On the limits of spreading activation in ACT-R: Predictions and testability

Christopher Fisher[1*] (christopher.fisher@cubic.com), Brent D. Fegley[2*] (bfegley@aptima.com), Christopher A. Stevens,[3] & Christopher W. Myers[3]
[1]Cubic Defense, [2]Aptima, Inc., USA, [3]Air Force Research Laboratory, USA

Abstract

In the fan effect, reaction time (RT) increases as a function of fan size (i.e. the number of associations of a fact). Spreading activation in ACT-R provides a good account of the fan effect at low fan size (i.e., 1–4). However, little is known about the predictions of ACT-R at ecologically valid scales. We developed a general guessing mixture model (GMM) within ACT-R in which a guessing process is triggered by retrieval failures, and analyzed the predictions for fan sizes much larger than those used in laboratory experiments. Our analysis revealed the following properties of the GMM: RT increased as a function of fan size, but stays within a plausible range (< 2 seconds) as long as the retrieval threshold is not excessively low, and, in the limit, accuracy asymptotes at the value of the guessing bias parameter. We discuss practical challenges with testing the predictions at larger fan sizes.

Keywords: ACT-R; spreading activation; fan effect; simulation study; declarative memory; retrieval threshold

Introduction

One goal of cognitive architectures is to develop unified theories of cognition that scale to complex tasks in realistic environments (Newell 1990). Part of this larger goal is identifying memory processes and representations that support the retrieval of information from an extensive knowledge base. To achieve this goal, it is necessary to stress test existing theories and identify boundary conditions where the predictions may breakdown. Confidence in a theory will invariably increase if it survives rigorous stress testing. However, a failure provides an opportunity to revise the theory or develop alternatives. In either case, pushing the limits of a theory can provide important scientific insights and serve as a catalyst for scientific progress.

One theoretical question concerning the ACT-R cognitive architecture (Anderson 2007) is whether there are limits in the ability of spreading activation to account for the classic fan effect as the fan size increases. The fan effect is a phenomenon whereby retrieval time increases as the number of associations with a fact i.e. the fan size increases (Anderson 1974). For example, it takes longer to verify whether the hippie was located in the house if he or she was known to be in three places rather than one place. According to ACT-R, the fan effect arises through spreading activation in which a fixed quantity of activation, evenly distributed among associations in memory, spreads through a semantic network. As the fan size increases, the amount of activation distributed to each memory decreases, leading to slower retrieval time.

In a typical fan experiment, the fan size ranges from 1 to 4 (Anderson 1974; Sohn et al. 2004). ACT-R provides an accurate description of the fan effect within this limited range of fan size. Whether the fan effect increases with larger fan size and whether ACT-R continues to provide an accurate account remain open questions. From a theoretical standpoint, these questions are interesting because spreading activation may greatly inhibit the retrieval of requested information at large fan sizes, leading to low accuracy. Nonetheless, humans seem to retrieve information effectively even though the fan size in certain knowledge domains might be large, such as autobiographical memory. From a practical standpoint, this question is interesting for modeling human knowledge in applied domains. Given that human knowledge is extensive and associations among some facts may be high, what are the implications for predicting retrieval time and accuracy? Our goal is to analyze ACT-R's predictions at these boundary conditions.

Overview

The remainder of the paper is organized as follows. First, we describe how the fan effect is typically studied in a paired associates recognition memory task. Next, we present a general model of the fan effect and analyze several submodels including the model presented in the ACT-R tutorial (ACT-R Research Group n.d., Tutorial Unit 5). We compare the submodels at fan sizes much larger than have been examined in the laboratory. Finally, we detail some practical limitations in testing the predictions of spreading activation at scale. We conclude with a discussion of the theoretical and practical implications of our findings.

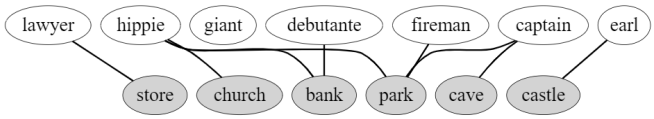

Figure 1: A bipartite graph of person-place pairs in a typical fan experiment (ACT-R Research Group n.d., Tutorial Unit 5). Nodes represent persons or places and edges represent associations between nodes. The number of edges connecting to a node represents the fan size.

