Hogenesch, Provencio, & Kay, 2002).

There are a variety of circadian rhythms that influence physiological functioning in humans. Among them are rhythms in the sleep-wake cycle, glucose uptake, core body temperature, neurotransmitter function, heart rate, and circulating hormones (Folkard, 1983; Hrushesky, 1994; Monk, 1989; Moore-Ede et al., 1982). Circadian rhythms also play an important role in the treatment of many diseases. Circadian variations are found in disease symptoms as well as in the ability of tissues to absorb drugs. In many cases circadian patterns play a role in disease severity. For example, the great majority of asthma attacks take place between two and six o'clock in the morning and myocardial infarctions strike twice as often in the morning as at other times of day (e.g. Hardin, 2000; Hasher & Goldstein, 2001). Against this background then, it is not surprising to find that there is a circadian rhythm that impacts on intellectual and physical performance, with better performance at some times of day than at others.

The arousal rhythm that impacts on cognition can be measured with a simple paper and pencil task developed by Horne and Ostberg (1976; 1977) called the Morningness-Eveningness Questionnaire (MEQ). The test consists of 19 questions, among which are the following: "Considering only your own 'feeling best' rhythm, at what time would you get up if you were entirely free to plan your day?" and "If you went to bed at 11PM, at what level of tiredness would you be?"

The test sorts people into five categories that range from "definitely morning type" through "neutral" to "definitely evening type." This test has been translated into many languages and used to assess circadian patterns around the world (e.g., Adan & Almirall, 1990). Psychometric assessments show that the questionnaire has good reliability and that scores on this test correlate with variation in body temperature, sleep-wake cycles, and periods of perceived alertness (Tankova, Adan, &

Buela-Casal, 1994; Vitiello, Smallwood, Avery, & Pascualy, 1986).

DEVELOPMENTAL AND INDIVIDUAL DIFFERENCES IN CIRCADIAN RHYTHMS

Using the MEQ, a substantial literature has accumulated confirming observations that children (8 to 16 years old), university students (18 to 25 years old), and older adults (50 years and older) have different time of day preferences (e.g., Adan & Almirall, 1990; Intons-Peterson, Rocchi, West, McLellan, & Hackney, 1998; Hasher & Goldstein, 2001; Kerkhof, 1985; Kim, et al., 2002; May & Hasher, 1998; May, Hasher, & Stoltzfus, 1993; Tankova, et al., 1994; Vitiello, Smallwood et al., 1986; Yoon, 1997). In general, most children prefer morning times, most younger adults prefer afternoon or evening times, and most older adults once again prefer morning times for both intellectual and physical activities.

In addition, within these broad age-based tendencies, there are substantial individual differences. As a result of these age and individual differences, general conclusions about the "optimal" time of day - conclusions that characterize most people regardless of age - can rarely be made.

Children and Adolescents

A small but growing literature has examined children's morningness-eveningness preferences (e.g., Bearpark & Michie, 1987; Carskadon, Vieira, & Acebo, 1993; Ishihara, Honma, & Miyake, 1990; Kim et al., 2002). For instance, Carskadon et al. (1993) reported a phase delay in children's sleep and wake-up time and concluded that biological rather than psychosocial (e.g., birth order and peer group) factors are causes of this sleep phase delay. Also, Bearpark and Michie (1987) examined the relation between the morning/evening preferences of 350 children aged 10 to 17 years and their sleep disturbances using a modified version of the MEQ. They reported that MEQ scores significantly decreased with age, moving towards an evening preference, and also reported that sleep disturbances (e.g., restless sleep) were related to evening preferences. Ishihara et al. (1990) examined changes in morningness-eveningness by Japanese females aged 9 to 15. They reported a similar finding, namely that with advancing grades, students changed their preference toward eveningness. Additionally, they argued that this circadian phase shift seemed to be established by around 12 years of age.

Until very recently, however, there was little normative data available on time of day preferences in children. However, Kim et al. (2002) provided such norms using a larger and more diverse sample than previously available: 900 US children ranging from 8 to 16 years of age, including boys and girls as well as children from five racial/ethnic groups (Asian, African American, Caucasian, Hispanic and Native American). This investigation used a children's version of the MEQ, adapted from Carskadon et al. (1993), called the Children's Morningness-Eveningness Preferences (CMEP) scale. This scale has 11 questions of the following sort: "Is it easy for you to get up in the morning?" and "Guess what! Your parents decided to let you set your own bedtime. What time would you pick?"

The Kim et al. findings on the relation between age and children's time of day preference are consistent with those of others (Bearpark & Michie, 1987; Carskadon et al., 1993; Ishihara et al., 1990) in that younger children's time of day preference was more toward morningness whereas that of older children's was more toward eveningness (Fig. 9.1). In particular, this shift toward eveningness appears to occur around the age of 13. The shift was seen for both boys as well as girls, and so cannot be attributed in any simple way to the onset of puberty (which occurs about two years earlier on average for girls than for boys). In addition, the shift was seen for students examined under a variety of educational and social conditions - specifically, for both summer, private school students and academic year, public school students - and so cannot easily be attributed to changes associated with time of year or social circumstances. This age shift finding is also similar to that of Ishihara et al. (1990) who argued on the basis of data collected on Japanese children that the circadian phase shift is established by the first year of junior high school (around 12 years).

It is worth re-emphasizing that older children start school even earlier than younger children in North America. Thus, it is ironic that as older children are moving away from being "early birds", their school day is shifting towards earlier start times. Of course, if these preference rhythms are associated with intellectual efforts and outcomes, then the school day is structured such that serious achievement problems may be created for some children, particularly those who are most shifted towards eveningness. Data collection relevant to this point is currently in progress (Hahn, Goldstein, Ralph, Hasher & Zelazo in progress).

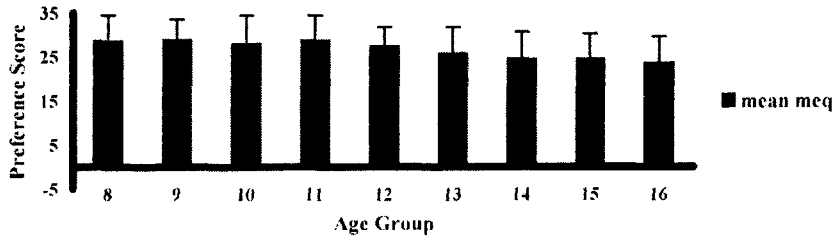


FIG. 9.1. Preference scores for children aged 8 to 16 years. (from Kim et al., 2002 adapted with permission from the publisher/authors).

In addition to evidence of an age–related shift away from morningness, the Kim et al. data (Fig. 9.1) also reveal that there is variability within each age group. So, while most young children tend towards being "early birds," not all of them are. While most older children tend away from being early birds, not all of them do so. It remains to be seen what the academic and social consequences are for those children who do not fit the "early bird" pattern. It is not inconceivable that for some children and young adolescents, especially those who are already vulnerable to academic failure, the consequences of being a "night owl" in an educational environment structured for "early birds" could be severe.

Younger and Older Adults

Comparable individual difference data in adults using the MEQ (Fig. 9.2) were collected by May et al. (1993, Study 1) who examined 210 university undergraduates between the ages of 18 and 22. Of these, 94% were either "Definitely Evening," "Moderately Evening," or "Neutral" types; only 6% were "Moderately Morning" types and *none* were "Definitely Morning" types. The "early bird" is a rare breed indeed on North American campuses.

May et al. also reported data from 91 older adults between the ages of 66 and 78. These data, also shown in Fig. 9.2, demonstrate that there is individual variability within this age group as well. However, in marked contrast to the data on university students, there are very few older adults who are evening types (less than 2%). Instead, nearly 75% of the older adults tested were morning types, falling into either the moderate or the definite category. These patterns have been confirmed with larger samples of younger and older adults (e.g., Intons-Peterson et al., 1998; May & Hasher, 1998; Yoon, May, & Hasher, 1999).

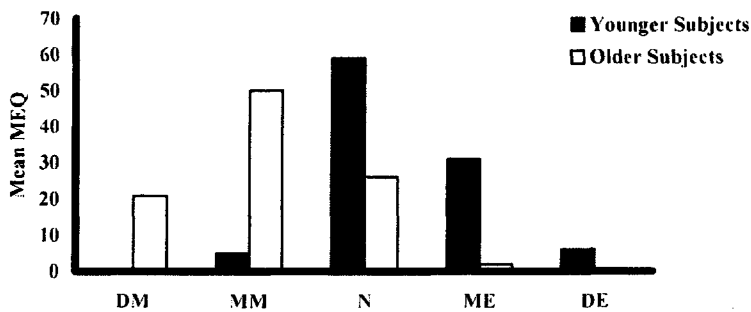


FIG. 9.2. Morningness-eveningness distributions. (from Yoon et al., 1999 adapted with permission from the publisher/authors).

The life-span trend, then, in time of day preference comes full circle - young children are primarily morning types, there is a fairly dramatic shift that begins at about the age of 12 away from morningness, and then at least by late adulthood the preference for morningness is re-established.

IMPLICATIONS OF AGE DIFFERENCES IN CIRCADIAN AROUSAL FOR COGNITIVE FUNCTIONING

The data reviewed above, then, clearly indicates that there is a major shift in arousal patterns from childhood to young adulthood and from young adulthood to old age. In addition, there are comparable age-related phase shifts that have been seen for older nonhuman animals as well (see below). In light of this shift in arousal patterns with age, two key questions arise: (1) What difference does this shift make for conclusions about development and cognition? (2) To what degree do age differences in circadian arousal contribute to observed differences in performance? Because data on children are currently unavailable, we address these questions using data that compare older and younger adults. We think it reasonable to presume that similar patterns will be found when comparable studies are done with children.

The key questions regarding circadian arousal and age were brought into focus by our attempts to test people across the day from early morning to late afternoon. A practical problem we faced was the difficulty we had in hiring undergraduates to test the older adults early in the morning when the latter actually wanted to be tested. Because virtually no undergraduates wanted to work at 8:30 or 9 AM, we found that we, like other investigators, were testing most participants after 11AM (May et al.). Thus, if arousal preference patterns influence cognitive functioning, we might be accidentally biasing our data to inflate age differences artificially.

This concern motivated our initial study on this topic (May et al., 1993, Study 2). Its participants were university students ranging in age from 18 to 24 and who fell into the evening-type range on the MEQ. The older adults were well educated and ranged in age from 60 to 76 and all fell into the morning-type on the MEQ. Participants read a series of stories followed by a recognition test that consisted of a mixed series of old and new sentences. Participants were to say "old" or "new" to each sentence as they saw it. The data (Fig. 9.3) are shown for corrected recognition, or hits (old sentences called old) minus false alarms (new sentences called old). The data for younger and older adults tested in the afternoon may be seen in the right panel of Fig. 9.3. Note that the young adults show a 35 percentage point advantage over the older adults. This advantage in favor of younger participants is not atypical. However, when recognition performance was tested in the morning, there were no age differences at all! Needless to say, this is a far from typical finding in the cognitive aging literature.

Looking at the full set of findings - based on younger and older adults who were tested early in the day (8 or 9AM) or late in the afternoon (4 or 5PM) - it is clear that younger adults' scores improve from the morning testing times to the afternoon times, while those of older adults decline. The 35% advantage for younger adults over older adults tested in the afternoon is reduced to a 20% advantage when both age groups were tested at near their peak times, morning for most older adults and afternoon for most young adults. Taken together with the fact that the majority of North American older adults are morning types, these data suggest the very important possibility that age-related differences can be exaggerated in the research literature whenever time of testing is uncontrolled.

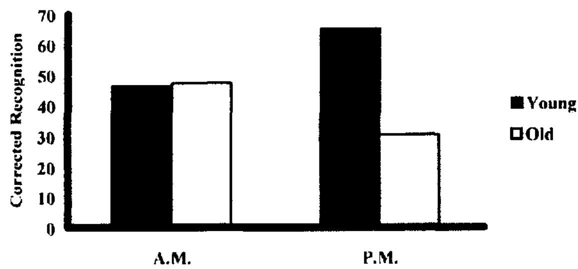


FIG. 9.3. Sentence recognition by testing time. (from May et al., 1993 adapted with permission from the publisher/authors).

As well, the data suggest that performance is better when it is assessed at an optimal time rather than at a non-optimal time. May and Hasher (1998) called the beneficial effect of a match between task demands and preferred time of day *the synchrony effect.*

THE SYNCHRONY EFFECT

The purpose of the first set of studies of the synchrony effect was to document declines in performance from optimal to non-optimal times of day. In most of these experiments, the participants are older adults who are morning types and younger adults who are evening types. Everyone is tested either first thing in the morning (8 or 9AM) or late in the afternoon (4 or 5PM).

One investigation of the synchrony effect (May, 1999) focused on people's ability to resist distraction in the context of a verbal problem solving task in which people solved a series of word problems, called remote associates. Each problem presents three words which can only be related to each other by generating a missing fourth word. The participant's task was to come up with that missing word. For example, the following three words can be related to each other by the word SICK: SEA, HOME, STOMACH. The three words were either presented alone or in combination with some distractors which were visually different from the targets. With respect to the solution word, the distraction could either be leading (i.e., helpful) - nausea, lonely, ache, or misleading (i.e., harmful) - horse, house, liver. Everyone was instructed to ignore the distraction.

Performance on the basic problem-solving task (presented with no distraction) did not vary with age or with time of testing - everyone got about a third of the problems correct. The focus of the analysis was on the costs and benefits of distraction: The difference between the proportions of items solved when no distraction was present versus when distraction was present. These data are presented in Fig. 9.4. Consider first the data for young adults: They are clearly bothered by distraction in the morning but, late in the afternoon, it is as if the distraction were invisible to them. The data for older adults show the opposite pattern with respect to time of testing: They show smaller distraction effects (both costs and benefits) in the morning than in the afternoon. These data are consistent with the proposed synchrony effect – better performance at optimal times of day. Further, if older and younger adults had been compared only in the morning, the obvious conclusion would have been that younger adults are more distractible than older adults.

On the other hand, if older and younger adults had been compared only in the afternoon, the data would suggest what we now know to be an exaggerated distraction effect for older adults compared to younger adults. Keeping in mind that everyone in this study was instructed to ignore the distraction, these data are theoretically interesting because they suggest that the goal driven regulation of attention can vary across the day. Based on the attentional regulation theory proposed by Hasher and Zacks (1988; see also Hasher, Zacks, & May, 1999), the data suggest that what gets encoded into memory is going to vary across the day, with larger, more cluttered "bundles" of information being encoded at non-optimal times than at optimal times. These larger bundles of information will include information that is irrelevant to the momentary goals of the participant, along with the relevant information.

What impact might this lapse of control have on "downstream" tasks like retrieval? One clue comes from the well established "fan effect" (e.g. Anderson & Bower, 1973). This effect shows that the larger the number of items that need to be searched through to find a target to recall, the slower and less accurate recall will be. The distraction findings from the remote associates task suggests that because memory bundles are likely to include more irrelevant information at non-optimal times than at optimal times (Carlson, Hasher, Connelly, & Zacks, 1995; Connelly, Hasher, & Zacks, 1991; Li, Hasher, Jonas, Rahhal, & May, 1998), memory performance will also be poorer at non-optimal times of day (Hasher, Zacks, & May, 1999). Indeed, the recognition study described above is consistent: It shows poorer memory at non-optimal times of day.

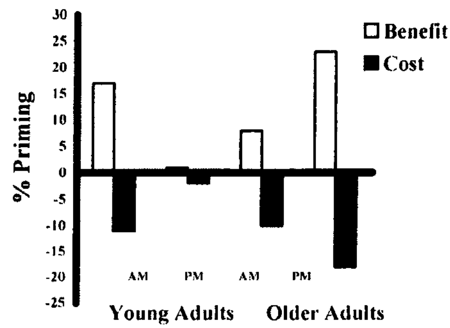


FIG 9.4. RAT performance: costs and benefits. (from May, 1999 adapted with permission from the publisher/authors).

We turn now to a series of studies on memory tasks; all show poorer performance at nonoptimal times of day. We begin with data from a simple word span task (Yoon et al., 1999). The focus of this study was the measurement of immediate memory using a verbal version of the classic digit span task, in which one starts with a short series of numbers and asks for serial order recall and then goes on to increasingly longer series of numbers. The data may be seen in Fig. 9.5. Note that young and old adults do not differ in the morning, with the younger adults improving their span across the day and the older adults showing a reduced span across the day.

The next set of data come from a long term memory task (reported in Winocur & Hasher, 2002 based on unpublished data from May) in which participants read two very brief stories (taken from the Logical Memory Test) and then recalled them immediately and again twenty minutes later. The recalls were scored for the essence of the facts in the original stories.

Fig. 9.6 presents the number of facts that were forgotten (based on how many were remembered on the immediate recall test) in the morning versus in the afternoon. Again, the data indicate that young and old participants do not differ much in the morning when both age groups forget very few facts. However, there is a substantial increase in forgetting across the day shown by older adults - they forget about 5 facts in the morning and just fewer than 14 in the afternoon, almost tripling the size of the forgetting effect.

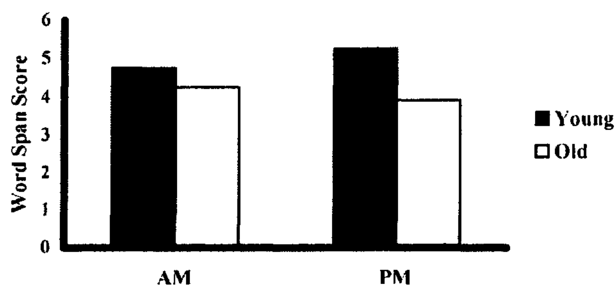


FIG. 9.5. Word span. (from Yoon et al., 1999 adapted with permission from the publisher/authors).

A final retrieval study was a cued recall test that used the first three letters of words as cues (May & Hasher, in preparation). A list of words was presented initially for pleasantness ratings, followed by a filled interval. At test, people were given the first three letters of words that had appeared on the list presented earlier in the study session. They were asked to use the cues to remember words on that list. Once again, as Fig. 9.7 shows, there are minimal differences between younger and older adults in the morning with substantial differences in the afternoon (the young improving while the older adults decline from morning to afternoon).

Thus, across a series of memory tasks, both recognition and free, cued and serial recall, and for several types of materials, young adults tend to improve their performance across the day, while older adults show a decline, conforming to a synchrony effect.

The memory findings are consistent with the notion that the reduced regulation of attention at the time of encoding results (e.g., May, 1999; Carlson et al., 1995) in larger bundles of information being stored (Hasher et al., 1999) which in turn reduces retrieval. The evidence clearly suggests that attentional regulation over distraction is impacted by circadian arousal cycles.

