



A model-based analysis of the impairment of semantic memory

Holly A. Westfall¹ · Michael D. Lee¹

Accepted: 28 December 2020 / Published online: 8 April 2021
© The Psychonomic Society, Inc. 2021

Abstract

We use cognitive models to evaluate three theories of the change in semantic memory caused by Alzheimer's disease. We use data from 14,096 clinical assessments of 3602 Alzheimer's patients and their caregivers. Each patient completed a semantic memory task involving the odd-one-out comparison of animal names. Each patient was also independently evaluated to determine their level of impairment. Our cognitive models assume a feature-based representation of the animals and odd-one-out choice probabilities based on common-feature similarities. We find no evidence for the restructured representation hypothesis, which claims that impairment causes changes in the features used to represent stimuli. We also find no evidence for the attention change hypothesis, which claims that impairment causes greater attention to be given to concrete features at the expense of more abstract features. We do find evidence for the noisy access hypothesis, which claims that odd-one-out choices become less determined by semantic similarity and more prone to the simple response strategy of choosing the last option. We conclude that the noisy access hypothesis provides a simple account of odd-one-out choice behavior throughout the progression of Alzheimer's disease. More elaborate theories involving changes to underlying mental representations and attention processes need to provide evidence they are superior to the noisy access account.

Keywords Semantic memory · Alzheimer's disease · Odd-one-out task · Representation change · Attention change · Noisy access

Introduction

Alzheimer's disease is a neurodegenerative disease that, according to the World Health Organization (2019), is the leading cause of dementia, with about 10 million new cases worldwide each year. Memory decline is one of the first noticeable symptoms of Alzheimer's disease (McKhann, Knopman, Chertkow, Hyman, Jack, Kawas, & Phelps, 2011), and many cognitive measures used to diagnose and monitor this disease rely on tests of episodic memory, such as tests of free and cued recall (Buschke, 1984; Grober & Buschke, 1987) and autobiographical memory (Kopelman, Wilson, & Baddeley, 1989; Levine, Svoboda, Hay, Winocur, & Moscovitch, 2002). However,

other diagnostic cognitive measures rely on semantic memory, such as category fluency tasks (Newcombe, 1969), verbal fluency tasks (Benton & Hamsher, 1983), and picture naming tasks (Kavé, 2005). People diagnosed with Alzheimer's disease tend to do poorly on semantic memory tasks and have semantic memory deficits over and above those associated with normal cognitive aging (Nebes, 1989).

One way to study semantic memory is by testing people's odd-one-out choices in a triadic comparison task. This task is part of some well-established Alzheimer's testing batteries (e.g., Shankle, Mangrola, Chan, & Hara, 2009; Trenkle, Shankle, & Azen, 2007). One of the most common tests involves odd-one-out choices between animal names. In this task, on every trial, people are verbally presented with three animal names and must choose the one that is the least like the other two. For example, out of the names "giraffe", "elephant", and "cow", a person might choose "cow" as the odd one out. This task does not have correct answers but, from a person's odd-one-out choices, it is possible to make inferences about their semantic representation of the animals. If a person chooses "cow" as the odd one out, the implication is that they believe "giraffe" and "elephant" are more similar to each other than either is to "cow".

✉ Holly A. Westfall
hwestfal@uci.edu

Michael D. Lee
mdlee@uci.edu

¹ Department of Cognitive Sciences, University of California Irvine, Irvine, CA 92697-5100, USA

Previous work has examined changes in stimulus similarity between healthy controls and Alzheimer's patients using triadic comparisons of odors and colors (Razani, Chan, Nordin, & Murphy, 2010), line drawings of common objects (Au, Chan, & Chiu, 2003), and pictures of animals and tools (Chan, Salmon, & De La Pena, 2001). While it is clear that perceptions of similarities do change, the underlying reason for the change is debated. This debate can be summarized in terms of three competing hypotheses. The first, which we call the *restructured representation* hypothesis, is that the underlying semantic representation has changed with the onset of Alzheimer's disease. In an influential body of work, Chan and colleagues (e.g., Chan, Butters, Paulsen, Salmon, Swenson, & Maloney, 1993a, Chan, Butters, Salmon, & McGuire, 1993b, Chan, Butters, Salmon, Johnson, Paulsen, & Swenson, 1995) argue that patients with Alzheimer's disease have semantic networks where the associations between items differ from those of elderly healthy controls. One way to quantify people's latent semantic networks is by using scaling and clustering techniques, such as multidimensional scaling (MDS; Shepard, 1980). In an MDS analysis, the similarity within a set of items is represented as distance between points in k -dimensional space, so that items that are judged to be more similar are located closer together in the space. In elderly healthy controls, an MDS analysis showed a smaller distance between similar animals (e.g., cat and dog) and a larger distance between less similar animals (e.g., cat and sheep). In contrast, for Alzheimer's patients, the typical clustering of similar items became more diffuse, and MDS analyses revealed an increased distance between similar animals and decreased distance between less similar animals (Chan et al., 1993a).

A related hypothesis, which we call the *attention change* hypothesis, is that Alzheimer's patients use different information about the animals to make odd-one-out choices, when compared to healthy controls. In particular, Chan and colleagues argue that Alzheimer's patients have a tendency to focus on concrete over abstract information. For example, elderly healthy controls are more likely to make similarity judgments of animals based on domesticity as compared to predation or size. In contrast, Alzheimer's patients are more likely to make similarity judgments based on size (Chan et al., 1993b; 1995). The attention change hypothesis can be viewed as a special case of the restructured representation hypothesis. Both involve changes in the way people represent stimuli, but the attention change hypothesis assumes that the underlying representation remains the same, and it is only the selective attention to the components of that representation that are impacted by impairment.

A final hypothesis, which we call the *noisy access* hypothesis, is that differences in the triadic comparison task are not due to a change in the underlying semantic

representation. Instead, the atypical response pattern seen in Alzheimer's patients is due to an increasing loss of regularity in their choice behavior, rather than any fundamental representational change. The hypothesis is simply that Alzheimer's patients have trouble accessing the information their semantic representations provide (Nebes & Brady, 1990). Adopting this position, some researchers (Elvevåg & Storms, 2003; Storms, Dirikx, Saerens, Verstraeten, & De Deyn, 2003; Voorspoels et al., 2014; White, Voorspoels, Storms, & Verheyen, 2014) argue that the conclusions drawn about semantic reorganization from clustering and scaling techniques such as MDS may need to be tempered. Their argument is that differences in MDS output indicate only that the elderly healthy controls and Alzheimer's patient groups are different in *some* way, but not *how* they are different. It could be that the patient data are simply noisier than the control data, and this noise could have a number of causes, including problems in accessing representations, a failure to understand instructions, or differing response strategies.