Fan Effect

The fan effect is typically studied in a paired associates recognition task (Anderson 1974; Anderson and Reder 1999). During the learning phase, subjects study a series of word pairs that vary in fan size. For example, consider the network diagram in Figure 1. Each node represents either a person or a place, each edge represents connections between two nodes, and the number of edges connected to a node corresponds to its fan size. In Figure 1, *hippie* has a fan size of $f = 3$, whereas *earl* has a fan size of $f = 1$. During the test phase,

subjects must indicate "yes" if the word pair was studied or "no" if the word pair was not studied. Half of the test trials are *targets* in which the person and place were studied together as a pair, such as (earl, castle) in Figure 1. The remaining test trials are *foils* formed by switching person and place values across studied pairs such that the person and place in the new pair were not studied together. An example of a foil based on Figure 1 is (earl, cave).

Typically, fan size is manipulated factorially across a small range of values for the person and place attributes. Considering that we are interested in the predictions at large fan sizes, we will simplify the design by setting f equal for both attributes. At minimum, this design requires two sets of pairs, with f^2 pairs in each set for a total of $2 \cdot f^2$. Having two sets of pairs ensures that sufficient pairs exist for creating foils with equal f.

General Model

Our analysis of ACT-R is organized around a general model of the fan effect which we term the guessing mixture model (GMM). In the GMM, responses are determined by a mixture of a retrieval process and a guessing process. As described below, the fan model presented in ACT-R Tutorial 5 is a special case of the GMM. Figure 2 depicts the structure of the GMM as a processing tree in which each node represents a state and each branch represents a transition between states. Each path—defined as a series of branches—terminates in a "yes" or "no" response. The probability of traversing a path is the product of branch probabilities within the path. The marginal probability of a response is computed as the sum of all branch probabilities that map to the response. For example, the probability of responding "yes" on a target trial is $\Pr(\text{yes} \mid \text{target}) = t_m + (1 - t_m - t_{\text{mm}}) \cdot g$, which is composed of two paths: a path in which the matching chunk is retrieved and a path in which a retrieval failure occurs and the response "yes" is produced through a guessing process. Below, we will show how the transition probabilities in Figure 2 can be expressed in terms of ACT-R's memory retrieval mechanisms.

Knowledge Representation

In ACT-R, declarative memory consists of a set of chunks $M = \{\mathbf{c}_1, \mathbf{c}_2, \ldots \mathbf{c}_n\}$. A chunk is a basic unit of declarative knowledge. For a given fan size f, we assume that declarative memory consists of a minimum required $2 \cdot f^2$ chunks corresponding to each studied pair. Formally, a given chunk m is a collection of slot-value pairs denoted as $\mathbf{c}_m = \{(s_i, v_i)\}_{i \in I_m}$, where s_i and v_i are the slot and value of pair i, and I_m is the index set for the elements (slot-value pairs) of chunk m. An example of a chunk in a typical fan experiment is $\mathbf{c}_m = \{(\text{person}, \text{hippie}), (\text{place}, \text{park})\}$, which indicates the hippie is in the park. We will represent the mapping from slots to values as $c_m(s) = v$, which is empty or null if slot s is not in $\mathbf{c}_m$. Continuing with the example above, we can express the mapping between place and park as $c_m(\text{place}) = \text{park}$.

