Interpreting one’s subjective experience: Heuristics and biases

Robert A. Bjork
University of California, Los Angeles

Festschrift in Honor of Douglas L. Nelson
University of South Florida
Tampa, Florida
March 21-22, 2008
Interpreting subjective indices of learning and performance

- Perceptual fluency;
- Retrieval fluency;
- Fluency of induction;
- Fluency of association;
  - Foresight bias
  - Inflation of conditional predictions
- Experience-based vs. theory-based predictions of one’s own forgetting or learning;
- Egocentrism in social communication
  - Concluding comments on our subjective experience as teachers.
Important qualification

Subjective experience is not *always* misleading, it is sometimes even the best basis for judgments.

Example: Jacoby and Kelley (1987)
Jacoby & Kelley (1987)

- FSCAR ?????
- vs.
- FSCAR SCARF
“Subjective experience, like the public media, is unavoidable, serves useful functions, and is not to be fully trusted”
Prelude: Interpreting objective indices of performance

- **Learning versus performance**
  - What we can **observe** is performance;
  - What we must **infer** is learning;
  - …and the former is an unreliable guide to the latter.
    - Conditions of instruction that make performance improve rapidly often fail to support long-term retention and transfer, *whereas* Conditions of instruction that appear to create difficulties for the learner, slowing the rate of *apparent* learning, often optimize long-term retention and transfer

- We are susceptible to being fooled by our own current performance
  - Example: Interleaved versus blocked practice
Blocked versus random/interleaved practice

Simon & Bjork (2001)

<table>
<thead>
<tr>
<th>Keys</th>
<th>Color</th>
<th>Goal MT</th>
<th>Keyboard Pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td>9-5-1-2-3</td>
<td>Green</td>
<td>900 ms</td>
<td><img src="image" alt="Keyboard Pattern" /></td>
</tr>
<tr>
<td>3-6-5-8-4</td>
<td>Red</td>
<td>1200 ms</td>
<td><img src="image" alt="Keyboard Pattern" /></td>
</tr>
<tr>
<td>4-2-5-8-9</td>
<td>White</td>
<td>1500 ms</td>
<td><img src="image" alt="Keyboard Pattern" /></td>
</tr>
</tbody>
</table>
A Typical Trial

4 seconds
(first 3 trials of each pattern)

Your goal movement time for the green pattern is 900 milliseconds.

3 seconds

participant responds plus

2-second delay

5 seconds

keys were correct
MT (ms): 840
You were 60 ms too fast

1 second
Rohrer & Taylor (2007)

Wedge $V = \frac{r^2 h \pi}{2}$

Spheroid $V = \frac{4r^2 h \pi}{3}$

Spherical Cone $V = \frac{2r^2 h \pi}{3}$

Half Cone $V = \frac{7r^2 h \pi}{3}$
Writing Capital Letters

Say the name of each capital letter below. Now trace over the letters with your crayon. Next, practice writing each letter 3 more times.

W  W  W  X  X

Y  Y  Y  Z  Z

Interpreting and misinterpreting perceptual fluency or familiarity

- Heuristic value

- Misattributions and illusions
  - Misinterpreting the cause of perceptual fluency
  - Misinterpreting the meaning of perceptual fluency
  - Illusions of competence
    - Example: Reder (1987, 1988)
“What is the term in golf for scoring one under par?”
Interpreting and misinterpreting *retrieval fluency*

- Heuristic value
- Misattributions and illusions
  - Interpreting performance as learning; illusions of competence
  - Egocentrism in instruction and social communication
  - Incomplete/faulty models of ourselves as learners/remembers
    - Example: Benjamin, Bjork, & Schwartz (1998)
Benjamin, Bjork, & Schwartz (1998)

- Phase 1: 20 (easy) general-knowledge questions
  - E. g., “Who was the first president of the United States?”
  - Participants asked to:
    - (a) hit ENTER as soon as the answer “came to mind” (latency recorded);
    - (b) say the answer;
    - (c) predict the likelihood they would be able--at the end of the experiment--to free-recall having given that answer.

- Phase 2: Distracting activity (spatial/map task)

- Phase 3: Final test
  - Free recall: Write down as many of the 20 answers you gave earlier as you can;
  - (Original questions were not shown again)
Interpreting fluency of induction/learning

- Kornell & Bjork (in press).
  - Abstracting concepts and patterns: Is “spacing the enemy of induction”? 
  - Lesson on the reliability, or lack thereof, of researchers’ intuitions.
Interleaving.spacing versus blocking/massing

Interleaved/spaced: items re-studied after other items

Blocked/massed: items studied in succession
Gentoo
Where’s the Gentoo?
Inductive learning

- The ability to generalize concepts and categories through exposure to multiple exemplars.
Hypothesis

• Blocking/massing allows the learner to notice characteristics that unify a category

• Interleaving/spacing makes doing so difficult
“Spacing is the friend of recall but the enemy of induction.”