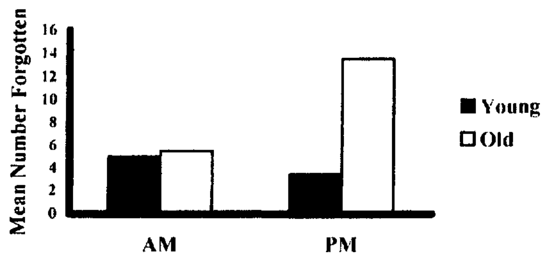


FIG. 9.6. Forgetting of story materials (from Winocur & Hasher, 2002 adapted with permission from the publisher/authors).

The memory literature also provides evidence of another form of attentional regulation difficulty – this occurring at the time of retrieval rather than at encoding. At non-optimal times, there is less monitoring for goal consistent responses, and less control over dominant responses. As a result, more "schema driven" behaviors will be seen than at optimal times (Bodenhausen, 1990). Strong, highly accessible responses will come to mind (and body) and will not get carefully examined for their appropriateness. Hence, errors tied to dominant responses that are not filtered will be heightened at non-optimal times of day. For example, Intons-Peterson et al. (1998) have shown that people make more false memory errors at non-optimal than at optimal times of day.

This failure to monitor what is produced at recall can also be seen in the "Moses Illusion" task in which people fail to monitor their output in response to queries. In this task, people are presented with a long series of questions. Consider, for example, the following questions:

"Who did Clark Kent turn into when he went into a telephone booth?"

"Who was the first president of the US?"

Mixed in with such questions are false items such as:

"How many animals of each type did Moses take on the ark?"

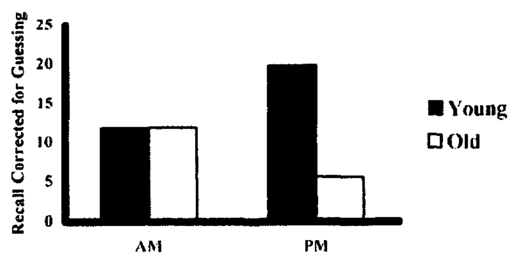


FIG. 9.7. Stem cued recall of words (May & Hasher, in preparation).

The apparent answer is two, but the question is false since it was Noah, not Moses, who was involved with the ark. Participants are forewarned that there will be trick questions and when such questions occur, they should not answer them. However, as one might imagine, people do indeed answer them. Fig. 9.8 shows the time of day effects for answering such questions for both older and younger adults. Once again a familiar pattern can be seen, with increased errors at non-optimal times (Yoon et al., 1999).

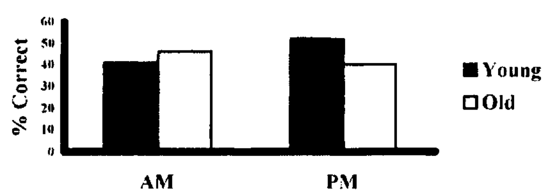


FIG. 9.8. Moses illusion errors (from Yoon et al., 1999 adapted by permission of the publisher/author).

Other work shows that at non-optimal times, people are more likely to use easily accessible stereotypes to judge individuals than they are at optimal times (Bodenhausen, 1990). These errors of thought can all be called "slips" of thought to relate them to the "slips" of action literature which shows that strong motor responses are less controllable at non-optimal times. By way of illustration of such slips, consider data from a stop signal task in which people have to occasionally withhold a response that they otherwise make most of the time. The response is a button press to a visual cue and, after this response is well established, an occasional tone sounds that signals people not to respond on that particular trial. See Fig. 9.9. Young adults make more errors in the morning than in the afternoon and older adults show the reverse pattern making more errors in the afternoon (May & Hasher, 1998).

Attentional regulation of strong responses, like attentional regulation over distraction, also appears to vary with circadian arousal and these variations are not limited to humans. Winocur and Hasher (1999) have shown a very similar pattern for old rats tested on a go/no-go task at the beginning versus the end of their activity cycle. Although their "go" response times do not change across the day, their ability to not respond is reduced at the end of their cycle. As well, the ability of old rats to perform a delayed non-matching to sample test (on which they have to reverse a previously produced response) is also reduced late in the activity cycle relative to early in the cycle (Winocur & Hasher, 2004).

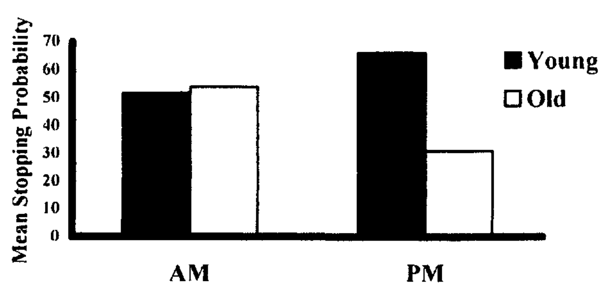


FIG. 9.9. Performance on stop signals trials (from May & Hasher, 1998 adapted by permission of the publisher/author).

At non-optimal times, individuals have less control over attentional processes, reducing their ability to ignore distraction and to evaluate responses for their appropriateness. As a direct consequence (from the Hasher et al., 1999 theoretical perspective), people are more likely to have difficulty remembering, and they are more likely to use highly accessible decision and retrieval routes rather than more difficult ones that involve analysis and evaluation. Thus, strong responses in both thought and action are more likely to be observed at off peak than at peak times of day.

Because strong responses are preserved at off peak times of day, any time a strong response is the correct one, no time of day differences will be found. Indeed, there are several reports in the literature showing such performance for very well established knowledge, such as vocabulary tests and the accuracy and speed of category membership decisions. For university educated adults, these are all highly over-learned responses and since the first response is correct, changes across the day are not seen (e.g., May & Hasher, 1998).

If retrieval of overlearned information is not impacted by arousal cycles, this suggests the possibility that experts can perform at very high levels, independent of the synchrony between their arousal levels and the time of performance. To our knowledge, there are no such data in the literature.

CONCLUSION

A few generalizations are suggested by the results of the studies presented here. The research reviewed here confirms the earlier observations by Bodenhausen (1990) and by May et al. (1993) that synchrony between individual preferences and the time of testing is a powerful effect – even within an age group of individuals with similar arousal preferences. The data suggest that only highly practiced responses are invariant across the day – all others will be impacted. We emphasize that attentional regulation over both incoming information and outgoing responses are particularly vulnerable to time of day effects.

Second, there are large differences in circadian cycles between young and older adults, and a real assessment of the extent and nature of cognitive declines requires that investigators attend to when they are testing their participants, particularly because most older adults are morning people. A casual perusal of the aging literature suggests that time of testing is not an important feature of most studies. Our own work suggests that failing to attend to such differences may lead to an overestimate of age differences. Further, to the degree that people are tested on tasks and materials that are not in their domains of expertise, the single exception we have found for time of testing effects, there is every reason to expect that the cognitive gerontology literature exaggerates actual age differences.

The developmental literature, too, possibly gives us an inaccurate picture of age differences in cognitive functioning since even more than the aging literature, there is no acknowledgement that there are age and individual differences in arousal patterns. In our own work, we are especially interested in the possibility that academic achievement may be disrupted by a mismatch between the individual's arousal cycle and the school day pattern. Finally, there may be real world consequences for those people whose natural rhythms put them out of synchrony with their environment. These include evening type university students forced to take organic chemistry at 8AM and evening type high school students taught math at 9AM.

REFERENCES

Adan, A., & Almirall, H. (1990). Adaptation and standardization of a Spanish version of the morningness eveningness questionnaire: Individual differences. *Personality and Individual Differences*, *11*, 1123-1130.

Anderson, J. R., & Bower, G. H. (1973). *Human Associative Memory*. Washington, DC: Winston.

Baddeley, A. D., Hatter, J. E., Scott, D., & Snashall, A. (1970). Memory and time of day. *Quarterly Journal of Experimental Psychology*, *22*, 605-609.

Bearpark, H., & Michie, P. (1987). Changes in morningness-eveningness scores during adolescence and their relationship to sleep/wake disturbances. *Chronobiologia*, *14*, 151.

Bergstrum, F. G. (1894). An experimental study of some of the conditions of mental activity. *American Journal of Psychology*, *6*, 247-274.

Blake, M. J. F. (1967). Time of day effects on performance in a range of tasks. *Psychonomic Science*, *9*, 349-350.

Bodenhausen, G. V. (1990). Stereotypes as judgmental heuristics: Evidence of circadian variations in discrimination. *Psychological Science*, *1*, 319-322.

Carlson, M. C., Hasher, L., Connelly, S. L., & Zacks, R. T. (1995). Aging, distraction, and the benefits of predicable location. *Psychology and Aging*, *10*, 427-436.

Carskadon, M. A., Vieira, C., & Acebo, C. (1993). Association between puberty and delayed phase preference. *Sleep*, *16*(3), 258-262.

Carskadon, M. A., Wolfson, A. R., Acebo, C., Tzischinsky, O., & Seifer, R. (1998). Adolescent sleep patterns, circadian timing, and sleepiness as a transition to early school days. *Sleep*, *21*(8), 871-881.

Connelly, S. L., Hasher, L., & Zacks, R. T. (1991). Age and reading: The impact of distraction. *Psychology and Aging*, *6*, 533-541.

Dunn, R., Dunn, K., Primavera, L., Sinatra, R., & Virostko, J. (1987). A timely solution: Effects of chronobiology on achievement and behavior. *The Clearing House*, *61*, 5-8.

Folkard, S. (1983). Diurnal variation. In G. R. J. Hockey (Ed.), *Stress and Fatigue in Human Performance* (pp. 245-272). New York: Wiley.

Gates, A. I. (1916). Diurnal variation in memory and association. *University of California Publications in Psychology*, *1*, 323-344.

Hardin, P. E. (2000). From biological clock to biological rhythms. *Genome Biology*, *1*(4), 1023.1-1023.5.

Hasher, L., & Goldstein, D. (2001). Aging, cognition, and circadian rhythm. *Geriatrics and Aging*, *4*(6), 12, 33.

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In G. H. Bower (Ed.), *The Psychology of Learning and Motivation Vol. 22*, (pp. 193-225). New York: Academic Press.

Hasher, L., Zacks, R. T., & May, C. P. (1999). Inhibitory control, circadian arousal, and age. In D. Gopher & A. Koriat (Eds.), *Attention and Performance XVII; Cognitive Regulation of Performance: Interaction of Theory and Application* (pp. 653-675). Cambridge, MA: MIT Press.

Horne, J., & Ostberg, O. (1976). A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythm. *International Journal of Chronobiology*, *4*, 97-110.

Horne, J., & Ostberg, O. (1977). Individual differences in human circadian rhythms. *Biological Psychology*, *5*, 179-190.

Hrushesky, W. J. M. (1994). Timing is everything. *The Sciences*, *3*, 32-37.

Intons-Peterson, M. J., Rocchi, P., West, T., McLellan, K., & Hackney, A. (1998). Aging, optimal testing times, and negative priming. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *24*, 362-376.

Ishihara, K., Honma, Y., & Miyake, S. (1990). Investigation of the children's version of the morningness-eveningness questionnaire with primary and junior high school pupils in Japan. *Perception and Motor Skills*, *71*, 1353-1354.

Kerkhof, G. A. (1985). Inter-individual differences in the human circadian system: A review. *Biological Psychology*, *20*, 83-112.

Kim, S., Dueker, G. L., Hasher, L., & Goldstein, D. (2002). Children's time of day preference: Age, gender and ethnic differences. *Personality and Individual Differences*, *33*, 1083-1090.

Li, K. Z. H., Hasher, L., Jonas, D., Rahhal, T. A., & May, C. P. (1998). Distractibility, circadian arousal, and aging: A boundary condition? *Psychology and Aging*, *13*, 574-583.

Lombard, W. P. (1887). The variations of the normal knee jerk, and their relation to the activity of the central nervous system. *American Journal of Psychology*, *1*, 5-71.

May, C. P. (1999). Synchrony effect in cognition: The costs and a benefit. *Psychonomic Bulletin and Review*, *6*, 142-147.

May, C. P., & Hasher, L. (1998). Synchrony effects in inhibitory control over thought and action. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(2), 363-379.

May, C. P., Hasher, L., & Stoltzfus, E. R. (1993). Optimal time of day and the magnitude of age differences in memory. *Psychological Science*, *4*, 326-330.

Monk, T. H. (1989). Circadian rhythm. *Clinics in Geraitric Medicine*, *5*(2), 331-346.

Moore-Ede, M. C., Sulzman, F. M., & Fuller, C. A. (1982). *The clocks that time us*. Boston: Harvard University Press.

Panda, S., Sato, T., Castrucci, A., Rollag, M., DeGrip, W., Hogenesch, J., Provencio, I., & Kay, S. (2002). Melanopsin (Opn4) Requirement for Normal Light-Induced Circadian Phase Shifting. *Science*, *298*, 2213-2216.

Tankova, I., Adan, A., & Buela-Casal, G. (1994). Circadian typology and individual differences. A review. *Personality and Individual Differences*, *16*(5), 671-684.

Vitiello, M. V., Smallwood, R. G., Avery, D. H., & Pascualy, R. A. (1986). Circadian temperature rhythms in young adult and aged men. *Neurobiology of Aging*, *7*, 97-110.

Winocur, G., & Hasher, L. (1999). Aging and time-of-day effects on cognition in rats. *Behavioral Neuroscience*, *113*, 991-997.

Winocur, G., & Hasher, L. (2002). Circadian rhythms and memory in aged humans and animals. In L. R. Squire & D. L. Schacter (Eds.), *Neuropsychology of Memory*, (3rd ed.) (pp. 273-285). New York: The Guilford Press.

Winocur, G., & Hasher, L. (2004). Age and time-of-day effects on learning and memory in a non-matching-to-sample test *Neurobiology of Aging*. Forthcoming.

Yoon, C. (1997). Age differences in consumers' processing strategies: An investigation of moderating influences. *Journal of Consumer Research*, *24*, 329-342.

Yoon, C., May, C. P., & Hasher, L. (1999). Aging, Circadian Arousal Patterns, and Cognition. In N. Schwarz, D. Park, B., Knauper, & S. Sudman (Eds.), *Cognition, Aging, and Self-Reports* (pp. 117-143). Philadelphia, PA: Psychology Press.

This page intentionally left blank

Chapter 10

Fuzzy-Trace Theory: Memory

C. J. Brainerd
University of Arizona, U.S.A.

Fuzzy-trace theory (FTT) concerns the interface between memory processes and higher reasoning operations in adults and the developmental relations between memory and reasoning: Memory is the bridge between the domains and the memory-reasoning interface is a promising place to initiate integrative theorizing. The research program on this focus consists of three categories of studies: First, studies that revolve around on-line or developmental relations between memory content and reasoning accuracy; second, those concerned with reasoning per se; and third, those concerned with memory per se, especially aspects of memory that savor of reasoning and inference. In this chapter, I sketch contributions that fall in the last category. The other two categories figure in the companion chapter by Reyna (Chap. 11).

FIVE EXPLANATORY PRINCIPLES

In the memory literature, FTT is usually called a dual-process theory (Seamon, Lee, Toner, Wheeler, Goodkind, & Birch, 2002; Seamon, Luo, Schwartz, Jones, Lee, & Jones, 2002) , but we consider it more accurate to call it an opponent-processes theory. The key ideas deployed to explain false-memory effects, and also to gain predictive leverage on such effects, are *parallel storage*, *dissociated retrieval*, *opponent judgments about false-memory items*, *different time courses of verbatim and gist memory*, and *developmental variability*. These principles are defined in Table 10.1. I now elucidate them with supporting data.

TABLE 10.1

Explanatory Principles

Principle	Definition
Parallel storage	Verbatim and gist traces of experience are deposited in parallel, rather than gist traces being extracted from previously stored verbatim traces.
Dissociated retrieval	Memory performance involves dissociated retrieval of verbatim and gist traces, so that observed levels of false memory depend upon the mix of verbatim and gist processing.
Opponent judgments	Verbatim and gist retrieval both support true memory for experienced events, but they have opposite effects on false memory for events that preserve the meaning of experience.
Different time courses	Over time, the accessibility of verbatim traces declines more rapidly than that of gist traces, yielding a net shift towards reliance of memory performance on the latter.
Developmental variability	Between early childhood and adult, there are developmental changes in both verbatim and gist storage, in subjects' ability to access these representations on memory tests, and in subjects' long-term retention of these representations.

Principle 1: Parallel Storage

Experimental evidence points to the conclusion that subjects store separate verbatim and gist traces of experience. Because verbatim and gist traces contain information that originates in the same events, it is natural to assume that their storage would be serially dependent; that verbatim traces would be stored first and then be processed to extract their gist. It turns out, however, that data on subliminal semantic activation have challenged this common-sense model (e.g., Abrams & Greenwald, 2000): Adults begin to store the meaning content of target events within 30-50 milliseconds after onset, long before targets' surface forms can be fully processed. There are several effects that have been reported that seem to be consequences of this fast parallel extraction of meaning. One example is the word-superiority effect (Ankrum & Palmer, 1989), wherein familiar words are presented at very fast exposures (e.g., less than 100 ms) and subjects are then given either a probe word to identify (e.g., "Was the word TABLE?") or a probe letter to identify (e.g., "Did the word contain the letter B?"). Whole words can be recognized earlier than their constituent letters.

FTT assumes that the initial encoding of targets' surface features initiates meaning access and elaboration, with verbatim and gist storage running to completion in parallel as encoding continues. A crucial implication is that subjects may preserve considerable information about the meaning of experience in memory even if they fail to completely process its surface form. Another important consideration is that because targets participate in multiple meanings, multiple gist traces can be stored on the basis of a single target, and those gist traces targets may vary in their level of specificity. For instance, the words ANGUS, POODLE, and ROBIN would likely all cause subjects to access the concept "animal," as well as the concepts "cow," "dog," and "bird."

Principle 2: Dissociated Retrieval of Verbatim and Gist Traces

If subjects store dissociated verbatim and gist traces of experience, the types of representations that are accessed on memory tests will depend upon the retrieval cues of the tests. We have found that as long as verbatim traces are still accessible, target probes are better retrieval cues for verbatim than for gist traces on recognition tests and both hits (on recognition tests) and target recall (on recall tests) are based predominately upon the retrieval of verbatim traces (Reyna & Kiernan, 1994; Reyna & Brainerd, 1995). We reported two parallel findings about false memory: first, semantically-related distractors are better retrieval cues for gist than for verbatim traces on recognition tests and, second, both semantic false alarms (on recognition tests) and intrusions (on recall tests) are based predominately upon retrieval of gist rather than verbatim traces (Brainerd & Reyna, 1998a, 1998b). "Retrieval dissociation" is used to explain a seemingly paradoxical finding about false memory that is sometimes called the *coexistence constraint*--namely, that subjects can provide both true- and false-memory reports about the same event. Although this troubles our intuition, FTT says that it is a natural by-product of the first two principles, according to which memory is of two minds--minds that are not well integrated with each other, either when memories are first stored or when they are subsequently retrieved.