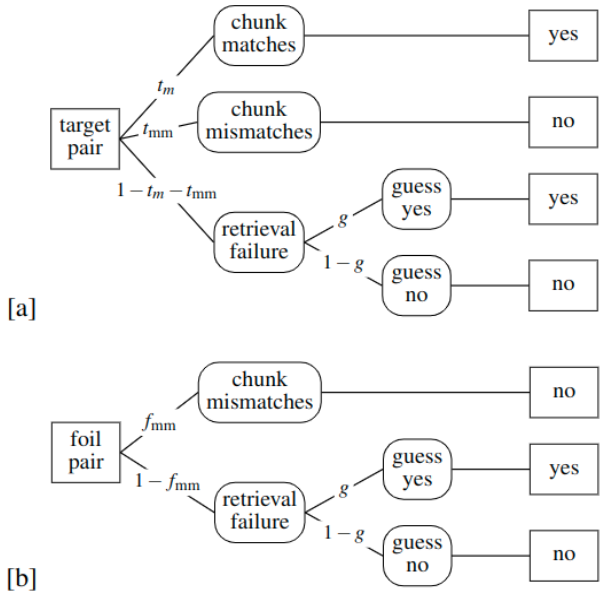

Figure 2: A tree diagram of the guessing mixture model. Panel [a]: process tree for target trials where t_m is the probability of retrieving matching chunk on target trial, t_{mm} is the probability of retrieving mismatching chunk on target trial, and g is the probability of guessing "yes". Panel [b]: process tree for foil trials where f_{mm} is the probability of retrieving mismatching chunk on a foil trial

Retrieval Process

Upon submitting a retrieval request $\mathbf{r}$ to declarative memory, a set of matching chunks R compete for retrieval and the chunk with the highest activation is retrieved if it exceeds the retrieval threshold, τ. A retrieval failure occurs if the highest activation is less than τ. The retrieval request is a mixture of retrieving from the person slot-value pair with probability w or the place slot-value pair with $1 - w$. Although in the tutorial $w = .5$, in our design, the value of w does not matter because f is equal for both attributes. On this basis, we can simplify the model by setting the mixture probability to $w = 1$ so that the person slot-value pair is always used as the retrieval request. We represent a retrieval request similarly to a chunk, which is defined as $\mathbf{r} = \{(\text{person}, v)\}$ where v is the value associated with the person slot. Upon submitting the retrieval request to declarative memory, a set R of candidate chunks compete for retrieval:

$$R = \{\mathbf{c}_m \in M : c_m(\text{person}) = r(\text{person})\}$$

The number of chunks matching the retrieval request is f. On a target trial, 1 chunk in R matches the stimulus on the person and place slot-value pairs. The remaining $f - 1$ chunks match only on the person slot-value pair. On foil trials, each of the f chunks match only on the person slot-value pair.

Activation

In ACT-R, each chunk has a memory activation value representing the log odds it will be encountered or needed (Anderson 2007). As activation increases, the probability and speed with which the chunk is retrieved also increases. Activation for chunk m is defined as

$$a_m = \beta + \mathrm{SA}_m + \varepsilon_m \tag{1}$$

where β is the base level constant, SA is the spreading activation term, $\varepsilon \sim \mathrm{normal}(0, \sigma)$ is activation noise, and σ is the standard deviation. We will use β to represent activation associated with relatively stable, asymptotic learning. Based on the assumptions we introduced, we can simplify the spreading activation term for the following two cases. The spreading activation term for chunk c which matches the stimulus on both the person and place slot value pairs is defined as

$$\mathrm{SA}_c = \frac{1}{2}[\gamma - \ln(f+1)] + \frac{1}{2}[\gamma - \ln(f+1)] = \gamma - \ln(f+1) \tag{2}$$

where γ is the maximum association parameter. The other case occurs when the chunk only matches on one slot-value pair of the retrieval request, which is given by:

$$\mathrm{SA}_i = \frac{1}{2}[\gamma - \ln(f+1)] \tag{3}$$

According to the ACT-R documentation, SA is truncated at zero by default, stating that undesirable behavior may occur with negative values (Bothell 2020, December 21, p. 287). Negative values occur when $f > e^{\gamma} - 1$. However, to be consistent with the theoretical interpretation of activation as log odds, which ranges between $-\infty$ and ∞, we do not impose any restrictions on SA. In ACT-R, retrieval time is the following inverse function of activation: $t_m = Fe^{-a_m}$ where F is the latency factor parameter with a default value of 1.

Response Mapping

As shown in Figure 2, the GGM uses the following response mapping: if the retrieved chunk matches the stimulus, the model responds "yes"; if the retrieved chunk does not match the stimulus, the model responds "no"; if a retrieval failure occurs, the model guesses "yes" with probability g.