Figure 1 provides a conceptual overview of how these three hypotheses relate to understanding odd-one-out choices. The example involves choosing between the animals "cow", "elephant", and "giraffe". The left side of the figure shows a cognitively healthy person choosing "cow" by relying on a feature-based representation. Each of the animals is represented in terms of whether or not they are fat, are a draught animal, have spots, have fur, come from Africa, or are commonly found in zoos. Different features receive different levels of attention, and the cognitively healthy person is shown as paying the most attention to the African and zoo features. The similarity between each pair of animals is assumed to correspond to how many features they have in common, with the features weighted by their level of attention. For the cognitively healthy person, elephant and giraffe are the most similar, and so cow is chosen as the odd one out.

The remainder of Fig. 1 shows how this choice can change according to the three hypotheses. Under the restructured representation hypothesis, the features involved in representing animals are different. In the example presented, the African and zoo features are no longer used. Based on the four features that remain, the cow and giraffe are the most similar, and the elephant is chosen as the odd one out. Under the attention change hypothesis, all of the features continue to be used, but the patterns of attention are different. In the example, the fat feature gains attention, making cow and elephant the most similar, and leading to giraffe being chosen as the odd one out. Finally, under the noisy access hypothesis, no aspect of the representation is changed. Instead, access to the information becomes less precise, and the odd-one-out choice becomes less based on similarities and closer to random responding.

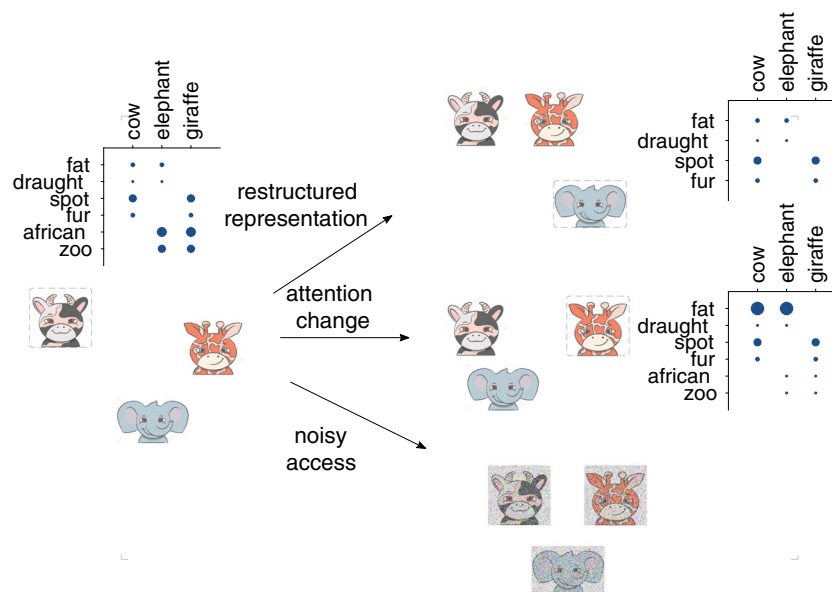


Fig. 1 Overview of three hypotheses for changes in odd-one-out choices. The left side shows the representation of “cow”, “elephant”, and “giraffe” in terms of a set of six features, where the area of circles corresponds to the weight given to each feature. The cow is chosen as the odd one out, as indicated by the encompassing *dashed line*. The right side of the figure shows how this choice changes according to the

three hypotheses. At the *top*, the representational restructuring hypothesis involves only a subset of the features, and the elephant is chosen. In the *middle*, the attention change hypothesis involves different feature weights, and the giraffe is chosen. At the *bottom*, the noisy access hypothesis involves degraded access to the representation and more random choice

The goal of this article is to evaluate these three hypotheses as accounts of the change in odd-one-out performance caused by the progression of Alzheimer’s disease. In the next section, we detail clinical data that assess patients with different levels of impairment. We then introduce a basic cognitive model of odd-one-out choice behavior based on simple assumptions about how stimuli are represented and decisions are made. The model is then extended to capture the specific assumptions of the restructured representation, attention change, and noisy access hypotheses, and each extension is evaluated against the clinical data. We find no evidence for restructured representation and no evidence for the sorts of changes in attention that have previously been proposed. Instead, we find that the noisy access model provides a good account of the changes in triadic choice behavior. We conclude by discussing the implications and limitations of these findings.

Behavioral data

Our behavioral data come from the animal triadic comparison task of the Mild Cognitive Impairment Screen (MCIS; Shankle et al., 2009) at a cognitive disorders clinic, as part of routine cognitive assessment of patients and their caregivers. The task draws from a pool of 21 animals: antelope, beaver, camel, cat, chimpanzee, chipmunk, cow, deer, dog, elephant, giraffe, goat, gorilla, horse, lion, monkey, rabbit,

rat, sheep, tiger, and zebra.¹ For each specific assessment, nine animals are chosen from the pool, and each is presented in a triad with every other animal over a total of twelve trials. The triads are presented in accordance with a λ -2 balanced incomplete block design (Burton & Nerlove, 1976). On each trial, patients are verbally presented with the animal names and have to respond by choosing the animal that is least similar to the other two.

The data include 14,096 assessments of 3602 patients and their caregivers, (52% female, age range 16–103 years, mean age 76 years). At the time of assessment, patients were classified by a physician with the Functional Assessment Staging Test (FAST; Reisberg, 1988). The FAST stages describe the severity of Alzheimer’s disease symptoms in terms of people’s ability to perform daily living tasks, such as managing finances, cooking, and grooming. Table 1 provides a summary of the number of patients and total number of assessments by FAST stage, and a description of each stage. Patients in stage 1 have no discernible deficits, and those in stage 2 have only a subjective functional deficit. These stages are considered cognitively normal. Patients in stage 3 have symptoms of mild cognitive impairment, while patients in stages 4, 5, and 6 have mild, moderate, and moderately severe dementia, respectively. We did not

¹ Animal names are chosen based on work in cultural anthropology that provides evidence for the universality of the semantic structure of the domain (Romney & Moore, 1998).

