Special Issue: Cognitive Conflict Control

Interactive Activation and Competition Models and Semantic Context: From Behavioral to Brain Data

Markus J. Hofmann*[12] and Arthur M. Jacobs[13]

doi://10.1016/j.neubiorev.2014.06.011

Accepted Manuscript: This version will not exactly match the published manuscript

1. Experimental and Neurocognitive Psychology, Free University Berlin

2. General and Biological Psychology, University of Wuppertal

3. Dahlem Institute for the Neuroimaging of Emotion (D.I.N.E.)

* Correspondence to: M. J. Hofmann, Department of Psychology, General and Biological Psychology, Room Z.01.11, Max-Horkheimer Str. 20, 42119 Wuppertal, Germany

e-mail: mhofmann@uni-wuppertal.de). Tel. +49 202 4392340; Fax: +49 202 439 2926

Highlights

- Interactive Activation and Competition models (IAMs) address implicit memory
- Associative Read-Out Model (AROM) simulates semantic processes in explicit memory
- IAMs can account for neurocognitive data in occipital, temporal and frontal cortices
- Semantic cohesion accounts for positive, but not negative valence

Abstract

Interactive activation and competition models (IAMs) can not only account for behavioral data from implicit memory tasks, but also for brain data. We start by a discussion of standards for developing and evaluating cognitive models, followed by example demonstrations. In doing so, we relate IAM representations to word length, sequence, frequency, repetition, and orthographic neighborhood effects in behavioral, electrophysiological, and neuroimaging studies along the ventral visual stream. We then examine to what extent lexical competition can account for anterior cingulate cortex (ACC) activation and the N2/N400 complex. The subsequent section presents the Associative Read-Out Model (AROM), which extends the scope of IAMs by introducing explicit memory and semantic representations. Thereby, it can account for false memories, and familiarity and recollection - explaining why memory signal variances are greater for studied than non-studied items. Since the AROM captures associative spreading across semantic long-term memory, it can also account for different temporal lobe functions, and allows for item-level predictions of the left inferior frontal gyrus' BOLD response. Finally, we use the AROM to examine whether semantic cohesiveness can account for effects previously ascribed to affective word features, i.e. emotional valence, and show that this is the case for positive, but not for negative valence.

Key words: word recognition, associative spreading, Multiple-Read Model (MROM), semantic process model, episodic memory.

Contents

1. Introduction

Back in 1981, the interactive activation and competition model (IAM) was a major step ahead in cognitive modeling for several reasons. Sternberg's (1969) seminal model of verbal working memory, or Morton's (1969) logogen model already had zoomed into the blackbox between stimulus and response, breaking it up into specific serial or parallel stages of information processing (i.e., the famous boxes and arrows of 'boxological models'; cf. Jacobs and Grainger, 1994). The IAM combined features of previous formal word recognition models by Broadbent (1967), Morton (1969), Rumelhart and Siple (1974), or Treisman (1978) with pioneering "neural models" of the time (e.g., Anderson et al., 1977; Grossberg, 1980). It was the first model that really zoomed into (the dynamics of) those 'boxes' and allowed to simulate the time course of information processing in several parallel layers (i.e., feature, letter, and word unit layer; Figure 1).

---

insert Figure 1

The IAM also implemented two neurally plausible features, connectivity (excitation and inhibition allowing within-level competition) and interactivity (top-down feedback allowing between-level memory effects on perceptual processing). Both features were disputed at the theoretical level by main stream cognitive modelers favoring modular cognitive architectures at the time (e.g., Massaro, 1988; Massaro and Cohen, 1991; Paap et al., 1982), but they were also experimentally testable (Jacobs and Grainger, 1992; McClelland, 1991). By making the top-down feedback algorithmically concrete, the IAM succeeded in elegantly simulating the word superiority effect (Figure 2). This corresponded to Helmholtz's idea of *unconscious inferences,* expressing his belief that sensory data are modified by previous experience via ideas/concepts, before they become a true perception (Boring, 1950;

Grossberg, 1980). In the more modern words of Grossberg (1980) „sensory data activate a feedback process whereby a learned template, or expectancy, deforms the sensory data until a consensus is reached between what the data are and what we expect them to be. Only then do we perceive anything“. In Friston's (2010) unifiying principle of brain function, such feedback processes (mathematically formulated within the frameworks of free energy and predictive coding) play a central role and there is now ample evidence for its neural plausibility (e.g., Price and Devlin, 2011).

---

insert Figure 2

The IAM was perhaps the first model in this field that made all information processing steps between input and output fully transparent, thus providing a comprehensive description of information processing at the micro level, and -- achieving what can be considered a gold standard of model evaluation criteria (Jacobs and Grainger, 1994) -- it predicted a new phenomenon which had previously not been observed: the neighborhood frequency effect (cf. Jacobs et al., 1998). This effect was experimentally confirmed (Grainger et al., 1989) and thereby fired further developments leading to many successful extensions or variants of the basic interactive activation architecture, e.g. the model of the Stroop task (Cohen et al., 1990), the Dual Read-Out Model (DROM; Grainger and Jacobs, 1994), the Multiple Read-Out Model (MROM; Grainger and Jacobs, 1996) and its extension including phonological processing units (MROM-p; Jacobs et al., 1998), the conflict monitoring theory (CMT; Botvinick et al., 2001), the dual-route cascaded model (DRC, Coltheart et al., 2001), the connectionist dual-process model (CDP++ ; Perry et al., 2007, 2010), or the recent AROM including an implemented semantic layer (Hofmann et al., 2011), to name only a few.

While during the 80s and 90s IAMs were very successful in predicting behavioral data such as error rates, or response time means and distributions in many different tasks, in 1995 only Jacobs and Carr

(1995) speculated how they could be applied to neuroimaging data, and how functional neuroimaging could be used to constrain computational models of cognition in general: (1) by providing information about the neuroanatomical loci of different subprocesses and hence system decomposability and (2) by delineating the temporal dynamics of the cognitive process(es) under investigation (cf. Barber and Kutas, 2007). It took a few years until Botvinick et al. (2001) attempted to indirectly connect the output of IAM simulations to neuroimaging data (hypothetical ACC activation; see 3.1 below). With regard to brain-electrical data, Braun et al. (2006) were the first to account for N400 amplitudes from the output of an IAM. However, in contrast to other cognitive models (e.g., Anderson et al., 2004), there still appear to be no direct IAM simulations of hemodynamic activation functions (but see Taylor et al., 2012, for a viable proposal in this direction).

In this review, we will address possibilities and limitations of IAMs as heuristic tools for better understanding the functioning of mental processes at both the cognitive and neuronal levels. We understand the term ‘heuristic' here in the sense of "useful for generating experiments, ideas, or explanations" (e.g., Grainger and Jacobs, 1998; Massaro, 1989). As exemplified by the classical definition of probability (Popper, 1935), or the first DNA models of Crick and Watson, hypotheses and models can indeed be of great heuristic value. In terms of Popper's logic of research, heuristic hypotheses and models are not meant to represent the truth, but they are a means to approximate it. Therefore, as is the case for many computer models in general, and connectionist models in particular, we see them as providing a sufficiency analysis (for an early appraisal of simulations as sufficiency analyses see Miller, Galanter, and Pribram, 1960). That is, they show how something could function, but not how it necessarily functions. In terms of the sufficiency of such computer models, Jackendoff (1987) proposed the following hypothesis: ‘‘Every phenomenological distinction is caused by/supported by/projected from a corresponding computational distinction’’ (cf. Sun et al., 2005). In other fields, computer-simulation models have become one of the most prevalent ways of discovering

new insights. For example, they have yielded insights into how galaxies form spiral arms, or how solids, liquids, and gases vibrate, flow, and change state. Wherever complexity makes exact analytical solutions impossible, computer simulations can show us the way (Estes, 1975; Jacobs and Grainger, 1994). Our aim here is to show that this also holds for IAMs in cognitive neuroscience. As recently demonstrated impressively by Taylor et al. (2012), cognitive models are relevant for interpreting neuroimaging studies. Neuroscientific studies in turn can provide data relevant for advancing cognitive models. However, for this enterprise to be really promising, section two presents a general framework for building, testing, and evaluating computational models (Jacobs et al., 1998; Jacobs and Grainger, 1999) and connecting them meaningfully with neuroimaging data(Jacobs and Hofmann, 2013). This framework will be demonstrated in the subsequent sections. Section three discusses the connectability of IAMs to brain data, reviewing studies from posterior to anterior regions along the ventral visual stream. Concerning cognitive conflict control, we then also test the hypothesis that competition between lexical units can account for the N400. This reveals a major limitation of previous IAMs, which lies in their failure to represent semantic information in an algorithmically concrete fashion. Therefore, section four reviews work on semantic processes in word recognition and presents recent progress in simulating these. As parsimony is an important model feature, we finally test whether effects of the emotional valence of words can actually be explained by semantic cohesion in section five.

2. Towards standards for developing and evaluating neurocognitive models

"How would we look for a new law?

...

First, we guess it. ...

Then, we compute the consequences of the guess. ...

And then we compare these computation results to ... experiment ...

If it disagrees with experiment, it's wrong.

And that simple statement is the key to science."

Lecture by Richard Feynman

(e.g., http://amiquote.tumblr.com/post/4463599197/richard-feynman-on-how-we-would-look-for-a-new-law)

In the following, we elaborate on a general framework for building, testing, and evaluating computational models, and then illustrate it using examples of IAMs. A rough guide to model development is the five-step framework inspired by psychometrics and test theory proposed by Jacobs et al. (1998; Jacobs and Grainger, 1999; cf. Collyer, 1985).

In *step 1*, the modeler examines the global appropriateness of the architectural assumptions. In classic modeling approaches to behavior, this initial phase of model construction serves to uncover fundamental flaws in the model architecture. This can be done by parameter tuning studies that identify parameter configurations that, for example, produce catastrophic behavior. The global architecture of

IAMs of reading, as exemplified in Figures 1 and 2, has been supported by numerous behavioral studies, and we can safely continue to use it until a major falsification is published. Neurocognitive data can provide additional information about the neurobiological plausibility of this architecture (Price and Devlin, 2003; Taylor et al., 2012). Section 3 addresses the question whether this architecture fits also with neuroimaging and electrophysiological findings.

In *step 2*, *estimator set studies* estimate model parameters from "important" data such that the point in parameter space is identified for which the model is "true." In *step 3*, *criterion set studies* test model predictions with parameters fixed (in step 2) against fresh data, using an explicit criterion (Jacobs and Grainger, 1994; Perry et al., 2007). In *step 4*, *strong inference studies* involve formal, criterion-guided comparisons of alternative models against the same data sets to select the best model. In this step, the modeler actually computes the potential consequences, i.e. what empirical data should follow from each model if it was correct. Particularly if the model predictions disagree with experiments, *step 5* is necessary and the model has to be refined or replaced.

Though computing the logical consequences of a model provides many advantages, a computational model is an implementation of a verbal-level theoretical rationale. The key hypotheses of the model can often be tested without explicit computational simulations. Therefore, any empirical work that uses the model to interpret data can be seen as a part of step 5. We refer to such a verbal-level theorizing as a prequantitative model, because the next step should lie in refining the computational model (Jacobs and Grainger, 1994). To illustrate the architecture of the model to researchers that focus their efforts on the empirical work rather than the development of algorithmic models, it is also useful to represent the verbal-level description of a computational model by boxes and arrows (Figure 1). However, models that additionally provide a computationally concrete form can better be compared formally.

Among such computational or quantitative models, two major types can be discriminated: Mathematical and algorithmic ones. The critical distinction between these two types of models is that

the former consist of commonly accepted mathematical operations, only (+, -, *, / …). Using only laws in such a closed-form expression allows to guide the selection of the best model by the Akaike Information Criterion (AIC), or the Bayesian Information Criterion (BIC; cf. step 3 to 5). They provide a quantitative heuristic to select the simplest model that accounts for a maximum of variance in the data. Therefore, they penalize model complexity in terms of free parameters and reward the fit between model and data. Minimum Descriptive Length (MDL) as the most recent mathematical model evaluation criterion additionally takes into account the complexity of the functional form (MDL, Pitt et al., 2002). This can be demonstrated best by an artificial example, in which we define a 'true' model, that generates the data: Assume that a predictor x accounts for the data as a function f(x)=a*x + b + error. When testing whether this model can better account for data generated by itself than by an alternative model, say $f(x) = x^a + b*x + error$, it is likely that the latter model accounts for more variance, though both models have the same amount of free parameters (a, b). The reason for this is that when x is taken to the power of a free parameter, the model can represent linear as well as several types of curvilinear relations of x and f(x). Therefore, it can be more flexibly adopted to the data. Thus, it more likely fits a considerable amount of error variance of the original model. This has been termed 'overfitting'; and overfitting makes it unlikely that the explained variance can be reproduced in another test of the model (Pitt et al., 2002). Therefore, the functionally more complex model is less likely to be generalized.