Response Probabilities

Although the results we report below are based on Monte Carlo simulations of ACT-R, we will express the model in terms of approximate equations to provide a deeper understanding of the factors that determine the predictions. Using μ_c and μ_i as the expected activation for the cases based on Equations (2) and (3), the probability of correctly responding "yes" on a target trial can be approximated with the following softmax function (Weaver 2008):

$$\Pr(\text{yes} \mid \text{target}) = \frac{e^{\mu_c/\sigma}}{e^{\mu_c/\sigma} + (f-1)\cdot e^{\mu_i/\sigma} + e^{\tau/\sigma}} + \frac{e^{\tau/\sigma}}{e^{\mu_c/\sigma} + (f-1)\cdot e^{\mu_i/\sigma} + e^{\tau/\sigma}} \cdot g \tag{4}$$

where s is the logistic scalar parameter, $\sigma = s\sqrt{2}$, and τ is the retrieval threshold. Accuracy initially decreases as f increases because the the preponderance of chunks eligible for retrieval ($f-1$ out of f) do not match the target. However, in the limit, responding is driven entirely by guessing because activation becomes much lower than τ. We can see this behavior in Equation (4) where the term for retrieval failures, $e^{\tau/\sigma}$, has the largest exponent and thus determines the limit. As f increases, the first term on the right approaches zero whereas the second term approaches g. Setting Equation (4) to $h(f)$, we can state: $\lim_{f\to\infty} h(f) = g$.

On foil trials, the probability of a correctly responding "no" is given by:

$$\Pr(\text{no} \mid \text{foil}) = \frac{f \cdot e^{\mu_i/\sigma}}{f \cdot e^{\mu_i/\sigma} + e^{\tau/\sigma}} + \frac{e^{\tau_i/\sigma}}{f \cdot e^{\mu_i/\sigma} + e^{\tau/\sigma}} \cdot (1-g) \tag{5}$$

Setting Equation (5) to $z(f)$, the limiting behavior of the GMM is $\lim_{f\to\infty} z(f) = 1-g$ based on the same logic used for target trials.

Simulation Study

In this section, we analyze the predictions of two special cases of the GMM: the fan effect model from the tutorial, and an extension of the tutorial model with noise added to memory activation and the retrieval threshold. Table 1 lists the parameter values used for both submodels. For each combination of parameters, we repeated the simulation 5,000 times to ensure that stable predictions were generated.

Table 1: Parameter symbols, their descriptions, and values used in the submodels: tutorial model (TM), tutorial model plus noise (TM+N).

parameter	description	TM	TM+N
β	base level constant	0	1.5
τ	retrieval threshold	0	0, 2
γ	max associative strength	1.6	1–4
s	activation noise	0	0.2
F	latency factor	0.63	1
g	guess yes	0.5	0.5
t_{er}	non-retrieval time	.845*	.845*

*0.050 seconds is added for guessing.

Tutorial Model (TM)

The ACT-R Tutorial Unit 5 model for the fan effect (TM) is a special case of the GMM with parameter values specified in Table 1. One important characteristic of the TM is that retrieval time is deterministic because $s = 0$. In the tutorial, specifying a guessing process was unnecessary because the maximum fan size of 3 ensured that all $a_m > \tau$. However, a guessing process must be incorporated into the model

to handle retrieval failures at larger fan sizes, which incurs an additional overhead of .050 seconds. Incorporating the guessing process requires adding two production rules with noisy utility values selected to produce the desired guessing probability. We assume that guessing is unbiased (i.e., $g = .50$).

The TM predicts an instantaneous drop in accuracy from 100% to 50% when fan size forces activation below the retrieval threshold. Rounding to the next integer, this occurs at $f = 4$ with the specific parameters in Table 1. In general, the shift in accuracy occurs on target trials when

$$f > e^{\gamma+\beta-\tau} - 1 \tag{6}$$

and on foil trials when

$$f > e^{\gamma+2(\tau-\beta)} - 1 \tag{7}$$

RT predictions are also affected by an abrupt shift from retrieving chunks to retrieval failures. In Figure 3, RT increases with fan size on both target and foil trials until activation decreases below τ at $f = 4$. When $f \geq 4$, retrieval failures trigger a guessing process that produces the same constant RT for correct and incorrect responses regardless of increases in f. In general, the switch to guessing occurs when Equation (6) and Equation (7) are true, in which case the predicted RT becomes $Fe^{-\tau} + t_{er} + .05$ seconds regardless of increases in f. In addition, when $f \leq 4$, RTs for correct responses on foil trials are greater than the RTs for the corresponding responses on target trials. The reason is that only one source of spreading activation contributes to retrieved chunks on foil trials whereas two sources contribute to retrieved chunks on target trials.

