

Blending:
An ACT-R Mechanism for
Aggregate Retrievals

Christian Lebiere
Psychology Department
Carnegie Mellon University
Pittsburgh, PA 15213
<http://act.psy.cmu.edu>

Plan

- Background
- Basic mechanism
- Spectrum of applicability
- Application: similarity judgment
- Application: magnitude estimate
- Details
- Activation
- Related IA mechanisms
- Conclusions

Background

- Goal: “ACT-R doesn’t have a mechanism for producing continuously varying answers like confidence ratings, similarity judgments, or magnitude estimates.”
Atomic Components of Thought p. 459
- Intuition: Just as partial matching provided some generalization capability by allowing the retrieval of closely matching chunks, blending can produce continuously varying answers by retrieving not just one but an aggregate of a set of related facts.
- Qualities:
 - Produce continuous answers, including values not yet generated
 - Reflect the entire state of knowledge rather than individual fact
 - Express implicit subsymbolic knowledge such as similarities

Basic Mechanism

- The idea is to retrieve the best compromise value for all possible answers weighted by their probability of retrieval.

- Formally, that means the value V that satisfies:

$$V = \min_i \sum P_i \cdot \left(1 - \text{Sim}(V, V_i)\right)^2 \quad \text{Blending Equation}$$

$\text{Sim}(V, V_i)$: similarity between compromise value V and actual value V_i returned by chunk i

P_i : probability of retrieving i as a function of match score M_i :

$$P_i = \frac{e^{M_i/t}}{\sum_j e^{M_j/t}} \quad \text{Boltzmann equation}$$

Spectrum of Applicability

If values are:

- Chunks without similarities: as currently the most active chunk wins **except** chunks with identical values pool their strengths.
- Chunks with some similarities: some other chunk than the retrieved values **can** win if it is the best compromise. The set of chunks considered is those of the same type as the values.
- Chunks implementing a regular scale (e.g. integers): if the similarities between chunks are set according to a difference (ratio) scale, then the winning value is the one closest to the arithmetic (geometric) **mean** (easy proof from Blending Equation)
- Actual values such as integers or reals: since the number of possible values is infinite, the values are **averaged** together according to **blending-hook-fn** (default: arithmetic mean).

Application: Similarity Judgment

- Task: given stimuli (hypothetically here, sounds), rate them on a scale from 1 to 9, with 1 for low (quiet) and 9 for high (loud).
- Assumption: the similarities between chunks reflect underlying similarities between stimuli in the environment (linear scale).

```
(clear-all)
(sgp :bll 0.5 :ga 0.0 :pm t :mp 10.0 :ans 0.5 :rt -10.0
      :bln t :blt t :v nil) ;; turns on blending and trace
```

```
(chunk-type stimulus)
```

```
(chunk-type association stimulus rating)
```

```
(add-dm
```

```
;; define the stimuli with the proper similarities
```

```
(whisper isa stimulus)
```

```
(home isa stimulus)
```

```
...
```

```
(jackhammer isa stimulus)
```

```
(airplane isa stimulus)
```

```
;; encode the instructions linking stimuli to scale
```

```
(whisper-1 isa association stimulus whisper rating 1)
```

```
(airplane-9 isa association stimulus airplane rating 9))
```

```
;; Basic pattern-matching production
```

```
;; Given a stimulus, retrieves the best rating
```

```
(p stimulus-rating
```

```
=goal>
```

```
isa association
```

```
stimulus =stimulus
```

```
=fact>
```

```
isa association
```

```
stimulus =stimulus
```

```
rating =rating
```