Principle 3: Opponent Judgments About False-Memory Items

FTT assumes that retrieval of verbatim traces can induce a vivid form of remembering in which subjects consciously re-experience targets' prior occurrence in particular contexts; targets echo in the mind's ear or flash in the mind's eye. This remembering phenomenology was first studied by

Strong (1913) and is usually called *recollection* nowadays (e.g., Jacoby, 1991). On recognition tests, verbatim retrieval is thought to support mental comparisons between retrieved memories and test probes in which memories and probes are perceived to exactly match, and on recall tests, it is thought to support simple readout of targets' surface forms from consciousness. Because verbatim retrieval predominates with targets, this is the principal basis for true-memory responses. However, targets may instead provoke gist retrieval, which usually induces a more global and inchoate remembering phenomenology that was also studied by Strong and is called familiarity nowadays. On recognition tests, gist retrieval supports mental comparisons between retrieved memories and test probes in which probes' meaning content is perceived to overlap with meanings that were processed during target experiences, and on recall tests, it supports meaning-based regeneration of targets. Note that verbatim and gist retrieval are *convergent* processes in true memory because both lead to recognition and recall of targets.

However, when it comes to false-memory responses, false alarms and intrusions, FTT assumes that verbatim and gist retrieval are *opponent* processes, with gist retrieval supporting false-memory responses and verbatim retrieval suppressing them. We know from Principle 2 that gist retrieval predominates with false-but-meaning-preserving items, so that such items will customarily induce global feelings of meaning, overlap with target experiences, which supports semantic false alarms and intrusions. However, such items may sometimes produce retrieval of verbatim traces of the occurrence of the corresponding target experiences (e.g., the recognition probe "ate a hot dog at the baseball game" may provoke retrieval of verbatim traces of eating a hamburger, or reconstruction of the false event "drank a Coke" while recalling events from a baseball game may provoke retrieval of verbatim traces of drinking 7-Up). Simultaneous processing of two related items (one is *known* to have been experienced because verbatim traces are available that induce recollective phenomenology and the other is similar in meaning but different in surface form) can generate mismatches at the level of verbatim detail. On the one hand, hot dogs and hamburgers have similar meanings because both are sandwiches that are commonly consumed at baseball games, and Coke and 7-Up have similar meanings because both are soft drinks that are commonly drunk at baseball games. But on the other hand, hot dogs mismatch with hamburgers in palpable ways and so does Coke with 7-Up. Such verbatim mismatches provide principled grounds for rejecting distracter probes in recognition and for suppressing semantic intrusions in recall because they supply a compelling events that were not experienced could nevertheless seem very familiar.

An important qualification is in order with respect to the phenomenology that accompanies gist retrieval. FTT assumes that on recognition tests, targets predominately prompt verbatim retrieval, which induces recollective phenomenology, whereas false-but-gist-consistent items predominately prompt gist retrieval, which typically includes familiarity phenomenology. This assumption is consistent with the results of numerous studies using Tulving's (1985) remember/know methodology (e.g., Conway, Collins, Gathercole, & Anderson, 1996; Donaldson, 1996). However, FTT also posits that there are circumstances in which gist retrieval induces illusory vivid recollective phenomenology that resembles the true recollective phenomenology that is induced by verbatim retrieval (e.g., the prior "presentation" of a false-but-meaning-preserving item echoes in the mind's ear or flashes in the mind's eye). This is the phenomenon of *phantom recollection* (Brainerd, Wright, Reyna, & Mojardin, 2001). Experiments that have been conducted to date demonstrate that phantom recollection can occur at high levels when (a) a familiar meaning (e.g., "medicine," "furniture") has been repeatedly cued by target experiences and (b) the false-but-gist-consistent item is an especially good instance of that meaning. Examples of some materials that produce phantom recollection of false memories very consistently, which were originally developed by Deese (1959; see also Roediger & McDermott, 1995), are shown in Table 10.2.

TABLE 10.2
Sample Lists that Produce High Levels of Phantom Recollection of Unpresented Stimulus Words (Based on Deese, 1959; Roediger & McDermott, 1995)

List Themes	Targets
Furniture words	table, sit, legs, seat, couch, desk, recliner, sofa, wood, cushion, swivel, stool, sitting, rocking, bench (unpresented stimulus word: *chair*)
Medical words	nurse, sick, lawyer, medicine, health, hospital, dentist, physician, ill, patient, office, stethoscope, surgeon, clinic, cure (unpresented stimulus word: *doctor*)

The key point about phantom recollection is that some false events are accompanied by an illusory vivid phenomenology that emulates the true recollective phenomenology that is induced by retrieval of verbatim traces.

Consequently, it is very difficult to weed those particular events out of memory reports via retrieval of verbatim traces.

Principle 4: Different Time Courses of Verbatim and Gist Memory

In addition to Principle 2, relative retrievability must also be influenced by differences in the tendency of verbatim and gist traces to be preserved and to remain accessible in memory. In this connection, numerous findings (e.g., Murphy & Shapiro, 1994) converge on the conclusion that over time, the accessibility of verbatim traces declines more rapidly than that of gist traces. The straightforward implication is that the tendency for true recognition or true recall to be based predominately upon verbatim retrieval will depend on how much time has passed since the target experience. This means that the high levels of statistical independence and experimental dissociation initially observed between true- and false-memory responses ought to be replaced by dependency and association as time passes, an outcome that has been reported by several investigators (e.g., Brainerd, Reyna, & Kneer, 1995; Reyna & Kiernan, 1994, 1995). Two other implications are: (a) False-memory responses will be fairly stable over time because the accessibility of gist memories declines slowly and (b) subjects' ability to avoid false recognition and false recall via verbatim mismatches will be compromised by the passage of time because the verbatim traces that yield such mismatches are rapidly becoming inaccessible.

Principle 5: Developmental Variability

Based upon findings from classical developmental work, FTT posits age changes in both verbatim and gist storage, in the ability to access such traces on memory tests, and in long-term retention. Relevant findings come from studies whose designs provide separate estimates of the strengths of verbatim and gist memory. The two standard patterns in such studies are, first, estimates of verbatim and gist memory both improve between early childhood and young adulthood, and second, these improvements are independent of each other. These patterns are illustrated by two familiar results from developmental studies of free recall. First, recall of individual targets (a measure of verbatim memory) and the clustering together of meaning-sharing targets during recall (a measure of gist memory) both increase during the preschool to young adult age years (Bjorklund & Muir, 1988). Second, between early childhood and early adolescence, the age range during which most of the improvement in both measures occur, subjects' levels of performance on the two measures are statistically independent and experimentally dissociated (e.g., DeMarie-Dreblow, 1991).

The traditional view of developmental variability in false memory is that susceptibility to false recognition and false recall decrease steadily between early childhood and young adulthood and then increase again late in life (e.g., Ceci & Bruck, 1993). Although there are several experimental findings that are consistent with this scenario, FTT posits more complex patterns of developmental variability. It assumes, to begin, that there is not a monolithic developmental trend in false recognition/recall that holds for all memory tasks. If verbatim and gist abilities both contribute to false alarms and intrusions, albeit in opposite ways, and if both vary with age, it follows that age differences in measured levels of false memory will be highly task dependent. In particular, it is easy to see that such age differences will turn on whether a false-memory task poses greater obstacles to verbatim memory than to gist memory, or whether conversely, it poses greater obstacles to gist memory than to verbatim memory.

Suppose that the former is the case; that a false-memory task is hard from a verbatim point of view but easy from a gist point of view. To illustrate, the targets might consist of a series of statements with very similar surface forms, making it difficult to store distinctive verbatim traces of each sentence, but simple meanings that are familiar to young children, making it easy to store gist traces that support false-memory responses (e.g., John is taller than Jim. Jim is taller than Sam. Sam is taller than Stan. Stan is taller than Fred. Fred is taller than Mike. Mike is taller than Ron. Ron is taller than Dave.) Here, the relevant gist memory (a linear ordering with respect to height) ought to be easy to store, but the verbatim side of the information is confusing. If this task were administered to elementary schoolers, one would expect developmental decreases in false recognition or false recall of statements such as "Stan is taller than Ron" and "Jim is taller than Mike" because age improvements in verbatim memory contribute more to performance than developmental improvements in gist memory (see Reyna & Kiernan, 1994). In contrast, suppose that the task is easy from a verbatim point of view (e.g., a narrative consisting of a series of highly distinctive sentences) but hard from a gist point of view (e.g., the narrative revolves around a mature theme that is unfamiliar to most young children). Here, one expects the opposite developmental trend, increasing levels of false memory, because age improvements in gist memory (resulting from increased acquaintance with the theme) will be more pronounced in this task than age improvements in verbatim memory (Brainerd & Reyna, 2001).

PREDICTIVE CONTROL OF FALSE MEMORIES

Beyond the job of explaining data patterns, the other task of a theory of false memory is to control its occurrence by being able to forecast new aspects of the phenomenon. The most interesting predictions are those that seem surprising when judged by the yardstick of common-sense or the principles of other currently-accepted theories. Such predictions have been confirmed in applications of FTT to false memory. Often, predictions are formulated and tested because they contrast with those of two early accounts of false memory, constructivism (Bransford & Franks, 1971) and the source-monitoring framework (Johnson, Hashtroudi, & Lindsay, 1993). These accounts are both *one-process* models in the sense that they posit that false recognition and false recall are based on a common memory code (constructivism) or on a common retrieval process (source monitoring). In contrast, FTT is a dual opponent-processes explanation of false memory, and this fact makes it possible to generate a number of differential predictions. In the remainder of the chapter, I will describe findings on five such contrasting predictions.

Dissociations Between True and False Memories

Positive associations between true- and false-memory responses are basal predictions of any theory which assumes that the two types of responses are based on common representational or retrieval mechanisms (Reyna & Lloyd, 1997). Under FTT's opponent-processes analysis, however, the expectation is that there will be true-false dissociations in the presence of certain conditions, but true-false associations in the presence of others. Such predictions fall out of the first three principles above. It follows that true-false dissociations should be observed in situations that encourage reliance on verbatim traces as the basis for true responses, such as the administration of immediate tests for memorable material, so that true and false responses will be based on different representations. Early data confirmed this prediction (Reyna & Kiernan, 1994, 1995) in experiments on memory for narrative material. See Fig. 10.1. The subjects listened to short statements (with hits being the measure of true memory), unpresented statements that preserved narrative meaning (with false alarms being the measure of false memory), and unpresented statements that violated narrative meaning (with false alarms being the measure of response bias). The individual true- and false-memory statements produced different levels of recognition, of course. Fig. 10.1 shows the pattern of statistical association between the two types of statements. The values plotted on the abscissa are unconditional probabilities of false alarms to the false-memory statements, and the values plotted on the ordinate are conditional probabilities of false alarms to the same false-memory statements *given that the corresponding true memory statements had also been accepted.* One-process theories predict that the two sets of values should be positively related, which means that the plotted points ought to fall in the upper half of the graph. However, FTT predicts that the two sets of values will be stochastically independent and that the points therefore ought to fall in the center of the graph as a flat line, which they did, indeed.

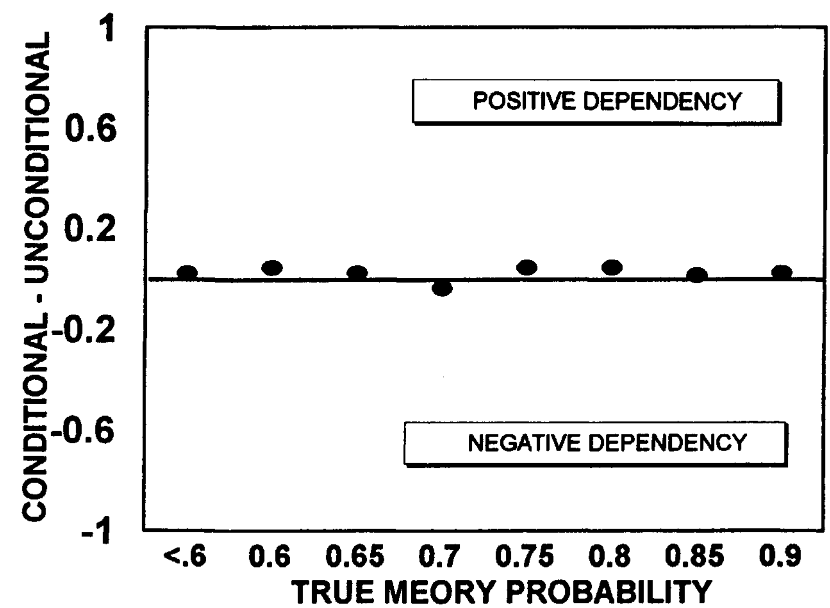


FIG. 10.1. Stochastic independence of hits and semantic false alarms in Reyna and Kiernan's (1994) studies of false recognition of narrative sentences

On the other hand, the theory says that associations between the same measures of true and false memory should result if conditions are imposed that enforce reliance on gist traces, rather than verbatim traces, as the basis for true-memory responses. Here, the obvious manipulation involves the fourth principle above: Administration of memory tests after a delay that is long enough so as to allow the differential forgetting rates for verbatim and gist traces to take their course. The same types of materials and tests (on which the data in the last two figures are based) have been administered following delays of one to three weeks. The consistent finding has been that the true-false dissociations that are obtained on immediate tests are replaced by true-false associations on delayed tests (Reyna & Kiernan, 1994, 1995; Brainerd & Mojardin, 1998).

Persistence of False Memories

What would common-sense expect about the relation between false-memory responses and the passage of time? Because such responses refer to events that were never experienced, common-sense says that false-memory responses lack the genuine memorial support that true-memory responses enjoy. So, false-memory responses ought to be highly unstable, and even though the stability of true-memory responses is imperfect, it should greatly exceed that of false-memory responses. Indeed, these two ideas are so self-evident that they are enshrined in the law, where they form what is known as the consistency rule of testimony. When different witnesses testify about their memories of the same target events, it is not uncommon for them to give different accounts of crucial features of the events, so that triers of fact must decide which version to accept. Fisher and Cutler (1992) and others have found that the criterion that is most commonly applied is the consistency of witnesses' memory reports. Specifically, it is assumed that events that are inconsistently reported are less likely to be true than events that are consistently reported.

Contrary to these common-sense expectations, FTT's opponent-processes distinctions lead to the predictions that false-memory reports can be quite stable over time, and that under certain conditions, they can be more stable than true-memory reports. These two predictions, which are collectively known as *false-persistence effects,* are straightforward consequences of the notions that initial false-memory responses are rooted in gist memory, that initial true-memory responses are slanted towards verbatim memory, and that over time, verbatim memories become inaccessible more rapidly than gist. Obviously, the long-term stability of false-memory responses falls out of these considerations, but the prediction that they can sometimes be more stable than true-memory responses does too: If the same (stable) memorial content supports initial and subsequent false-memory reports while the initial (verbatim) basis for true-memory reports becomes rapidly inaccessible, initial false reports will be more stable in situations in which they are overwhelming gist-based (rather than due to guessing or prevarication) and initial true-memory reports are overwhelmingly verbatim-based.

The first prediction was originally investigated in a series of experiments that were conducted in our laboratory (Brainerd, Reyna, & Brandse, 1995). The level of false-memory consistency can be assessed by simply computing correlations between individual false-memory responses on initial and subsequent memory tests. If these responses are ephemeral, the fact that some false event A is erroneously reported on an initial test will not increase the chances that it is reported on a subsequent test. In our experiments, we evaluated this possibility using a design in which subjects first studied fairly lengthy word lists and then responded to a recognition test containing targets, semantically-related distractors (the names of categories whose exemplars had been presented as targets), and unrelated distractors. One week later, the same recognition test was repeated. The measure of false-memory persistence was the relation between the *unconditional* probability of a false alarm to a related distracter on the delayed test and the *conditional* probability of a delayed false alarm given an initial false alarm to that same distracter. False memories are not persistent if the values of these two measures are equal, but they are persistent if the conditional probability is greater than the unconditional probability. In our experiments, the conditional probability was always the greater of the two by a wide margin, so that false alarms to meaning-preserving distractors were not only persistent, they were highly persistent.

The other prediction, that the survival of false memories over time can sometimes be superior to the survival of true memories, has typically been studied with a two-session independent-groups design. All subjects are exposed to the same set of target materials. One group of subjects then receives an immediate memory test, while another group receives a delayed memory test a few days or weeks later. The true- and false-memory performance of subjects receiving delayed tests is compared to that of subjects receiving immediate tests, and the amount of decline between the immediate and delayed sessions is computed for both types of performance.

An early experiment that implemented this design and produced the pattern predicted by FTT was reported by Payne, Elie, Blackwell, and Neuschatz (1996). The target materials were from Deese (1959) lists shown earlier. Since Payne et al. reported this pattern of delayed declines in true-memory responses, without corresponding declines in false-memory responses, similar data have often been reported (e.g., Seamon, Luo, Kopecky, Price, Rothschild, Fung, & Schwartz, 2002; Thapar & McDermott, 2001; Toglia, Neuschatz, & Goodwin., 1999).

The Creation of False Memories by Mere Memory Testing

There is another important assumption that the law makes about true and false memory that conflicts with the predictions of FTT and with data. In certain cases, the law recognizes that false-memory responses can be implanted by investigative interviews that suggest events to witnesses that support charges made against defendants. However, the law also assumes that as long as interviewers do not offer specific suggestions and merely provide recall prompts (e.g., "Tell me about the robber.") and recognition probes (e.g., "Did the robber have a gun in his hand?"), investigative interviews will not elevate levels of false reporting during subsequent interviews and sworn testimony. Further, on the contrary, their principal effect will be a beneficial one--specifically, to inoculate true memories against forgetting (Brainerd & Ornstein, 1991). However, FTT predicts, on the basis of Principles 2 and 3, that neutral, nonsuggestive memory questions can elevate later false-memory levels under certain conditions. Those conditions are ones in which subjects are apt to respond to questions by retrieving gist traces. False-memory levels ought to rise under such conditions, because subjects are becoming practiced at accessing and processing the very types of representations that support false-memory responses.