Table 1 A description of the number of assessments, patients, and identifying characteristics for each FAST stage in the data set

FAST stage	Description	Patients	Assessments
1 & 2	no deficit or subjective deficit	576	2032
3	objective deficit in complex tasks	908	3685
4	mild dementia evident in daily living tasks	1169	5352
5	moderate dementia	430	1313
6	moderately severe dementia	519	1714

include patients in FAST stage 7, who have severe dementia, because their cognitive function has degenerated to the point of an inability to understand simple instructions (Reisberg, 1988). The FAST stage assessments of impairment were made independent of memory test performance and so provide an external measure for grouping patients in order to study changes in the odd-one-out task behavior.

Some basic analyses of the data make clear that odd-one-out choices change as impairment progresses. As an example of change within a specific triad, consider the earlier cow, elephant, and giraffe example. For this triad, cow was chosen 53% of the time by patients in stages 1 and 2. Patients in stage 3, however, chose giraffe more often than cow, choosing giraffe 42% of the time and cow 37% of the time. Patients in stage 5 showed yet another pattern, choosing elephant as often as cow, with both accounting for 42% of all choices. The overall probability of different animals being chosen also often changes. For example, zebra was chosen as the odd one out 22% of the time by patients in stages 1 and 2, but 32% of the time by patients in stage 6, whereas rat was chosen 53% of the time by patients in stages 1 and 2, but only 43% of the time by patients in stage 6.

As a more thorough analysis of change, we examined the changes in similarities between pairs of animals across the FAST stages. Specifically, we compared the rates with which neither of a pair of animals was chosen, when both were presented. The Bayes factor (Lee & Pope, 2006) favored a change in this rate for 5% of all pairs moving from stages 1 and 2 to stage 3, in 26% of pairs moving from stage 3 to 4, in 18% of pairs from stage 4 to 5, and in 7% of pairs from stage 5 to 6. These differences make it clear that there is widespread change in odd-one-out choices across the FAST stages.

Modeling analysis

A basic model of odd-one-out comparison

We begin the modeling analysis by developing a basic model of choice behavior in the odd-one-out tasks, which serves as the foundation for evaluating the competing hypotheses. The model has two core components: one for

representing the stimuli and their similarities, and one for the decision-making processes that act on the similarities to produce choice probabilities.

Common-features similarity

Previous research modeling the change in semantic memory with impairment (Chan et al., 1993a), including modeling focused specifically on animal odd-one-out comparisons (Lee, Abramyan, & Shankle, 2016), has relied on spatial representations like MDS. These representations assume that stimuli can be represented in terms of values on a small number of underlying psychological dimensions. An alternative to the dimension-based representational assumptions of MDS is to assume stimuli are represented in terms of features (Goldstone, 1999; Shepard, 1980). We think this is a more natural assumption for the representation of conceptual stimuli like animal names, and it aligns better with the set of hypotheses we aim to evaluate. In particular, the representational restructuring hypothesis assumes that components of the representation are added or deleted with impairment. This hypothesis seems far more plausibly expressed in terms of the gain or loss of a few features, which typically apply to only a few stimuli each, rather than the gain or loss of an entire dimension, which always apply to every stimulus.

To represent the animal stimuli in terms of features, we assume that the similarity, s_{ab} , of animal a and animal b is equal to the sum of the weights of their shared features (Shepard & Arable, 1979; Tversky, 1977),

$$s_{ab} = \sum_k w_k f_{ak} f_{bk}, \quad (1)$$

where f_{ak} is a binary indicator variable that determines whether or not animal a has feature k , and w_k represents the salience or weight of feature k . We used a truncated Gaussian² for the priors on the feature weights,

$$w_k \sim \text{Gaussian}(0, 1) \text{ T}(0,). \quad (2)$$

²We parameterize the Gaussian distribution in terms of the mean and precision, consistent with the JAGS software we use to implement the models. The precision is the reciprocal of the variance.

This common-features model makes animals similar only to the extent that they share features. Animals do not become more similar by both not having a feature. There is good evidence this is usually a reasonable assumption (Navarro & Lee, 2004), and the common-features model is the basis of widely used additive clustering and related methods in similarity data analysis (Shepard & Arabie, 1979; Navarro & Griffiths, 2008; Peterson, Abbott, & Griffiths, 2018).

Rather than using additive clustering methods to infer the features, we used the Leuven concept database as a set of possible features (De Deyne et al., 2008). This database includes a total 288 features for the domain of animals, and lists whether a set of animals has each feature based on four independent raters. We adapted the database for our modeling goals by including animals in the MCIS not originally listed, and eliminating features that had identical patterns of presence or absence across the 21 animals. These changes resulted in a final set of 118 possible features.

Luce choice rule

The probability, π_a of choosing animal a as the odd one out in a triad of animals a , b , and c is determined by their relative similarities. Intuitively, it is more likely animal a will be chosen if animals b and c are the most similar to each other. This intuition can be formalized by the Luce (1959) choice rule, which defines the choice probabilities as

$$\begin{aligned}\pi_a &= \frac{s_{bc}}{s_{ab} + s_{ac} + s_{bc}} \\ \pi_b &= \frac{s_{ac}}{s_{ab} + s_{ac} + s_{bc}} \\ \pi_c &= \frac{s_{ab}}{s_{ab} + s_{ac} + s_{bc}}.\end{aligned}\quad (3)$$

Given these probabilities, the observed choice of participant i on trial t , which we denote y_{it} , is modeled as

$$y_{it} = \text{categorical}(\pi_a, \pi_b, \pi_c). \quad (4)$$

This basic model provides an account of how the features used to represent stimuli, and the attention weights for those features, combine to produce similarities, as well as an account of the decision-making processes by which the similarities produce choice probabilities.

Restructured representation analysis

The restructured representation hypothesis assumes that the features to which people attend may change as memory impairment progresses. To create a model consistent with this hypothesis, we extend the basic model in several ways. First, we assume that people in any FAST stage use only a subset of the available features. Following Zeigenfuse and

Lee (2010), this assumption is implemented by introducing a latent binary indicator parameter z_k^t that determines whether feature k is considered in the similarity judgments made by people in stage t , so that

$$s_{ab}^t = \sum_k z_k^t w_k f_{ak} f_{bk}. \quad (5)$$

The feature-inclusion parameters are given the prior $z_k^t \sim \text{Bernoulli}(\phi_t)$ with a base-rate for each stage $\phi_t \sim \text{beta}(1, 5)$.