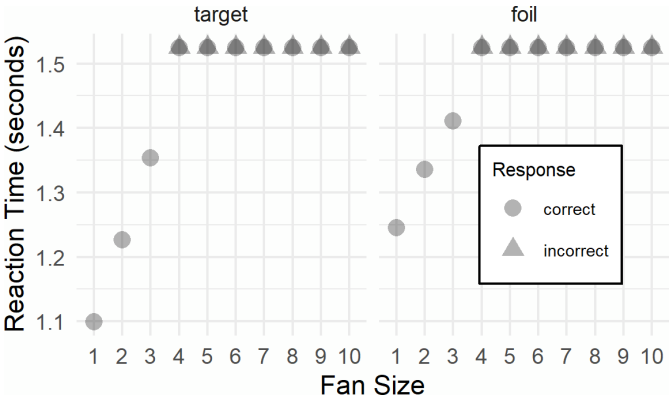

Figure 3: RT predictions of the TM model paneled by target and foil trials.

Tutorial Model Plus Noise (TM+N)

The TM suffers from the following limitations: (1) RTs are unrealistically deterministic, and (2) accuracy drops instantaneously from 100% to 50% once activation decreases below τ. In light of these limitations, we investigate a less restrictive special case of the GMM that we term the Tutorial Model Plus Noise (TM+N). The TM+N differs from the TM in one important way: noise is added to both memory activation and the retrieval threshold. Adding noise to activation and the retrieval threshold improves the model in two ways. First, the TM+N predicts a distribution of times for retrievals and retrieval failures rather than a deterministic time. Given that human RTs are variable, some have argued that adding noise to the retrieval threshold makes the model more plausible (Weaver 2008; Nicenboim and Vasishth 2018). Second, the TM+N predicts a gradual decrease in accuracy as a function of fan size rather than an immediate drop from 100% to 50%.

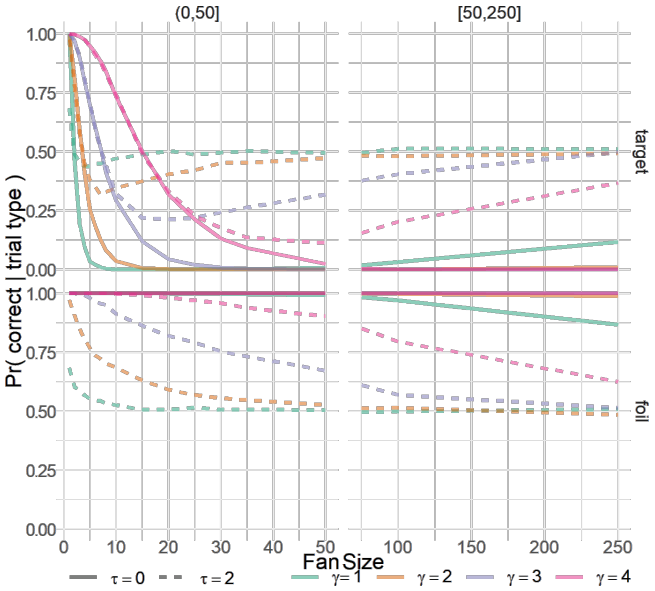

Figure 4: The probability of a correct response for trial types across fan sizes for TM+N as a function of τ and γ pair. Note that the x-scale from 1-50 is stretched to prevent overplotting.