In 1996, our laboratory reported some experiments confirming this prediction for neutral, nonsuggestive recognition tests, and Payne et al. (1996) confirmed this prediction for neutral, nonsuggestive recall tests. In our experiments, the subjects first studied a long list of familiar words, such as DIAMOND, TABLE, COLLIE, and OAK. Some of the words on the list were presented once (low repetition) and some were presented three times (high repetition). Next, the subjects responded to a standard recognition test composed of half target probes and half unpresented distracter probes. The false-memory items were distractors that were the labels of categories to which presented words belonged (e.g., the distracter JEWEL might be substituted for the target DIAMOND or the distractor FURNITURE might be substituted for the target TABLE) or they were exemplars that belonged to the same categories as presented words (e.g., the distractor POODLE might be substituted for the target COLLIE or the distractor PINE might be substituted for the target OAK). One week later, the subjects returned for a second recognition test. On the second test, half of the false-memory items had also been tested earlier and half had not. We compared the false-alarm rates for these semantically-related distractors on the second test for items that were previously tested versus previously untested. The resulting pattern in Fig. 10.2 was quite clear. Although the first test was nonsuggestive, indeed subjects were carefully instructed that half of the items on that test would be distractors, false-alarm rates on the second test were substantially higher for distractors that had appeared on the first test. Interestingly, false-alarm rates were elevated on the second test *regardless of whether subjects had responded correctly or incorrectly to these items on the first test.*

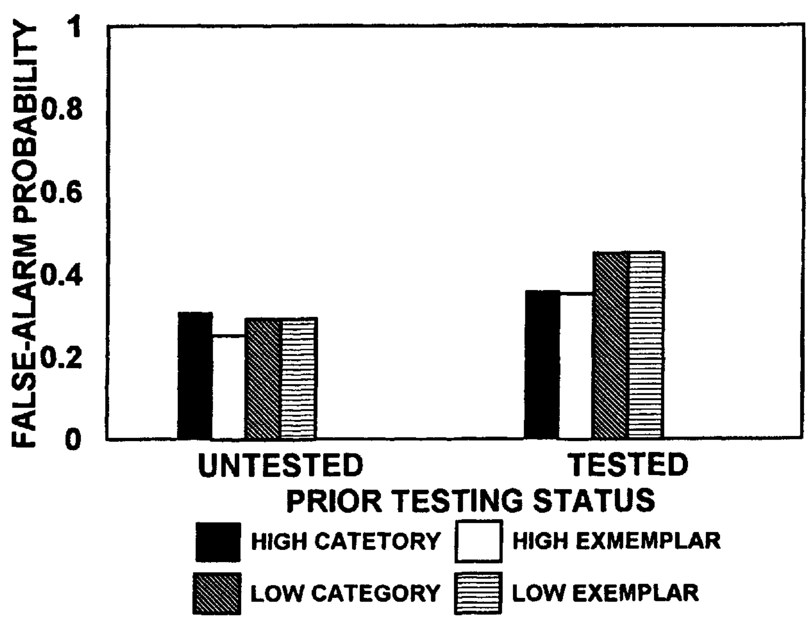


FIG. 10.2. Effects of prior recognition tests for semantically-related distracters on subsequent false-alarm rates for these distracters (based upon Brainerd & Reyna, 1996).

Developmental Increases in Semantic False Alarms and Intrusions

Another familiar generalization about false memory in the law as well as the memory literature is that young children are more prone to make false-memory responses than adolescents or adults. However, FTT predicts that, in fact, there is no single, monolithic age trend and that, indeed, there are some important forms of false memory that ought to *increase* with age. Specifically, it is not difficult to see that a straightforward implication of the third and fourth principles above is that depending upon the particular gist memories that support a specific type of false-memory response and the age levels that are being compared, the incidence of that response may be found to increase, decrease, or remain invariant with age. For instance, suppose that we are dealing with false-memory responses that are supported by gist memories of individual items whose meanings are well understood by young children (e.g., the names of familiar objects, such as WATER, TABLE, SUN, and SNAKE, or pictures of familiar objects). Developmental improvements in meaning storage will be negligible, while developmental improvements in verbatim memory will be substantial, and such responses should therefore decline with age because improvements in verbatim suppression will swamp improvements in gist support.

Suppose, in contrast, that we are dealing with false-memory responses that either (a) are supported by gist memories of individual items whose meanings are unfamiliar to young children (e.g., words such as BUZZARD, JUGGLER, and HYENA, or narratives that involve unfamiliar meanings, such as adult emotional relationships) or (b) are supported by gist memories that are formed by connecting meaning across items whose individual meanings are understood (e.g., simple declarative sentences). Here, both of the mechanisms that are pertinent to age variability in false-memory responses, the gist mechanism that supports them and the verbatim mechanism that suppresses them, may be expected to undergo substantial developmental improvement. It follows that whether false-memory responses increase or decrease with age (or even remain unchanged), must turn on the relative amounts of improvement in the supportive and suppressive mechanisms, but less obviously, it will also turn on the mix of the mechanisms that are tapped by particular memory tasks.

In light of the traditional hypothesis of developmental decreases in false-memory responses, the prediction of greatest interest is that of possible developmental *increases*, and it is that particular prediction that we have concentrated on in our own developmental studies.

It is evident that FTT's opponent-processes analysis would forecast such developmental increases for memory tasks that satisfy two general conditions: First, the gist abilities that support false-memory responses are ones that are known to improve considerably with age, so that memory support for false recognition and false recall is growing, and second, the memory tests that are administered make it difficult for subjects of all ages to access verbatim traces of experience, so that it is difficult for underlying age differences in verbatim-suppression abilities to express themselves. Two tasks that satisfy these constraints are recall and recognition of the Deese (1959) lists in Table 10.2.

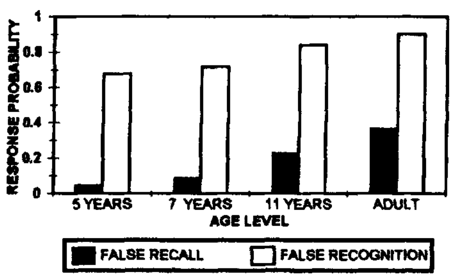


FIG 10.3. Developmental variability in false recall and recognition rates for Deese/Roediger/McDermott lists (based upon Brainerd, Reyna, & Forrest, 2002).

Research with adults suggests that connecting meaning across the words on such a list is critical to false-memory responses, an ability that develops slowly during childhood. The same research suggests a strong inclination to access gist memories of the meaning content of these lists rather than verbatim traces of individual target words. These considerations imply that the high rates at which adults falsely recall and falsely recognize unpresented stimulus words will not be obtained in children. Of late, this prediction has been evaluated in several developmental studies, with subject samples spanning the kindergarten to mid-high school age ranges (e.g., Brainerd, Reyna, & Forrest, 2002). As expected on theoretical grounds, false recall and false recognition of unpresented but meaning-preserving words have been found to increase between early childhood and young adulthood. Developmental trendsin false - recall and false recognition are illustrated in Fig. 10.3, which contains composite data from several studies that have been conducted in our laboratory. Note that these particular types of false-memory responses develop gradually and that there are increases during adolescence as well as childhood.

Inverted-U Relations Between False Memory and Repetition of Target Experiences

Repetition is one of the oldest manipulations in the scientific study of memory, the first systematic data on it was reported by Ebbinghaus (1885/1913). Ebbinghaus investigated the relation between true memory performance and the number of presentations of a list--typically, a list of nonsense syllables--using savings in relearning as his performance measure. He found that savings was proportional to the number of prior list repetitions. Since Ebbinghuas, a great many experiments have shown that levels of performance on measures of true memory improve as a negatively accelerated function of repetition.

How should recognition and recall of false-but-meaning-preserving events react to repetitions of the corresponding true events? Underwood (1965), who proposed an early one-process theory of false memory, thought that repetition of target words would increase false alarms to their synonyms and antonyms because hits and false alarms are both supported by a shared memory-strength variable. However, Seamon, Luo, Schwartz et al. (2002) pointed out that under FTT's opponent-processes analysis, the expected relation between repetition and false memory must be an inverted U, with such responses increasing following initial repetitions but decreasing following later repetitions. This ought to be the overall relation becomes trivially obvious when one considers that zero presentations of target events counts as a "repetition"--specifically, as the lowest possible number of repetitions. Levels of false memory for meaning-preserving distractors must increase as one moves from zero presentations to one presentation. Beyond this, however, one must allow for the possibility that there could be further increases as a consequence of additional repetitions if, for some reason, meaning processing was not complete following a single presentation. To investigate this proposal, Seamon, Luo, Schwartz et al. presented 9 Deese-type lists to their subjects at a rate of 2 seconds per item. In one experiment, 3 lists were presented once, 3 lists were present 5 times, and 3 lists were presented 10 times. In a second experiment, 3 lists were presented once, 5 times, and 25 times, respectively. The lists were followed by a recognition test composed of target probes, unpresented stimulus words for the 9 lists, and comparable words for 9 other Deese-type lists that had not been presented (0 repetitions). The pattern that emerged in each experiment is shown in Fig. 10.4. As predicted, the overall relation between repetition and false recognition was an inverted U: In both experiments, false recognition increased as the number of list presentations increased from 0 to 5, but it declined sharply as a consequence of further repetitions.

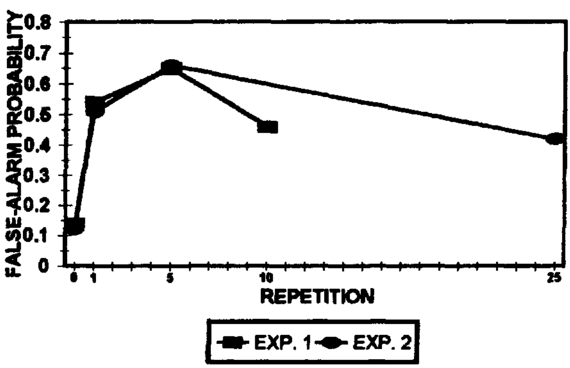


FIG. 10.4. False alarms to meaning-preserving words for Deese-type lists following different numbers of list repetitions (based upon Seamon, Luo, Schwartz, Jones, Lee, & Jones, 2002c).

SUMMARY

Although psychologists have long been interested in illusions of memory, systematic scientific exploration of false memory is a comparatively recent phenomenon. Because the field is still in search of new research paradigms, deciding which paradigms are the most productive ones and which findings to investigate intensively, it is particularly desirable to impose some theoretical order. I have attempted to do this by applying principles from FTT, principles that originally evolved from studies of higher reasoning rather than from studies of memory, to the task of explaining the most common family of false-memory responses--namely, false recognition and false recall of events that preserve the meaning content of experience.

Thus, so much of this chapter describes examples of new theoretical predictions and summarizes illustrative findings. The good news is that FTT's opponent-processes distinctions can generate predictions that conflict with those of common-sense, with older psychological theories of false memory, and with the law's view of false memory. Also on the positive side, confirmatory data have been reported for some major examples of such predictions. However, because false-memory research is at an early stage of development, theoretical interpretation is also in its infancy. In a decade, when our empirical footing is more secure, the theoretical landscape may have a very different appearance.

ACKNOWLEDGMENTS

Preparation of this paper and some of the research described herein were supported by a National Science Foundation grant (SBR-9730143), by National Institutes of Health grants (RO1MH61211-01and P50AT00008), by a Department of Health and Human Services grant (HRSA 1D34MB02077-01), and by an Academic Medicine and Managed Care Forum grant.

REFERENCES

Abrams, R. L., & Greenwald, A. G. (2000). Parts outweigh the whole (word) in unconscious analysis of meaning. *Psychological Science, 11,* 118-124.

Ankrum, C., & Palmer, J. (1989, November). *The perception and memory of objects and their parts.* Paper presented at Psychonomic Society, Atlanta, GA.

Bjorklund, D. F., & Muir, J. E. (1988). Children's development of free recall memory: Remembering on their own. *Annals of Child Development, 5,* 79-123.

Brainerd, C. J., & Mojardin, A. H. (1998). Children's false memories for sentences: Long-term persistence and mere testing effects. *Child Development, 69,* 1361-1377.

Brainerd, C. J., & Ornstein, P. A. (1991). Children's memory for witnessed events: The developmental backdrop. In J. Doris (Ed.), *The suggestibility of children's recollections.* Washington, D.C.: American Psychological Association.

Brainerd, C. J., & Reyna, V. F. (1996). Mere memory testing creates false memories in children. *Developmental Psychology, 32,* 467-476.

Brainerd, C. J., & Reyna, V. F. (1998a). Fuzzy-trace theory and children's false memories. *Journal of Experimental Child Psychology, 71,* 81-129.

Brainerd, C. J., & Reyna, V. F. (1998b). When things that never happened are easier to remember than things that did. *Psychological Science, 9,* 484-489.

Brainerd, C. J., & Reyna, V. F. (2001). Fuzzy-trace theory: Dual processes in reasoning, memory, and cognitive neuroscience. *Advances in Child Development and Behavior, 28,* 49-100.

Brainerd, C. J., Reyna, V. F., & Brandse, E. (1995). Are children's false memories more persistent than their true memories? *Psychological Science, 6,* 359-364.

Brained, C. J., Reyna, V. F., & Forrest, T. J. (2002). Are young children susceptible to the false-memory illusion? *Child Development, 73,* 1363-1377.

Brainerd, C. J., Reyna, V. F., & Kneer, R. (1995). False-recognition reversal: When similarity is distinctive. *Journal of Memory and Language, 34,* 157-185.

Brainerd, C. J., Wright, R., Reyna, V. F., & Mojardin, A. H. (2001). Conjoint recognition and phantom recollection. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 307-327.

Bransford, J. D., & Franks, J. J. (1971). The abstraction of linguistic ideas. *Cognitive Psychology, 2,* 331-380.

Ceci, S. J., & Bruck, M. (1993). Suggestibility of the child witness: A historical review and synthesis. *Psychological Bulletin, 113,* 403-439.

Conway, M. A., Collins, A. F., Gathercole, S. E., & Anderson, S. J. (1996). Recollections of true and false autobiographical memories. *Journal of Experimental Psychology: General, 125,* 69-95.

Deese, J. (1959). On the prediction of occurrence of certain verbal intrusions in free recall. *Journal of Experimental Psychology, 58,* 17-22.

DeMarie-Dreblow, D. (1991). Relation between knowledge and memory: A reminder that correlation does not imply causality. *Child Development, 62,* 484-498.

Donaldson, W. (1966). The role of decision processes in remembering and knowing. *Memory & Cognition, 24,* 523-533.

Ebbinghaus, H. (1885/1913). *Memory: A contribution to experimental psychology.* New York: Teachers College, Columbia University.

Fisher, R. P., & Cutler, B. L. (1992, September). *The relation between consistency and accuracy of eyewitness testimony.* Paper presented at the Third European Conference on Law and Psychology, Oxford, England.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language, 30,* 513-541.

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

Murphy, G. L., & Shapiro, A. M. (1994). Forgetting of verbatim information in discourse. *Memory & Cognition, 22, 85-94.*

Payne, D. G., Elie, C. J., Blackwell, J. M., & Neuschatz, J. S. (1996). Memory illusions: Recalling, recognizing, and recollecting events that never occurred. *Journal of Memory and Language, 35,* 261-285.

Reyna, V. F. (1998). Fuzz-trace theory and false memory. In M. Intons-Peterson & D. Best (Eds.), *Memory distortions and their prevention* (pp. 15-27). Mowah, NJ: Erlbaum.

Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and Individual Differences, 7,* 1-75.

Reyna, V. F., & Brainerd, C. J. (1998). Fuzz-trace theory and false memory: New frontiers. *Journal of Experimental Child Psychology, 71*, 194-209.

Reyna, V. F., & Kiernan, B. (1994). The development of gist versus verbatim memory in sentence recognition: Effects of lexical familiarity, semantic content, encoding instructions, and retention interval. *Developmental Psychology, 30*, 178-191.

Reyna, V. F., Kiernan, B. (1995). Children's memory and interpretation of psychological metaphors. *Metaphor and Symbolic Activity, 10*, 309-331.

Reyna, V. F., & Lloyd, F. (1997). Theories of false memory in children and adults. *Learning and Individual Differences, 9*, 95-123.

Roediger, H. L., III, & McDermott, K. B. (1995). Creating false memories: Remembering words not presented on lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 803-814.

Seamon, J. G., Lee, I. A., Toner, S. K., Wheeler, R. H., Goodkind, M. S., & Birch, A. D. (2002a). Are false memories dependent on mental rehearsal: Thinking out loud in the DRM procedure. *Psychological Science, 13*, 526-531.

Seamon, J. G., Luo, C. R., Kopecky, J. J., Price, C. A., Rothschild, L., Fung, N. S., & Schwartz, M. A. (2002b). Are false memories more difficult to forget than accurate memories? The effect of retention interval on recall and recognition. *Memory & Cognition, 30*, 1054-1064.

Seamon, J. G., Luo, C. R., Schwartz, M. A., Jones, K. J., Lee, D. M., & Jones, S. J. (2002c). Repetition can have similar or different effects on accurate and false recognition. *Journal of Memory and Language, 46*, 323-340.

Strong, E. K. (1913). The effect of time-interval upon recognition memory. *Psychological Review, 20*, 339-371.

Thapar, A., & McDermott, K. B. (2001). False recall and false recognition induced by presentation of associated words: Effects of retention interval and level of processing. *Memory & Cognition, 29*, 424-432.

Toglia, M. P., Neuschatz, J. S., & Goodwin, K. A. (1999). Recall accuracy and illusory memories: When more is less. *Memory, 7*, 233-256.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychologist, 26*, 1-12.

Underwood, B. J. (1965). False recognition produced by implicit verbal responses. *Journal of Experimental Psychology, 70*, 122-129.

Chapter 11

Fuzzy-Trace Theory (FTT), Judgment, and Decision-Making: A Dual-Processes Approach

Valerie F. Reyna
University of Arizona, U.S.A.

This chapter reviews traditional assumptions about relations between memory and higher cognition, recent research that challenges these assumptions, and implications of this research for theories of reasoning, judgment, and decision-making. The analysis supports an intuitionist approach to judgment and decision-making. In contrast to either heuristics-and-biases or adaptive-ecological approaches, the fuzzy trace theory (FTT) accounts for inconsistencies in human reasoning by assuming dual (gist-based versus verbatim-based) processes. These assumptions explain how reasoners exhibit different *degrees* of rationality across contexts. Examples from framing, availability, hindsight bias, syllogistic reasoning, base-rate neglect, adolescent risk-taking, and conjunctive/disjunctive probability judgment are discussed. Key explanatory principles are (1) that multiple gist and verbatim representations provide cognitive flexibility, but (2) reasoning gravitates to the least precise level of gist that the task allows, increasingly so as expertise develops. Thus, despite underlying conceptual similarities across tasks, rationality is contextually and developmentally dependent.

Memory Limitations and Higher Cognition

The first traditional assumption to be considered is that constraints of short-term or working memory shape the accuracy of reasoning, judgment, and decision-making. Information-processing theorists (e.g., Simon, 1956) argued that rationality is bounded by the limitations of memory. Humans *would* process all relevant information in making judgments or decisions, but memory limitations encourage shortcuts or "satisficing," which is highly adaptive despite falling short of optimality. Recent tests seemed to confirm this long-held assumption (Payne, Bettman, & Johnson, 1992). Careful examination of the data and more direct tests of the hypothesis, however, reveal that interference and dual processes, rather than memory "load," account for extant findings.