Changes in the features z_k across FAST stages would provide evidence in favor of the restructured representation hypothesis. To measure this evidence, we introduce a change process by which the features used to represent the animals can change across FAST stages. This is formalized by binary parameters τ_k^t that indicate whether the inclusion of feature k is the same or different between stage t and stage $t + 1$, such that

$$z_k^{t+1} = \begin{cases} z_k^t & \text{if } \tau_k^t = 0 \\ 1 - z_k^t & \text{if } \tau_k^t = 1. \end{cases} \quad (6)$$

The feature-change parameters are given a prior $\tau_k^t \sim \text{Bernoulli}(\psi^t)$ with a base-rate $\psi^t \sim \text{uniform}(0, 1)$ for the transition from stage t to stage $t + 1$.

We implemented the restructured representation model, and all of the models considered in this article, as graphical models in JAGS (Plummer, 2003). JAGS provides a high-level scripting language for implementing probabilistic cognitive models that allows for computational Bayesian analysis using Markov-chain Monte Carlo sampling methods.

Results

Figure 2 shows the marginal posterior expectations of the feature-inclusion parameters z_k^t . Most features have posterior probabilities of inclusion close to 0 or 1. Only for a very few features is the inference uncertain. This provides good evidence that participants use about 45 of the 118 features in all of the stages. What remains to be determined is whether these 45 features are the same across the stages.

Figure 3 shows, in blue, the posterior distributions of the base-rates ψ for each of the transitions between successive FAST stages. One way to understand this result is in model-selection terms, comparing a “null” model that assumes a base-rate of change less than 1% against an alternative model that allows for greater change. Comparing the prior and posterior densities in the interval $0 < \phi_k < 0.01$ allows the Bayes factors between these two models to be estimated using the Savage–Dickey method (Wetzels, Grasman, & Wagenmakers, 2010). These Bayes factors are

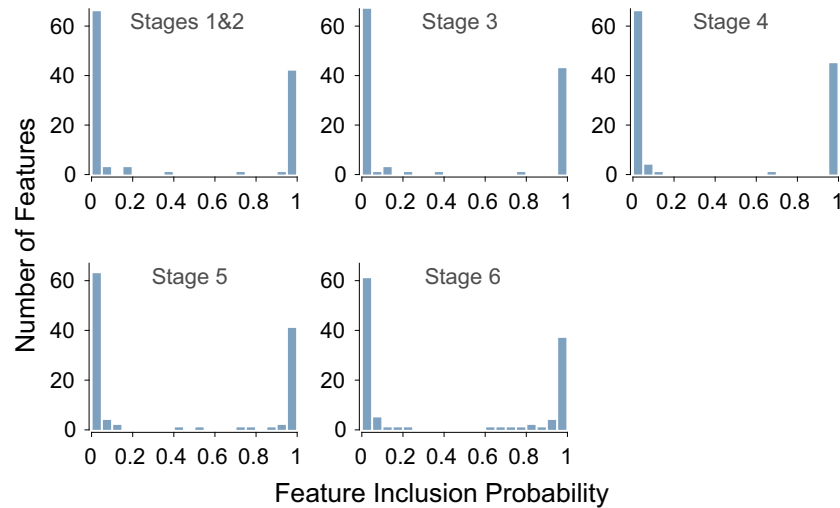


Fig. 2 The distributions of marginal posterior expectations of z_k^i for all features, within each FAST stage

about 55, 6, 28, and 23 for the four transitions, all favoring the null model that assumes there is negligible change. The other interpretation of the results in Fig. 3 is in parameter-estimation terms, treating the posterior distributions as measuring the extent of change that is assumed to exist. From this perspective, it is clear that the probability of a feature changing—either being added to a representation, or dropped from a representation—across FAST stages is very small, and almost certainly below 5% in every case.

From either perspective, the modeling results provide clear evidence against the restructured representation hypothesis. While participants use a subset of the possible features to make odd-one-out decisions, they use very close to the same subset in all of the FAST stages.

Attention change analysis

The attention change hypothesis assumes that the features used to make odd-one-out choices do not change with impairment, but the weights given to the features do change. The best theoretically developed version of this hypothesis is provided by Chan and colleagues (1993b, 1995), who argue that one type of features, called physical features, are given more attention as impairment progresses, while other types of features, called thematic and abstract, are given less attention. Physical features are those related to the animal’s appearance, such as “is fat”, “is specked”, and “has horns”. Abstract features are those related to the animal’s behavior, such as “can swim”, “eats nuts”, and

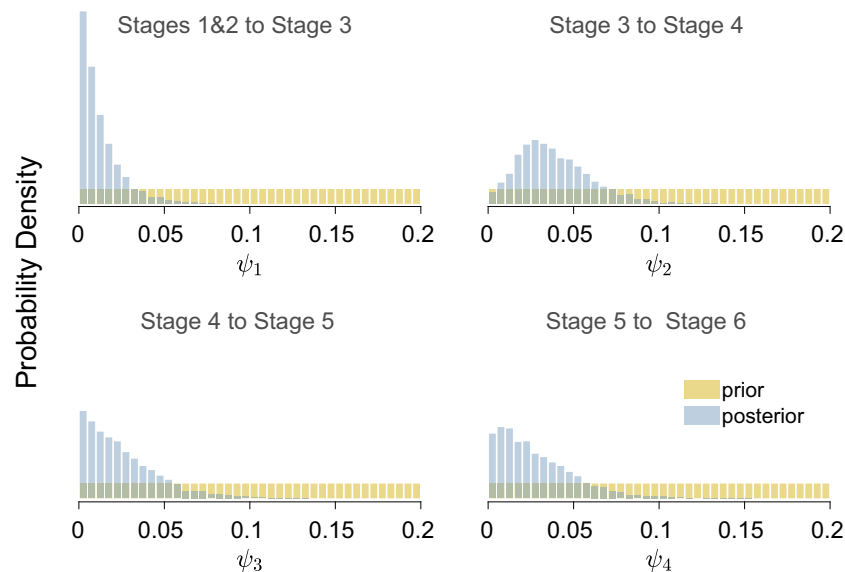


Fig. 3 The prior (yellow) and posterior (blue) distributions for the feature transition base-rate ψ for all transitions between successive FAST stages