Four noteworthy patterns for accuracy can seen in Figure 4. First, all other things being equal, accuracy is higher for larger values of γ. Second, on target trials, accuracy is a non-monotonic function of fan size, beginning above the asymptote at $g = 0.50$ and decreasing below $g = 0.50$ before increasing to $g = 0.50$. If activation is sufficiently larger than τ, accuracy will decrease to zero before converging on the asymptote. Third, on foil trials, accuracy starts high and decreases towards the asymptote at $1 - g = .50$ as fan size increases. Fourth, the speed with which the trends change increases with smaller differences between γ and τ. In other words, guessing dominates responding sooner when activation begins closer to the retrieval threshold.

Several important trends for RTs are present in Figure 5. First, RT increases as a function of fan size, but stays within a plausible range of approximately .90 to 1.6 seconds. Second, unlike the TM, the TM+N shows smooth curves for correct and incorrect RTs rather than an abrupt change from correct to incorrect RTs. Third, as expected, RTs were faster for higher values of γ. Fourth, RTs are faster when $\tau = 2$ than when

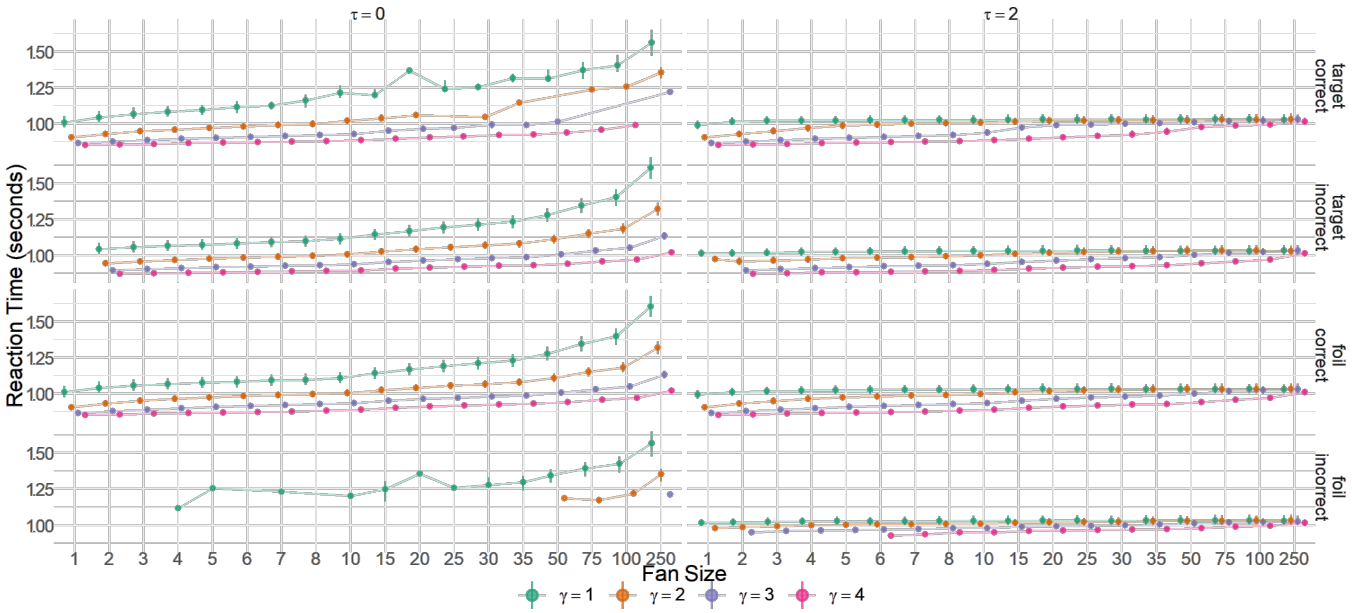

Figure 5: Median RT by fan size, trial type (target or foil), and response for TM+N as a function of τ and γ. Vertical lines are interquartile ranges. Horizontal lines depict the trajectory of median RT. Note that x-scale is nonlinear to prevent overplotting.

$\tau = 0$.