Specifically, when the hypothesis that memory limitations constrain performance was directly tested, three surprising findings emerged: (1) simplified processing typically does not result from memory demands, (2) tasks that produce errors make few memory demands, and (3) when tasks demand memory, reasoning performance is independent of those demands (Reyna, Lloyd, & Brainerd, 2003; Shafir & LeBoeuf, 2002). For example, when asked to judge whether it is more likely that a word begins with "k" or has "k" as its third letter, most subjects choose the former, incorrect answer. The traditional explanation for this bias is that it is easier to retrieve words that begin with a specific letter, and people use ease of retrieval as an indication of frequency of occurrence (Kahneman, Slovic, & Tversky, 1982). This "availability heuristic," that more easily retrieved items are more frequent, is a memory strategy. However, the subsequent literature has failed to support the availability hypothesis for judgments of frequency. Manis, Shedler, Jonides, and Nelson (1993) showed that judgments of frequency were independent of availability of instances in memory (see also Shedler & Manis, 1986). Interestingly, the availability heuristic does account for judgments of categories, as opposed to instances. For example, Manis et al. gave subjects lists of names and either biased the list with more famous male names (Cond. 1) or more famous female names (Cond. 2). Availability in memory (as measured by proportion of male names that were recalled) was related to judgments of category size, such as the number of male names that had been presented on the study list, but not to judgments of frequency, such as the number of times individual names had been presented. Although these data refute the original memory availability hypothesis for frequency judgments, they support *dual* processes in memory and judgment. Gist memory supports qualitative memory for categories, but verbatim memory supports memory for individual instances.

The hindsight bias is another example of a memory explanation for judgment biases (Kahneman et al., 1982). In hindsight, outcomes (such as the outcome or length of a war) that have occurred seem more likely in retrospect than they seemed prior to knowing the outcomes. It is argued that the hindsight bias is irrational because (1) the same information is judged to be predictive of different, opposite outcomes and (2) recollections of earlier likelihood estimates are colored by subsequent outcome information. Using traditional global memory ideas, the hindsight bias was described as resulting from updating cue values in memory, "remembering as a process of reconstruction" (Gigerenzer, Todd, & ABC Group, 1999, p. 208). This traditional explanation for hindsight bias is similar to that for misinformation effects, once described as destructive updating of memory by misleading information (Reyna, Holliday, & Marche, 2002). For both misinformation and hindsight effects, dual-process explanations have superseded reconstructive updating models.

To illustrate, Erdfelder and Buchner (1998) tested a series of multinomial models of the hindsight bias in which opposing processes of recollection of original judgments were pitted against reconstructive processes that biased judgments in the direction of later outcomes' feedback. The model that best fit the data incorporated assumptions about opposing memory processes, similar to those of FTT, and accounted for the fact that, as Hasher, Attig, and Alba (1981) had observed, "subjects can, under some circumstances, gain access to their original knowledge" (p. 93). The point is not to claim, as had some reconstructive memory theorists, that memory is reconstructive except when it isn't, but to predict when memory judgments would be a gist-based reconstruction versus a verbatim recollection of original knowledge (Reyna & Brainerd, 1995).

The standard heuristics-and-biases view is that inconsistencies such as the hindsight bias are troubling. If events that occurred are judged to have been more predictable than they really were, then it is difficult to judge the past or to learn from it. The standard view has been misunderstood as characterizing human reasoning as *generally* biased or fallacious: When heuristics are used, they can lead to errors, but that does neither imply that they must lead to errors, nor that they typically do. Contrary to widespread characterizations, Kahneman et al. (e.g., 1982, p. 164) argued that heuristics are *adaptive*—because they conserve information-processing resources—and that they ordinarily do not produce errors: "Availability is an ecologically valid cue for the judgment of frequency because, in general, frequent events are easier to recall or imagine than infrequent ones." Contemporary theorists have reiterated this ecological approach (Gigerenzer, 1994), and agree with the standard view that rationality is bounded by information- processing constraints (they disagree about whether certain behaviors are errors).

Evidence for "fast and frugal" simplified processing (Payne et al., 1992; Shafir & LeBoeuf, 2002), however, does not indicate that memory limitations are the source of such processing. Furthermore, simplified processing does not occur only in complex tasks, but also under conditions in which memory capacity cannot easily explain performance. Most heuristics-and-biases effects are obtained in tasks involving small amounts of written information and ample processing time.

Research shows that framing effects, for example, are mainly due to *qualitative* processing of gross contrasts between some and none along the dimensions of outcomes and of probabilities (Reyna & Brainerd, 1991, 1995). In contrast, analogous framing problems with virtually identical memorial and processing constraints produce evidence of precise, *quantitative* processing. For example, choosing between winning $300 versus a 1/3 chance of winning *$903*—as opposed to $300 versus a 1/3 chance of winning *$900* as in standard problems—produces evidence of *quantitative* processing (i.e., subjects choose the richer gamble). The shift from risk aversion to risk seeking when outcomes are changed slightly indicates that subjects are sensitive to subtle quantitative differences in expected value (Reyna & Brainerd, 1995). Reasoners generally avoid quantitative processing even when the memory capacity or effort required is low (Fischer & Hawkins, 1993; Reyna & Brainerd, 1994). The change in expected value, not in memory load, produces the shift in behavior.

When tasks are memory-based (Hastie & Park, 1986), the accuracy of judgment and decision-making performance is independent of memory load. In tasks ranging from covariation estimation to probability judgment, single and double dissociations between memory and judgment have been demonstrated (Fisher & Chandler, 1991; Reyna & Brainerd, 1995; Shedler & Manis, 1986). Conditions that impair memory for essential problem information do not necessarily affect judgment accuracy. Conversely, increasing memory accuracy for task information (e.g., pictorial rather than verbal presentation) decreases reasoning accuracy under specific conditions (Brainerd & Reyna, 1993). These results and others challenge traditional conceptions of working memory in explaining reasoning and judgment (Dempster, 1992). Information-processing resource limitations have little to do with the quality of judgment and decision-making, removing the main rationale behind both heuristics-and-biases and fast-and-frugal approaches: cognitive economy. What, then, is responsible for the tendency to engage in simplified processing?

Alternatives to Fast and Frugal Processing: Dual Processing Models

Traditional memory models assume a generic resource, "one" global memory that can be used for information representation and processing. Although assuming one as opposed to dual memories seems parsimonious, these models relinquish parsimony to the extent that additional assumptions must be introduced to account for data demonstrating dual processes in memory (cf. Anderson, Budiu, & Reder, 2001). For example, single-process accounts of recognition can account for some dual-process patterns in timed recognition judgments, namely, a non-monotonic pattern of recognition of related lures across time. Given the studied word "dog," presenting a related word such as "poodle" on a recognition test initially evokes false alarms at shorter lags but recollection rejection ("dog" is recollected and is used to reject "poodle") at slightly longer lags (Brainerd, Reyna, Wright, & Mojardin, in press). Global memory models add assumptions about the priority of certain features in recognition, assuming that *overlapping* feature information is processed quickly whereas other details are processed later, in order to accommodate this result.

FTT and global-memory (e.g., source monitoring) approaches differ in their treatment of different kinds of features. In FTT, differences between semantic gist and surface features are central and have implications for predictions (gist is simpler, faster, and endures longer), whereas there is no reason a priori to treat semantic features differently in the source-monitoring framework and in the familiarity-recollection approach (Jacoby, 1996). Thus, in order to account for both monotonic patterns of recognition for presented targets and for non-monotonic patterns for never-presented related lures, global and dual-process approaches must have a way of distinguishing among stimulus features and *predicting* which ones are processed more quickly.

There are four main *dual-process* approaches to memory: task-based (Roediger, Weldon, & Challis, 1989), process dissociation (Jacoby, 1996), ROC (Rotello, 2001), and conjoint recognition (Brainerd et al., in press). In the task-based approach, explicit memory (e.g., cued recall) is contrasted with implicit memory in indirect tasks (e.g., fragment completion in which subjects unconsciously complete partially spelled words with words from a prior "unrelated" task). Although demonstrating task-based differences was once viewed as prima facie evidence for separate memory systems, "task impurity" and other criticisms of the ability to cleanly separate memory systems has led to the use of new methods, such as process dissociation and conjoint recognition. Process dissociation relies on instructions to include and exclude items from different lists or contexts, with recollection identified with conscious memory and familiarity with automatic or unconscious memory, for example, for to-be-excluded items (Jacoby, 1996).

The third approach, ROC analyses, draws on concepts of both global and dual-memory models (but data have tended to support dual processes; Rotello, 2001). Signal detection theory is generalized by generating ROC curves for confidence ratings of related versus unrelated distracters, as opposed to targets versus unrelated distracters. Changes in the geometry of ROC curves are used to diagnose recognition processes. The fourth approach, conjoint recognition, incorporates distinctions from FTT. A mathematical model is defined over recognition memory designs in which subjects respond to three different instructions (recognize only verbatim items, both verbatim and related items, and only related items) to measure identity and similarity processes (Reyna, Holliday, & Marche, 2002). FTT distinguishes between verbatim and gist representations as well as between all-or-none identity matches versus graded similarity judgments.

Data on false-memory effects bear directly on dual-memory assumptions (see Reyna & Kiernan, 1995 for experimental manipulations of false-memory acceptance and rejection). In recognition memory, a paradigm from the verbal learning tradition involving studying semantically related word lists (the Deese-Roediger-McDermott or DRM paradigm) is used to reliably induce false memories. A striking finding is the tendency to erroneously "recognize" words that were never studied but are semantically related to studied words. After people are presented with several related words, such as "door," "glass," "pane," and "ledge," they are likely to falsely recognize a highly associated, critical lure item such as "window." The false alarm rate for critical lure items is about as high as the hit rate for presented words. According to FTT, false recognition of critical lures is supported by memory for the "gist of the list," whereas true recognition of presented items (especially on immediate tests) is supported by verbatim memory for individual words (Reyna & Lloyd, 1997). McEvoy, Nelson, and Komatsu's (1999) and Roediger, Watson, McDermott, & Gallo's (2001) data support FTT's opposing-processes account: Activation of targets produces acceptance of targets and recollection rejection of lures, but activation of semantic gist produces acceptance of both targets and lures.

Using the same word-list paradigm, it is possible to obtain seemingly contradictory data patterns, which are resolvable by invoking dual processes in memory (Payne & Elie, 1998; Toglia, Neuschatz, & Goodwin, 1999). Consider the well-known beneficial effects on memory performance of either repetition or deep (meaningful) processing of studied words. Two studies showed opposite consequences of improving memory for studied items. Payne and Elie (1998) found that repeating words at study increased later recognition of presented words and *decreased* acceptance of never-presented related words. Toglia et al. (1999), using the same materials, showed that deep processing of words at study increased recognition of presented words, but it also *increased* acceptance of related words. Global memory models have difficulty accommodating both sets of results simultaneously; memory improvements should either increase acceptances of both targets and lures because of reconstructive, familiarity, or similarity processes or, alternatively, increase acceptance of targets but decrease that of lures because of discrimination processes. Efforts to combine different global memory models, such as source monitoring and activation approaches, address the need for opposing processes (Roediger et al., 2001), although it is difficult to predict from the hybrid models when one or the other global process should be invoked.

FTT incorporates dual opposing reconstructive and reproductive memory processes (see Reyna & Brainerd, 1995, for integration of conflicting data from verbal-learning and reconstructive-memory traditions). Assumptions about opposing processes explain both the Payne-Elie and Toglia-et-al. data, but they also explain other observations, such as patterns of false recognition versus false-recognition reversal (Reyna & Brainerd, 1995; Reyna & Kiernan, 1995). Specifically, exact repetition of presented items primes verbatim memory for those items, but semantic processing primes gist. Enhancing verbatim memory increases target acceptance and lure rejection, as in the Payne-Elie study, and enhancing gist memory increases acceptance of both targets and lures, as in the Toglia et al. study (Reyna, 2000; Reyna & Lloyd, 1997). Furthermore, FTT predicts a pattern that cannot be derived from the global memory models or from their hybrids, that acceptance of related lures could *exceed* that of presented targets under specific conditions. This pattern was predicted and obtained for both unrelated word lists and the DRM paradigm (Brainerd & Reyna, 1998).

Recent work combines use of the DRM paradigm with the study of time and conscious strategies. If gist processing is quicker than verbatim, false recognition should be greater for speeded judgments, which has been observed (see Heit, Brockdorff, & Lamberts, in press). Heit et al. varied time lags over a wide range (200-1000ms) and warned some subjects about the false-memory phenomenon to examine strategic processes. (Warned subjects were told that one associated word would not be presented, and to avoid mistakenly judging it as old.) Heit et al. confirmed that the false-memory effect would be strong early (and reduced later). Warning instructions did not change the false memory effect, consistent with unconscious gist-based acceptance of related lures.

In addition, Heit et al. used an inclusion condition in which subjects were instructed to say "old" to the critical lures (similar to conjoint recognition). The prediction here of FTT: Subjects would less likely engage in conscious processing—i.e., recollection rejection—that would otherwise reduce false recognition. Models that implement these assumptions provide consistently good fits to the data. The results of the inclusion condition confirmed that subjects could withhold recollection rejection that would otherwise reduce the false memory effect, showing greater acceptance of related words compared to the standard and warning conditions. This difference was greater at later response times, when recollection rejection would be expected to kick in.

The strategic use of distinctiveness heuristic should be separated from conscious deployment of verbatim memories in recollection rejection, as posited in FTT. It is possible to recollect *none* of the presented stimuli and yet use the distinctiveness heuristic, for example, rejecting furniture words if all of the stimuli were names of fruits. The Heit et al. study found evidence for recollection rejection, rather than distinctiveness strategies. Their study demonstrates that there are conscious strategies that reduce false memories, such as recollection rejection, and conscious strategies that do not, such as warnings.

Dual Processes in Memory and Judgment: A FTT Analysis

Overall, a gist basis for judgment has advantages because, as noted, gist is quicker, simpler, and endures over time. Rather than linear logic or precise computation, FTT offers an intuitionist alternative that applies to mature reasoners. FTT holds that reasoning progresses from verbatim to gist processing, that is, toward intuition with increasing experience (for children, Haines & Moore, 2003; Reyna & Ellis, 1994; for adult experts, Reyna, in press; Reyna, Lloyd, & Whalen, 2001; Reyna & Adam, 2003).

Other dual-process approaches have also contrasted intuitive with analytic modes. Epstein (1994) observed "that people apprehend reality in two fundamentally different ways, one variously labeled intuitive, automatic, natural, non-verbal, narrative, and experiential, and the other analytical, deliberative, verbal, and rational" (p. 710). Given the question, is the Pope a bachelor, the *analytic* academic would answer "yes" but the *intuitive* and socially appropriate response is "no."

The conflict between analysis and emotion is not new, but contemporary theorists emphasize the wisdom in emotion. They suggest that emotion is seldom wrong, but analysis without emotional cues is treacherous. Although FTT has characterized intuition as advanced, it contrasts with other contemporary views. Unthinking emotion is the source of great pain, especially for young people (Reyna, Adam, LeCroy, Muller-Poirer, & Brainerd, in press, for applications of FTT to adolescent risky sexual decision making). Although emotion provides important social cues (and so brain-damaged patients who lack this information have problems coping), emotional cues can also be misleading (Klaczynski & Fauth, 1997). The habit of mind to see through irrelevant details to the core intuitive gist is advanced, but some emotionally salient gists are inappropriate. The dilemma is to distinguish emotions that are misleading from those that signal relevant gist. A task analysis is required to sort out the appropriate gist representations and reasoning principles.

Types of Errors: Rationality by Degrees

The FTT analysis of rationality builds on prior work concerning such criteria as internal coherence (internal consistency according to semantics, logic, or probability, assuming their relevance) and consistency with reality (good medical decisions are associated with good patient outcomes in the aggregate; sound business decisions do not lead to recurrent bankruptcy). Evidence for rationality also comes from developmental data concerning the order of emergence of specific errors (Reyna & Brainerd, 1994, 1995). In order for judgment-and-decision-making research to be relevant to real-world issues, behaviors must be judged. Rationality goes beyond ratifying existing behavior as adaptive or uncritically accepting all goals, no matter how inconsistent with reality or self-destructive. Judging the quality of reasoning requires a process analysis that distinguishes types of errors (Reyna et al., 2003).

According to FTT, successful reasoning involves "selecting from among many relationships given as background facts, retrieving some among many principles that could be applied to such relationships, and, finally, applying the principles coherently" (Reyna & Brainerd, 1993, p. 105). Most studies have not identified lack of knowledge of key facts or of the appropriate reasoning principle as a major source of errors (but see Reyna & Adam, 2003, for exceptions related to health). For example, first graders possess the rudiments of probability concepts (Haines & Moore, 2003; Reyna & Brainerd, 1994). Reasoning errors often occur despite knowledge of relevant facts and the competence to solve problems correctly. These results support an intuitionist view of reasoning as dynamic, parallel, and uncertain as opposed to analytic, logical, and precise.

Although there is growing acceptance of this view of reasoning, its interpretation differs. One interpretation is that dynamic variability is an issue of performance but reasoning competence is sound. This interpretation arbitrarily recognizes good reasoning as diagnostic of true competence rather than equally relevant bad reasoning that involves the same concepts. Correctly answering the class-inclusion question, "Are there more children in your class or in the school?" does not make the class-inclusion error in the standard paradigm any less wrong: If there are 10 animals, 7 cows and 3 horses, there are not more cows than animals, as children claim until they are about ten years of age (Reyna, 1991). However, answering reasoning questions correctly that require subtle knowledge demonstrates that reasoners possess basic knowledge. According to FTT, errors in different versions of the same problem are really errors, but reasoning can be placed on a dimension of *degrees of rationality* above the level of absolute ignorance of the reasoning principle but below the level of reliable implementation of that principle. The idea of degrees of rationality recognizes that there are many ways to get a problem wrong, and some errors are worse than others.

Ignorance of the reasoning principle, knowledge competence, is the most basic error and is present at the earliest stages of development. The next level of error is the representational mistake of relying on verbatim representations in a gist task, such as literal interpretation of metaphors (The prison guard was a rock interpreted as his having hard muscles; Reyna, 1996; Reyna & Kiernan, 1995) or slavish rule-following in moral reasoning (Authority figures should be obeyed regardless; Reyna & Brainerd, 1995), or mindless adherence to exactly what was said in an inference task (A child remembers that the experimenter said that the bird was in the cage and that the cage was under the table, but denies that the bird was under the table because "you didn't say that"; Brainerd &

Reyna, 1993).

Most judgment-and-decision-making tasks require the processing of meaning. At younger ages or earlier stages in the development of expertise, better memory for relevant information can be detrimental to reasoning performance: Preschoolers who better remember input sentences are more likely to reject true inferences that follow from those input sentences; improving memory for those sentences lowers reasoning performance (Brainerd & Reyna, 1993). Novice physicians who memorize lists of possible diagnoses are often paralyzed with possibilities and unable to reason coherently about patients' problems (Reyna et al., 2003). Verbatim interference occurs when reasoners substitute rote retrieval for thinking.