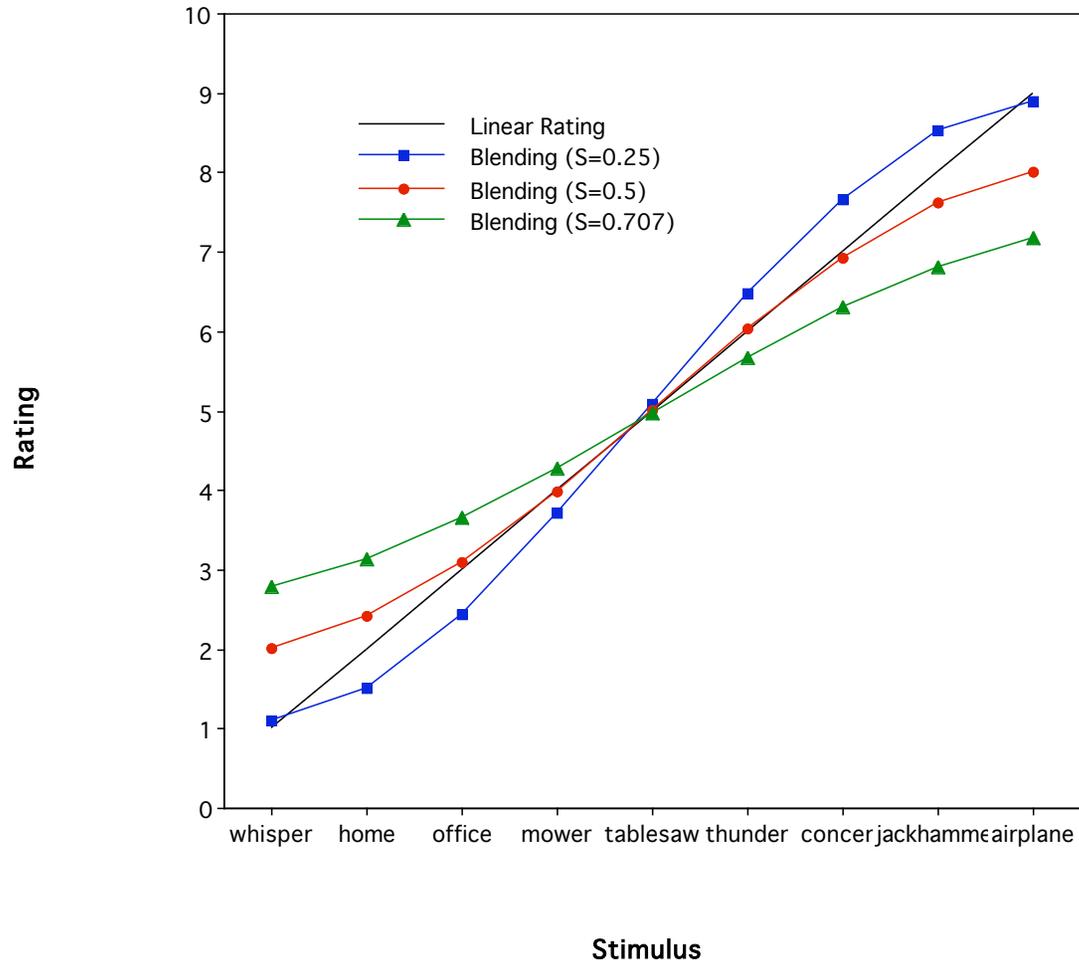
```
==>
```

```
!output! (=stimulus rates a =rating)
```