IAMs are an example of algorithmic models. This format allows not only closed-form expressions, but everything that can be programmed down. This provides a more unconstrained expression of creative theoretical ideas. On the downside, the aforementioned criteria can not always the straightforwardly applied to algorithmic models, which makes model comparisons more difficult. We propose that for algorithmic models, model evaluation should focus on the following four criteria (Jacobs and Grainger, 1994), expressing the very same principles of model evaluation (Pitt et al., 2002).

(i) *Descriptive adequacy.* This provides a quantitative benchmark of how well the model fits the data. The amount of explained item-level variance is the classic benchmark for algorithmic models (Spieler and Balota, 1997), e.g. the model predicts which letter string elicits which mean response latency or ERP amplitude, when averaged across participants (Spieler and Balota, 1997; Hofmann et al., 2008b; Rey et al., 2009).

(ii) *Generality*. *Vertical* generality indicates a model's ability to generalize across different scales of the modeled process (Jacobs and Grainger, 1994). This is particularly important for models accounting for neurocognitive data, since they may not only account for the behavioral end product of a cognitive process (e.g., response time), but also simulate the dynamics of stimulus processing to predict brain-electrical or neuroimaging data (e.g., Braun et al., 2006; Hofmann, et al., 2008b; section 3.1). *Horizontal* generality describes how well the model can be generalized across different tasks or novel stimuli. However, the probability of overfitting the model to error variance in the estimator-set studies of step 2 increases with model complexity. Therefore, it is more likely that simpler models are generalized to novel stimuli, for instance (Pitt et al., 2002).

(iii) *Simplicity and falsifiability*. A first rough approximation to this criterion is a count of the number of boxes and arrows of the boxological, verbal-level representation of the whole model. This also allows to quickly check the plausibility of the model, as illustrated by the following question (Jacobs, 2008): how many pure neuropsychological disorders predicts a boxological model with nine boxes and fourteen arrows (a typical prequantitative word recognition model, as illustrated in Jacobs and Grainger, 1994)? Answer: 9+14 = 23! Even if the clinical neuropsychological literature provided as many fitting case descriptions, the question is whether a model that contains all logically possible cases or outcomes is not a tautology. Note, however, that the boxological representation of the very same model can be simpler or more complex (see Figures 1 and 4 below). Therefore, counting the number of free parameters is a better heuristic. What makes model comparisons in this regard somewhat tricky,

though, is that some parameters have a theoretical impact, and therefore they are less free to vary. Such critical parameters can also turn out to be the strongest falsificator of a model, or to test the necessity of particular theoretical assumptions in the simulation. As an example, the top-down excitation of the orthographic word layer to the letter layer is scaled by a parameter, which was viewed to be necessary to account for the word superiority effect, i.e. letter activation depends on orthographic word-level activation (McClelland and Rumelhart, 1981). Jacobs and Grainger (1992) tested whether this top-down excitation improves the prediction of lexical decision performance by setting the corresponding parameter to zero. They found that the simpler model could equally well account for the data, thus questioning the necessity of this parameter and the underlying assumption. Another way to determine the simplicity of an algorithmic model is to take minimum descriptive length literally (Pitt et al., 2002). The less words it takes to make readers understand how the whole model works, the better is the model. Keep in mind that the reason for building algorithmic models is to understand how complex processes interact with each other. The simpler the model, the more researchers can –in principle- evaluate the model on a prequantitative level and therefore contribute to step 5 of model development. When many researchers understand the model and use it to interpret their data, falsification probability should increase (not withstanding extreme forms of confirmation bias). And, of course, falsifiability is a major virtue of any model. The more constraints a model puts on observations consistent with the model, the easier it is to falsify and learn from it (Pitt et al., 2002).

(iv) *Explanatory adequacy.* A tentative definition is: how close does the model come to reality (cf. section 1 in Grainger and Jacobs, 1998). This is closely related to the success of a model to predict novel phenomena. Feynman would perhaps define it as the amount of' computed consequences that follow from the rationale of your model, which have successfully been 'compared to experiment'. For instance, the model prediction of an inhibitory effect of a higher-frequency orthographic neighbor on word recognition has been confirmed experimentally (Grainger et al., 1989; Jacobs et al., 1998).

Explanatory adequacy is also closely related to the amount of ad-hoc assumptions, i.e. features of the model architecture like a free parameter that only describe the empirical phenomenon for which they were introduced, without any independent evidence calling for or other predictions following from it. For example, the unequal variance model of recognition memory makes two ad-hoc assumptions (Glanzer et al., 1999): First, the memory signal is greater for words learned in a previous study phase, as compared to non-studied items (cf. section 4.3). Second, the memory-signal variance of studied items is greater than the memory signal variance of non-studied items. This can describe two empirical phenomena of recognition memory: First, more studied items are more likely to be classified as 'old'. Second, studied items have a greater memory signal variance than non-studied ones.

However, there are no other predictions that can be derived from this model that would allow to falsify it, resulting in a low explanatory adequacy. The ad-hoc assumption of a greater memory signal in studied items describes only the empirical phenomenon for which is was introduced. In section 4.3, we will describe how ad-hoc assumption one and a few well-proven IAM assumptions can account for both empirical phenomena (Hofmann et al., 2011; cf. Appendix A).

While it is still hard to offer an exhaustive operational definition of explanatory adequacy, a tentative definition of the term '*explanatory value*' could be useful: the number of empirical phenomena (significant experimental effects) predicted by the model divided by the number of ad-hoc assumptions of the model. Following this, the IAM has a higher explanatory value than the unequal variance model of recognition memory, because implementing one novel assumption into the IAM accounts for two novel phenomena (Hofmann et al., 2011).

More importantly, we can derive a test of how adequate the IAM's explanation of recognition memory phenomena is. For instance, a longer study phase results in even greater memory signals, and therefore the variance of longer-studied items is even greater than shorter-studied items (Glanzer et al., 1999; Hofmann et al., 2011). We are optimistic, though, that the IAM's account of recognition memory

phenomena will be falsified one day, because „All models are wrong, but some are useful“ for making successful predictions (Box, 1979, p. 202; quoted from Wagenmakers et al., 2007). When finding out why the previous model failed, the new model comes closer to reality.

3. Model-to-brain-data connections of IAMs

The issue of this section is to what extent IAMs can be connected to neuroanatomical or neuroimaging data in a way that is fruitful for both the computational modeling and the mind mapping initiatives (Jacobs and Carr, 1995; Jacobs and Rösler, 1999). We start with a review about brain regions that can be associated with the basic visual-orthographic processes of IAMs. As the CMT played a pioneering role in this respect, we then address the issue whether orthographic and/or semantic competition can account for the hypothesis that the N400 can be considered a special case of an ACC-driven N2 generator (Yeung et al., 2004).

3.1. IAMs and the ventral visual stream

Ungerleider and Mishkin (1982) proposed that the ventral visual stream from occipital to temporal regions is specialized for object perception and serves for identifying 'what' a visual object is. The IAM framework can illustrate the cognitive processes leading to object recognition using the example of visual word recognition (Norris and Kinoshita, 2012).

---

insert Figure 3 about here

The visual feature layer represents information in the visual cortex, while more anterior regions are sensitive to the combination of such features (Figure 3). Letters are an example of such a greater chunk

of information in more anterior regions (Miller, 1956). Finally, neurons that respond to coincidences of letters have been located in the left posterior fusiform gyrus – in a region also termed 'visual word form area' (Cohen et al., 2000; Petersen et al., 1988; but see Dehaene et al., 2002). Overall, the more anterior a region is in this stream, the greater appears its level of abstraction (Vinckier et al., 2007), and IAMs can well account for this based on three levels of grain size: visual features, letters and orthographic word forms (Ziegler and Goswami, 2005; cf. Figure 1). This model-to-brain data connection hypothesis can account for five effects in neuroimaging, electrophysiological, and behavioral data (Table 1).

---

insert Table 1 about here

The IAM can first account for word length effects in the ventral visual stream (Ziegler et al., 2001). The longer a word is, the greater is the amount of its visual features. Therefore, occipital activation is greater for longer letter strings (Mechelli et al., 2000; Schurz et al., 2010). While visual information has been estimated to reach the occipital cortex about 50 ms after stimulus presentation (Jeffreys and Axford, 1972), ERP and MEG studies suggest that occipital word length effects start around 60 ms post-stimulus and can extend up to 125 ms (Assodollahi and Pulvermüller, 2001; Hauk and Pulvermüller, 2004). For nonwords only, fMRI length effects extend to more anterior regions such as the fusiform gyrus (Schurz et al., 2010). This suggests that top-down excitation from semantic processes can take effect on these representations in word stimuli, and thus their activation is hardly predictable by word length only.

Second, the IAM can account for cross-trial sequence effects that can cause top-down driven excitation of the lowest levels of visual feature processing. For example, Kuchinke et al. (2011) found that items

following word stimuli yield greater occipital activation during lexical decision than items following nonwords. Words apparently engage expectations about other words. In another study, Dambacher et al. (2009) found that word predictability from sentence context can affect ERPs as early as 50 ms after stimulus presentation (cf. Penolazzi et al., 2007; Sereno and Rayner, 2003; Skrandies, 1998). A recent optical imaging study using the same experimental design confirmed the hypothesis that this results from occipital cortex activation (Dambacher et al., 2009; Hofmann et al., 2014). A potential explanation is that expectations generated at a semantic level feed activation back to the lowest level of feature representations in the occipital cortex (cf. McClelland, 1993; Price and Devlin, 2011; Rauss et al., 2011; Rey et al., 2009).

A third effect that can be explained by IAMs is the word frequency effect (McClelland and Rumelhart, 1981). The model predicts that words are lexically accessed faster when they represent high-frequency words, because orthographic units of a high-frequency word have a greater resting-level activation in the orthographic word layer than the units coding low-frequency words. Thus, high-frequency words reach a critical level of activation earlier and are responded to faster than low-frequency words (Grainger and Jacobs, 1996; Kronbichler et al., 2004; McClelland and Rumelhart, 1981). This early access to orthographic representations occurs between 100 ms and 200 ms post-stimulus, as indicated by greater ERP amplitudes to low than to high-frequency words in this time frame (Dambacher et al., 2006; Hauk and Pulvermüller, 2004; Sereno et al., 1998). Likewise, visual information reaches the fusiform gyrus around 100 ms (Hofmann et al., 2009), as also indicated by the finding that transcranial magnetic pulses start to disrupt lexical decision performance in a region including the fusiform gyrus in a time frame of 80-120 ms (Duncan et al., 2010).

A fourth key effect concerns the orthographic overlap between a prime and target, which facilitates word recognition. This behavioral effect that can be simulated by an IAM (Grainger and Jacobs, 1993).

In accordance with our proposal that orthographic layer activation is reflected by fusiform gyrus activation, orthographic priming reduces the activation in this region (Devlin et al., 2006), and results in early ERP effects (Huber et al., 2008). Such priming effects, however, vanish as soon as the orthographic similarity is accompanied by a semantic relation between the prime and the stimulus. This lead Devlin et al. (2006) to conclude on the primary function of the fusiform gyrus as a hub to higher-order processes (Hofmann et al., 2009; Price and Devlin, 2003).

In an IAM, however, the seemingly contradictory views concerning the fusiform gyrus to either represent orthographic word forms or to be a hub to higher-order functions can be reconciled (Cohen and Dehaene, 2004; Price and Devlin, 2003): Orthographic representations are computed by help of this region, but the interactivity principle commands that it interfaces with other regions subserving higher order functions (McClelland and Rumelhart, 1981; McClelland, 1993).

A fifth key effect that was successfully simulated by IAMs at the behavioral level is the orthographic neighborhood effect (Grainger and Jacobs, 1996; Jacobs and Grainger, 1992). Orthographic neighbors differ from the target word with respect to one letter (cf. Figure 4 below). IAMs predict that such orthographically similar items are co-activated by the stimulus. This increases the total activation in the orthographic lexicon and thus facilitates word recognition (Grainger and Jacobs, 1996). While there were successful quantitative IAM simulations of the N400 for nonword but not for word stimuli (Braun et al., 2006; Hofmann et al., 2008b; cf. 4.2 below), Binder et al. (2003) tested whether an increase in hypothetical lexical activation is accompanied by greater hemodynamic responses in the fusiform gyrus. They found no effects in this region – a zero finding that has been replicated (Fiebach et al., 2007). This seems to falsify the idea that the fusiform gyrus straightforwardly reflects activation in the orthographic word layer of the IAM (Grainger and Jacobs, 1996; McClelland and Rumelhart, 1981). However, for the temporal cortex, Binder and colleagues (2003) found an inverted orthographic

neighborhood effect: Words with fewer neighbors elicited greater activation in the temporal cortex. They suggested that a harder identification at the orthographic level leads to compensatory semantic activation, and thus temporal cortex activation is greater for words with fewer neighbors (cf. Klonek et al., 2009). Thus, Binder et al.'s (2003) results suggest a model revision of the classic three-layered IAM, which is in line with Price and Devlin's (2011) proposal that the fusiform gyrus acts as a hub to higher-order processes: IAMs need to take semantic processes into account.