At the next level of sophistication is failing to encode the appropriate gist. These errors are more sophisticated because reasoners are operating on a meaningful representation of problem information, but it is the wrong meaning for the task. For example, patients with significant blockage of their carotid artery were told that they have a 22% chance of a stroke, but surgery to correct that problem (carotid endarterectomy) carries a 2% risk of stroke (Reyna & Hamilton, 2001). Despite surgeons informing patients using the same numbers, patients reported widely different estimates of risk. Most had the gist right: having surgery was less risky than not having surgery. However, informed consent requires that patients understand the gist that surgery has some risk. Paradoxically, patients who reported a 10% risk of a stroke during surgery were less wrong than those who reported a 0% risk, although 0% is numerically closer to the correct answer.

In many situations, the reasoner has encoded the appropriate gist representation of the problem, but fails to use that representation. Instead, reasoning is hijacked by a competing conceptualization of the problem. Is the question of risk better answered with estimates of relative or absolute risk, one-time or cumulative risk (Stone, Yates, & Parker, 1994)? What conditional probability addresses the question of which playground equipment is more dangerous: the probability that children have an accident given that they are playing on that equipment or the probability that children are playing on that equipment given that they have an accident (Reyna & Brainerd, 1993; Reyna, in press)? Each representation could be "accurate" but inappropriate for some tasks.

Retrieval failure becomes apparent later in development, after representational errors decrease (Reyna 1991; Reyna & Brainerd, 1995). If reasoners think that the appropriate representation involves numerosity, they will tend to retrieve reasoning principles that involve numerosity. A problem that should be about the logic of inclusion relations (cows are animals) becomes a problem about comparing the numbers of cows and horses. (Linguistic misunderstandings do not explain most errors; Reyna, 1991). Although adults do not commit errors in the Piagetian task, they show unusually long response times and they commit errors in other versions of class-inclusion tasks such as the conjunction fallacy (Reyna, 1991). For example, adults rank "feminist bank teller" as a more likely description than "bank teller" of a stereotyped liberal activist named "Linda," although the class of bank tellers includes the feminist ones. In more transparent versions of class-inclusion tasks, children and adults retrieve the correct principle that more inclusive classes must be more numerous and more probable: Adults readily affirm statements that the next person who walks through a door is more likely to be a bank teller than to be a feminist bank teller. Children perform much better on class-inclusion problems after the correct principle is cued. Not retrieving a relevant principle in context is a failure of insight, but it is easily remedied by reminding reasoners of the principle and providing opportunities to practice recognizing cues.

The most advanced error, one that persists into adulthood, is incoherence in the application of retrieved reasoning principles, especially when classes overlap. Consider the classes of people who have genetic mutations that have been linked to breast cancer and the people who have breast cancer. Genetic counseling is particularly complicated for breast cancer patients and their physicians because partially overlapping classes make understanding conditional probabilities difficult (Reyna et al., 2001). For example, patients with the mutations may not get cancer, patients without the mutations may still get cancer, those who get cancer probably do not have the mutations, and those who do not get cancer may still have the mutations. Hemochromatosis, which causes iron to build up in the body, is cognitively easier because most people with the disease have the mutation although people with the mutation do not necessarily have the disease. Finally, Huntington's disease is the easiest to understand. People with the mutation develop the disease, and vice versa.

FTT has been applied to many class-inclusion tasks that are subject to such processing interference, including fractions, proportionality concepts, syllogistic arguments, conjunctive fallacies, disjunctive fallacies, conversion errors in conditional probability, and Bayesian updating of conditional probabilities (Reyna, 1991; Wolfe, 1995). In each of these tasks, reasoners lose track of denominators or marginals, and errors increase as overlapping relations increase (Reyna & Brainerd, 1995). Despite initial enthusiasm, properly controlled studies have shown that frequency formats do not reliably reduce errors. However, providing a notational system that allows reasoners to keep classes distinct improves performance, such as using diagrams or labels for super-ordinate classes (Lloyd & Reyna, 2001; Reyna, 1991).

How Memory Processes Explain Risky Decision Making

Although global memory models have been used to explain judgment and decision-making (Dougherty, Gettys, & Ogden, 1999), dual processes in memory and reasoning are consistent with a broader array of phenomena. Dual processes also account for task variability (Reyna & Brainerd, 1994, 1995), the intriguing observation that the same person can exhibit different levels of competence in different situations. Adequate theory should encompass these human dualities, that reasoning is quantitative and qualitative, self-regulated and impulsive, fair-minded and self-serving, logical and emotional (Haines & Moore, 2003). Memory processes explain some of these countervailing tendencies, illustrated in the following example involving sexual decision-making.

The gist of a decision, such as whether to engage in sexual activities, is the individual's semantic representation of the decision situation—its meaning—which reflects knowledge, worldview, culture, emotional attitude, and developmental level. Knowledge may be lacking, for example, that females are more biologically susceptible to sexually transmitted infections than males or that human papillomavirus is more prevalent than previously estimated, producing underestimation of risks (Reyna & Adam, 2003). Simple retrieval cues, such as enumerating examples of sexually transmitted infections, significantly reduced underestimation, showing that decision makers have knowledge about risks that they do not retrieve when cues are not present. This cuing effect was obtained for all subject groups, including expert physicians, nurses, health educators, and adolescents.

According to decision analysis, hundreds of potential risks and benefits could be adduced for an adolescent facing a choice between hanging out at the mall with friends or going to an unsupervised party where sexual activity is likely to take place. Rationally, the mall versus party decision depends on the amount of fun and degree of risk to be had at the party compared to the mall. Most parents disagree. No amount of fun can compensate for the risks, contrary to a strictly rational cost-benefit analysis.

The representation of the decision for adolescents is that the amount of fun at the mall is considerably less than that available at the party; thus, it is "worth it" to take a calculated, objectively small risk and have much more fun at the party. The general principle that is retrieved is that more fun is better than less fun, and so going to the party is preferred. (What is "fun" changes with development, for example, from playing with toys to going to parties, and with culture: hanging out at malls is an American pastime.) The representation of the decision for parents is that fun can be had at the mall or at the party, but the party is risky. In Russian roulette, the exact amount of money to be won or the exact number of bullets in the chamber are immaterial; the gist of the decision to mature reasoners is that there is no amount of money that makes dead better than not dead (Reyna et al., in press). The operative principle is to avoid catastrophic outcomes. According to FTT, the crude categorizations of adults are more advanced than the nuances that adolescents consider. Global categorical policies (avoid risk) exist on a higher plane of rationality, rendering details about amounts of risk and reward irrelevant.

Adolescents and adults should be similarly subject to processing interference, however, in combining probability information. To test this hypothesis, adolescents and adults including physicians, nurses, and experts attending a national workshop on HIV prevention were presented with the same problem (Reyna & Adam, 2003): Suppose that 10% of people were infected with a disease. A test is given to one person, and it has 80% sensitivity (80% of the people who have the disease test positive) and 80% specificity (80% of the people without the disease test negative). This person tests positive. Is the probability of disease closer to 30% or 70%? Only 32% of physicians selected the correct response (about 30%), and the other adult groups were comparable, except for national experts who managed to achieve an accuracy level that was equal to chance, 55% correct. Adolescents did not differ from physicians; their accuracy level was 33%. Thus, significant processing interference in judging conditional probabilities, which involve overlapping classes, remained late into development.

SUMMARY

A variety of experiments show that dual-process assumptions in memory and reasoning provide a better account of judgment and decision-making biases, such as framing, availability, hindsight bias, and base-rate neglect. They also explain the dominant finding that judgment and decision-making is typically adaptive. Developmentally, research on many tasks in cognition and social decision-making indicates that development progresses from verbatim-based analytical processing to gist-based intuitive processing, contrary to traditional models. Task analyses reveal that judgment and decision-making errors undergo evolution; although later errors are still wrong, they represent more superficial mental bookkeeping mistakes rather than fundamental flaws in understanding. The order of emergence of errors exhibits some regularity across tasks, which suggests an approach to rationality that differentiates degrees of competence: Errors of knowledge are the most basic, followed by literal capture of reasoning by verbatim processes, and then more subtle failures to encode the appropriate gist or to resist interference from competing but inappropriate gists. Retrieving knowledge, values, and reasoning principles in context when they are relevant appears later, and processing interference is the last, most refractory source of error.

The most compelling criterion of rationality is internal consistency; this is why the hindsight bias and framing effects are troubling. The same information should not be treated differently because of differences in phrasing or the presence of irrelevant features. Gist representations allow reasoners to go beyond the phrasing of problems or packaging to discover the underlying meaning. By abstracting across problems that differ only superficially, gist representations support rationality.

REFERENCES

Anderson, J. R., Budiu, R., & Reder, L. (2001). A theory of sentence memory as part of a general theory of memory. *Journal of Memory & Language, 45*, 337-367.

Brainerd. C. J., & Reyna, V. F. (1993). Memory independence and memory interference in cognitive development. *Psychological Review, 100,* 42-67.

Brainerd, C. J., & Reyna, V. F. (1998). When things that were never experienced are easier to "remember" than things that were. *Psychological Science, 9,* 484-489.

Brainerd, C. J., Reyna, V. F., Wright, R., & Mojardin, A. H. (in press). Recollection rejection: False-memory editing in children and adults. *Psychological Review*

Dempster, F. N. (1992). The rise and fall of the inhibitory mechanism: Toward a unified theory of cognitive development and aging. *Developmental Review, 12,* 45-75.

Dougherty, M. R. P., Gettys, C. F., & Ogden, E. E. (1999). MINERVA-DM: A memory processes model for judgments of likelihood. *Psychological Review, 106,* -159.

Epstein, S. (1994). Integration of the cognitive and psychodynamic unconscious. *American Psychologist, 49,* 709-724.

Erdfelder, E., & Buchner, A. (1998). Decomposing the hindsight bias: A multinomial processing tree model for separating recollection and reconstruction in hindsight. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 387-414.

Fischer, G. W., & Hawkins, S. A. (1993). Strategy compatibility, scale compatibility, and the prominence effect. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 580-597.

Fisher, R. P., & Chandler, C. C. (1991). Independence between recalling interevent relations and specific events. *Journal of Experimental Psychology: Learning, Memory and Cognition, 17,* 722-733.

Gigerenzer, G. (1994). Why the distinction between single-event probabilities and frequencies is important for psychology (and vice versa). In G. Wright & P. Ayton (Eds.), *Subjective probability* (pp. 129-161). New York: Wiley.

Gigerenzer, G., Todd, P. M., & ABC Group. (1999). *Simple heuristics that make us smart.* New York: Oxford University Press.

Haines, B. A., & Moore, C. (2003). Integrating cognitive and social cognitive development into the study of judgment and decision-making. In S. L Schneider & J. Shanteau (Eds.), *Emerging perspectives on judgment and decision research* (pp. 246-283). New York: Cambridge University Press.

Hasher, L., Attig, M. S., & Alba, J. W. (1981). I knew it all along--or did I? *Journal of Verbal Learning and Verbal Behavior, 20,* 86-96.

Hastie, R., & Park, B. (1986). The relationship between memory and judgment depends on whether the judgment task is memory-based or on-line. *Psychological Review, 93,* 258-268.

Heit, E., Brockdorff, N., & Lamberts, K. (in press). Strategic processes in false recognition memory. *Psychonomic Bulletin & Review.*

Jacoby, L. L. (1996). Dissociating automatic and consciously controlled effects of study/test compatibility. *Journal of Memory and Language, 35,* 32-52.

Kahneman D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases.* New York: Cambridge University Press.

Klaczynski, P. A., & Fauth, J. (1997). Developmental differences in memory-based intrusions and self-serving statistical reasoning biases. *Merrill-Palmer Quarterly, 43,* 539-566.

Lloyd, F. J., & Reyna, V. F. (2001). A Web exercise in evidence-based medicine using cognitive theory. *Journal of General Internal Medicine, 16*, 94-99.

Manis, M., Shedler, J., Jonides, J., & Nelson, T. E.. (1993). Availability heuristic in judgments of set size and frequency of occurrence. *Journal of Personality and Social Psychology, 65*, 448-457.

McEvoy, C. L., Nelson, D. L., & Komatsu, T. (1999). What is the connection between true and false memories? The differential roles of interitem associations in recall and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 1177-1194.

Payne, D. G., & Elie, C. J. (1998, November). *Repeated list presentation reduces false memories for pictures and words.* Paper presented at Psychonomic Society, Dallas, TX.

Payne, J. J., Bettman, J. R., & Johnson, E. J. (1992). Behavioral decision research: A constructive processing perspective. *Annual Review of Psychology, 43*, 87-131.

Reyna, V. F. (in press). How memory and reasoning processes explain risk. *Current Directions in Psychological Science.*

Reyna, V. F. (1991). Class inclusion, the conjunction fallacy, and other cognitive illusions. *Developmental Review, 11*, 317-336.

Reyna, V. F. (1996). Meaning, memory and the interpretation of metaphors. In J. Mio and A. Katz (Eds.), *Metaphor: Implications and applications* (pp. 39-57). Hillsdale, NJ: Erlbaum.

Reyna, V. F. (2000). Fuzzy-trace theory and source monitoring: A review of theory and false-memory data. *Learning and Individual Differences, 12*, 163-175.

Reyna, V. F., & Adam, M. B. (2003). Fuzzy-trace theory, risk communication, and product labeling in sexually transmitted diseases. *Risk Analysis, 23*, 325-342.

Reyna, V. F., Adam, M. B., LeCroy, C.W., Muller-Poirier, K., & Brainerd, C.J. (in press). The development of judgment and decision-making from childhood through adolescence. In J. Jacobs & P. Klacynski (Eds.), *The development of children's and adolescents' judgment and decision-making.*

Reyna, V. F., & Brainerd, C. J. (1991). Fuzzy-trace theory and framing effects in choice: Gist extraction, truncation, and conversion. *Journal of Behavioral Decision Making, 4*, 249-262.

Reyna, V. F., & Brainerd, C. J. (1993). Fuzzy memory and mathematics in the classroom. In R. Logie & G. Davies (Eds.), *Memory in everyday life* (pp. 91-119). Amsterdam: North-Holland.

Reyna, V. F., & Brainerd, C. J. (1994). The origins of probability judgment: A review of data and theories. In G. Wright & P. Ayton (Eds.), *Subjective probability* (pp. 239-272). New York: Wiley.

Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and Individual Differences, 7*, 1-75.

Reyna, V. F., Lloyd, F. J., & Brainerd, C. J. (2003). Memory, development, and rationality: An integrative theory of judgment and decision making. In S. L. Schneider & J. Shanteau (Eds.), *Emerging perspectives on judgment and decision research* (pp. 201-245). New York: Cambridge University Press.

Reyna, V. F. & Ellis, S.C. (1994). Fuzzy-trace theory and framing effects in children's risky decision making. *Psychological Science, 5,* 275-279.

Reyna, V. F., & Hamilton, A. J. (2001). The importance of memory in informed consent for surgical risk. *Medical Decision Making, 21*, 152-155.

Reyna, V. F., Holliday, R., & Marche, T. (2002). Explaining the development of false memories. *Developmental Review, 22,* 436-489.

Reyna, V. F., & Kiernan, B. (1995). Children's memory and metaphorical interpretation. *Metaphor and Symbolic Activity, 10*, 309-331.

Reyna, V. F. & Lloyd, F. L. (1997). Theories of false memory in children and adults. *Learning and Individual Differences, 9,* 95-123.

Reyna, V. F., Lloyd, F. L., & Whalen, P. (2001). Genetic testing and medical decision making. *Archives of Internal Medicine, 161*, 2406-2408.

Roediger, H. L., Watson, J., McDermott, K., & Gallo, D. (2001). Factors that determine false recall: A multiple regression analysis. *Psychonomic Bulletin & Review, 3,* 385-407.

Roediger, H. L., Weldon, M. S., & Challis, B. H. (1989). Explaining dissociations between implicit and explicit measures of retention: A processing account. In H. L. Roediger & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 3-41). Hillsdale, NJ: Erlbaum.

Rotello, C. M. (2001). Recall processes in recognition memory. In D. L. Medin (Ed.), *The psychology of learning and motivation* (Vol. 40, pp. 183-221). San Diego, CA: Academic Press.

Shafir, E., & Leboeuf, R. A. (2002). Rationality. *Annual Review of Psychology, 53,* 491-517.

Shedler, J., & Manis, M. (1986). Can the availability heuristic explain vividness effects? *Journal of Personality and Social Psychology, 51*, 26-36.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review, 63,* 129-138.

Stone, E. R., Yates, J. F., & Parker, A. M. (1994). Risk communication: Absolute versus relative expressions of low-probability risks. *Organizational Behavior and Human Decision Processes, 60,* 387-408.

Toglia, M. P., Neuschatz, J. S., & Goodwin, K. A. (1999). Recall accuracy and illusory memories: When more is less. *Memory, 7,* 233-256.

Wolfe, C. (1995). Information seeking on Bayesian conditional probability problems: A fuzzy-trace theory account. *Journal of Behavioral Decision Making, 8,* 85-108.

Postscript

CLOSING REMARKS

The 4h Tsukuba International Conference on Memory (Tic 4) — Human Learning and Memory: Advances in Theory and Application

Tsukuba, Japan
13 January 2003

Chizuko Izawa
Tulane University, USA

Each sentence below in the Closing Remarks was immediately translated into Japanese by Bilingual Co-Organizer and Co-Chair Izawa:

What a magnificent International Conference on Memory this was! During the last three days, 225 of the best experts in the world participated in this conference program on Human Learning and Memory: Advances in Theory and Application. They represented 73 universities from 10 countries on four continents.

In addition to 11 invited speeches, 79 posters were presented, of which 13 were produced by 29 invited overseas memory specialists. Altogether, 116 colleagues registered to join us at Tsukuba, but quite a few more as unregistered ones attended as well. In fact, Tic4 recorded the largest participation and attendance of this sort in Tsukuba's history.

I am also happy to tell you that we were successful at this early stage in arranging for the publication of your contributions, having obtained a book contract from Lawrence Erlbaum Associations in New Jersey, USA. Speeches at Tic4 will be published in their own respective chapters.

Furthermore, I am delighted to tell you of several historical firsts for this conference. I will make certain that the program summary of Tic4 oral presentations will be provided as a postscript (Appendix PS.1). It will include not only the names of the speakers, but also those of the interpreters and their affiliations. The same is true for the two substitute readers (for absent speakers). Similarly, the titles of all of the poster presentations will be listed along with all authors and affiliations (see Appendix PS.2). So, to all those who contributed to Tic4 in one capacity or another, please look forward to its publication, as evidence of your productivity that may be shown to Departmental Chairs, University libraries, colleagues, students, and family members.

William K. Estes (seen in Slide 1), my mentor and a winner of the US National Medal of Science, will grace our volume with a foreword, provided that all of the speakers write first-rate pieces. So, speakers, you've got awesome responsibilities!

All of our presentations have now been completed. The only regrettable fact is that even this most productive conference must end now. Sincere thanks are due to all those who took the time to attend from early morning till late afternoon for these three days. You made this Conference an unprecedented success.

We now leave with profound gratitude to the government of Japan, the *Monbusho* that underwrote this international conference on memory, and the Institute of Psychology of the University of Tsukuba that administered it.