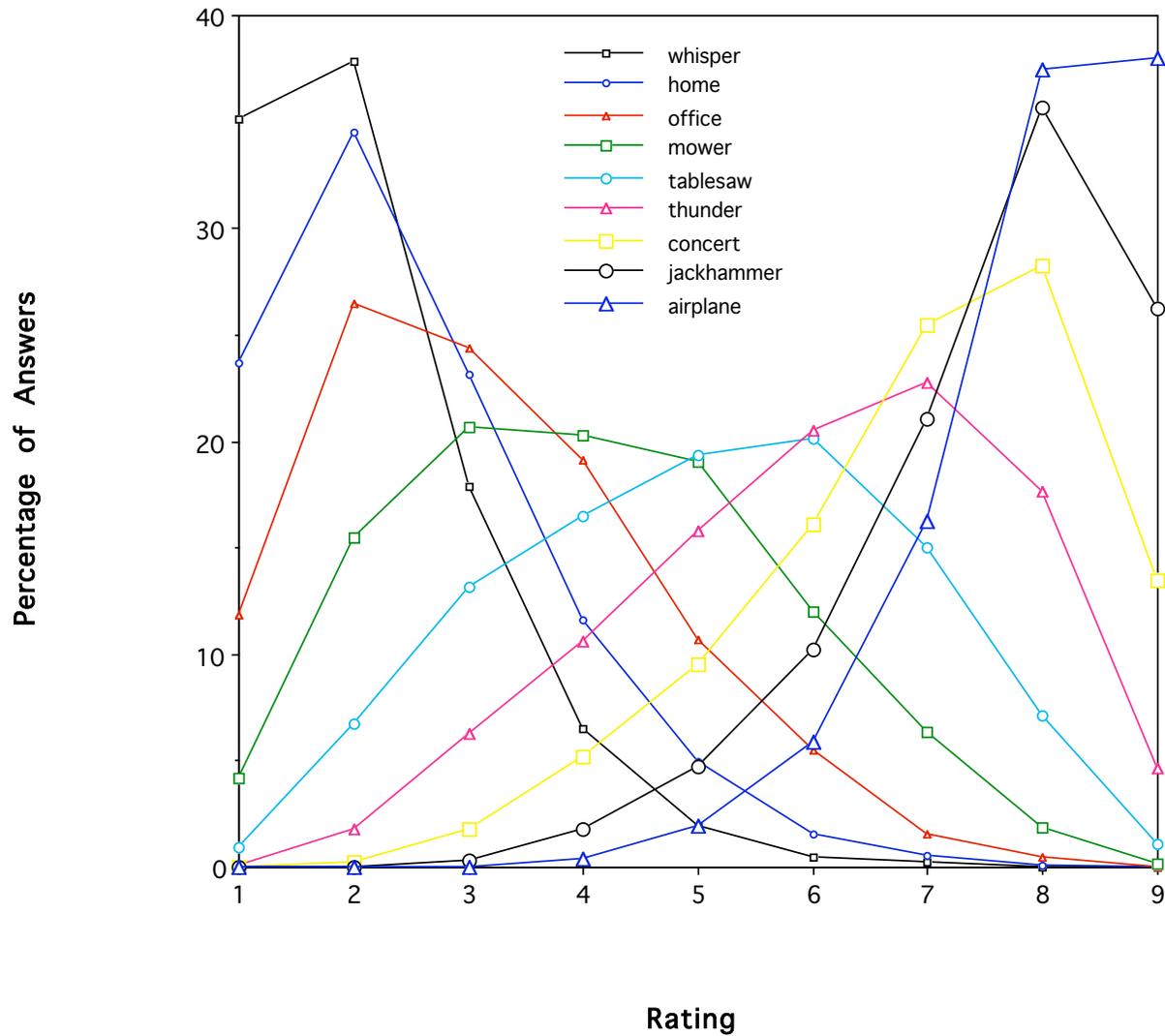
```
=goal>
```

```
rating =rating)
```

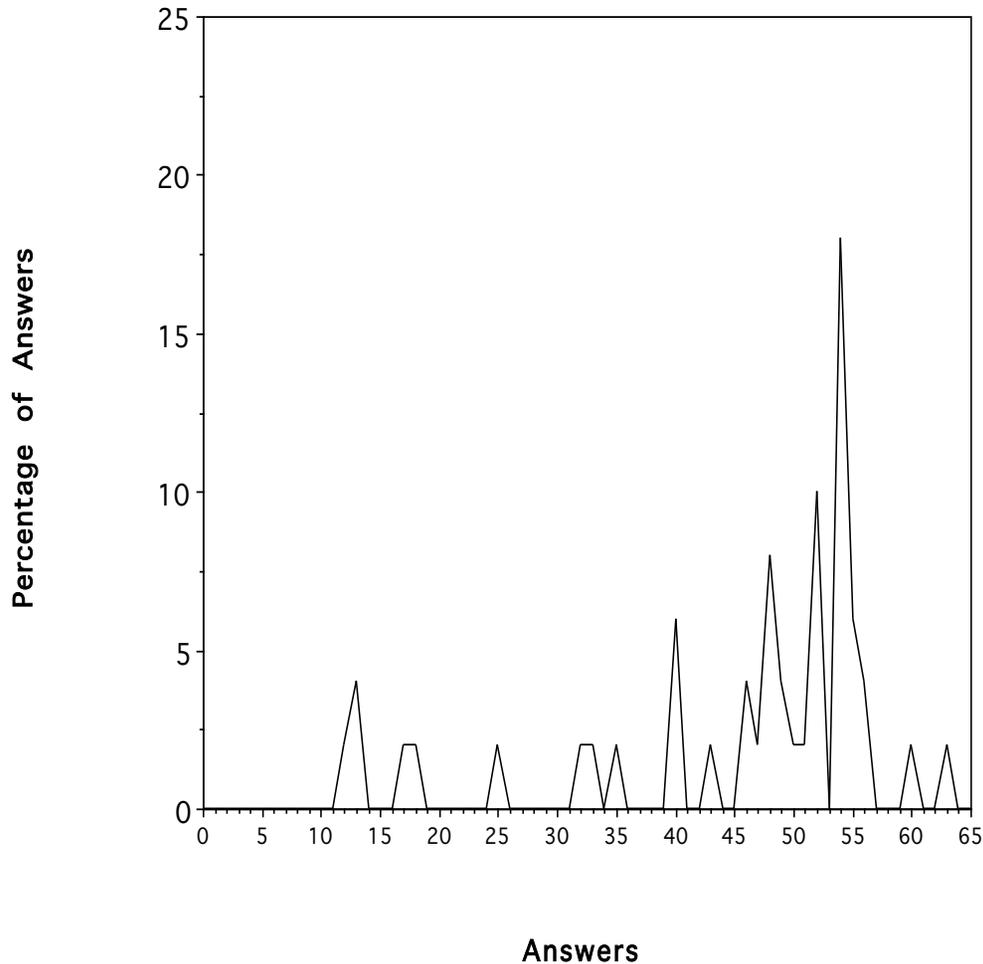
Average Ratings



Results: Ratings Distribution

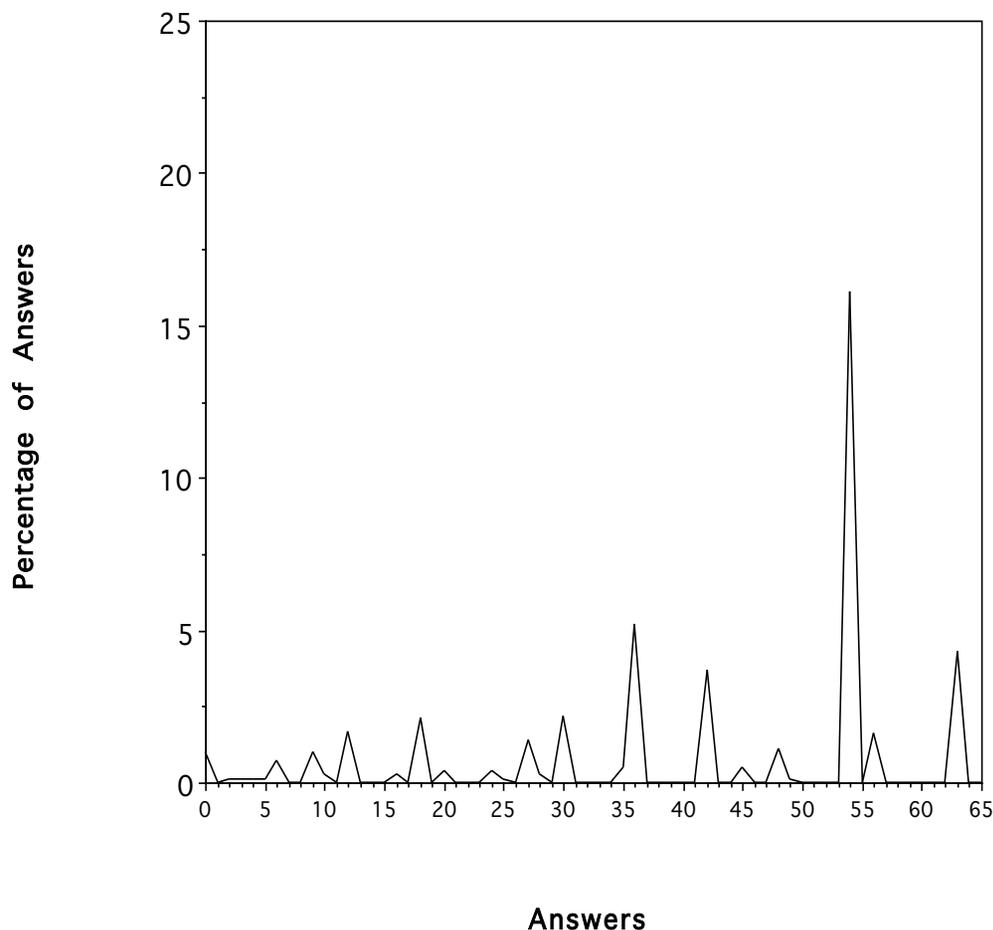


Application: Magnitude Estimate



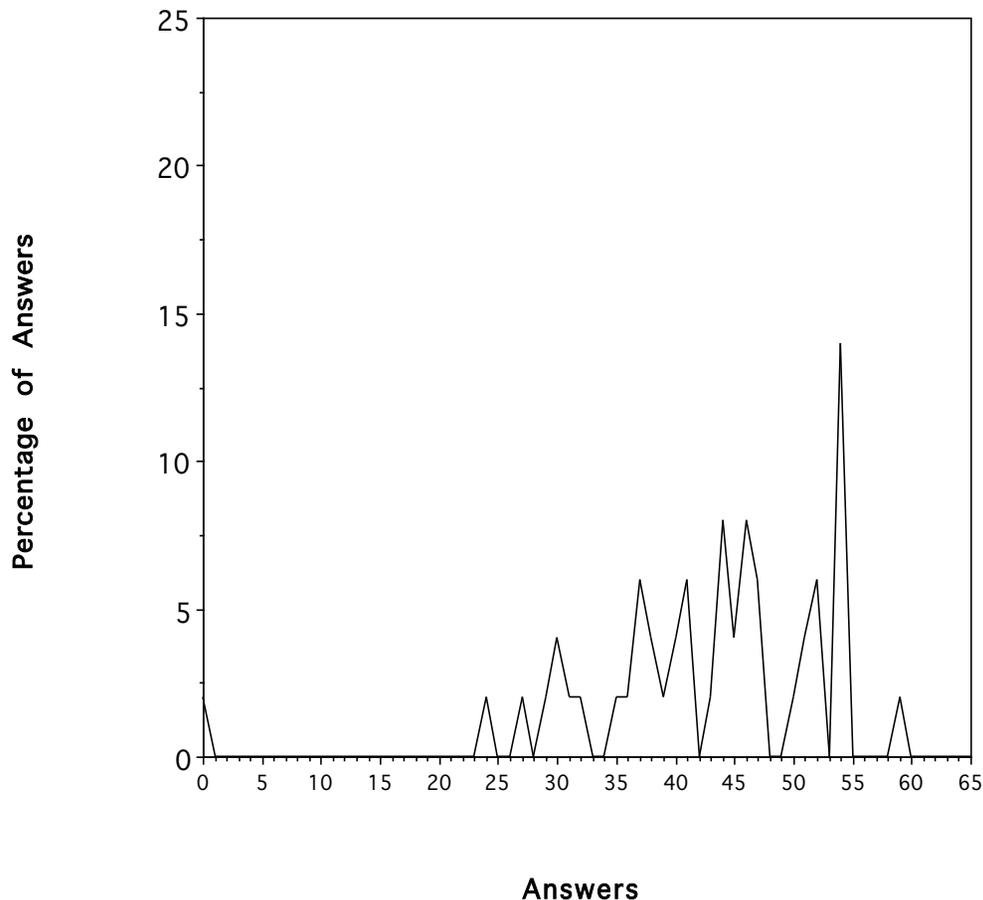
- Task: 4th-graders retrieving $6*9=54$ (Siegler, 1988).
- Many errors (>80%)
- Most errors are smaller than the correct answer
- Most errors are close to correct answer, with percentage decreasing with distance
- Most errors are not table errors, i.e. answers to related facts such as $6*7=42$

Lifetime Simulation: Retrieval



- Lifetime simulation of arithmetic (Lebiere, 1998)
- Many errors (>80%)
- Most errors are smaller than the correct answer
- Most errors are **not** close to correct answer
- Most errors **are** table errors, i.e. partial matching of related facts
- About 50% errors are retrieval failures
- Missing is the human ability to estimate answer

Magnitude estimate: Blending



- Blending: retrieve a mixture of facts when retrieval fails (**sgp :bln rt**)
- Many errors (>80%)
- Most errors are smaller than the correct answer: most (strong) answers are
- Most errors are close to correct answer: mismatch penalty keeps them close
- Most errors are not table errors because the blending process does not favor any specific table answers.