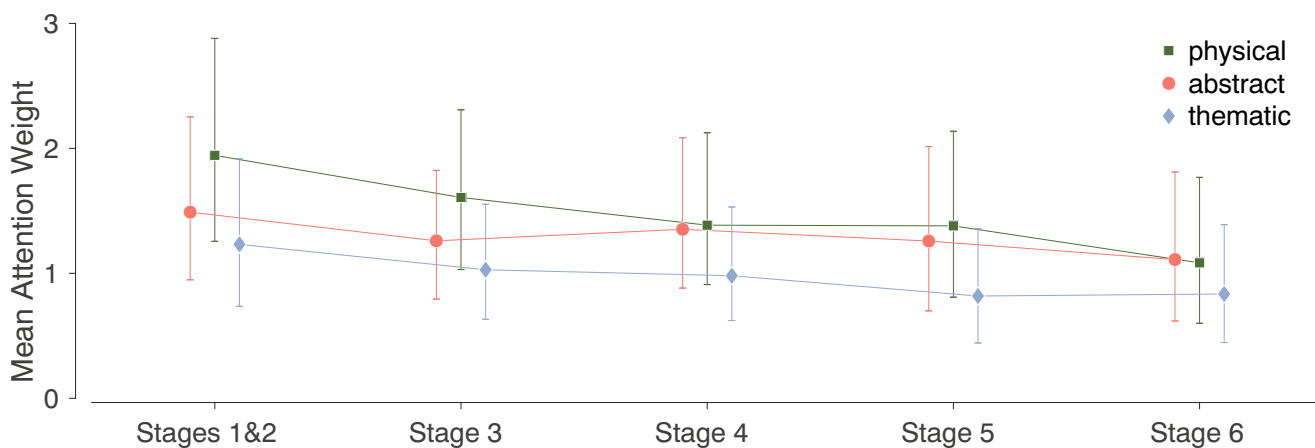


Fig. 4 The posterior mean and 95% credible intervals for the average attention weight for the physical, abstract, and thematic feature types across FAST stages

“crawls up trees”. Thematic features are those describing the animal’s role, such as “is a pet”, “is popular among children”, and “is a cartoon figure”. We classified all of our 118 possible features into these three types, using two independent judges to determine the classifications, and a third judge to resolve disagreements.

To create a model consistent with this specific attention change hypothesis, we extend the basic model to include an account of how the weights of the different feature types change across FAST stage. This is accomplished by a hierarchical extension that assumes the weight of each feature is sampled from an over-arching Gaussian distribution that depends on both its type and the stage. Specifically, the feature weight w_k^t in FAST stage t , for a feature k is given by

$$w_k^t \sim \text{Gaussian} \left(m_{\kappa(k)}t + c_{\kappa(k)}, 1/\sigma_{\kappa(k)}^2 \right) \text{T}(0, \cdot). \quad (7)$$

The function $\kappa(\cdot)$ assigns each feature to its appropriate type, so that $\kappa(k) = 1$ means feature k is a physical feature, $\kappa(k) = 2$ means it is an abstract feature, and $\kappa(k) = 3$ means it is a thematic feature. The parameters $m_1, m_2, m_3 \sim \text{Gaussian}(0, 1)$ are slopes for the physical, abstract, and thematic feature types respectively, measuring how the average attention to features of that type increases or decreases with changes in the FAST stage. Similarly, the parameters $c_1, c_2, c_3 \sim \text{Gaussian}(1, 1) \text{T}(0, \cdot)$ are intercepts that measure the absolute level of attention. Finally, the parameters $\sigma_1, \sigma_2, \sigma_3 \sim \text{Gaussian}(1, 1/2^2) \text{T}(0, \cdot)$ are standard deviations that measure the heterogeneity in the attention weights between different features of the same type in the same FAST stage.

These hierarchical assumptions allow every feature to have its own attention weight in every FAST stage, but provide a measure of the average attention given to the physical, abstract, and thematic feature types. Critically, the model also provides a measure, via the slope parameters, of

the change in the mean feature weights for each type as the FAST stages progress. The attention change hypothesis can be expressed concisely in terms of these slopes: attention to physical features increases, so $m_1 > 0$, while attention to abstract and thematic features decreases, so $m_2, m_3 < 0$.

Results

Figure 4 presents the average attention weights for each feature type by FAST stage. The error bars represent 95% credible intervals. The average attention weight for each of the physical, abstract, and thematic features all decrease with impairment. This is inconsistent with the pattern of change predicted by the attention change hypothesis. To quantify this result, we calculated the Bayes factor comparing the specific inequality-constrained attention change hypothesis $m_1 > 0, m_2, m_3 < 0$ against the alternative hypothesis without any constraints. The Bayes factor is about 35 in favor of the alternative hypothesis. Collectively, these results provide strong evidence against the attention change hypothesis. While there is a change in the average attention given to different types of features, there is not a systematic increase in the attention given to more concrete features at the expense of more abstract ones.

Noisy access analysis

The noisy access hypothesis assumes that all participants attend to the same features with the same attention weights, but that the ease of access decreases with the progression of impairment. To create a model consistent with this hypothesis, we focus on the decision-making assumptions. In particular, we extend the basic model to allow for response determinism and a recency bias in the choice rule.

Response determinism measures how closely choice behavior is determined by the underlying similarities, and

can be modeled by extending the Luce choice rule through exponentiation (Lee et al., 2016). Pairwise similarities are now raised to a power, in the form s_{ab}^γ , and different values of the parameter γ determine how consistently people make odd-one-out choices. If $\gamma = 1$, the original Luce choice rule is maintained. If γ decreases towards zero, the probability of choosing animals a , b , and c are all reduced towards $\frac{1}{3}$, and choice behavior becomes more random. As γ increases to values greater than one, responses become more deterministic, and the animal that is least similar to the other two will be more consistently chosen.

The second decision-making extension is to include a recency bias, corresponding to the strategy of choosing the last animal name presented in the sequential verbal presentation. An increase in the recency bias could be explained by a deficit to working memory typically seen in Alzheimer’s patients (see Huntley & Howard, 2010, for a review). To allow for this possibility, we assume simply that that last animal is favored by a bias measured by the parameter β . Combining the response determinism and recency bias extensions gives a set of choice probabilities defined as

$$\begin{aligned} \pi_a &= \frac{1 - \beta}{2} \left(\frac{s_{bc}^\gamma}{s_{ac}^\gamma + s_{bc}^\gamma + s_{ab}^\gamma} \right) \\ \pi_b &= \frac{1 - \beta}{2} \left(\frac{s_{ac}^\gamma}{s_{ac}^\gamma + s_{bc}^\gamma + s_{ab}^\gamma} \right) \\ \pi_c &= \beta \left(\frac{s_{ab}^\gamma}{s_{ac}^\gamma + s_{bc}^\gamma + s_{ab}^\gamma} \right). \end{aligned} \tag{8}$$

We assume the prior $\gamma \sim \text{gamma}(2, 1)$ for response determinism. This distribution has a mode at 1, corresponding to the original Luce choice rule and probability matching, allows for greater values corresponding to deterministic choice behavior, and allows for lesser values corresponding to more random choice behavior. We assume a prior $\beta \sim \text{uniform}(1/3, 1)$ for recency bias. This choice of prior

allows for the possibility of unbiased choices corresponding to $\beta = \frac{1}{3}$, but captures our assumption that any bias will be in favor of the last item.