Practical Challenges

Testing the boundary conditions of spreading activation is fraught with several practical challenges. One challenge is that the time complexity for completing the study phase is quadratic (i.e., $O(f^2)$) because the fan experiment requires at minimum $2 \cdot f^2$ pairs. Consequently, the duration of the study phase will quickly become impracticable as fan size increases. For example, suppose subjects must complete b practice blocks to reach a target learning criterion. Suppose further that each pair will require t seconds on average to study. Thus, the learning phase will require $t_{\text{total}} = t \cdot b \cdot 2 \cdot f^2$ seconds to complete. Figure 6 shows the duration of the study phase as a function of f and b with $t = 2$ seconds. Depending on b, the duration of the study phase ranges between approximately .7 and 2.1 hours for $f = 25$, and quickly increases to a range of 11 to 33 hours for $f = 100$.

A second challenge is counteracting memory decay by increasing practice blocks. As shown in Figure 7, the time between consecutive presentations of the same word pair grows in a non-linear fashion with respect to fan size. With a fan size of 25, the time difference between presentations is nearly .70 hours which poses difficulties for learning. As fan size increases, so will the number of practice blocks required to counteract increasing amounts of memory decay between consecutive presentations of the same word pair.

The test phase, by contrast, offers more flexibility because it is not necessary to test the entire stimulus set. Instead, one could sample a random subset for testing. Although this would reduce the duration of the test phase and mitigate the effects of decay, it would come at the cost of lower statistical power.

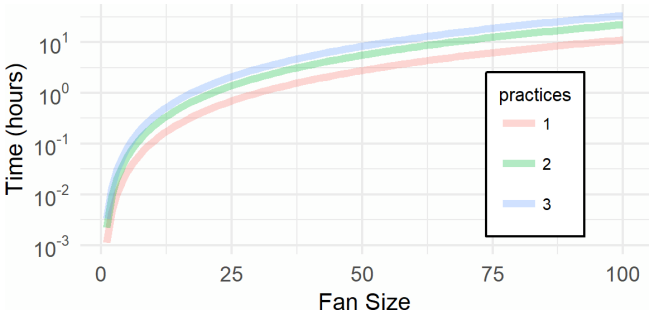

Figure 6: Study time as a function of fan size and the number of practice blocks.

Based on our analysis, it is clear that increasing fan size beyond 10–15 in a single experimental session would become prohibitively difficult. One way to increase fan size beyond 10–15 is to distribute practice across multiple sessions. Although using multiple practice sessions would make the time of a single session manageable, it suffers from inter-session decay effects and the potential for attrition. It is worth noting that researchers would likely vary fan size across several values to test the functional form of the fan effect, in which case the time demands would be even greater. For example, an experiment with fan sizes 2, 5, and 10 would require a one hour study phase assuming $t = 2$ seconds, $b = 3$, and an equal number of pairs per fan size: $\frac{3 \cdot 2f^2 \cdot t \cdot b}{60^2} = \frac{3 \cdot 2 \cdot 10^2 \cdot 2 \cdot 3}{60^2} = 1$ hour.

Discussion

Previous research has supported ACT-R's predictions for the fan effect within a small range of fan size. However, little is known about how ACT-R's predictions scale to ecologically

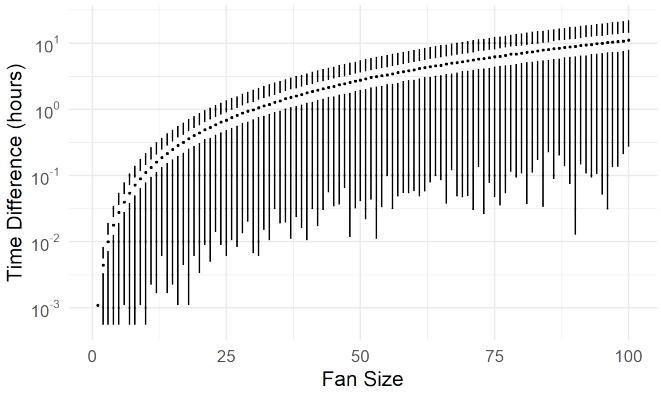

Figure 7: The difference in hours between two consecutive times a pair is studied. Medians are depicted as points, the interquartile ranges depicted as whitespace around the median, and minimum and maximum values at the outer extremes of the bounding lines.

valid domains in which fan size is likely large. In light of this gap in the literature, we set out to accomplish two goals: (1) to analyze the predictions of ACT-R at an ecologically valid range of fan size, and (2) to assess the practical challenges with studying the fan effect within ecologically valid parameters.