-Ernst Rothkopf
Method: phases

1. Instructions
2. Study
3. Distractor
4. Test
5. Questionnaire

Kornell & Bjork (in press, Psychological Science)
Instructions

In this experiment you’re going to look at some beautiful paintings. To start, you’ll be shown 72 paintings for 3 seconds each. The paintings will be by twelve artists, with six pictures per artist. Try to learn to recognize which artist painted which picture based on their style.

Later, you’ll be shown 48 new paintings, which you haven’t seen before. You’ll have to identify who painted each one.

For example, if there were only two artists, named Al and Barb, you’d be shown paintings by Al and Barb, and later, you’d be shown new paintings and asked who painted them, Al or Barb.
Design

• Two within-subject conditions: massed & spaced
Massed block
Lewis
Lewis
Schlorff
Stratulat
Mylrea
Feedback

Lewis

Correct!
Final Questions

Which do you think helped you learn more, massed or spaced?

- Massed
- About the same
- Spaced

Actual Responses

Spaced

Responses

Spaced
Spacing vs. massing manipulation within participants

Spacing vs. massing manipulated between participants
Why a spacing effect?

• The test depended on recalling name-style associations
  – Spacing is the friend of recall

• Experiment 2
  – Final test: Yes/No recognition test for paintings by studied artists
Method

Instructions

Study

Distractor

Test

Questionnaire

Only one change: Recognition test
Recognition Test
Studied vs. Distractor Artists

Hawkins:

Hawkins Distractors:
Results (n=80)

- Massed
- Spaced

Hit Response

FA Response

- Massed
- Spaced

Number of participants: 80
### Final Questions

**Which do you think helped you learn more, massed or spaced?**

- **Massed**
- **About the same**
- **Spaced**

### In reality

<table>
<thead>
<tr>
<th></th>
<th>29%</th>
<th>24%</th>
<th>48%</th>
</tr>
</thead>
</table>

### Responses

|       | 80%  | 10%  | 10%  |
Differentiation hypothesis

- Original hypothesis: Blocking/massing highlights similarities
  Lewis  Lewis  Lewis  Lewis

- New hypothesis: Interleaving/spacing highlights differences
  Lewis  Schlorff  Hawkins  Lewis
Interpreting and misinterpreting fluency of association

- **Foresight bias**
  - Koriat and Bjork (2005, 2006)

- **Prediction-inflation biases**
  - Nelson, Dyrdal, and Goodmon (2005)
  - Koriat, Fiedler, and Bjork (2006)
  - Maki (2007a, 2007b)
Foresight bias (Koriat & Bjork, 2005)

- Likelihood the second word will be given as a free associate to the first?
  - Lamp: Light
  - Find: Seek
  - Sell: Buy
  - Cheese: Cheddar
  - Citizen: Tax
Foresight bias (Koriat & Bjork, 2005)

- Likelihood the second word will be given as a free associate to the first?
  - Lamp: Light (.71)
  - Find: Seek (.03)
  - Sell: Buy (.56)
  - Cheese: Cheddar (.03)
  - Citizen: Tax (.00)
Foresight bias (Koriat & Bjork, 2005)

Likelihood the second word will be given as a free associate to the first?

- Lamp: Light (.71)  Forward pair
- Find: Seek (.03)  Backward pair
- Sell: Buy (.56)  Forward pair
- Cheese: Cheddar (.03)  Backward pair
- Citizen: Tax (.00)  “Purely a-posteriori” pair
**Foresight bias (Koriat & Bjork, 2005)**

- The tendency of learners to experience an illusion of competence during learning, resulting in inflated predictions of later recall
  - Judgments of learning are made in the presence of information that is absent, but solicited, on a subsequent test
  - Such as the targets in cue-target paired associates (e.g., Cheese-cheddar)
    - We are unable to anticipate the test situation, when the answer will be absent, but required.
    - Thus, “Cheese ____?_____” will trigger other strong associates, such as “mouse,” “bread,” “wine,” etc., which will compete with “cheddar.”
<table>
<thead>
<tr>
<th>Pair type</th>
<th>Predicted recall</th>
<th>Actual recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forward</td>
<td>.78</td>
<td>.79</td>
</tr>
<tr>
<td>Backward</td>
<td>.76</td>
<td>.60</td>
</tr>
<tr>
<td>Purely a-posteriori</td>
<td>.65</td>
<td>.47</td>
</tr>
</tbody>
</table>

Koriat & Bjork (2005)
Nelson, Dyrdal, & Goodmon (2005)

QuickTime™ and a TIFF (LZW) decompressor are needed to see this picture.
What is preexisting strength? Predicting free association probabilities, similarity ratings, and cued recall probabilities (Nelson, Dyrdal, & Goodmon, 2005)

INSTRUCTIONS TO PARTICIPANTS:

- “Some word pairs are more related than others. For example, most people would say that CAT DOG is more strongly related than GUARD DOG. The purpose of this study is to collect ratings on a large number of people in order to determine whether such ratings can predict recall. If DOG were recently studied, would CAT be a better memory cue than GUARD? Such results will be important for understanding how memory works.