The hard work by Ohta *Sensei* and his staff, associates, and students who assisted all practical and technical aspects made this conference possible. Ohta *Sensei*, would you please stand up, and receive the first round of thunderous applause! Thank you Ohta *Sensei*! Then, our sincerest appreciation to his staff! They are now lined up along both left and right sides of this auditorium. Please give their efforts a big round of applause! Their names are listed in Appendix PS.6 of our volume!

Also, Tic4 for the first time provided Japanese translations of all speeches to allow easier access to the frontier of memory research by the Japanese audience. How did you like it? <Tremendous applause followed.> I am delighted to hear such sincere approval!

Our 12 devoted volunteer translators who worked diligently without compensation also made this possible. Their names are on the screen (Slide 2, and also in Appendix PS.1 in our volume). Translators, please stand up. This round of big applause is for you, for a job well done! Thank you so much for your valuable contributions! <Big applause.>--

Please be seated.

As with any venture of this nature, the heart of its success depends on the superlative quality of speakers, who shape the world's learning and memory research. Their names are on the screen in the order of presentation (Slide 3). They constitute a galaxy of stars in our universe! You have our deepest appreciation from the bottom of our hearts! Speakers, please stand up and be recognized again. Audience, one more time, please give them the biggest applause! <A thunderous applause followed.> Thank you, speakers. Please be seated.

It is our hope that all participants enjoyed partaking in the 4th Tsukuba International Conference on Memory, and that it further sharpened their appreciation of the emerging trends reflected in what we have been privileged to hear and see.

We close now in the expectation that they will continue to mold tomorrow's psychology of learning and memory and that all of us shall meet again soon to discuss cutting edge advances.

Until then, we would like to mark the success of the 4th Tsukuba International Conference on Memory, Human Learning and Memory: Advances in Theory and Application, in a Japanese way by three "*Banzai*" cheers signifying best wishes for Tic4 accomplishments and its participants. *Ban* (= *Man*) of *Banzai* means "10,000," and *Zai*, "Years." Thus, *Banzai* signifies "Best wishes for 10,000 years of a life full of accomplishments." By repeating *Banzai* three times, our best thoughts are for 30,000 years of your happiness! We have these *banzai* cheers for practically all happy occasions/celebrations, including weddings, completion of difficult projects, purchasing a new house, achievement/success of any kind, winning awards or elections, etc.

Our cheers will be led by Ohta *Sensei*. So, after every one of Ohta *Sensei*'s *Banzai*, you follow him with a louder cheer of *Banzai*, because it's your 10,000 years of life, and those of Tic4!

Please rise. Ohta *Sensei, Dozo*!

"For success of Tic4 and for its participants, *Banzai!*"

"*Banzai!*"

"*Banzai!*"

Thank you. Please be seated.

Then, Overseas Chair Izawa sang a Farewell Song verse by verse in Japanese, followed by English translation:

Farewell, my Friends,
Until We Meet Again, Godspeed,
By Keeping Fond Memories Deep
In our Bosoms Forever,
Sayonara, Sayonara, Sayonara!

APPENDIXES

Appendix PS.1 provides a summary of the entire three day Tic4 program, involving 11 scientific speakers of five nations from four continents; whereas Appendix PS.2 covers the 79 posters inclusive of 13 invited overseas poster presentations. Altogether, the Tic4 Volume was made possible by 225 outstanding human learning and memory experts. They represented 73 universities of 10 countries. Appendixes PS.3 and PS.5 provide abstracts for the speeches by Shiffrin and Kawaguchi *Sensei*s who could not participate in the Tic4 volume.

Appendix PS.5 by Rich Shiffrin was accepted for publication by Co-Editor Izawa prior to the conference. Appendix PS.6 was prepared by Overseas Organizer Izawa to acquaint visitors from abroad with the essentials of Japanese cultural history and traditions.

Our heartfelt thanks for a smoothly and professionally run conference go to the 10 Tic4 officers listed in Appendix PS.7. Preparations of Appendixes PS.1, PS.2, and PS.7 were under the direction of Co-Editor Ohta, ably aided by Editorial Assistant Althea J.E.K. Izawa-Hayden.

APPENDIX PS.1

Program Summary: Oral Presentations
The 4th Tsukuba International Conference on Memory
Human Learning and Memory:
Advances in Theory and Memory
11, 12, and 13 January 2003

Chairpersons:
Nobuo Ohta (University of Tsukuba, Japan)
Chizuko Izawa (Tulane University, USA)
Hiroshi Yama (Kobe College, Japan)

Saturday, 11 January 2003:

Opening Remarks: **Nobuo Ohta**, University of Tsukuba, Japan
(Self-Translation)
Chair: Chizuko Izawa

Speaker 1: **Jeroen G. W. Raaijmakers**, University of Amsterdam, The Netherlands
Modeling implicit and explicit memory
(Translator: Yayoi Miyaji, Kobe College)

Speaker 2: **Chizuko Izawa**, Tulane University, USA
In search of optimum learning: Psychophysiological similarities and differences between study (S) and test (T) trials and effects of study-test (S-T) presentation programs
(Self-Translation)

Speaker 3 in Absentia: **Charles Brainerd**, University of Arizona, USA,
Read by Takafumi Terasawa, Okayama University
Fuzzy-trace theory and memory
(Translators: Takafumi Terasawa and Tazuko Aoki, Okayama University)

Speaker 4 in Absentia: **Valerie Reyna**, University of Arizona, USA
Read by Terry Joyce, Tokyo University of Foreign Studies, Japan
Fuzzy-trace theory, judgment, and decision-making
(Translator: Shigeru Ono, Tokyo Metropolitan University)

Japanese Banquet and Cultural Presentations:
Nihon Odori [Classic Japanese Dances]

Sunday, 12 January 2003:

Speaker 5: **Michael S. Humphreys**, University of Queensland, Australia
Recollection and familiarity
(Translator: Ryuta Iseki, University of Tsukuba)
Speaker 6, Keynoter: **Richard M. Shiffrin**, Indiana University, USA
Modeling of memory and perception
(Translator: Hideaki Shimada, University of Tsukuba)

Poster Session:
See Appendix PS.2 in Postscript for All Poster Titles

Speaker 7: **Jun Kawaguchi**, Nagoya University, Japan
Interaction between memory and environment: Automatic and intentional processes
(Self-Translation)
Speaker 8: **Alice Healy**, University of Colorado, USA
Optimizing the speed, durability, and transferability of training
(Translator: Hama Watanabe, Nagoya University)

Monday, 13 January 2003:

Speaker 9: **Nelson Cowan**, University of Missouri, USA
Working-memory capacity limits in a theoretical context
(Translator: Satoru Saito, Kyoto University)

Speaker 10: **Douglas L. Nelson**, University of South Florida, USA
Implicit activated memories, the missing links of remembering
(Translator: Kazuo Mori, Shinshu University)
Speaker 11: **Lynn Hasher**, University of Toronto, Canada
It's about time: Circadian rhythms, memory, and aging
(Translator: Etsuko Harada, Hosei University)

Closing Statement: **Chizuko Izawa**, Tulane University, USA
(Self-Translation)
Chair: Nobuo Ohta

APPENDIX PS.2

Tic4 — Human Learning and Memory: Advances in Theory and Application

POSTERS

Listed alphabetically in terms of the last names of the first authors, the Tic4 posters' authors, their affiliations, and titles are entered below in bold, parentheses, and italics, respectively.

Invited Overseas Posters

Ciccarelli, L. (University of Rome, Italy), **& Job, R.** (University of Padova, Italy). *Phonological information and working memory in preschool children.*

Hockley, W. E., & Wells, M. (Wilfrid Laurier University, Canada). *The influence of study presentation on criterion changes and the revelation effect for words versus nonwords in recognition memory.*

Korsnes, M. S., & Magnussen, S. (Stanford University, USA). *Fast perceptual priming in recognition memory.*

Mak, B. S. K., & Lo, C. L. Y. S. (University of Hong Kong, China). *A developmental change in facial recognition between preschool children and adults.*

Mantyla, T. (Umea University, Sweden). *Future-oriented metamemory: Prospective memory complaint and impairment in middle-aged* *adults.*

Mayers, A. (Université de Sherbrooke, Canada). *Utility of episodic knowledge in MIACE architecture.*

McEvoy, C. L. (University of South Florida, USA). *Indirect activation: Preserving semantic priming in older adults?*

Meeter, M. (University of Amsterdam, the Netherlands), **Murre, J. M. J.** (University of Amsterdam and University of Maastricht, the Netherlands), **& Janssen, S. M. J.** (University of Amsterdam, the Netherlands). *Fitting a patient's remote memory: Tests, models, outcomes.*

Pate, D. S. (State University of New York, Potsdam, USA). *Implicit memory: A critical history of the concept.*

Schneider, V. I., Healy, A. F., Kole, J. A., & Barshi, I. (University of Colorado, USA). *The verbal representation of navigation instructions depends on the spatial representation.*

Soraci, S. A. (Tufts University, USA), **& Murata-Soraci, K.** (Emerson College, USA). *Generative learning and human memory: Projective understanding.*

Wagenmakers, E.-J., Ratcliff, R., Gomez, P., & McKoon, G. (Northwestern University, USA). *A diffusion model for strategy effects in visual word recognition.*

Wenger, M. J. (University of Notre Dame, USA). *On the possibility of parallel retrievals: Modeling and testing the alternatives.*

Posters

Amemiya, Y., & Sekiguchi, T. (Tokyo Gakugei University, Japan). *Involuntary recollection of autobiographical memories during a semantic differential test.*

Bai, C., & Iwasaki S. (Tohoku University, Japan). *Effect of viewing orientation on implicit memory for Kanji characters.*

Boku, M. (St. Andrew's University, Japan). *An interface between inference and interpretation in understanding silence: Applications to second language learning.*

Gamboz, N. (University of Trieste, Italy), **Russo, R.** (University of Essex, UK), **& Semenza, C.** (University of Trieste, Italy). *Evidence for equivalent directed forgetting effects in young and old adults under both item-by-item and list cueing method.*

Goto, Y. (Hokusei Gakuen University, Japan). *Computational model of* *the relation between the processes of music perception and evocation of emotion.*

Gotoh, F. (University of Tsukuba, Japan). *Negative words as attention grabbers in task switching.*

Habuchi, Y. (Hiroshima University, Japan). *Nonverbal context effect in abstract word translation: Examination of the asymmetry model of bilingual memory representations.*

Hanafusa, M., Minamidate, T., & Ikoma, S. (University of Tsukuba, Japan). *The effects of background music on cognitive activities.*

Harada, E. T. (Hosei University, Japan), **& Suto, S.** (Chuo University, Japan). *Aging and interferences from lures on the same screen: Examining spatial/temporal inhibition by a selection task.*

Hayashi, M., Ohwada, M., & Eto, M. (University of Tsukuba, Japan). *Hypermnesia in implicit memory on maze tasks.*

Hirata, C. (Institute of Physical and Chemical Research, Japan). *Learning the multiple maps in manual point predicting task.*

Horiuchi, T. (Tokai Women's University, Japan), **Nakai, Y., & Nakamura, H.** (Nagoya University, Japan). *Autobiographical encoding and self-reference effect.*

Ideno, T. (Waseda University, Japan). *Does the Stroop interference depend on stimulus exposure-duration?*

Iijima, M, (Matsue National College of Technology, Japan). *Which is the most vital for learning English? Memory, thinking, or information processing?*

Ikoma, S. (University of Tsukuba, Japan). *Implicit memory for novel melody.*

Imai, H. (Tokyo Woman's Christian University, Japan). *Automatic plus intentional attention enables whole-to-part repetition priming.*

Iseki, R. (University of Tsukuba, Japan). *Can the fan effect occur in stories?*

Itoigawa, Y. (Wisdon Inc., Japan). *Memory reinforced but possibility creating new stereotyping existed in the Internet.*

Joyce, T. (Tokyo University of Foreign Studies, Japan). *Frequency and verb-morphology effects for constituents of two-kanji compound words.*

Kaji, Y., & Naka, M. (Tokyo Metropolitan University, Japan). *Effects of handwriting movement on memory for numbers.*

Kato, K., Matsui, M., & Kurachi, M. (Toyama Medical and Pharmaceutical University, Japan). *Age-related changes in cognitive function of elderly healthy people: An examination through Mini*

Neuropsychological Scale.

Kawaguchi, J., & Watanabe, H. (Nagoya University, Japan). *Aging effect on explicit and implicit memory, and meta-memory in healthy adults using a brief memory test.*

Kawano, R., Ayabe-Kanamura, S., & Ohta, N. (University of Tsukuba, Japan). *The mere exposure effect on sweet taste.*

Kobayashi, T. (University of Tsukuba, Japan). *Evidence for functional dissociation of striatal and hippocampal cholinergic systems in spatial localization.*

Kobayashi, Y., & Kawasaki, E. (Kawamura Gakuen Woman's University, Japan). *The role of working memory capacity and vocabulary in processing garden-path sentences.*

Kosaka, K., & Tamaoka, K. (Hiroshima University, Japan). *Developing an auditory sentence comprehension test to screen language learning difficulties in Japanese preschool children.*

Matsuda, K. (Kyoto University). *A mere exposure effect for the concept formation2: The effect of duration on the typicality and the effective judgment.*

Matsui, M., Yuuki, H., Kato, K., & Kurachi, M. (Toyama Medical & Pharmaceutical University, Japan). *The nature of memory impairments in patients with schizophrenia spectrum disorder.*

Miyaji, Y., & Yama, H. (Kobe College, Japan). *Is false memory of CNWs' shapes created in encoding process or retrieval process?*

Miyawaki, K., Takahashi, M., Une, Y., Ueda, T., & Nishimoto, T. (Waseda University, Japan). *Japanese normative measures for 359 line drawings: Naming time, name agreement, age of acquisition, and familiarity.*

Mori, K. (Shinshu University, Japan). *"No, Mum. It was a white car:" What happens if mother and child dyads witnessed the same event differently?*

Morita, T. (Kansai University, Japan). *Reminders in event-based prospective memory tasks.*

Naka, M. (Tokyo Metropolitan University, Japan). *Lawyers' perception of the causes of false charges and mistrials: A review of two surveys by Japanese Bar Association and implications for future study.*

Nakajima, M., & Kikuchi, T. (University of Tsukuba, Japan). *Effect of different voices on repetition deafness.*

Nakamaru, S. (University of Komazawa, Japan). *Effects of repeating marks to degree of emotion, evaluation, and trust.*

Niki, C., & Ohigashi, Y. (Kyoto University, Japan). *Disinhibitional* *associative priming found in a left anterior temporal lobe damaged patient.*

Niki, K., & Jing, L. (National Institute of Advanced Industrial Science and Technology, Japan). *Hi-level cognition and hippocampus.*

Ochi, K. (Tokyo Kasei University, Japan), **& Sagara, Y.** (Teikyo University, Japan). *Flashbulb memories of the murder and injury of children at Ikeda Elementary School.*

Okumura, K. (Nagoya University, Japan). *Developmental change of the self-reference effect on free recall.*

Phillips, S., & Niki, K. (National Institute of Industrial Science and Technology, Japan). *Increased bilateral occipito-parietal activity for retention of binary versus unary indexed lists in pair recognition.*

Powell, S., & Butler, M. (United States Air force Academy, USA). *Learning and memory recall compared to spatial reasoning and verbal comprehension: Cognitive ability testing across the bell curve.*

Rapinett, G., & Rusted, J. (University of Sussex, UK). *The intention superiority effect: A real case of superiority?*

Saeki, E. (Nagoya University, Japan), & **Saito, S.** (Kyoto University, Japan & University of Bristol, UK). *Phonological loop contribution to task switching performance: Evidence from an articulatory suppression technique.*

Saito, S. (Kyoto University, Japan, & University of Bristol, UK), **Jarrold, C.** (University of Bristol, UK), **& Gunn, D. M.** (University of Stirling, UK) *Effects of memory load and sentence order in a reasoning span test: Testing a task-switching hypothesis of working memory span performance.*

Saitou, T., Yamamoto, Y., & Kamio, Y. (Kyushu University, Japan). *Developmental variations of social cognition I: Face preference.*

Sano, T. (University of Tsukuba, Japan). *Effect of non-target words on reading span task performance.*

Shigemori, M. (Railway Technical Research Institute, Japan). *The mechanism and classification of human errors.*

Shimada, H., & Iwata, Y. (University of Tsukuba, Japan). *Structure of representations of commutative problems in mental arithmetic: An examination by priming paradigm.*

Shimojima, Y. (Kyorin University, Japan), & **Conway, M. A.** (University of Durham, UK). *Subjective elapsed time of "observer" and "field" memories.*

Sugimori, E., & Kusumi, T. (Kyoto University, Japan). *Repetition* *effects on frequency judgments of source.*

Suzuki, Y. (University of Tsukuba, Japan). *Can retrieval-induced forgetting (RIF) be observed for categorized-nonsense-syllables?*

Tajika, H. (Aichi University of Education, Japan), **Neumann, E.** (University of Canterbury, New Zealand), **Hamajima, H.** (Nagoya University, Japan), **& Iwahara, A.** (Shoin-Higashi Junior College, Japan). *Eliciting false memories on implicit and explicit memory tests after incidental learning.*

Takahama, S., & Kumada, T. (National Institute of Advanced Industrial Science and Technology, Japan). *Memory-based Simon effect.*

Takahashi, M. (University of the Sacred Heart, Japan). *The powerful effect of spaced retrieval-learning on the memory of the sutra.*

Terasawa, T. (Okayama University, Japan), **Yoshida, T.** (Tokoha Gakuen University, Japan), **Maemoto, K.** (Okayama Higashi Commercial High School, Japan), **Murayama, K.** (University of Tokyo, Japan), **Katsube, A.** (Okayama University, Japan), **& Ohta, N.** (University of Tsukuba, Japan). *Does five-minutes exercise in learning second-language words improve learner's lexical ability?*

Toyota, H. (Nara University of Education, Japan). *Developmental changes in the self-choice elaboration effects on incidental memory.*

Uchikoshi, A. (Tokyo Metropolitan University, Japan), **Yama, H.** (Kobe College, Japan), **& Naka, M.** (Tokyo Metropolitan University, Japan). *The effect of imagery on suggestibility.*

Ueda, T. (Waseda University, Japan). *On the working memory interference in the categorization task.*

Uehara, I. (University of Tokyo, Japan). *Young children's false episodic reports: Consideration on development of episodic memory.*

Une, Y., Takahashi, M., Ueda, T., Miyawaki, K., & Nishimoto, T. (Waseda University, Japan). *A standardized set of nonsensical paired pictures (doodles) for use in experiments of memory and cognition.*

Watanabe, T. (Sendai Shirayuri Women's College, Japan). *Retrieval inhibition in the generation effect.*

Yama, H. (Kobe College, Japan), **Nishioka, M.** (Konan Women's University, Japan), **Horishita, T.** (Osaka University, Japan), **Miyaji, Y.** (Kobe College, Japan), **& Taniuchi, J.** (Osaka University, Japan). *Cultural differences in thought: A dual process theory explanation.*

Yamada, N. (Kyushu University, Japan). *Theoretical interpretation for superiority effect of auditory channel on impression formation.*

Yamamoto, Y., Saitou, T., & Kamio, Y. (Kyushu University, Japan). *Developmental variations of social cognition II: Self vs. others.*

Yamashita, M., Maruyama, M., & Itsukushima, Y. (Nihon University, Japan). *Recognition memory of cognitive interviewee.*

Yoshida, T. (Tokoha Gakuen University, Japan), **Terasawa, T.** (Okayama University, Japan), **Maemoto, K.** (Okayama Higashi Commercial High School, Japan), **Murayama, K.** (University of Tokyo, Japan), **Katsube, A.** (Okayama University, Japan), & **Ohta, N.** (University of Tsukuba, Japan). *Examining the relation between subjective evaluation and objective performance of word test in long-lasting English words learning.*

APPENDIX PS.3

Modeling of Memory and Perception

Speaker 6: Richard M. Shiffrin
Indiana University, USA

Abstract

I present a framework for modeling memory, retrieval, perception, and their interactions. The models are inspired by Bayesian induction to determine optimal decisions, in the face of a memory system with inherently noisy storage and retrieval. The origins of the modeling enterprise precede the Bayesian approach: They begin with the Atkinson and Shiffrin article in the 1960s, emphasizing the distinction between short- and long-term memory, and the control processes of short-term memory, and include the SAM modeling of Raaijmakers and Shiffrin at the start of the 1980s that highlighted retrieval from long-term memory. The starting point for the Bayesian modeling was the Retrieving Effectively from Memory (REM) model for episodic recognition (Shiffrin & Steyvers, 1997), but it should be noted that this model was a natural outgrowth of the earlier modeling efforts and remains largely consistent with them.