IAMs can quantitatively account for all but the sequence effects at a behavioral level (cf. Grainger and Jacobs, 1996; Huber et al., 2008; McClelland and Rumelhart, 1981; Ziegler et al., 2001). At the level of electrophysiological data, there are at least three examples for a direct simulation by an IAM (e.g., Braun et al., 2006; Hofmann et al., 2008b; Huber et al., 2008). In contrast, so far IAMs have only 'qualitatively' accounted for neuroimaging findings. We can not yet compute a hypothetical fusiform gyrus activation at the item-level, which could then be directly compared to BOLD data. Testing for the descriptive adequacy of the model is not yet possible, because the expansion of the IAM that accounts for neurocognitive data is still at step 1 of model development, i.e. checking the global appropriateness of the architectural assumptions.

To gain neurobiological plausibility for this type of model, we have to ask how the BOLD response can be predicted by IAM units, which tentatively represent a neuron or an assembly of neurons in the brain (Hofmann et al., 2011). For instance, how can the temporarily greater orthographic unit activation of a high-frequency word be reconciled with the finding of a lower BOLD response in the ventral visual stream (Kronbichler et al., 2004)? A straightforward answer is provided by the *model-to-fMRI-data connection* assumption of the ACT-R (Anderson et al., 2004). It predicts greater BOLD responses to low-frequency words, suggesting that the longer a stimulus is processed, the greater is the corresponding hemodynamic response. We propose that in each processing cycle, the amount of model

activation is proportional to the measured relative hemodynamic activation (cf. Jacobs and Carr, 1995). As it takes less cycles to reach the response criterion for high-frequency words, a lower amount of hemodynamic activation would sum up. Thus, the hemodynamic response gradually decreases with the frequency of occurrence of a word in natural language (Kronbichler et al., 2004). This notion of model-to-data connection is compatible with a computational model of the hemodynamic response proposing that the longer a stimulus is processed, the greater is the hemodynamic response (Buxton et al., 2004). A second possibility for why hemodynamic responses are lower for high-frequency words is the earlier activation of high-frequency word units in IAMs, which also start to inhibit other word units earlier than low-frequency words. Thus, we would expect that overall less units become active for high-frequency words until a response is given. Low-frequency words, in contrast, take longer to become active and due to a lack of early inhibition, it is more likely that other (orthographically similar) orthographic units gain a chance to become active, thus increasing their BOLD response (Botvinick et al., 2001). We suggest that these two mechanisms - the rapid rise of a high-frequency word's activation and the resulting greater inhibition of other units - are well in line with Taylor et al.'s (2012) model-to-data connection proposal suggesting more effortful processing when stimuli do not match model representations.

However, future work still must provide explicit simulations of fMRI data to test which model-to-data connection option offers the best solution: If simulated reaction times predicted BOLD responses best, this would suggest that the time-based model-to-data connection is most appropriate in accounting for hemodynamic responses. If the competition in a representational layer, however, was a more suitable functional explanation, this could be formally addressed within the framework of Botvinick et al.'s IAM (2001), as will be addressed in the next section.

3.2. Lexical competition, ACC activation and the N400/N2 complex in language processing

The CMT proposed that a keener competition between representational units should lead to activation in the mediofrontal and anterior cingulate cortex (ACC; Botvinick et al., 2001). While most research on the CMT focused on competition between response units (Teodorescu and Usher, 2013), Botvinick et al. (2001) illustrated that also competition between lexical units leads to ACC activation. The typical IAM contains three implemented layers of different size and function (McClelland and Rumelhart, 1981; cf. Grainger and Jacobs, 1996; Figure 4; cf. also Figure 1). Each of these layers contains one type of representation. The first layer represents visual features, such as "|" as a straight line at a particular position, e.g. in the left part of the first letter. The second layer contains letters, which are activated by visual features. For instance, "F" is composed of two vertical and two horizontal features. When these become activated, they likely activate the letter representation of "F". The third layer represents orthographic word forms, such as "BLUR", which receive activation from certain letters at particular positions.

---

insert Figure 4 about here

Botvinick et al.'s (2001) IAM simulations predicted that the bigger the competition between orthographic word units is, the stronger should be ACC activation. Competition was quantified by Hopfield energy, i.e. the sum of the activation products of each possible activated word pair. For instance, when BLUR, BLUE and SLUR have been activated (Figure 4), Hopfield energy equals the activation of the orthographic word units BLUR * BLUE + BLUR * SLUR + BLUE * SLUR. A later extension suggested that the CMT can be tested using event-related potentials (ERP) and behavioral

data (Yeung et al., 2004): a greater amount of competition predicts not only a greater amount of errors and slower response times, but also a more negative N2 component, which is likely generated in the ACC.

Hofmann et al. (2008b) tested whether orthographic competition IAM can account for ERP responses: 300 nonwords were presented to the model, and the orthographic competition of the items was calculated by Hopfield energy. Using three levels of low, medium and high orthographic competition, they found that the keener the competition was, the larger was an N2 negativity. As predicted by Yeung et al.'s (2004) CMT extension, this negativity had a frontal maximum, and source localization suggested that ACC activation also gradually increased with the amount of conflict. Moreover, increasing levels of conflict resulted in longer RTs and more errors (Yeung et al., 2004). Since the N2 showed a striking similarity with the N400 component (Kutas and Hillyard, 1984), Hofmann et al. (2008b) proposed that the N400 reflects the competition of lexical information and therefore can be considered a special case of an N2 (cf. Polich, 1985). There are several reasons for this proposal. First, the N400 as a language-based N2 is obtained in a typical N400 time frame starting as early as 200 ms after stimulus exposure (Kutas and Federmeier, 2011). Second, the maximum difference at posterior, centroparietal electrodes speaks of a classic N400. Third, the maximum frontal negativity pointed at an FN400, which is functionally not distinct from the N400 (Kutas and Federmeier, 2011). Fourth, fMRI studies that investigated effects of orthographic neighborhood size support the conclusion that lexical competition drives ACC activation (Fiebach et al., 2007), which is the psycholinguistic variable that is most highly correlated with Hopfield energy ($r = 0.72$). And fifth, Hofmann et al. (2008b) tested whether the Hopfield energy elicited by each item could account for a significant portion of item-level variance in the N2/N400 amplitude. Since the simulated amount of conflict predicted ERP amplitudes of single items, the very same computational mechanism that accounted for conflict effects in Yeung et

al.'s (2004) averaged ERP amplitudes here predicted fine-grained gradual increases of the N2/N400. The computed amount of orthographic competition accounted for 12% of the ERP variance of the N2/N400, and thus descriptive adequacy can be quantified for brain-electrical data.

However, IAMs failed to account for ERP variance in word stimuli (Braun et al., 2006; Hofmann et al., 2008b; but see Holcomb et al., 2002), thus questioning the horizontal generality of the ERP predictions across different types of stimuli. Moreover, also low orthographic neighborhoods can show greater ACC activation (Binder et al., 2003; Fiebach et al., 2007). As the critical distinction between words and nonwords is that words carry semantic information, we suggest that decisions to words are also affected by a semantic level of representations, which was, however, not yet implemented in IAMs at that time.

---

insert Figure 5 about here

Higher-level features such as semantics had been proposed but not implemented in the broader, (prequantitative) theoretical framework of McClelland and Rumelhart (1981; Figure 5; Rumelhart and McClelland, 1982). Also, Coltheart et al. (2001) proposed a non-implemented semantic layer to account for deep dyslexia, in which semantic-associative confusions between words are frequent (Coltheart et al., 2001). Botvinick et al. (2001) also discussed that a verb generation task can elicit competition at a semantic-associative level, which leads to ACC activation. To define semantic associates, Thompson-Schill et al. (1997) used human performance in a free association task: participants named the first words that come into mind in response to a target word. When a later task required other participants to generate a verb, given such a noun stimulus, ACC activation was greater when two verb responses were prevalent in comparison to a single predominant verb response.

To address the question whether the 'classic' N400 reflects semantic competition during sentence processing, semantics is typically defined by cloze completion probabilities (Kutas and Hillyard, 1984). For instance, 'He posted the letter without a …' is completed by nearly all participants with 'stamp'. In contrast, 'The police had never seen a man so …' allows for plenty of viable completions (Bloom and Fischler, 1980). Thus, much as Botvinick et al. (2001) took free association performance to define the amount of completions during verb generation (Thompson-Schill and Botvinick, 2006), we suggest that sentence completions can be used to address the issue of competition between pre-activated 'semantic' candidates. If the N400 was a conflict-N2, access to long-term knowledge of an unequivocal and expected word should lead to a low N400, while for an unexpected word more lexical candidates may have become active, giving rise to a greater N2/N400 complex. And indeed, the greater the amount of typical completions of a sentence fragment, the greater is the N400 (Dambacher et al., 2006; Kutas and Hillyard, 1984). The idea that competition may be a functional constituent of the N2/N400 complex is also supported by an fMRI study in which semantically anomalous words elicited greater ACC activation than expected words (Kuperberg et al., 2003).

While competition between orthographic units could be simulated quantitatively by IAMs (Hofmann et al., 2008b), however, these IAM accounts of verb generation and N400-sentence-level effects were still at a prequantitative level of theorizing (Thompson-Schill and Botvinick, 2006). However, we think that the IAM's algorithmically concrete approach to orthographic processing provided some advantages. Most importantly, no pre-experiments were required to define the units of competition. Second, a performance-free definition of associations may better account for the functional overlap in verb generation and sentence processing than defining semantics either by human free-association or cloze-completion performance (Grainger and Jacobs, 1996). Third, cloze completion probabilities may not only represent semantic processes. Rather, they may also contain morpho-syntactic factors constraining

the words to be completed, as hypothesized in the unification model (Hagoort, 2003). Fourth, a performance-based definition of "association" can serve well for proof-of-concept studies, but it can hardly reflect all potential associative relations. Horizontal generality across different stimulus sets, however, is necessary when descriptive adequacy at the item-level should be testable for any stimulus set (Perry et al., 2007; Spieler and Balota, 1997). Consider that each word pair can have a semantic relation. So checking for the potential associations between 100 words means testing 100 * 100 = 10,000 potential associations, which is practically impossible by performance-based association measures. Therefore, one big remaining challenge for IAMs was to ubiquitiously define semantic representations, which have been postulated, but not yet implemented (e.g., Coltheart et al., 2001; Perry et al., 2007).

This theoretical deficit of IAMs becomes even more apparent when considering other, less successful tests of the hypothesis that the N400 is a special case of an N2 during language processing. While IAMs predicted N2/N400 amplitudes to nonwords (Braun et al., 2006; Hofmann et al., 2008b), they failed to predict electrophysiological responses elicited by an undefined mixture of orthographic and semantic sources of information (cf. Briesemeister et al., 2009). This can be demonstrated in a stem completion task, in which participants are exposed to a word stem, such as 'com...', and are required to complete it to a whole word, such as 'completion'. Botvinick et al. (2001) predicted "that stem completion should engage the ACC more strongly when the stem presented is associated with several completions than when the stem is associated with one strongly preferred response" (Botvinick et al., 2001, p. 633). However, falsifying this hypothesis, Klonek et al. (2009) observed a stronger N400 negativity to word stems with only one possible completion as compared to word stems with multiple completions (Kutas and Federmeier, 2011). They interpreted this result by the following mixture of processes: Word stems with many completions are orthographically much more familiar, which can

well be simulated by the MROM (Grainger and Jacobs, 1996; Klonek et al., 2009); and this available orthographic information makes it easier to semantically associate appropriate completions allowing to perform the task.