- “What we want you to do today is rate each pair of words that you are given on a scale of 1-7 in terms of their degree of mutual association or relatedness. If the first word readily calls the second word to mind or if you can find an easy way to relate the two words to each other, then give the pair a higher rating. Conversely, give the pair a lower rating when they seem weakly associated or related.
Nelson, Dyrdal, & Goodmon (2005): Results and discussion

- “Forward and backward strength emerged as the only significant predictor of the ratings.”
  - “… both backward (r = .30) and forward (r = .20) strength were correlated with the ratings;
  - “… the ratings explained 8.8% … of the variance in cued recall”

- “… associative similarity ratings are not … influenced by several predictors that influence cued recall, including mediated strength, target activation strength, target competitor strength, and target frequency.”
JAM = 50.3 + 0.36 FSG
$R^2 = 0.884$
**Inflation of conditional predictions**  
(Koriat, Fiedler, & Bjork, 2006)

- **Conditional prediction:** “What is the probability that event $Y$ will occur under condition $X$?
  - E.g., “What is the likelihood that if Iran stops supplying crude oil, the price of a barrel will exceed $110$”?

- **Prediction-inflation bias:**
  - Owing to a backward activation process, conditional predictions will tend to be inflated,
    - The stated outcome, whose likelihood is to be assessed, brings to the fore aspects of the condition that support the occurrence of the outcome.
    - Example: Predicting free-association probabilities
**Inflation of conditional predictions**  
*(Koriat, Fiedler, & Bjork, 2006)*

- **INSTRUCTIONS TO PARTICIPANTS:**
  - “In a word-association ask, a person is presented with a word and is asked to say as fast as he can the first word that comes to mind in response to the presented word.
  
  - “You will be presented with pairs of words. For each pair we would like you to estimate the percentage of people who would say the second word (on the right) as the first response to the stimulus word (on the left). Hebrew is written from right to left).
  
  - “You will find a blank line next to each word pair. Write on that line a number between 0 and 100 that reflects the percentage of people who, according to your opinion, would say the second word (the right one) in response to the first word (the left one).”
Koriat, Fiedler, & Bjork (2006)
Experiment 5
Koriat, Fielder, & Bjork (2006)
Experiment 6
How are memory judgments made?

- **Theory-based**
  - Deliberate
  - In general…
    - “… forgetting happens.”
    - “… studying causes learning.”

- **Experience-based**
  - Automatic
  - E.g.,
    - Encoding fluency
    - Retrieval fluency
    - Cue familiarity
Objective and subjective learning curves

- Koriat, Bjork, Sheffer, & Bar (2004): Predicting one’s own forgetting

- Kornell & Bjork (in progress): Predicting one’s own learning
Koriat, Bjork, Sheffer, & Bar (2004)
Predicting One's Own Forgetting:
The Role of Experience-Based and Theory-Based Processes.

“… judgments of learning (JOL), if governed by processing fluency during encoding, should be insensitive to anticipated retention interval.”

- **Subjective experience**: “Various mnemonic cues contribute directly to produce an immediate feeling of knowing that can serve as the basis of judgments. Thus, for example, encoding and retrieval fluency may foster a feeling of competence that can serve as [an experience-based basis for judgments of learning].”

- **Domain-specific knowledge retrieved from memory**: “Theory-based judgments, in contrast, rely on the deliberate use of specific beliefs and information to form an educated guess about one's own knowledge. Thus, JOLs may utilize such rules as “memory performance should be better on recognition than on recall memory test.”
Predicting one’s own forgetting (Koriat, Bjork, Sheffer, & Bar, 2004)

- **Experiment 1:**
  - 60 paired associates
    - 30 related;
    - 30 unrelated
  - Participants judge, pair by pair, the likelihood they will remember that pair on a later cued-recall test
  - Retention interval to the final test (between-subjects):
    - Immediately after the study phase;
    - One day;
    - One week

- **Experiment 2:**
  - To-be-predicted retention interval manipulated within-subjects.
Figure 1. Predicted recall (judgments of learning [JOL], dotted lines) and actual recall (solid lines) as a function of retention interval for related and unrelated paired associates in Experiment 1. Immed = immediate. Error bars represent 95% confidence intervals.