Details

- Blending switch (sgp **:bln** <value>) with <value> as:
 - NIL: default is off, i.e. the current system.
 - T: turned on all the time. Blending occurs for each retrieval.
 - RT: turned on conditionally: it applies when retrieval fails (guess).
- Blending trace (sgp **:blt** <value>) with <value> NIL or T.
 - Prints all chunks i retrieved with probability P_i and value V_i .
 - For all possible values V , prints the total error over all chunks
 - Prints the winning value and its resulting activation
- Blending hook function (setf *blending-hook-fn* '<fct>') used in averaging numerical non-chunk values with <fct>:
 - blending-arithmetic-mean: default, good for linear similarities
 - blending-geometric-mean: for ratio similarities ($\text{sim}(i,j) = i/j$)
 - whatever other function can be defined to reflect similarities

Issues

- How about learning of base levels and associative strengths?
 - Strengthen the single chunk with the best match score (bound to the variable) as before: might be correct in a probabilistic sense but...
 - Strengthen all chunks in proportion to their probability of retrieval or contribution to the answer: almost unworkably complex
 - Default: do not strengthen any chunk. Implication: subsymbolic declarative learning only happens when a chunk is popped and merged
- If multiple values are bound in the same retrieval:
 - Default: blending works on all values separately and in parallel
 - Alternative: satisfy all values at once in a single process (expensive)
- Is temperature used to compute probabilities the same as noise?
 - Perhaps noise reflects other external factors in modeling than actual one
 - Could get the best of both worlds with low noise and high temperature
 - To decouple, set temperature separately (**sgp :tmp <val>**). Default: nil.

Activation

What is the resulting activation for thresholding and latency?

A number of related alternatives are possible:

- Decrease each match score by the fit to V and sum over all:

$$M = \ln\left(\sum_i e^{M'_i}\right) \quad M'_i = M_i - (1 - \text{Sim}(V, V_i))$$

- Weigh each match score by the similarity to V and sum over all:

$$M = \ln\left(\sum_i e^{M_i} \cdot \text{Sim}(V, V_i)\right)$$

- Compute the probability of V and use the log-odds definition:

$$M = \ln\left(\frac{P}{1 - P}\right) \quad P = \sum_i P_i \cdot \text{Sim}(V, V_i)$$

Related AI Mechanisms

- Bayes optimal classifier
 - Standard Maximum A Posteriori: most likely hypothesis given data
 - Bayes optimal classifier: most likely outcome weighed over all hypotheses
 - Simplest case of blending but not the similarity-based generalization
- Locally weighted regression
 - Partial-matching as nearest-neighbor algorithm in similarity space
 - Weighted regression minimizes squared error between curve and data
 - Each data point weighted as a function (neg. exp.) of the distance from query to data (mismatch penalty) over the kernel size (temperature)
 - Special case for numerical values, not arbitrary similarity spaces
- Neural networks
 - Ability to generalize from distributed representation (similarities) and to reflect in each output value the consensus of the entire training set
 - The network runs quickly but training takes very many repetitions

Conclusions

- Blending seems to work well to provide continuous answers for similarity judgments and magnitude estimates.
- Also applied to:
 - a control task (Broadbent's Transportation) to provide interpolation-like capacities over learning instances. See Dieter Wallach's talk.
 - a real-time dynamic decision-making task (Pipes) to provide robust instance-based estimates of decision quality. See Cleo Gonzalez talk.
- Blending can be viewed as a generalization of
 - Partial-matching: it allows to retrieve not only one fact that is close to the perfect match but the collective results of a number of them.
 - Merging: a dynamic version of merging identical chunks (equivalent if $T=1$) that works as well for merely similar chunks as for identical ones.
- When is a mechanism part of the architecture?
 - Robust and general
 - Apply to a wide variety of tasks
 - Supported by empirical data