In sum, conflicts at the level of orthographic representations can be quantitatively simulated in an IAM for nonwords, while effects hypothetically due to semantic representations can be addressed by exemplary associations from free association performance, and not yet by a fully quantitative approach for all possible associations. We suggest that N400 evidence can contribute to the discussion about conflicting representations, or a general role of the ACC in violated predictions (Alexander and Brown, 2011; cf. paper in this special issue), or expected reward or value (cf. 4.; Holroyd and Coles, 2002; Shenhav et al., 2013). However, as soon as multiple sources of information mix-up to generate a non-trivial decision, ERP amplitudes are hard to predict, at least by a simple lexical conflict. Which levels of representation contribute to what extent to which task performance remains under strategic control (Dilkina, McClelland and Plaut, 2010). Therefore, we agree with Taylor et al. (2012) that the activation of particular brain regions can inform about this relative reliance on a particular code. However, we think that there is a sufficient amount of falsifying evidence against a generally applicable IAM without a computational definition of the most important feature of a word: its meaning.

4. Associative Read-Out Modeling of semantic information

While the visual-orthographic processes of IAMs could account for some neurocognitive data, their predictive power was generally limited, because of their lack of implemented semantic units. Another problem was that they accounted for tasks in which memory processes contributed implicitly, but not for explicit memory performance. How to bridge the gap between modeling single-word features and semantic associations between all possible words in a hypothetical lexicon, and likewise capture major

findings of studies on explicit memory?

4.1. Quantifying semantics in IAMs

Most research on semantics used the results of pre-experiments to define word associations. For instance, in the free association task a stimulus word is presented - e.g., chair - and participants name the first words that come to mind - e.g., table, sit, legs (Jung, 1905). In an episodic memory task, another group of participants typically learns these associates and when they erroneously remembered 'table' as having been studied (Deese, 1959), this was termed the 'false memory effect' (Roediger and McDermott, 1995). Already Jung (1905) conceptualized free associations as a type of human performance elicited by cognitive and neural processes. Thus, using such performance measures (i.e., dependent variables) as an independent input variable to the model is somewhat circular, when trying to explain how neurocognitive processes leading to free associations. It also undermines a clear strength of IAMs: They provide a computational definition of cognitive processes that is independent from human task performance (Hofmann et al., 2011; McClelland and Rumelhart, 1981; Perry et al., 2007). IAMs typically use non-subjective measures of word corpora to define cognitive processing as a chain of causes that elicit responses as an effect. A second reason against using free associations for operationalizing lexico-semantics was provided by McKoon and Ratcliff (1992): They found that processing of a target word can be facilitated by a preceding prime word that is not among the first ones produced in a free association task. Rather, words that occur often together can define even weak associations (McKoon and Ratcliff, 1992). Third, such coocurrence measures can nicely account for free association performance (Rapp and Wettler, 1991). Finally, the selection of words controlled for all sorts of single-word features is limited by using ‘performance to account for performance’. Using coocurrence measures, in contrast, makes no such constraints on factorial experimental designs. Therefore, they are already widely applied in models of semantics (Andrews et al., 2009; Griffiths et

al., 2007; Landauer and Dumais, 1997; Shaoul and Westbury, 2006). These models compute latent variables to distribute the meaning representation of a word across several non-symbolic variables, much as in PDP models (e.g., Harm and Seidenberg, 2004; Seidenberg and McClelland, 1989). The IAM, however, as the prototype of a localist connectionist model uses local variables to represent one concrete real-world entity (Page, 2000), as in the examples of Figure 4.

Representing one meaningful entity by one representation variable allows to simulate human performance in a deterministic and fully transparent fashion. Thus, rather than targeting "semantics" by some abstract latent dimensions that do not refer to a single meaningful entity, local representations offer an epistemically translucent approach: During each step in the computations, users can evaluate whether the processes that act on the symbolic representations provide face validity. There are no hidden or abstract factor representations that have no meaning by themselves. This allows users to subjectively evaluate whether the temporarily activated symbolic units reflect a phenomenologically plausible experience (Ranganath, 2010), that has elicited the empirical effects. To keep the IAM's strength of fully-transparent symbolic representations, Hofmann et al. (2011) used the direct cooccurrence of two words instead of dimension-reduced latent variables as an index of association in the AROM (Figure 6). While it is also not clear whether dimension reduction provides a significant advantage in predicting human performance (Bullinaria and Levy, 2007; Griffiths et al., 2007) or brain activation patterns elicited by them (Bullinaria and Levy, 2013; Mitchell et al., 2008), this approach also saves a large portion of the computational resources required, and is therefore applicable to larger and more representative word corpora (Gamallo and Bordag, 2010). Finally, such symbolic representations can be directly related to specific brain regions: For instance, words that are associated with motor representations may be represented in motor regions of the brain in an 'embodied' fashion (Pulvermüller and Fadiga, 2010; Schrott and Jacobs, 2011; but see Mitchell et al., 2008).

Each cooccurrence measure requires a higher-order entity in which the words cooccur. Previous

cooccurrence-based approaches selected the common occurrence of words in documents (Landauer and Dumais, 1997). Following visual features, letters, and orthographic word forms, Hofmann et al. (2011) chose sentences as the critical level for defining meaning relations between words. To allow for a maximum theoretical compatibility of the AROM with models in computational linguistics, associations were defined by the log-likelihood based standard cooccurrence measure (Evert, 2005): Two words were defined 'associated' when they occur significantly more often together in a sample of 43 million sentences than to be expected from their single occurrence frequency (Quasthoff et al., 2006). This statistical approach reflects a symbolic version of Hebbian-learning to define associations: Items that occur often together are likely to be associated (Hebb, 1949).

---

insert Figure 6 about here

In the AROM, the semantic unit for a word obtains activation from its corresponding orthographic unit. As soon as a semantic unit crosses the activation threshold of zero, it excites all associated units. This excitation is scaled by a cooccurrence-based connection weight, which represents the relative strength of association between two words. Such semantic associations represent two important functions. The first is representing the meaning of a word. Though this is often referred to by Firth's quote "you shall know a word by the company it keeps" (Firth, 1957, p. 11; quoted from Andrews et al., 2009), this idea can probably be traced back even further (cf. Biemann and Riedl, 2013). According to Harris' (1951, pp. 15) 'distributional hypothesis', the meaning of a word can be defined by "the total of all environments in which it occurs". Therefore, each associate can be considered a semantic feature of the word. There can be two types of semantic relation within this structure. Either two words are associated directly, which is also termed *first-order cooccurrence*; or, two words may have many common associates, i.e. *second-order cooccurrence*. The amount of common semantic features thus can be

defined as the meaning overlap of two items. In computational linguistics, the amount of common associates is used to find paradigmatic relations (Rapp, 2002). This implies that these words can substitute each other (e.g., synonyms like 'wedding' and 'marriage'; Biemann and Riedl, 2013; de Saussure, 1959). The direct association or first-order cooccurrence, in contrast, reflects the syntagmatic relation of two words that can occur together in a sentence ('inventor' and 'fame'; Hofmann et al., 2011; Rapp, 2002). If this was the whole story, the AROM would be just another representational model of semantics – though it would be the first computational model implementing all potential associative long-term associations between symbolic units (cf. Griffiths et al., 2007). Steyvers and colleagues (2006) suggested, however, that future models of semantic memory should additionally define the cognitive processes that act on these representations. Therefore, the second equally important function of the semantic layer is to algorithmically capture the cognitive processes that occur between these units while solving a particular psycholinguistic task. This function of a process model has been the traditional role and strength of IAMs since their initial proposal. In sum, the novel level of semantic representations in the AROM extends IAMs of visual-orthographic processing by the mechanism of associative spreading of activation between semantic word units (Anderson, 1983; Collins and Loftus, 1975): If there is enough excitation from associated items or common associates in the experimental context, a word unit can become pre-activated (Hofmann et al., 2014).

4.2. False and veridical recognition in explicit memory

An initial test of the AROM's cooccurrence-based definition of association replicated McKoon and Ratcliff's (1992) finding that the association to an immediately preceding word facilitates lexical decisions (Kuchinke et al., 2010; Hofmann et al., in prep.; Lucas, 2000). Thus the AROM can account for semantic priming effects in an implicit memory task, which is the classic domain of IAMs: They

were successfully used to account for perceptual identification, lexical decision, naming, stem completion or even natural reading (Grainger and Jacobs, 1996; Klonek et al., 2009; McClelland and Rumelhart, 1981; Perry et al., 2007, 2010; Reilly and Radach, 2006). Because IAMs were not able to account for explicit memory, however, the AROM was designed to answer the challenge set by Gallo's (2006) exhaustive overview of the false memory literature: He suggested that the major challenge consists of a theory that bridges both semantic priming in implicit and false memories in explicit memory tasks (Chen et al., 2008) – a challenge that would best be taken by a process model (Gallo, 2010). In our initial approach, we included one associative unit for each word appearing in the context of a study-test recognition memory task (Hofmann et al., 2011). While an increased resting level activation at cycle 0 represented the experimental manipulation that an item had been learned at study, the activation in cycle 1 becomes larger for items with many strong semantic associations to the rest of the stimulus set. Though associative interactions between context words can occur at this time, bottom-up information from the feature layer feeds to the letter layer in cycle 2, to the orthographic word layer in cycle 3, and reaches the associative layer in cycle 4 (Figure 7). From thereon, the presented item interacts with the associations to the stimuli available from the context. Therefore, the AROM predicts that the more associated items are in the context, the stronger becomes the activation of the presented item, thus accounting for the false memory effect: Associated items in the context elicit false responses to non-studied items. For studied items, in contrast, the model also successfully predicts a memory boost that is elicited by many associates in the stimulus set (Hofmann et al., 2011) – a finding that was reproduced with other stimuli (Kuchinke et al., 2013).

Since predicted amount of item-level performance is a major evaluation criterion of descriptive adequacy (Spieler and Balota, 1997; Perry et al., 2007), we also tested the AROM's ability to simulate the recognition probabilities of the individual items. In studied words, the associative layer's activations accounted for 10% of the variance in yes-response probabilities of a recognition memory task, while in

non-studied words, the AROM accounted for 14% of the false memory variance. Thus, in studied and in non-studied words, the AROM can predict which word is remembered with which probability. Moreover, the activations in the orthographic layer accounted for an additional 14% of unique variance. For simplicity, the AROM was implemented without top-down excitation from the semantic to the orthographic word layer (see Jacobs and Grainger, 1992). Therefore, the lower three layers of the AROM consisted of an unchanged MROM (Grainger and Jacobs, 1996). Since we also showed that the MROM can predict human performance in an episodic memory task, the horizontally more general expansion of the MROM is confirmed for exactly those processes for which it was designed: Predicting word recognition to non-re-studied words (Grainger and Jacobs, 1996). In sum, the AROM is the first computational model that can answer Roediger, Balota, and Watson's (2001) challenge to predict false memories by both, orthographic *and* semantic similarities to the other words in the context (cf. Hofmann et al., 2011; Jacobs and Grainger, 1996; Perry et al., 2007, 2010).

---

insert Figure 7 here

Finally, localist representations allow to test whether the AROM's associative processes are plausible at an intuitive level of phenomenological analysis (Ranganath, 2010). Figure 7 shows exemplary association functions for the word 'wedding' supporting the AROM’s face validity (Hofmann et al., 2011). The AROM suggests that this word elicited more false memories because of strongly co-activated items such as 'marriage', 'throne', and 'widow'. Such associates increase the activation of the non-studied item 'wedding', and therefore increase the probability for a false memory of this word. Moreover, 'marriage' and 'wedding' can be considered as synonyms. This suggests that the most strongly co-activated items have a tendency to reflect a semantic relation to the presented items in the narrowest sense of the word.

4.3. Familiarity and recollection in IAMs

Yonelinas (1994) determined familiarity by assuming that normally distributed memory signals of equal variance are apparent in studied and non-studied items, while a high-threshold signal increases the 'yes' response probability in studied items due to recollection. Because Receiver Operating Characteristics (ROC) contrast the distribution of 'yes'-response probabilities for non-studied items on the x-axis with those of studied items on the y-axis, the slope of the z-transformed ROC (z-ROC) indexes the relative contribution of familiarity and recollection: Pure familiarity can be characterized by a slope of 1, while additional recollection should tilt the z-ROC's slope down (Hofmann et al., 2011; Jacobs et al., 2003, Yonelinas, 1994). The alternative unequal-variance model assumes Gaussian distributions with greater signal variance for studied in comparison to non-studied items (Glanzer et al., 1999; Green and Swets, 1966). However, recent trends in neurocognitive theorizing suggest that such measurement models and their relation to neural observables is of limited use (Malmberg, 2008; Ranganath, 2010). Rather than only measuring them, neurocognitive models should better predict which stimuli and experimental factors elicit which cognitive processes, which in turn cause familiarity-like or recollection-like data patterns. Although additional model-to-data connections allow to relate the behavioral measures of recollection and familiarity to neurocognitive data (Kuchinke et al., 2013; Wixted and Mickes, 2013, Yonelinas et al., 2005), process models such as IAMs can more easily be falsified than both of these measurement models (Jacobs and Grainger, 1994; Malmberg, 2008). When aiming to constrain the notions of familiarity and recollection by processing assumptions, Gillund and Shiffrin's (1984) 'Search of Associative Memory Model' proposed that familiarity can be defined as the sum of evidences of all memory trace activations. In accordance with this, familiarity in the MROM was determined as the sum of all orthographic word unit activations (Jacobs et al., 2003). To design a task in which lexical decisions are based on familiarity only, Jacobs et al. (2003) introduced a process-purity assumption (cf. Wixted, 2007). They assumed that very short stimulus

exposures make the full recollection of a particular stimulus very unlikely. This was confirmed by the resulting z-ROC slope of 1. The defining feature of familiarity was simulated by the MROM: An increase in familiarity for words as compared to nonwords indeed increased the total familiarity signal strength level, but not its variability (Glanzer et al., 1999; Paap et al., 1999; Yonelinas, 1994). Jacobs et al. (2003) thus defined the experimental conditions under which familiarity-like responses can be modeled as a response to particular stimuli, rather than just measuring its contribution.