Exp 1: Between-subjects

Exp 2: Within-subjects
Koriat et al. (2004): Experiment 4c

“This experiment was an attempt to bring the matter to an absurdity: Would participants fail to take into account a year’s retention interval in comparison to participants who are asked to predict performance under immediate testing?”
## Results: Experiment 4c

<table>
<thead>
<tr>
<th>Measure</th>
<th>Ten minutes</th>
<th>One week</th>
<th>One year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicted recall: (Exp. 4c)</td>
<td>35.2</td>
<td>40.4</td>
<td>37.1</td>
</tr>
</tbody>
</table>
Koriat et al. (2004): Experiment 7

- Same as Experiment 4c, but participants asked to estimate number of items forgotten:
  - “In a previous experiment that we conducted, students were presented with a list of 60 word pairs such as “table-chair” and “girl-eagle” one after the other. Each word pair was presented for 4 seconds. The students’ task was to study these pairs so that when presented later with the first word, they would be able to recall the second word. The memory test took place ten minutes/one week /one year later.”
  - “In the test, the first word was presented with a blank line next to it, for example:
    - Table _________
    - Girl  _________
    and the students were asked to recall the second word and to write it down on the line. “
  - “We would like you to estimate how many of the word pairs the students forgot after ten minutes/one week/one year. Your estimate can range from 0 to 60 pairs. ... Write down your estimate at the appropriate space at the bottom of the next page.”
## Results: Experiment 7

<table>
<thead>
<tr>
<th>Measure</th>
<th>Retention Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ten minutes</td>
</tr>
<tr>
<td>Predicted recall: (Exp. 4c)</td>
<td>35.2</td>
</tr>
<tr>
<td>Predicted recall: (Exp. 7)</td>
<td>43.4</td>
</tr>
</tbody>
</table>
Koriat et al. (2004) conclusions

- “... participants can access their knowledge about forgetting...”

- but only when theory-based predictions are made, and then only when the notion of forgetting is accentuated ... 
  - either by manipulating retention interval within individuals, or
  - by framing recall predictions in terms of forgetting rather than remembering.”
Experiment 1: Predicting one’s own learning

- 24 paired associates
  - 12 related (Hill-Valley)
  - 12 unrelated (Clemency-Idiom)

- Number of study/test cycles (between-subject):
  - ST
  - STST
  - STSTST
  - STSTSTST

- Participants judge, pair by pair, the likelihood they will remember that pair on the first, second, third, or fourth cued-recall test cycle
  - Between-subjects
  - “Complete this sentence: When I make my 0-100% rating, I will do it about test number ___.”
Hill
Valley
Hill Valley

Enter the chance that you'll remember this on the final test (0-100%).
Clemency
Idiom
Clemency
Idiom

Enter the chance that you’ll remember this on the final test (0-100%).
Methodological reasons why predicting one’s own forgetting and learning might differ

- First, there is a match between learning and "remembering"
  - Analogous to Koriat et al.’s Experiment 7 in which forgetting and “forgetting” match
  - Versus Koriat et al.’s other experiments in which forgetting and “remembering” mismatch

- Second, when predictions were made for the second, third, and fourth test cycle, it was clear that there could be multiple study/test cycles, which should then trigger the concept of learning and activate theory-based judgments
Exp. 1 (between-subjects)
Exp. 1 (between-subjects)  Exp. 2 (within-subjects)
Concluding comments: 
Predicting one’s forgetting versus predicting one’s learning

- *Like* Koriat et al.’s results …
  - participants were sensitive to item differences
- *Unlike* Koriat et al.’s results …
  - within-subject manipulation of number of study-test cycles did *not* appear to activate the concept of learning
- Why?
  - Perhaps the concept of learning *was* activated, but people have a faulty and/or incomplete mental model of learning:
    - the benefits of study are underappreciated;
    - and the benefits of tests as learning events are not understood (e.g., Kornell & Son, 2006)
  - Or perhaps people are captured by item differences, which blocks access to what they know (or don’t know) about learning:
    - Tentative conjecture: Among the hierarchy of experience-based and theory-based processes, the concept of learning falls below the concept of forgetting.
Concluding comments on our subjective experience as teachers

- Egocentrism in social communication
  - Newton (1990) as a parable of teaching;
  - Piaget (1962) quote
  - Calvin & Hobbes
NEWTON (1990)

Percent songs identified

Listeners' performance predicted by tappers

Listeners' actual performance
NEWTON (1990)

- Listeners' performance predicted by tappers
- Listeners' actual performance
“Every beginning instructor discovers sooner or later that his first lectures were incomprehensible because he was talking to himself, so to say, mindful only of his point of view. He realizes only gradually and with difficulty that it is not easy to place one’s self in the shoes of students who do not yet know about the subject matter of the course.”
QuickTime™ and a TIFF (LZW) decompressor are needed to see this picture.
QuickTime™ and a TIFF (LZW) decompressor are needed to see this picture.
The End
References