In service of the first goal, we analyzed the properties of a general guessing mixture model (GMM), with an emphasis on the two special cases: the tutorial model (TM) and the tutorial model plus noise (TM+N), an extension of the TM with noise in the retrieval process. Across a broad range of conditions, three findings emerged: (1) RT stayed within a plausible range (.9 - 1.6 seconds) despite low memory activation at high fan size, (2) RT decreased with increases in the maximum association parameter, and (3) accuracy eventually reaches an asymptote equal to the guessing parameter, g, on target trials, and $1 - g$ on foil trials. The TM suffered from two limitations due to its assumption that memory retrieval is deterministic: (1) unrealistically deterministic RTs, and (2) instantaneous switching from perfect accuracy to guessing. Adding noise to the retrieval process to produce the TM+N eliminated these limitations. Interestingly, the TM+N can produce non-monotonic behavior where accuracy drops below the asymptote—sometimes as low as 0% accuracy—before increasing to the asymptote.

Our analysis revealed that the time complexity for running a fan effect experiment is quadratic, meaning that the time demands quickly become prohibitive as fan size increases. This is exacerbated by the fact the time between consecutive presentations of the same study pair also grows quickly with fan size, leading to substantial decay. Additional study blocks would be needed to counteract memory decay during increasingly long study sessions, putting tests at large fan size even further out of reach.

One of the most interesting findings from our analysis is that accuracy drops to guessing levels somewhat quickly under a wide range of parameter settings. The primary factor in determining how quickly accuracy drops is the difference between activation and the retrieval threshold. As this difference decreases, accuracy decreases more quickly. Given that the predictions depend on this relationship, it is necessary to empirically test ACT-R at large fan sizes. As our analysis revealed, doing so will be challenging and there are practical limits to the maximum fan size that can be tested. Nonetheless, testing ACT-R at larger fan sizes—even if only as large as 10 or 20—will be important in assessing ACT-R's robustness, and determining ACT-R's scalability in practical situations with large knowledge domains.

Acknowledgments

The opinions expressed herein are solely those of the authors and do not necessarily represent the opinions of the United States Government, the U.S. Department of Defense, the U.S. Air Force, or any of their subsidiaries or employees. The contents have been reviewed and deemed Distribution A. Approved for public release. Case number: AFRL-2022-2020

References

ACT-R Research Group. (n.d.). *ACT-R » Software* (7.21.6-<3099:2020-12-21>). Retrieved December 31, 2021, from http://act-r.psy.cmu.edu/software/

Anderson, J. R. (1974). Retrieval of propositional information from long-term memory. *Cognitive Psychology*, *6*(4), 451–474. https://doi.org/10.1016/0010-0285(74)90021-8

Anderson, J. R. (2007). *How can the human mind occur in the physical universe?* Oxford University Press.

Anderson, J. R., & Reder, L. M. (1999). The fan effect: New results and new theories. *Journal of Experimental Psychology: General*, *128*(2), 186–197. https://doi.org/10.1037/0096-3445.128.2.186

Bothell, D. (2020, December 21). *ACT-R 7.21+ Reference Manual*. https://act-r.psy.cmu.edu/actr7.x/reference-manual.pdf

Newell, A. (1990). *Unified theories of cognition*. Harvard University Press.

Nicenboim, B., & Vasishth, S. (2018). Models of retrieval in sentence comprehension: A computational evaluation using bayesian hierarchical modeling. *Journal of Memory and Language*, *99*, 1–34. https://doi.org/10.1016/j.jml.2017.08.004

Sohn, M.-H., Anderson, J. R., Reder, L. M., & Goode, A. (2004). Differential fan effect and attentional focus. *Psychonomic Bulletin & Review*, *11*(4), 729–734.

Weaver, R. (2008). Parameters, predictions, and evidence in computational modeling: A statistical view informed by ACT-R. *Cognitive Science*, *32*(8), 1349–1375.