That familiarity plays a role in lexical decision tasks was also proposed by Yonelinas (2002, p. 445), who suggested that "familiarity is assumed to support not only recognition memory performance, but also performance on implicit memory tasks". As "familiarity and recollection (...) support memory for perceptual and semantic (or meaning-based) information, respectively" (Yonelinas, 2002; p. 444), we suggested that activation in the semantic layer elicits a recollection-like data pattern (Hofmann et al., 2011). Gillund and Shiffrin (1984, p. 55; Mandler, 1980) simulated the recollection-like search process by the associatively cueable activation of a single memory trace (Yonelinas, 2002, p. 445). In accordance with Gillund an Shiffrin (1984), the AROM assumes that the recollection-inducing source of information is the semantic activation of a single item.

For bridging the gap between process models in which memory is only required implicitly (Berry et al., 2008) and models of explicit memory, however, an additional assumption was required that decides whether an item was studied or not. This assumption was borrowed from the unequal variance model (Glanzer et al., 1999). To describe a z-ROC, this measurement model made two ad-hoc assumptions (e.g., Green and Swets, 1966): First, the memory signal is greater for studied than for non-studied items, and, second, the memory signal variance is greater for studied words.

When implementing this first assumption into the semantic layer, the second ad-hoc assumption of unequal variances automatically resulted (Hofmann et al., 2011). Irrespective of the choice of the free parameters, the semantic activation variance was greater for studied than for non-studied items. It was

then discovered why the AROM's IAM-based architecture was capable of predicting unequal item variances without additional assumptions (McClelland and Chappel, 1998; Shiffrin and Steyvers, 1997): This is demonstrated in Figure 8 and Appendix A.

—-

insert Figure 8 here

—-

The AROM provides the advantage of a falsifiable mechanism that predicts particular z-ROC slopes dependent on task and stimulus features, in contrast to Glanzer et al.s, (1999) and Yonelinas' (1994) models (Malmberg, 2008). For instance, the assumption of stronger memory signals for studied items can be used to test the rationale of the model: The stronger the memory signal of the studied items is, the greater should be the variance when compared to the weaker memory signals of non-studied items. Thus, when we assume that longer study time further increases the memory signal of studied items, even greater variances should result. This assumption was confirmed by Glanzer et al. (1999), who repeatedly showed that the slope of the z-ROC was even lower when words were studied for a longer time (Shiffrin and Steyvers, 1997; McClelland and Chappell, 1998). A stronger initial resting level activation and the resulting greater memory signal variance for longer-studied items can explain this: Greater episodic pre-activation increases the variances of particular memory traces that are recollected from semantic long-term memory.

In sum, the AROM can define the experimental conditions and computational mechanisms that lead to the data patterns of familiarity and recollection, rather than just measuring them. Therefore, the AROM can more easily be falsified than measurement models such as the model of recollection and familiarity or the unequal variance model.

4.4. Semantic processes in the temporal lobe

The semantic layer of the AROM represents associations in a long-term memory structure. Complementary learning systems theory suggests that such outlearned associations are distributed across the cerebral cortex and therefore can not always be associated with a particular neural region (McClelland et al., 1995). A general issue, however, is the embodiment proposal that words are represented in those neural regions, which also represent their referent (Pulvermüller and Fadiga, 2010). Thus, for instance, an arm-related word may activate the same brain region in the sensorimotor cortex that is also activated by arm movements. However, these embodied constituents of word meaning usually form a network with temporal cortex regions, where semantic information converges (Pulvermüller and Fadiga, 2010).

For instance, lesion evidence suggests that particularly the posterior medial and inferior temporal cortex represent semantic knowledge (Binder et al., 2009). Thus, semantic activation may lead to a temporally increased neural activation in the posterior middle temporal gyrus (Hofmann et al., 2009). A lower BOLD response to associatively primed items can be explained by the fact that these are recognized faster (Rossell et al., 2003). Thus, when summing up the required neural energy over the whole time frame of recognition, the energy demand is lower (section 3.1.). In addition, when semantic activation crosses a response criterion early, the associatively triggered search is also terminated early. Therefore, fewer other associates become activated. When a presented word matches an associative prediction, reduced temporal cortex activation may index lower semantic competition, which may also contribute to a decrease of the N400 (Lau et al., 2008; section 3.2).

But also the anterior temporal pole seems to be engaged in processing semantic representations. Patterson, Nestor und Rogers (2007) propose this region to be a hub that integrates the semantic features of a concept. This is supported by results from intracranial studies that point at the role of anterior temporal regions responding to semantically unexpected words during lexical decision and in

sentence context (Bohrn et al., 2012; Nobre and McCarthy, 1995; McCarthy et al., 1995). In terms of the AROM: if the unit corresponding to an expected word was activated by a previously presented associated one, temporal cortex activation drops with the amount of semantic pre-activation (cf. Rossell et al., 2003; Wible et al., 2006).

Complementary learning systems theory also proposes a second associative mechanism, which can be found deep inside the medial part of the temporal lobe and particularly the hippocampus (McClelland et al., 1995). In 'recall mode', the hippocampus may re-activate consolidated long-term knowledge (Kumaran and McClelland, 2012). Therefore, greater hippocampal activation in words with a greater amount of long-term associations to the items in the episodic context can be expected. Kuchinke et al., (2013) confirmed this assumption: Hippocampus activation was greater for items with many associated words in the stimulus set of an episodic memory task. They also tested the proposal of Yonelinas et al. (2005) that recollection leads to hippocampus activation, which can also be explained by strong memories eliciting a high confidence that the item was studied (Squire et al., 2007). This prediction was confirmed for words with a high amount of associations in the stimulus set, but not for items with only a few associations to the stimulus set. In this case, hippocampus activation was equally high for items eliciting a no-memory decision of high confidence that the item was not studied (Kuchinke et al., 2013). This finding is difficult to reconcile with the Glanzer et al. (1999) and Yonelinas et al. (2005; Squire et al., 2007).

A complementary learning systems approach to hippocampus function, on the other hand, likely suggests that in surely non-studied items with few associations, novel associative representations are generated in the hippocampus (Kumaran and McClelland, 2012). While greater hippocampus activation in words with many associations suggests that the AROM can be considered a special case of this hippocampus model in the 'recall mode' (Kuchinke et al., 2013), it can not learn novel representations or their connections. Extending the AROM by Kumaran and McClelland's (2012) fully localist

conjunction layer, however, would enable it to do just this, thus implementing a neurobiologically plausible model feature, given that new neurons can be generated in the hippocampus (Eriksson et al., 1998).

Eichenbaum, Yonelinas und Ranganath (2007) proposed that the hippocampus binds perirhinal item-information to the context represented in the parahippocampal gyrus. Because Kuchinke et al.'s (2013) result of greater activation to associatively wired words in the hippocampus was accompanied by stronger parahippocampal activation, this also supports the AROM which proposes that false memories are elicited by associations to the items in the experimental context.

In sum, different regions in the temporal lobes may represent different recency levels of information with an inside-to-outside temporal gradient: While semantic-associative long-term knowledge is represented in the outer cerebral cortex (e.g., McClelland et al., 1995), parahippocampal regions may represent an episodic context and thus temporally more adjacent events (Eichenbaum et al., 2007). The hippocampus, however, either re-activates this knowledge, or it generates new representations that may wander to the outer regions when often re-activated in time (Kumaran and McClelland, 2012).

4.5. Association strength predicts left inferior frontal gyrus activation

For presently activated memory traces, Thompson-Schill et al.'s (1997) lexical selection hypothesis states that the primary function of the left inferior frontal gyrus (IFG) is the selection of an appropriate semantic representation from multiple, pre-activated long-term representations. The more representations are active, the larger is semantic competition and thus IFG activation (cf. 2.1; Thompson-Schill and Botvinick, 2006; but see Martin and Byrne, 2006). This can account for the effect of word frequency in the IFG (e.g., Fiebach et al., 2002; Hofmann et al., 2008a), because low-frequency words are identified more equivocally. Thus, selection demands would be higher when many units are active and in competition. Since we mentioned above that previous IAMs failed to predict

brain responses to word stimuli (3.1) because of the lack of semantic representations, here we would like to test whether IFG activation can reliably inform about the activity in the theoretically postulated semantic layer (Hofmann et al., 2011). We assume that the weaker association strength between two words is, the greater should be semantic competition and thus IFG activation (Wagner et al., 2001). To test this hypothesis, we performed a re-analysis of recent neuroimaging data by Forgács et al. (2012; cf. Appendix B for details). In this study, participants identified noun-noun compound words and indicated whether they are familiar or not. The concurrent presentation of two words was optimally suitable to address the influence of each single association. In particular, we tested whether the AROM's association strength parameter of the 48 associated noun-noun compounds was a reliable predictor of IFG activation (cf., Hofmann et al., 2011, formula 2). This prediction was confirmed: Even after a full Bonferroni correction for more than 90.000 statistically independent samples, 109 IFG-voxels surpassed a significance threshold of $p < 0.05$ ($t(39) > 5.95$), and the most significant voxel was significant at a full Bonferroni-corrected level of $P < 0.005$ ($t(39) = 8.23$; Talairach x/y/z = -40/22/-5; cf. Figure 9).

---

insert Figure 9 about here

In sum, the AROM responds successfully to the challenge of providing an "explicit framework in which the effects of manipulations such as association strength can be (...) formally assessed." (Thompson-Schill and Botvinick, 2006, p. 402). While a previous computational model did so for experimentally induced associations (Danker et al., 2008), the AROM allows to define long-term associations between each possible word pair in a steadily growing amount of languages (230 at present), because the measures used by the AROM were taken by the multilingual Leipzig Wortschatz Project (http://corpora.informatik.uni-leipzig.de; Quasthoff et al., 2006). The evaluation of the

descriptive adequacy of computational models at the item-level became a major criterion in 1997 for behavioral data (Spieler and Balota, 1997; Perry et al., 2007), and was applied to ERP data in 2008 (Hofmann et al., 2008b; Rey et al., 2009). Finally, the present paper shows that fine-grained quantitative predictions at the item-level are also possible for fMRI data.

4.6. Functional, neurobiological and phenomenological analyses of brain connectivity

The prediction of IFG data by parameters of the AROM appears to be based on the idea of cognitive modules in the brain (Anderson et al., 2004; Fodor, 1983; Newell, 1990). Since the IAM is based on the principle of interactivity, however, the reverse is the case. The IAM layers are connected and the excitation and inhibition parameters were often ‘handwired' to account for behavioral data (McClelland and Rumelhart, 1981). This leads to a continuous exchange of information between the different levels of representation. Since the activation of the IAM layers can be associated with regions in the ventral visual stream (section 3.1), and the associative layer can be associated with temporal gyrus and the IFG (sections 4.4. and 4.5.), these information exchange parameters can be taken as a theoretical proposals of connectivity between these brain regions.