Results

Figure 5 presents the posterior distributions for the response bias and response determinism parameters by FAST stage. It is clear that response determinism decreases as impairment progresses across the stages. The Bayes factors testing whether γ is the same or different across successive stages are all greater than 100 in favor of there being a difference. It is also clear that the recency bias generally increases across the FAST stages. The Bayes factor is only 2 in favor of a difference between stages 1 and 2 compared to stage 3, but is greater than 25 for all of the other comparisons. Interestingly, there is a significant recency bias even for cognitively healthy participants in stages 1 and 2, since the Bayes factor is more than 100 in favor of β being greater than $\frac{1}{3}$.

These results are very consistent with the noisy access hypothesis. The natural interpretation is that changes in choice behavior can be explained by progressively impaired use of the underlying similarities between animals. The choices made are less well determined by the similarities as impairment increases, and there is a compensatory greater reliance on the simple strategy of choosing the last animal name presented.

Modeling conclusions

Our final analysis of all of four models—the basic model, the restructured representation model, the attention change model, and the noisy access model—involves their descriptive adequacy. Being able to describe the data is a basic requirement for a model to be useful, and an important part of their evaluation. To measure descriptive adequacy, we considered how well model choice probabilities matched behavior. On

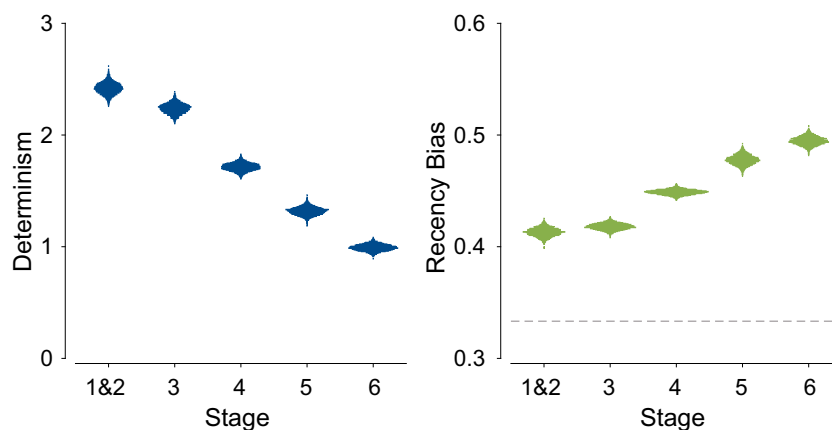


Fig. 5 Posterior distributions for response determinism γ and recency bias β for each FAST stage

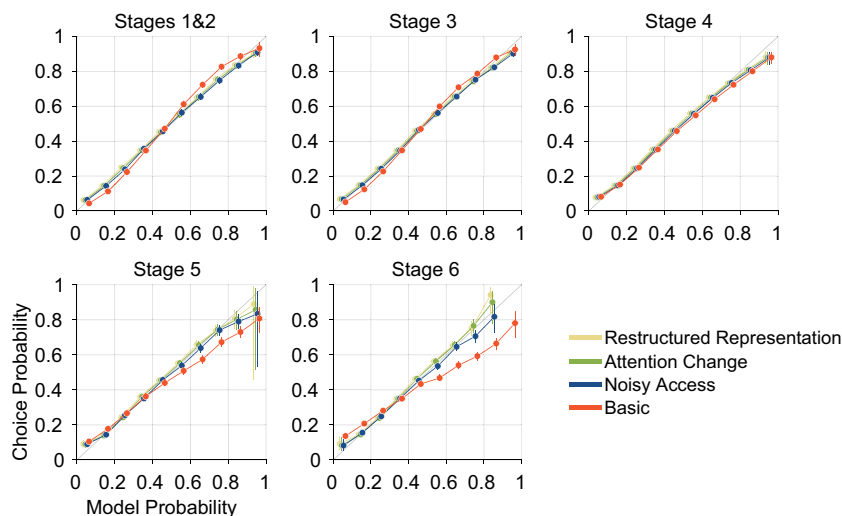


Fig. 6 Posterior predictive analysis of the descriptive adequacy of the restructured representation, attention change, noisy access, and basic model for all FAST stages

each trial, a model generates probabilities π_a , π_b , and π_c for the three alternatives. If these probabilities describe the data well, they should match the frequency with which the alternatives are actually chosen. For each of the models, we binned the choice probabilities in increments of 0.1, and calculated the frequency with which the alternative was chosen.

The results of this analysis are shown in Fig. 6. Panels correspond to FAST stages, the lines show the mean observed choice proportions, and error bars represent 95% credible intervals. The basic model is the worst performed. It fails to describe the data in FAST stages 5 and 6, and also has systematic deviations in the other stages. This finding is not surprising, since the basic model assumes there is no change in odd-one-out behavior across the stages, but our basic data analysis found clear evidence of change. The three other models all describe the data well, and are all very similar to each other.

These results show that the restructured representation model, the attention change model, and the noisy access model all meet the basic requirement of descriptive adequacy. Accordingly, our overall evaluation of the models focuses on *how* they achieve descriptive adequacy, which amounts to asking how they explain the changes in odd-one-out choice behavior. It is clear that the restructured representation model does not capture the changes in the data by changing the features used to represent animals. The attention change model also fails to explain the data in a way consistent with the guiding theory, which requires increases in attention to physical features at the expense of abstract ones. The noisy access model, in contrast, achieves descriptive adequacy in exactly the way predicted by the theory. Response determinism decreases with impairment and is accompanied by an increasing use of the simple response strategy of choosing the last item. Thus, our

conclusion is that the noisy access model provides the best account of the data. It is the psychologically and statistically simplest model, and accounts for the data in an interpretable way consistent with its theoretical assumptions.