The general REM framework describes: 1) the storage of episodic traces, the accumulation of these into knowledge (e.g. lexical/semantic traces in the case of words), and the changes in knowledge caused by

learning; 2) the retrieval of information from episodic memory and from general knowledge; 3) decisions concerning storage, retrieval and responding. I give examples of applications to episodic recognition, and episodic cued and free recall, perceptual identification (naming, yes-no, and forced choice), lexical decision, and long-term and short-term priming.

APPENDIX PS.4

The following invited address was delivered by Richard M. Shiffrin, a guest of honor at the 19 November 1999 testimonial dinner-symposium, organized and chaired by Chizuko Izawa, at the 1999 Psychonomic Society Meeting in Los Angles, California, USA, in celebration of the 30th Anniversary of the Atkinson-Shiffrin Model (1968), one of the most cited references in the field of psychology, not to mention in experimental and cognitive psychology in the world.

On the Atkinson-Shiffrin Chapter

Richard M. Shiffrin
Indiana University, USA

When I promised Chizuko Izawa to speak on the Atkinson-Shiffrin chapter, I had certain trepidations, because I didn't remember our chapter particularly well. Of course, I wasn't overly worried because this situation could be remedied with a quick refresher course. So I turned to my files, where a thorough search failed to turn up even one copy of our chapter. Requests to my colleagues were even less fruitful. I turned to the IU library, but the staff also failed to locate a copy; they ordered one through interlibrary loan, and informed me that, as usual, the expected waiting time would be: 'indefinite.'

At this point, I decided to revert to a back-up plan: I would do as everyone else does and read about our chapter in secondary and tertiary sources. I began this process but then a startling insight came to me: Our chapter had over the years migrated from the category of 'Science' to that of 'Folklore,' or perhaps 'Myth,' joining other famous articles, like "The Magic Number Seven," "Toward a Statistical Theory of Learning," and others whose names I am sure you can supply, writings that no one reads in their original form. In fact, the Atkinson-Shiffrin article, like these others, needs not actually ever have existed for it to serve its present purposes.

Please do not misquote me here. I am not claiming that these articles, and in particular the Atkinson and Shiffrin chapter, were never written. In fact, I am virtually sure that at least one or two may have truly seen it in print. The critical point, however, is that it no longer matters today whether this is the case. The first slide lays out the role of such articles in present-day scientific discourse.

Publication Policy

Editors of books and journals, and reviewers as well, somewhat peevishly insist on filling a few pages at the end of any publication with a particularly arcane part of the scientific tradition known as citations. I don't want to start a rant here about how citations serve as the barter of academic commerce, and in reality I can certainly see a purpose behind reference sections. So don't take my comments the wrong way--a few references to some particularly relevant writings would undoubtedly be justifiable scientifically. However, authors quickly learn that this practice is fraught with danger. Most editors take advantage of the situation and send a submitted article to the cited authors for review. In turn, these authors make heated demands that their theories be presented fairly, their data presented accurately, and their conclusions represented correctly. At best, the submitted article is returned for extensive revision, and more commonly rejected outright. Under these circumstances it becomes almost impossible to get anything published, and authors in our field quickly learn the only adequate defense: The 'mythic citation,' such as the Atkinson-Shiffrin chapter.

Citing a myth has many beneficial effects. For one example, publication probability is increased. Why? Because the authors of mythic citations seldom play a negative role in the reviewing process. First, they are seldom asked to review articles citing their articles. Second, their fame allows them to return submitted articles unreviewed. Third, they have only the faintest remembrance of the original article, and hence only the foggiest notion of the possible relation between it and the submission.

Perhaps of greatest utility, almost anything can be claimed to have been said in such mythic articles, since no one has read them in their original form; it is especially safe to make dubious claims in those cases where the article never actually existed, as verification is correspondingly harder.

Intellectual Acumen

When papers are submitted with mythic citations, the reviewers are generally quite pleased and immediately predisposed toward acceptance. For one thing, the reviewers' own theories are protected by such ancient historical reference. For another, such citations are familiar and comfortable, and any article lacking the same would be returned forthwith with requests for more extended literature reviews. Most important, such citations lend an air of scholarship and intellectual historicity to any submission. Many studies have shown that the estimated scientific excellence of a submission rises in direct correspondence to the number of mythic citations (these studies, I might add, have large and appropriately filled reference sections).

Theoretical Antecedents

It is often extraordinarily difficult to publish a totally new theory, based only on the defense that the theory predicts the data. Psychology is inherently conservative, and new ideas are treated with the caution they deserve. The utility of noting the pre-existence of one's new theory in prior writings is therefore self-evident. The example of William James springs to mind, but this particular reference has been overdone to the point of fatigue. Better are plausible sounding claims that one's theory was presaged in one or another mythic article like that of Atkinson and Shiffrin, articles that cannot actually be located, and that have not been read. Incidentally, it is quite helpful if one's mythic reference has a large page count, in order that it has the potential to bear up under the weight of serving as scientific progenitor to so many models.

Theoretical Foil

Conversely, not to be overlooked is the role of such mythic articles as foils for whatever new theory is presently under examination. Too many researchers in our field demand that a new theory outdo some old one in order to reach publication criterion. Unfortunately, finding a theory that your new model can beat out can be a fiendishly difficult exercise, if one restricts oneself to actual articles, chapters, and books.

Mythic articles are perfect for this purpose, as even the authors, if they are still active, cannot remember what they once said, or even whether they actually wrote it. Also, as distinct from recent research, an author of a mythic article feels no disgrace in anyone pointing out deficiencies in what presently seems more like an act of history than science. The Atkinson-Shiffrin article has served honorably in this capacity for many years now. Our published defenses against such attacks have been, as is almost always the case with attacks on such articles, quite lackluster and uninspired, an entirely expected outcome, given that they were prepared on the basis of secondary and tertiary summaries.

Final Remarks

It will not fail to escape the notice of the discerning listener that becoming a mythic citation, or author, is a delicate matter, depending as it is on a mathematical catastrophe in psychological hypothesis space. At first, an article is cited accurately for its content, until in rare cases a critical mass of citations takes place; from that point onwards, authors feel compelled to include such citations whether or not they have any relevance for the scientific enterprise, for all the reasons I have laid out.

This argument brings us to the final mystery, which can be stated thusly: If mythic articles begin as real articles, is this not a proof that all such articles must have a basis in reality? The explanation of this seeming conundrum lies in Heisenberg's Uncertainly Principle, and the paradoxes of Quantum Mechanics, as laid out so elegantly in the Atkinson and Shiffrin 1968 article, I think.

Thank you all for your attention.

APPENDIX PS.5

Interaction between Memory and Environment: Automatic and Intentional processes

Speaker 7: Jun Kawaguchi
Nagoya University, Japan

Abstract

People use a variety of strategies and tools to help memorize certain things and events to facilitate everyday activities. For example, people may try to keep the name of a casual acquaintance by the method of voluntary imagery, or may take notes on items/events to be remembered or those of significance. This implies that both internal and external memory aids may work well jointly. The purpose of this study is to elucidate the relationship between internal process (e.g., memory strategies) and external environment (e.g., memory aids). First, I will show how people use these strategies via questionnaire data. This survey shows people mainly depend on external memory aids rather than internal ones. Furthermore, the way of using these strategies may change, as people get older. Second, I will cover the experiment on memory for schedule. Data suggested that the retrieval of schedule was influenced by the condition of encoding environment (calendar format). Because of an incidental nature of the experiment, the encoding of environment (calendar) might be considered to be automatic. In closing, I comment on the interaction between internal memory processes and external environment.

APPENDIX PS.6

Discover Japan, the Host: Enduring Cultural Heritages

Chizuko Izawa
Tulane University, USA

Japan, the host of the Tsukuba International Conferences on Memory, is the country of Sony, Toyota, Honda, NEC, Mitsubishi, Toshiba, and many other modern high tech firms known around the world. It is fitting, indeed, that Tsukuba (about 80 km northeast of Tokyo) is an high tech city specifically developed to be a part of the University of Tsukuba's infrastructure. University-industry relationships similar to those of, for example, Stanford University and Silicon Valley have developed. It is the product of the arrangements that also reflects Japan's curiosity and willingness to adopt and refine advanced foreign technology/culture, as it has done throughout its long history. At the same time, Japan also proudly maintains its time-honored and unique traditions and culture. It is a country of old and new, foreign and domestic, all harmoniously blended and coexisting.

This Postscript Appendix is devoted to two examples of Japan's unique cultural heritages easily missed by most tourists who may appreciate, without difficulties, the physically perceptible such as natural wonders, objects d'art, monuments, gardens, music, and cuisines.

Timeless Literary Masterpieces and Culture

Eleven hundred years ago in 905 (nearly six centuries before Columbus discovered America), Japan produced a supreme treasure of poetry, publishing its first official collection of *Waka*s [Japanese poems] in 20-volumes, the *Kokin Wakashu* (*Kokinshu* for short) [*the Collection of Japanese Poems from Ancient and Modern Times*]. The *Waka* includes several forms of poetry unique to Japan. Of them, *Tanka,* is so predominant that it has become synonymous with *Waka,* each consisting of 31 (5-7-5-7-7) syllables (shorter *Haiku* with 17 (5-7-5) syllables came many centuries later, e.g., Basho, 1716-1784). This collection commissioned by Emperor Daigo was the first in a series and was the most significant collection of Japanese poems along with *Manyoshu* [*the Collection of Ten Thousand Leaves*][1] completed in A. D. 759.

---

[1] This earliest extant extensive collection of Japanese poetry consists of 20 volumes (4,500+ *Waka*s composed in or before 759), a major literary accomplishment of the *Nara* period (710-794, the capital, in Nara). Given the magnitude of its poetic accomplishment, the *Manyoshu* was widely understood that its publication could not be possible without royal blessings. However, no written records seem to exist that it was formally commissioned by an imperial command as was the case with the *Kokinshu*, completed nearly 1.5 centuries later than the *Collection of Ten Thousand Leaves* [*Manyoshu*].

After Kyoto became the capital of Japan (for 1,074 years), following the *Nara* period, there was spectacular advancement in every aspect of Japanese unique culture, unadulterated by foreign influences. This was especially pronounced during the *Heian* Period (794-1185). The *Kokinshu,* the fruit of a-decade-long efforts, contains 1,111 poems by over 120 known men and women poets, plus over 450 poems of unknown authorship, the largest single group in the entire collection. These Japanese poems [*Waka*s] were composed by citizens of nearly all walks of life, from Imperial families and aristocrats to *samurai/bushi* [warriors], farmers, fishermen, or others who mastered and produced superb art (the same was true of the *Manyoshu,* 759). Surely, to be included in such a collection was the highest poetic honor.

Quite intriguingly, some of the most attractive *Waka* poems were found among the unknown authors. Indeed, the most famous of all *Waka*s, then and ever, came from the *Kokinshu* (905), later known as *Kimigayo,* by an anonymous poet.

This particular *Waka*'s unparalleled beauty and artistry extolling eternal hope/prosperity touched the hearts of the Japanese: It appealed to the sentiments of nature loving Japanese. Therefore, its immediate popularity was no small wonder: Many recite it with admiration repeatedly. My regret is that it is nearly impossible to appreciate it fully via translation (an English translation is provided in Chapter 1, nonetheless). The poem inspired a variety of widely popular songs including *Joruri*, *Kouta*, and *Biwa,* which are often heard, in musical and theatrical performances.

Literary masterpieces in the *Heian* Period (794-1185) were not limited to poetry. The richness of the Japanese language (after all, which other language can express the first person "I" in scores of different ways, and one is reserved for the Emperor/Empress alone? And similarly so are pronouns for the other persons as well!) enabled Lady Murasaki *(Murasaki Shikibu,* a 970s–1020s lady-in-waiting/advisor/ confidant to Empress Akiko/*Fujiwara-no-Shoshi*), the greatest master of the narrative prose in the history of Japanese literature, to write 54 volumes/Chapters of the *Tale of Genji* [*Genji Monogatari*] at a time when practically no prose literature existed in the West. Indeed, the *Tale of Genji* is the first great novel in world literature (Seidensticker, 1983), continuously read for over the last 1,000 years by Japanese students and the public, not to mention scholars and specialists worldwide. This world classic, a supreme masterpiece for its beauty, artistry, dramas, and the human condition still enchants readers.

Lady Murasaki started writing this novel before she entered imperial court service at a young age and continued writing throughout her life, in the process creating several hundreds of characters, and at least 50-60 major ones! The protagonist of the first 44 volumes is *Shining Genji* (a prince in the imperial household), who lost his mother shortly after his birth. It is said his many romantic encounters reflected a search for his mother's identity. The complexities of court life prevented him from becoming an Emperor, and the last 10 volumes are about *Genji*'s son.

The unique literary success came from not only just dramas so exquisitely created, but also from her ability to make the events so realistic (stemming from her own experiences in the imperial court). As she grew mature, wiser and experienced in life, so did her characters. Every installment/volume of the literary treasure, the *Tale of Genji,* was a sensation, eagerly awaited by the admiring readership for its elegantly poetic beauty in Lady Murasaki's novel, meticulously written in superb calligraphy. It is still appreciated by every high school student throughout Japan today, as has been for the last 10 centuries.

Lady Murasaki's success can, in part, be attributable to her high intellect, creativity, a rich artistic imagination, spiritual attainments, superb knowledge and mastery of the Japanese language and history, and an unlimited (so seems) memory for these characters and the daily events she involved them in (she would have made a spectacular participant in human learning and memory experiments!), human significance, her life philosophy, religious (inclusive of karma) and political sophistications, and ability to utilize readers' imagination.

It is most intriguing indeed that some 600 years later, another literary genius, William Shakespeare (1564-1616) also wrote dramas about a prince (Hamlet), an emperor (Julius Caesar), many English kings, and stories/lives of noble families (which, however, are often too violent/bloody for Japanese taste, notwithstanding their high artistic values).

The *Tale of Genji* has, of course, greatly impacted later literature and various other arts, as well as life philosophy, popular doctrines, and schooling/education. Theatrical productions inspired by the *Genji Monogatari* proliferated; practically all of these are still with us today. For example, the *Nō* plays (musical dance-dramas), performed in the oldest professional theaters of Japan that originated in mid-14th century (informal origins date back by several centuries earlier). The *Tale of Genji* was one of the principal sources for the *Nō* drama. After *Kabuki* started in the early 17th century, *Genji* was adapted to *Kabuki* stages.

Fictional recounting and adaptations of the *Tale of Genji* have continued throughout centuries, including recent and modern adaptations to movies and television productions. Of several renditions into modern Japanese, some have become best sellers. The *Tale of Genji* continues to be an essential and unparalleled source of inspirations for artists for over 1,000 years!

For another example, *the Genji Monogatari* generated a new theme of exquisite *Yamato-E* [native style Japanese painting] for illustrated hand scrolls [*E-Makimono*] portraying scenes of this novel more than 100 years after it was written, but still in the *Heian* Period (794-1185). These paintings utilize particular styles and techniques, richly colored, highlighted by gold and silver; lavish use of thick colors which were carefully selected to enhance the mood of each scene. Originally, there were 1–3 scrolls for each of the 54 volumes, but the oldest and finest examples are now found in a few museums in Japan and other collections. The *Tale of Genji* with accompanying illustrations in the 11^{th} -12^{th} centuries set standards for literary and pictorial illustrations of modern novels. These *Yamato-E*s also help to preserve visual images of imperial robes and headgear, as well as descriptions of imperial ceremonial practices.

The costumes representative of the *Heian* Period Court have been recreated for official functions in the modern Imperial Household for coronations/enthronements, imperial weddings (a recent example being the marriage of the current Crown Prince a decade ago), and similar functions.

It is remarkable indeed that the depictions of court life in literary works more than 1,000 years old became linked to the scientific presentations at the 4^{th} Tsukuba International Conference on Memory.

SPECIFIC REFERENCES

Seidensticker, E. G. (1983). Tale of Genji. *Kodansha Encyclopedia of Japan, 7*, 328-331. Tokyo: Kodansha.

GENERAL REFERENCES

Parts of general backgrounds/sources of the detailed historical facts for Appendix PS.6 came from the nine volumes of the *Kodansha Encyclopedia of Japan*, Tokyo: Kodansha (1983), including some of the sections by E. Minor, E.

G. Seidensticker (cf. Chap. 1), Y. Suzuki, and S. Thompson, among many others.

APPENDIX PS.7

Acknowledgements:
Officers of Tic4: Staff and Graduate Students
Institute of Psychology
University of Tsukuba, Japan

Fumiko Gotoh	Terry Joyce
Yoko Hara	Rie Kawano
Mitsuko Hayashi	Yasuyuki Sakuma
Satoshi Hoshino	Tetsuya Yoshida
Shinobu Ikoma	Emiko Yoshimura

This page intentionally left blank

Author/Name Index

A

B

C

D

E

F

G

H

I

L

M

N

O

P

Q

R

S

T

U

V

W

X

Y

Z

Subject Index

A

B

C

D

F

G

H

I

J

K

L

M

N

O

P

Q

R

U

V

W

Y