We propose that data-driven neuro-functional connectivity analyses provide evidence that allows to test for the appropriateness of the relative size of these parameters. For instance, when McClelland and Rumelhart (1981) simulated performance in the perceptual identification task, they proposed that the orthographic word layer delivers greater top-down excitation (0.3) than it receives bottom-up excitation (0.07). During word recognition, in contrast, Jacobs and Grainger (1992) found that lowering the top-down excitation from the orthographic word layer does equally well account for human performance. When testing this hypothesis by connectivity analyses of brain regions, several dynamical causal modeling (DCM; Friston et al., 2003) studies show that the bottom-up excitation of the fusiform gyrus is stronger than its top-down excitation during word recognition (Richardson et al., 2011; Schurz et al.,

2013). Therefore, they support the initial proposal of Jacobs and Grainger (1992), which was based on IAM simulations of behavioral data. To our knowledge, however, there is no study that would test McClelland and Rumelhart's proposal (1981). If their explanation of the word superiority effect was correct, we would expect a stronger top-down connectivity to words than to perceptual identification of letters in nonwords. Such an index of the amount of top-down excitation should also be lower during lexical decision than during perceptual identification – as proposed by Jacobs and Grainger (1992). While for neuroimaging, DCMs or Granger causality are two methods that provide empirical data allowing to evaluate parametric connectivity assumptions between IAM layers (Friston et al., 2003; Goebel et al., 2003; Smith, 2012; Woolrich and Stephan, 2013), the (lagged) phase synchrony of the electrophysiological responses of two brain regions may inform about the relative weight of connectivity between two representation layers (Bar et al., 2006; Fell and Axmacher, 2011; Jacobs and Carr, 1995; Lachaux et al., 1999). However, we do not know any study that addresses such IAM predictions by EEG data yet.

While previous approaches to the cognitive function of information exchange between differential representation layers were constrained to behavioral data, data-driven connectivity methods offer the possibility to test for the neurobiological plausibility of an IAM's theoretically proposed connectivity assumptions at the level of brain regions. At the level of single neurons, in contrast, the symbolic representations of IAMs seem to have a limited neurobiological plausibility (Bowers, 2009, 2010; Plaut and McClelland, 2010; Quiroga and Kreiman, 2010): there is probably no single neuron that exclusively codes the appearance of a single symbol such as a grandmother. On the other hand, IAMs allow to deterministically compute the consequences of a stimulus set presented to the model. These can be compared to experimental data. Therefore, they offer not only the advantage of a testable model-to-behavioral-data connection (Grainger and Jacobs, 1998), but also a model-to-brain-data connection (Jacobs and Hofmann, 2013). As a consequence, the neurobiological plausibility can be addressed in

terms of decriptive adequacy (4.5). In addition, symbolic representations allow to likewise test for the phenomenological plausibility (Figure 7, Ranganath, 2010), i.e. is it plausible that 'marriage' is more likely to be remembered erroneously, because 'wedding' was presented earlier?

In sum, we agree with Ranganath's (2010) proposal that the explanatory adequacy of a model can be critically evaluated at three levels of *plausibility*: the *functional*, e.g. by implementing the cognitive mechanisms of information exchange across differential representations, the *neurobiological*, e.g. by using this for testing computational predictions about the activation and connectivity of brain regions, and the *phenomenological*, by simulating which word units elicit associative activation in which other word units.

5. Can semantic cohesion account for emotional valence effects?

Maratos et al. (2000) proposed that much of the behavioral and neurophysiological variance accounted for by affective word features can actually be explained by the word's higher semantic-associative cohesiveness. This hypothesis can be traced back to a conference proceeding by Phelps and LaBar (1997), showing "that effects of valence on memory can be eliminated by varying the inter-item associations" (p. 534; Phelps et al., 1998). This hypothesis is testable by using the AROM. An affectively positive word like 'wedding' – when presented to the AROM – often elicits the strongest co-activation in word representations like 'marriage', which are semantically close to the stimulus (Figure 7). But do affectively loaded words have a larger amount of semantically related items in terms of the AROM? To answer that question, we checked whether there is a relationship between the valence of a word and its number of associates (NOA) in the 2901 unique word forms of the Berlin Affective Word List – Reloaded (Võ et al., 2009).

To examine whether a confound between valence and semantic cohesion can make the AROM explain

valence effects, we tested whether NOA can predict the absolute value of the emotionality of words, irrespective of its positive or negative sign (Kuchinke, 2007). We found a very small but significant influence ($R^2 = 0.0035$; $F = 10.23$; $P = 0.0014$; emotionality = 1,01 + 0,00039* NOA), which provides only moderate evidence for the semantic cohesiveness hypothesis in its original formulation. Testing whether signed emotional valence can be predicted by semantic cohesion revealed a somewhat stronger effect ($R^2 = 0.063$; $P < 0.001$; emotional valence = -0,315 + 0,0030 * NOA; Figure 10). Overall, these small effects make it unlikely that NOA in terms of the AROM can account for all of the variance accounted for by affective word features (e.g., Briesemeister et al., 2011a,b; 2012; 2014; Citron, 2012; Kuchinke et al., 2005; Recio et al., in press). The strongest relationship between semantic cohesion and signed valence was rather unexpected: The more positive a word is, the more associated items it has.

---

insert Figure 10 about here

In contrast, Maratos et al. (2000) found support for the semantic cohesiveness hypothesis of affective word processing in their ERP findings for negative words that revealed a remarkable similarity to false memory effects in a recognition memory task. Critically, they found the same increase in false alarm rate for negative words as for false memories (4.3). In a subsequent study, Windmann and Kutas (2001) controlled for semantic cohesion, and still found an increase of false alarm rate to negative words, thus rendering the explanation unlikely that negative valence variance can be totally absorbed by semantic cohesion. However, in Talmi and Moscovitch (2004), participants learned negative and neutral words, and an additional category of neutral words with a high semantic cohesion to the stimulus set. They found that both, semantic cohesion in neutral words, as well as negative valence increased false recall – while the neutral cohesive and the negative words did not differ from each other. They suggested that

associations play a similar role as the classic valence effect in memory tasks. Using the same stimulus categories in a recognition memory task, McNeely et al. (2004) found no increase in false alarm rate to neutral words of high semantic cohesion, but an increase for negative valence. Moreover, their ERP findings suggested differences between semantic cohesion in neutral and negative words. While arousal was not taken into account in this study, valence and arousal were confounded in Talmi and Moscovitch (2004). Therefore, it is questionable whether the observed effects were due to negative valence itself or to the confounded manipulation of arousal (Bayer et al., 2012; Hofmann et al., 2009; Maratos et al., 2000; Windmann and Kutas, 2001). None of these studies, however, included a control category of negative words of low semantic cohesion.

To test whether semantic cohesion and negative valence provide independent contribution to false memories, we next present the results of a study-test recognition memory experiment crossing the experimental factors of negative valence (neutral/negative), and NOA in the stimulus set (low/high; cf. Appendix C for experimental details), while keeping all sorts of single-word features constant, such as arousal, imageability, word length, number of orthographic neighbors, as well as letter, bigram, and word frequency.

---

insert Figure 11

We found significant effects of negative valence ($F(1,28) = 46.06$; $P < 0.001$), and NOA on false memories ($F(1,28) = 18.08$; $P < 0.001$; Figure 11). Moreover, both factors yielded a significant interaction ($F(1,28) = 5.92$; $P < 0.05$).

Since both factors increased false alarm rate, this suggests that both semantic cohesion and negative valence can cause false memories. Such negative valence effects occur even in the absence of a confounded manipulation of arousal (Windmann and Kutas 2001), and the semantic cohesion main

effect replicates the result that the AROM can predict false memory effects (Hofmann et al., 2011). The interaction, however, could best be explained in terms of the interplay between bottom-up driven affective evaluation (Kuchinke, 2007), and a top-down-driven activation boost from associations in the language context (Hofmann et al., 2011).

To further investigate a hypothesis suggested by our surprising corpus analysis finding - i.e. a relationship between positive valence and semantic cohesion – the experimental factors of positive valence (neutral/positive) and NOA (low/high) were crossed in a second recognition memory experiment (N = 34; cf. Appendix C for method details).

---

insert Figure 12

The NOA ($F(1,34) = 6.22$; $P < 0.05$), but not positive valence, increased false alarm rate ($F = 1.00$; Hofmann et al., 2011; Figure 13), and there was a nonsignificant trend towards an interaction between valence and NOA ($F(1,33)=3.51$; $P = 0.07$). In this experiment, a strong version of the semantic cohesiveness hypothesis was straightforwardly confirmed for positive words (Maratos et al., 2000; Talmi and Moscovitch, 2004): As positive valence and semantic cohesion are typically confounded, positive valence effects can thus be explained by semantic cohesion only. The stronger the associative-semantic connectivity to the other words of the experiment, the more likely was the erroneous recognition of a word as having been studied. False alarm rate did not further increase as a function of positive emotional valence, suggesting that the AROM offers a parsimonious explanation of effects of positive valence - without the need of an additional affective evaluation mechanism (Kuchinke, 2007). In sum, our corpus analysis provided no evidence in favor of the possible confound of negative valence with (a larger) amount of inter-item associations, at least not for the German language. Rather, our recognition memory findings suggest that negative valence engages an evaluative mechanism that

affects word recognition (Kuchinke, 2007); and mere semantic cohesion does not appear to be sufficient to explain false memories attributed to negative valence manipulations (Windmann and Kutas, 2001).

Effects previously ascribed to positive word valence, in contrast, may have actually been due to a confound between positive valence and associations: First, there is a confound between positive valence and the NOA; second, false memory effects due to positive valence can more parsimoniously be described by inter-item associations in the AROM (see 4.; (e.g., Kuchinke et al., 2006). Thus, not only the horizontal generality of the AROM was increased by showing that semantic cohesion can account for an effect of positive valence, but also its explanatory adequacy, by accounting for an additional effect by the same model mechanism of semantic cohesion.

6. Conclusions

The preceding analyses allow us to conclude that IAM-type models can favorably be evaluated by the criteria of *descriptive adequacy*, *generality*, *falsifiability and simplicity*, and that IAMs can *adequately explain* numerous phenomena at the *functional*, *phenomenological,* and *neurobiological* levels. The IAM layers of visual features, letters and orthographic word forms appear sufficient for simulating activations in the ventral visual stream from the occipital to the left posterior fusiform regions in implicit memory tasks. Competition between orthographic units can account for N2/N400 effects, while semantic competition effects can be addressed by cloze completion probabilities or free association performance. It is questionable, however, whether one can really 'explain' human performance by other human performance measures. The AROM remedies the disadvantages of previous IAMs due to the lack of a semantic layer by providing a convenient modeling framework that integrates implicit and explicit memory. In recognition memory tasks, it accounts for orthographic and

semantic influences on false memories. While familiarity was defined at the orthographic level, recollection likely resides in the semantic layer, where greater memory signals to studied items elicit greater signal variances than non-studied items, i.e. a z-ROC slope < 1. The AROM's long-term associations can be related to neurocognitive data elicited by the temporal cortex. Contextual associations can be related to the parahippocampus, which facilitates hippocampal associative processing in recall mode. When two words are presented together in a noun-noun compound, the AROM's association strength between the nouns further allows for item-level predictions in the IFG. Finally, false memory effects were replicated in two experiments in which negative and positive word valence was manipulated. While negative valence provided an additional increase in false memories, effects of positive valence on false memories could be explained by an increased number of associations in positive words.

Acknowledgments

Parts of this paper have been similarly outlined in the dissertation of the first author (Hofmann, 2011). This paper has been supported by grants from the Deutsche Forschungsgemeinschaft to Arthur Jacobs (JA 823/4-1 and 4-2) and to Markus Hofmann (HO 5139/2-1), as well as to the cluster of excellence "languages of emotion". We like to cordially thank an anonymous reviewer, e.g., for inspiring the focus on model evaluation.

Appendix A

Hofmann et al. (2011) implemented the first ad-hoc assumption of increased memory signals for studied items into the semantic layer. For this purpose, the initial *resting level activation a* for *studied* items *s* was increased in comparison to *non-studied* items *n* in the semantic layer:

$a_s > a_n$ (1)

To be neurobiologically plausible, each connectionist model requires an assumption of nonlinearity that scales activation between a maximum activation of 1 and a *minimum activation* of $m = -1$ (Grossberg, 1978). This assumption is neurobiologically plausible, because physical limitations of neurons that cause a maximum and minimum firing rate (McClelland, 1993), and implies that the summed excitatory and inhibitory input can not be translated into the activation change of a unit in a linear fashion. Rather, *net-input n* and *effect e* (activation change) must be described by a nonlinear, typically sigmoid activation function (McClelland, 1993). In the IAM, nonlinearity is implemented in a simple manner. If the excitatory and inhibitory activation change sum up to a negative inhibitory signal, the activation change of the unit can be calculated by the following formula (McClelland and Rumelhart, 1981, p. 381):

$$e = n * (a - m) \qquad (2)$$

After the learning phase of a recognition memory task, many items are active, i.e. many word units strive for activation, because they 'want' to be remembered. This causes a great amount of inhibition for each unit. Therefore, each unit gains an inhibitory net input, and thus formula 2 applies. To calculate one *distribution of all effect changes E* for all the *studied* items E_s and one distribution for all *non-studied items* E_n, this formula has to be applied repeatedly. For the simplicity of this demonstration, we keep the distribution of the net inputs N constant for non-studied and studied items.

$$E_s = N * (a_l - m); \text{ and } E_n = N * (a_n - m) \qquad (3)$$

As a result, a greater initial memory signal a_s for studied items than for non-studied items a_n (1) will automatically scale up the variance *var()* of the effect changes E_s for studied as compared to non-studied items E_n.

$$\text{var}(E_s) > \text{var}(E_n) \qquad (4)$$

It follows that the greater the resting-level difference between the two distributions, the greater is the variability of the studied as compared to the non-studied item activation variance. The resting level implements the memory signal strength increase to learning in the study phase.