Discussion

Using behavior in odd-one-out tasks, we conducted a model-based comparison of three hypotheses previously proposed as accounts of changes in semantic memory performance caused by Alzheimer's disease. We found no evidence for the idea that semantic representations themselves fundamentally change. We also did not find any evidence for the idea that the attention given to different types of features changes systematically, with more concrete physical properties becoming more prominent at the expense of more abstract features. Instead, we found that the differentiation between these types of features decreases as impairment progresses, suggesting a gradual loss of acuity in the use of the representations rather than a structured shift in attention. Consistent with this interpretation, we did find evidence favoring the idea that access to semantic information becomes noisier as impairment worsens. Choice behavior became less consistently linked to the underlying similarities between stimuli, and participants became more likely simply to repeat the last option. It is possible that patients in higher FAST stages do not fully understand the instructions of the task—and fail to do a search of semantic memory—but do understand that repeating an animal name back to the clinician is an acceptable response.

We do not believe these results eliminate the possibility that Alzheimer's disease does cause basic systematic

changes in semantic memory. The odd-one-out comparison task provides only one window onto semantic memory, and the clinical data we used considered only basic-level exemplars from the natural kind of animals. This limits the theories that can be tested. Some researchers, for example, have proposed a “bottom-up” degeneration of semantic memory (Henry, Crawford, & Phillips, 2004; Martin & Fedio, 1983; Tröster, Salmon, McCullough, & Butters, 1989), where Alzheimer’s patients have more difficulty generating specific exemplars (e.g., broccoli, orange) than categories (e.g., vegetables, fruits). There is also evidence that Alzheimer’s patients have a category-specific semantic memory deficit for living things (Chan et al., 2001; Whatmough & Chertkow, 2002). Our data cannot test either of these possibilities directly.

In addition to the limitations of our data, each of our models made a number of strong assumptions. Our models do not take into account any properties of the animal names themselves, such as their age of acquisition, word length, or word frequency. Word frequency may be particularly important, since previous research has shown that Alzheimer’s patients do not show typical effects of word frequency in tests of episodic memory (Balota, Burgess, Cortese, & Adams, 2002; Wilson, Bacon, Fox, Kramer, & Kaszniak, 1983). Another assumption is that, while we think our use of feature-based representations is appropriate, previous authors have often relied on dimensional representations (Chan et al., 1993a; Lee et al., 2016). The attention change and noisy access hypotheses probably could be formalized using dimensional representations, which would provide an alternative approach to evaluation. In addition, within the feature-based framework, it is possible to specify other models consistent with the broad idea of attention change. Our model focuses on one very specific influential proposal, but it is possible other systematic changes in attention do occur. What we do believe is that the noisy access model provides a psychologically and statistically simple and compelling account of our data, and serves as a theoretical safeguard that any more elaborate theory of change in semantic memory must outperform.

One avenue for stronger testing of the adequacy of the noisy access hypothesis, as compared to the possibility of more fundamental changes in representations, is to apply our models to different populations where there are clear expectations about whether and how semantic memory should change. Healthy aging provides a context in which we might expect noisy access to continue to provide a better account than restructured representation or attention change, and thus presents an opportunity for replication. Early child development, in contrast, provides a context in which we would expect to observe fundamental changes in the way children represent stimuli. A goal of future work

is to apply our models to children’s performance on the odd-one-out comparison task.

Understanding how Alzheimer’s disease affects semantic memory is a basic theoretical question with important societal implications. We have developed a model-based approach that is capable of expressing and evaluating competing theoretical accounts, and have demonstrated how Bayesian methods allow the models to be applied to a large real-world clinical data set. We hope that future work continues to expand and refine the models, and provides insights into how people’s semantic knowledge is impacted by memory impairment.

Acknowledgements There is a project page on the Open Science Framework associated with this article at <https://osf.io/xas7b/>. It contains code for the cognitive models and the feature data used in modeling. The archival clinical behavioral data are not available. This research was supported by funding from Medical Care Corporation. We thank Dennis Fortier, Junko Hara, Tushar Mangrola, and especially Rod Shankle and Jason Bock for their help in providing the data and their contribution to the research. We also thank Juliana Chhouk for help with the animal feature data, and members of the Bayesian cognitive modeling lab for helpful feedback. An earlier version of this research was presented at the 57th Annual Edwards Bayesian Research Conference, under the title “Can we create a model of triadic comparison of animal names? You bet giraffe we can!”.

References

- Au, A., Chan, A. S., & Chiu, H. (2003). Conceptual organization in Alzheimer’s dementia. *Journal of Clinical and Experimental Neuropsychology*, 25, 737–750.
- Balota, D. A., Burgess, G. C., Cortese, M. J., & Adams, D. R. (2002). The word-frequency mirror effect in young, old, and early-stage Alzheimer’s disease: Evidence for two processes in episodic recognition performance. *Journal of Memory and Language*, 46, 199–226.
- Benton, A. L., & Hamsher, K. (1983). Verbal fluency: Multilingual aphasia examination. *AJA Associates*.
- Burton, M. L., & Nerlove, S. B. (1976). Balanced designs for triads tests: Two examples from English. *Social Science Research*, 5, 247–267.
- Buschke, H. (1984). Cued recall in amnesia. *Journal of Clinical and Experimental Neuropsychology*, 6, 433–440.
- Chan, A. S., Butters, N., Paulsen, J. S., Salmon, D. P., Swenson, M. R., & Maloney, L. T. (1993a). An assessment of the semantic network in patients with Alzheimer’s disease. *Journal of Cognitive Neuroscience*, 5, 254–261.
- Chan, A. S., Butters, N., Salmon, D. P., & McGuire, K. A. (1993b). Dimensionality and clustering in the semantic network of patients with Alzheimer’s disease. *Psychology and Aging*, 8, 411–419.
- Chan, A. S., Butters, N., Salmon, D. P., Johnson, S. A., Paulsen, J. S., & Swenson, M. R. (1995). Comparison of the semantic networks in patients with dementia and amnesia. *Neuropsychology*, 9, 177–186.
- Chan, A. S., Salmon, D. P., & De La Pena, J. (2001). Abnormal semantic network for ‘animals’ but not ‘tools’ in patients with Alzheimer’s disease. *Cortex*, 37, 197–217.
- De Deyne, S., Verheyen, S., Ameel, E., Vanpaemel, W., Dry, M. J., Voorspoels, W., & Storms, G. (2008). Exemplar by feature