Appendix B

To determine whether the AROM's association strength can predict IFG activation, we selected the 48 compounds from Forgács et al. (2012) for which both nouns provided a significant cooccurrence in the German Leipzig Wortschatz Corpus (taken from Hofmann et al., 2011; Quasthoff et al., 2006). Moreover, all compounds existed as commonly used word forms in the corpus. In Forgács et al. (2012), they were taken from the experimental conditions of literal combinations (e.g., 'Taxifahrer'; i.e. taxi driver), or dead metaphors (e.g., 'Erfolgsrezept'; i.e. recipe for success). Please confer Forgasz et al. (2012) for further details. In the analysis provided in Figure 9, we used Brain Voyager to test whether the association strength of the AROM provides a significant predictor of IFG activation.

Appendix C

In the experiment crossing **negative valence** (neutral/negative) and **NOA** (low/high), 29 participants took part (average age = 25.62; SE = 0.89; 17 female). All had corrected or corrected-to-normal vision, had no known language disorders, were native speakers of German, received course credits or were paid. 128 words were learned in a study phase, which had to be discriminated from the 128 non-studied target words in a test phase. Three non-associated words were presented before and after the target stimuli in the study phase to prevent primacy and recency effects. In both phases, stimuli were presented in a random sequence and preceded by five non-associated practice stimuli. In the study and the test phase, a 500-ms fixation cross was followed by the stimulus for 1500 ms. In the study phase, five hashmarks appeared until a response was given. In the test phase, a 6-point rating scale appeared at

which 1 to 3 indicated 'sure non-studied' to 'unsure non-studied', and 4 to 6 indicated 'unsure studied' to 'studied'. The former and latter categories were collapsed to 'no' and 'yes' responses to analyze the recognition memory decision, while this assignment was reversed in a random amount of participants. The four stimulus categories consisted of 32 stimuli, respectively, and did not differ with respect to arousal, imageability, word length, number of orthographic neighbors, mean lemma-based letter and bigram frequency (token), and frequency class (Fs < 1; Hofmann et al., 2007; Quasthoff et al., 2006; Võ et al., 2009). High NOA words had at least 5 associates in the stimulus set, and low NOA words had 4 or less. Negative words had a valence score lower than -0.5, while neutral words ranged between 0.5 and -0.5. There was no significant valence difference between low (mean = -1.02; SE = 0.07) and high NOA (mean = -1.09; SE = 0.09; t < 1) *negative* words, but the number of associates differed (mean = 2.5; SE = 0.23 vs. mean = 7.75 ; SE = 0.5; t = 9.52, P < 0.001). Likewise, there was no significant valence difference between low (mean =-0.05 ; SE = 0.05) and high NOA (mean = -0.06 ; SE = 0.05 ; t < 1) *neutral* words, but NOA differed (mean = 2.13; SE = 0.23 ; vs. mean = 8.75 ; SE = 0.55; t = 11.13 , P < 0.001). Moreover, valence differend significantly between the neutral/negative cells within both NOA categories (both ts > 10; Ps < 0.001).

In the experiment crossing **positive valence** (neutral/positive) with **NOA** (low/high), 34 native speaking Germans participated (mean age: 28.16; SE = 1.84; 20 female). They had no known reading disorder, normal or corrected-to-normal sight, and participated voluntarily, received course credits, or were paid. In all, 216 stimuli were presented, while a quarter of these were contained in each condition. For each participant, half of these 54 stimuli were chosen randomly for the study phase. The remaining 27 served as non-studied target words in each stimulus category. Pupillary responses were concurrently measured using a video-based IView X Hi-Speed eyetracker (SensoMotoric Instruments, Germany), which revealed no significant effects. Each trial started with the presentation of a fixation cross that remained until pupil dilation varied for less than 0.01mm for 150, but maximally for 3 seconds. Stimuli

were presented for 1500 ms. In the study phase, five hashmarks appeared until a response was given. In the test phase, stimulus presentation was followed by a blank screen of 1500 ms. Participants were required to make a binary yes/no recognition decision within the 3000 ms after stimulus exposure that was used for the analyses. It was followed by a rating scale (cf. previous experiment). Before each phase, five practice stimuli were presented to familiarize the participants with the task. The four stimulus categories did not differ with respect to arousal, imageability, word length, number of orthographic neighbors, mean lemma-based letter and bigram frequency (token), and frequency class (Fs < 1; Hofmann et al., 2007; Quasthoff et al., 2006; Võ et al., 2009). High NOA words had at least 17 associates in the stimulus set, and low NOA words had 16 or less. Neutral words had a valence score 0.5 or lower, while positive words had a valence of at least 1. There was no significant valence difference between low (mean = 1.57; SE = 0.06) and high NOA (mean = 1.52; SE = 0.06; t < 1) *positive* words, but the number of associates differed (mean = 11.44; SE = 0.41; vs. mean = 23.87 ; SE = 0.95; t = 12.07, P < 0.001). Likewise, there was no significant valence difference between low (mean = 0.07; SE = 0.04) and high NOA (mean = 0.10 ; SE = 0.04 ; t < 1) *neutral* words, but NOA differed (mean = 11.35; SE = 0.48; vs. mean = 23.89; SE = 0.82; t = 13.27, P < 0.001). Moreover, valence differed significantly between the neutral/negative cells within both NOA categories (both ts > 10; Ps < 0.001). In both experiments, approximately 10 minutes passed between the study and test phases.

References

Alexander, W.H., Brown, J.W., 2011. Medial prefrontal cortex as an action-outcome predictor. Nat. Neurosci. 14, 1338–44.

Anderson, J.A., Silverstein, J.W., Ritz, S.A., Jones, R.S., 1977. Distinctive Features, Categorical Perception, and Probability Learning: Some Applications of a Neural Model. Psychol. Rev. 84,

413–451.

Anderson, J.R., 1983. A spreading activation theory of memory. J. Verb. Learn. Verb. Behav. 22, 261–295.

Anderson, J.R., Bothell, D., Byrne, M.D., Douglass, S., Lebiere, C., Qin, Y., 2004. An integrated theory of the mind. Psychol. Rev. 111, 1036–60.

Andrews, M., Vigliocco, G., Vinson, D., 2009. Integrating experiential and distributional data to learn semantic representations. Psychol. Rev. 116, 463–98.

Assadollahi, R., Pulvermüller, F., 2001. Neuromagnetic evidence for early access to cognitive representations. Neuroreport 12, 207–13.

Bar, M., Kassam, K.S., Ghuman, A. S., Boshyan, J., Schmid, A.M., Schmidt, A M., Dale, A.M., Hämäläinen, M.S., Marinkovic, K., Schacter, D.L., Rosen, B.R., Halgren, E., 2006. Top-down facilitation of visual recognition. Proc. Natl. Acad. Sci. U.S.A. 103, 449–54.

Barber, H.A., Kutas, M., 2007. Interplay between computational models and cognitive electrophysiology in visual word recognition. Brain Res. Rev. 53, 98–123.

Bayer, M., Sommer, W., Schacht, A., 2012. P1 and beyond: functional separation of multiple emotion effects in word recognition. Psychophysiol. 49, 959–69.

Berry, C.J., Shanks, D.R., Henson, R.N.A., 2008. A unitary signal-detection model of implicit and explicit memory. Trends Cogn. Sci. 12, 367–73.

Biemann, C., Riedl, M., 2013. Text : now in 2D ! A framework for lexical expansion with contextual similarity. Journal Lang. Model. 1, 55–95.

Binder, J.R., Desai, R.H., Graves, W.W., Conant, L.L., 2009. Where is the semantic system? A critical

review and meta-analysis of 120 functional neuroimaging studies. Cereb. Cortex 19, 2767–96.

Binder, J.R., Mckiernan, K.A., Parsons, M.E., Westbury, C.F., Possing, E.T., Kaufman, J.N., Buchanan, L., 2003. Neural Correlates of Lexical Access during Visual Word Recognition. Journal Cogn. Neurosci. 15, 372 – 393.

Bloom, P. A., Fischler, I. 1980. Completion norms for 329 sentences. Mem. Cogn. 8, 631– 642.

Bohrn, I. C., Altmann, U., Jacobs, A. M. 2012. Looking at the brains behind figurative language - A quantitative meta-analysis of neuroimaging studies on metaphor, idiom, and irony processing. Neuropsychologia, 50, 2669-2683.

Boring, E. G. 1950. A history of experimental psychology, second ed. Appleton-Century-Crofts, New York.

Botvinick, M.M., Braver, T.S., Barch, D.M., Carter, C.S., Cohen, J.D., 2001. Conflict Monitoring and Cognitive Control. Psychol. Rev. 108, 624–652.

Bowers, J. S. 2009. On the Biological Plausibility of Grandmother Cells : Implications for Neural Network Theories in Psychology and Neuroscience. Psychol. Rev., 116, 220–251.

Bowers, J. S. 2010. More on Grandmother Cells and the Biological Implausibility of PDP Models of Cognition : A Reply to Plaut and McClelland (2010) and Quian Quiroga and Kreiman (2010). Psychol. Rev. 117, 300–308.

Box, G. E. P. (1979). Robustness in scientific model building. In R. L. Launer and G. N. Wilkinson (Eds.), Robustness in statistics (pp. 201-236). New York: Academic Press.

Braun, M., Jacobs, A.M., Hahne, A., Ricker, B., Hofmann, M., Hutzler, F., 2006. Model-generated lexical activity predicts graded ERP amplitudes in lexical decision. Brain Res. 1073-1074, 431–439.

Briesemeister, B.B., Hofmann, M.J., Tamm, S., Kuchinke, L., Braun, M., Jacobs, A.M., 2009. The pseudohomophone effect: evidence for an orthography-phonology-conflict. Neurosci. Lett. 455, 124–8.

Briesemeister, B.B., Kuchinke, L., Jacobs, A.M., 2011a. Discrete emotion effects on lexical decision response times. PloS one 6, e23743.

Briesemeister, B.B., Kuchinke, L., Jacobs, A.M., 2011b. Discrete emotion norms for nouns: Berlin affective word list (DENN-BAWL). Behav. Res. Meth. 43, 441–8.

Briesemeister, B.B., Kuchinke, L., Jacobs, A.M., 2012. Emotional Valence: A Bipolar Continuum or Two Independent Dimensions? SAGE Open 2, 1–12.

Briesemeister, B.B., Kuchinke, L., Jacobs, A.M., 2014. Emotion word recognition: Discrete information effects first, continuous later? Brain Res., doi.org/10.1016/j.brainres.2014.03.045

Broadbent, D.E., 1967. Word-frequency effect and response bias. Psychol. Rev. 74, 1-15.

Bullinaria, J. a, Levy, J.P., 2007. Extracting semantic representations from word co-occurrence statistics: a computational study. Behavior Res. Meth. 39, 510–26.

Bullinaria, J. A., Levy, J.P., 2013. Limiting factors for mapping corpus-based semantic representations to brain activity. PloS one 8, e57191.

Buxton, R.B., Uludağ, K., Dubowitz, D.J., Liu, T.T., 2004. Modeling the hemodynamic response to brain activation. NeuroImage 23, S220–33.

Chen, J.C.W., Li, W., Westerberg, C.E., Tzeng, O.J.-L., 2008. Test-item sequence affects false memory formation: an event-related potential study. Neurosci. Lett. 431, 51–6.

Citron, F.M.M., 2012. Neural correlates of written emotion word processing: a review of recent

electrophysiological and hemodynamic neuroimaging studies. Brain Lang. 43, 211–26.

Cohen, J.D., Dunbar, K., McClelland, J.L., 1990. On the control of automatic processes: a parallel distributed processing account of the Stroop effect. Psychol. Rev. 97, 332–61.

Cohen, L., Dehaene, S., 2004. Specialization within the ventral stream: the case for the visual word form area. NeuroImage 22, 466–76.

Cohen, L., Dehaene, S., Naccache, L., He, M., 2000. The visual word form area Spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. Brain 123, 291–307.