- applicability matrices and other Dutch normative data for semantic concepts. *Behavior Research Methods*, 40, 1030–1048.
- Elvevåg, B., & Storms, G. (2003). Scaling and clustering in the study of semantic disruptions in patients with schizophrenia: A re-evaluation. *Schizophrenia Research*, 63, 237–246.
- Goldstone, R. L. (1999). Similarity. In Wilson, R. A., & Keil, F. C. (Eds.) *The MIT encyclopedia of the cognitive sciences*, (pp. 763–765). Cambridge: MIT Press.
- Grober, E., & Buschke, H. (1987). Genuine memory deficits in dementia. *Developmental Neuropsychology*, 3, 13–36.
- Henry, J. D., Crawford, J. R., & Phillips, L. H. (2004). Verbal fluency performance in dementia of the Alzheimer's type: A meta-analysis. *Neuropsychologia*, 42, 1212–1222.
- Huntley, J. D., & Howard, R. J. (2010). Working memory in early Alzheimer's disease: A neuropsychological review. *International Journal of Geriatric Psychiatry: A Journal of the Psychiatry of Late Life and Allied Sciences*, 25, 121–132.
- Kavé, G. (2005). Standardization and norms for a Hebrew naming test. *Brain and Language*, 92, 204–211.
- Kopelman, M., Wilson, B., & Baddeley, A. (1989). The autobiographical memory interview: A new assessment of autobiographical and personal semantic memory in amnesic patients. *Journal of Clinical and Experimental Neuropsychology*, 11, 724–744.
- Lee, M. D., & Pope, K. J. (2006). Model selection for the rate problem: A comparison of significance testing, Bayesian, and minimum description length statistical inference. *Journal of Mathematical Psychology*, 50, 193–202.
- Lee, M. D., Abramyam, M., & Shankle, W. R. (2016). New methods, measures, and models for analyzing memory impairment using triadic comparisons. *Behavior Research Methods*, 48, 1492–1507.
- Levine, B., Svoboda, E., Hay, J. F., Winocur, G., & Moscovitch, M. (2002). Aging and autobiographical memory: Dissociating episodic from semantic retrieval. *Psychology and Aging*, 17, 677–689.
- Luce, R. D. (1959). *Individual choice behavior*. Hoboken: Wiley.
- Martin, A., & Fedio, P. (1983). Word production and comprehension in Alzheimer's disease: The breakdown of semantic knowledge. *Brain and Language*, 19, 124–141.
- McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack, J. R. C. R., Kawas, C. H., & Phelps, C. H. (2011). The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia*, 7, 263–269.
- Navarro, D. J., & Lee, M. D. (2004). Common and distinctive features in stimulus similarity: a modified version of the contrast model. *Psychonomic Bulletin & Review*, 11, 961–974.
- Navarro, D. J., & Griffiths, T. L. (2008). Latent features in similarity judgments: A nonparametric Bayesian approach. *Neural Computation*, 20, 2597–2628.
- Nebes, R. D. (1989). Semantic memory in Alzheimer's disease. *Psychological Bulletin*, 106, 377–394.
- Nebes, R. D., & Brady, C. B. (1990). Preserved organization of semantic attributes in Alzheimer's disease. *Psychology and Aging*, 5, 574–579.
- Newcombe, F. (1969). *Missile wounds of the brain: A study of psychological deficits*. Oxford: Oxford University Press.
- Peterson, J. C., Abbott, J. T., & Griffiths, T. L. (2018). Evaluating (and improving) the correspondence between deep neural networks and human representations. *Cognitive Science*, 42, 2648–2669.
- Plummer, M. (2003). JAGS: A Program for analysis of Bayesian graphical models using Gibbs sampling. In Hornik, K., Leisch, F., & Zeileis, A. (Eds.) *Proceedings of the 3rd International workshop on distributed statistical computing*. Vienna, Austria.
- Razani, J., Chan, A., Nordin, S., & Murphy, C. (2010). Semantic networks for odors and colors in Alzheimer's disease. *Neuropsychology*, 24, 291–299.
- Reisberg, B. (1988). Functional assessment staging (FAST). *Psychopharmacology Bulletin*, 24, 653–659.
- Romney, A. K., & Moore, C. C. (1998). Toward a theory of culture as shared cognitive structures. *Ethos*, 26, 314–337.
- Shankle, W. R., Mangrola, T., Chan, T., & Hara, J. (2009). Development and validation of the memory performance index: Reducing measurement error in recall tests. *Alzheimer's & Dementia*, 5, 295–306.
- Shepard, R. N. (1980). Multidimensional scaling, tree-fitting, and clustering. *Science*, 210, 390–398.
- Shepard, R. N., & Arabie, P. (1979). Additive clustering representations of similarities as combinations of discrete overlapping properties. *Psychological Review*, 86, 87–123.
- Storms, G., Dirix, T., Saerens, J., Verstraeten, S., & De Deyn, P. P. (2003). On the use of scaling and clustering in the study of semantic deficits. *Neuropsychology*, 17, 289–301.
- Trenkle, D. L., Shankle, W. R., & Azen, S. P. (2007). Detecting cognitive impairment in primary care: Performance assessment of three screening instruments. *Journal of Alzheimer's Disease*, 11, 323–335.
- Tröster, A. I., Salmon, D. P., McCullough, D., & Butters, N. (1989). A comparison of the category fluency deficits associated with Alzheimer's and Huntington's disease. *Brain and Language*, 37, 500–513.
- Tversky, A. (1977). Features of similarity. *Psychological Review*, 84, 327–352.
- Voorspoels, W., Storms, G., Longenecker, J., Verheyen, S., Weinberger, D. R., & Elvevåg, B. (2014). Clustering techniques and their pitfalls Deriving semantic structure from category fluency: Clustering techniques and their pitfalls. *Cortex*, 55, 130–147.
- Wetzels, R., Grasman, R. P. P. P., & Wagenmakers, E. J. (2010). An encompassing prior generalization of the Savage–Dickey density ratio test. *Computational Statistics and Data Analysis*, 54, 2094–2102.
- Whatmough, C., & Chertkow, H. (2002). Category-specific recognition impairments in Alzheimer's disease. In Forde, E. M. E., & Humphreys, G. W. (Eds.) *Category specificity in brain and mind*, (pp. 181–210): Psychology Press.
- White, A., Voorspoels, W., Storms, G., & Verheyen, S. (2014). Problems of reliability and validity with similarity derived from category fluency. *Psychiatry Research*, 220, 1125–1130.
- Wilson, R. S., Bacon, L. D., Fox, J. H., Kramer, R. L., & Kaszniak, A. W. (1983). Word frequency effect and recognition memory in dementia of the Alzheimer type. *Journal of Clinical and Experimental Neuropsychology*, 5, 97–104.
- World Health Organization (2019). Dementia. <https://www.who.int/en/news-room/fact-sheets/detail/dementia>.
- Zeigenfuse, M. D., & Lee, M. D. (2010). Finding the features that represent stimuli. *Acta Psychologica*, 133, 283–295.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.