Collins, A.M., Loftus, E.F., 1975. A spreading-activation theory of semantic processing. Psychol. Rev. 82, 407–428.

Collyer, C. E., 1985. Comparing strong and weak models by fitting them to computer- generated data. Percept. Psychophysics 38, 476–81.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., Ziegler, J., 2001. DRC: a dual route cascaded model of visual word recognition and reading aloud. Psychol. Rev. 108, 204–56.

Dambacher, M., Kliegl, R., Hofmann, M., Jacobs, A.M., 2006. Frequency and predictability effects on event-related potentials during reading. Brain Res. 1084, 89–103.

Dambacher, M., Rolfs, M., Göllner, K., Kliegl, R., Jacobs, A.M., 2009. Event-related potentials reveal rapid verification of predicted visual input. PloS one 4, e5047.

Danker, J.F., Gunn, P., Anderson, J.R., 2008. A rational account of memory predicts left prefrontal activation during controlled retrieval. Cereb. Cortex 18, 2674–85.

Deese, J., 1959. On the prediction of occurrence of particular verbal intrusions in immediate recall, J.

Exp. Psychol. 58, 17–22.

Dehaene, S., Le Clec'H, G., Poline, J.-B., Le Bihan, D., Cohen, L., 2002. The visual word form area: a prelexical representation of visual words in the fusiform gyrus. Neuroreport 13, 321–5.

Devlin, J.T., Jamison, H.L., Gonnerman, L.M., Matthews, P.M., 2006. The role of the posterior fusiform gyrus in reading. J. Cogn. Neurosc. 18, 911–22.

Dilkina, K., McClelland, J.L., Plaut, D.C., 2010. Are there mental lexicons? The role of semantics in lexical decision. Brain Res. 1365, 66–81.

Duncan, K.J., Pattamadilok, C., Devlin, J.T., 2010. Investigating occipito-temporal contributions to reading with TMS. J. Cogn. Neurosc. 22, 739–50.

Eichenbaum, H., Yonelinas, A.P., Ranganath, C., 2007. The medial temporal lobe and recognition memory. Ann. Rev. Neurosci. 30, 123–52.

Eriksson, P.S., Perfilieva, E., Björk-Eriksson, T.B., Alborn, A.-M., Nordborg, C., Peterson, D.A., Gage, F.H., 1998. Neurogenesis in the adult human hippocampus. Nat. Med. 4, 1313–7.

Estes, W. K. 1975. Some targets for mathematical psychology. *J. Math. Psychol., 12,* 263-282.

Evert, S., 2005. The Statistics of Word Cooccurrences Word Pairs and Collocations. Thesis, Universität Stuttgart.

Fell, J., and Axmacher, N. (2011). The role of phase synchronization in memory processes. Nat. Rev. Neurosci., 12, 105–18.

Fiebach, C.J., Friederici, A.D., Cramon, D.Y. Von, 2002. fMRI Evidence for Dual Routes to the Mental Lexicon in Visual Word Recognition. J. Cogn. Neurosc. 14, 11– 23.

Fiebach, C.J., Ricker, B., Friederici, A.D., Jacobs, A.M., 2007. Inhibition and facilitation in visual word

recognition : Prefrontal contribution to the orthographic neighborhood size effect. NeuroImage 36, 901–911.

Firth, J. R. 1957. A synopsis of linguistic theory 1930–1955, in Studies in linguistic analysis. Oxford, England: Blackwell Publishers, pp. 1–32.

Fodor, J.A. 1983. Modularity of mind: An essay on faculty psychology. MIT Press, Cambridge.

Forgács, B., , I., Baudewig, J., Hofmann, M.J., Pléh, C., Jacobs, A.M., 2012. Neural correlates of combinatorial semantic processing of literal and figurative noun noun compound words. NeuroImage 63, 1432–42.

Friston, K. J., Harrison, L., and Penny, W., 2003. Dynamic causal modelling. NeuroImage, 19, 1273–1302.

Friston, K., 2010. The free-energy principle: a unified brain theory? Nat. Rev. Neurosci. 11, 127–38.

Gallo, D.A., 2006. Associative illusions of memory: False memory research in DRM and related tasks. Psychology Press, New York.

Gallo, D.A., 2010. False memories and fantastic beliefs: 15 years of the DRM illusion. Mem. Cogn. 38, 833–48.

Gamallo, P., Bordag, S., 2010. Is singular value decomposition useful for word similarity extraction? Lang. Resour. Eval. 45, 95–119.

Gillund, G., Shiffrin, R.M., 1984. A Retrieval Model for Both Recognition and Recall. Psychol. Rev. 91, 1–67.

Glanzer, M., Kim, K., Hilford, A., Adams, J.K., 1999. Slope of the receiver-operating characteristic in recognition memory. J. Exp. Psychol. Learn. Mem. Cogn. 25, 500–513.

Goebel R., Roebroeck A., Kim D.S., Formisano E. 2003. Investigating directed cortical interactions in

time-resolved fMRI data using vector autoregressive modeling and Granger causality mapping. Magn Reson Imaging. 21, 1251-1261.

Grainger, J., Jacobs, A.M., 1993. Masked Partial-Word Priming in Visual Word Recognition: Effects of Positional Letter Frequency. J. Exp. Psychol. Hum. Percept. Perform. 19, 951–964.

Grainger, J., Jacobs, A.M., 1994. A dual read-out model of word context effects in letter perception: Further investigations of the word superiority effect. J. Exp. Psychol. Hum. Percept. Perform. 20, 1158–1176.

Grainger, J., Jacobs, A.M., 1996. Orthographic processing in visual word recognition: a multiple read-out model. Psychol. Rev. 103, 518–65.

Grainger, J., Jacobs, A.M., 1998. Localist connectionist approaches to human cognition, Lawrence Erlbaum, Mahwah.

Grainger, J., O'Regan, J. K., Jacobs, A. M., Segui, J. 1989. On the role of competing word units in visual word recognition: The neighborhood frequency effect. Percept. Psychophys., 45, 189-195.

Green, D. M., and Swets, J. A. 1966. Signal Detection Theory and Psychophysics. Wiley, New York.

Griffiths, T.L., Steyvers, M., Tenenbaum, J.B., 2007. Topics in Semantic Representation Psychol. Rev. 114, 211–244.

Grossberg, S., 1978. A Theory of Visual Coding : Memory, and Development. in Leeuwenberg, M., E.L.J., Buffart, H.F.J. (Eds.), Formal Theories of Visual Perception, pp. 7–26.

Grossberg, S., 1980. How does a brain build a cognitive code? Psychol. Rev. 87, 1–51.

Hagoort, P., 2003. How the brain solves the binding problem for language : a neurocomputational model of syntactic processing. NeuroImage 20, 18–29.

Harm, M.W., Seidenberg, M.S., 2004. Computing the meanings of words in reading: cooperative division of labor between visual and phonological processes. Psychol. Rev. 111, 662–720.

Harris, Z. S. 1951. Methods in Structural Linguistics. University of Chicago Press, Chicago, http://archive.org/details/structurallingui00harr

Hauk, O., Pulvermüller, F., 2004. Effects of word length and frequency on the human event-related potential. Clin. Neurophysiol. 115, 1090–103.

Hebb, D., 1949. The Organization of Behavior. Wiley, New York.

Hofmann, M.J., 2011. Setting letters and words into context : Towards an Associative Read-Out Model. Auf der Suche nach dem Sinn des Lesens: Buchstaben und Wörter im Kontext. Thesis, Free University Berlin. http://www.diss.fu-berlin.de/diss/servlets/MCRFileNodeServlet/FUDISS_derivate_000000010214/Hofmann11DissPublication.pdf?hosts=local

Hofmann, M.J., Dambacher, M., Jacobs, A.M., Kliegl, R., Radach, R., Kuchinke, L., Plichta, M.M., Fallgatter, A.J., and Herrmann, M.J., 2014. Occipital and orbitofrontal hemodynamics during naturally paced reading. NeuroImage. doi:10.1016/j.neuroimage.2014.03.014

Hofmann, M.J., Herrmann, M.J., Dan, I., Obrig, H., Conrad, M., Kuchinke, L., Jacobs, A.M., Fallgatter, A.J., 2008a. Differential activation of frontal and parietal regions during visual word recognition: an optical topography study. NeuroImage 40, 1340–9.

Hofmann, M.J., Kuchinke, L., Biemann, C., Tamm, S., Jacobs, A.M., 2011. Remembering words in context as predicted by an associative read-out model. Front. Psychol. 2, 252.

Hofmann, M.J., Kuchinke, L., Tamm, S., Võ, M.L.-H., Jacobs, A.M., 2009. Affective processing within 1/10th of a second: High arousal is necessary for early facilitative processing of negative but not

positive words. Cog. Affect. Behav. Neurosci. 9, 389–97.

Hofmann, M.J., Stenneken, P., Conrad, M., Jacobs, A.M., 2007. Sublexical frequency measures for orthographic and phonological units in German. Behav. Res. Meth. 39, 620–9.

Hofmann, M.J., Tamm, S., Braun, M.M., Dambacher, M., Hahne, A., Jacobs, A.M., 2008b. Conflict monitoring engages the mediofrontal cortex during nonword processing. Neuroreport 19, 25–9.

Holcomb, P.J., Grainger, J., Rourke, T.O., 2002. An Electrophysiological Study of the Effects of Orthographic Neighborhood Size on Printed Word Perception. J. Cogn. Neurosc. 14, 938 – 950.

Holroyd, C.B., Coles, M.G.H., 2002. The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. Psychol. Rev. 109, 679–709.

Huber, D.E., Tian, X., Curran, T., O'Reilly, R.C., Woroch, B., 2008. The dynamics of integration and separation: ERP, MEG, and neural network studies of immediate repetition effects. J. Exp. Psychol. Hum. Percept. Perform. 34, 1389–416.

Jackendoff, R., 1987. Consciousness and the computational mind. Explorations in cognitive science, MIT Press, Cambridge.

Jacobs, A.M., Carr, T. H., 1995. Mind mappers and cognitive modelers: toward cross-fertilization. Behav. Brain Sci. 18, 362–363.

Jacobs, A.M., Graf, R., Kinder, A., 2003. Receiver operating characteristics in the lexical decision task: Evidence for a simple signal-detection process simulated by the multiple read-out model. J. Exp. Psychol. Learn. Mem. Cogn. 29, 481–488.

Jacobs, A.M., Grainger, J., 1992. Testing a semistochastic variant of the interactive activation model in different word recognition experiments. J. Exp. Psychol. Hum. Percept. Perform. 18, 1174–88.

Jacobs, A.M., Grainger, J., 1994. Models of visual word recognition: Sampling the state of the art. J.

Exp. Psychol. Hum. Percept. Perform. 20, 1311–1334.

Jacobs, A.M., Grainger, J., 1999. Modeling a theory without a model theory, or "computational modeling after Feyerabend". Behav. Brain Sci. 22, 46–47.

Jacobs, A. M., and Rösler, F. (1999). Dondersian dreams in brain-mappers minds, or, still no cross-fertilization between mind mappers and cognitive modelers. Behav. Brain Sci. 22, 293-295.

Jacobs, A.M., Hofmann, M.J., 2013. Neurokognitive Modellierung [Neurocognitive Modeling]. In: Schröger, E., Kölsch, S. (Eds.), Enzyklopädie der Psychologie. Affektive und Kognitive Neurowissenschaft. Hogrefe, Göttingen, pp. 431–447.

Jacobs, A.M. (2008). Kognitive Modellierung und Simulation / „diagram making" (Cognitive modeling and simulation / diagram making). In S. Gauggel and M. Herrmann (Hrsg.), Handbuch der Neuro- und Biopsychologie (54 - 60). Göttingen: Hogrefe.

Jacobs, A. M., Rey, A., Ziegler, J. C., Grainger, J. 1998. MROM-p: An interactive activation, multiple readout model of orthographic and phonological processes in visual word recognition, in Grainger J., Jacobs A. M. (eds.), Localist connectionist approaches to human cognition, Lawrence Erlbaum Associates Inc., Mahwah, pp. 147–188.

Jeffreys, D.A., Axford, J.G., 1972. Source Locations of Pattern-Specific Components of Human Visual Evoked Potentials. I. Component of Striate Cortical Origin. Exp. Brain Res. 16, 1–21.

Jung, C.G., 1905. Ueber das Verhalten der Reaktionszeit beim Assoziationsexperimente. Ambrosius Barth, Leipzig.

Klonek, F., Tamm, S., Hofmann, M.J., Jacobs, A.M., 2009. Does familiarity or conflict account for performance in the word-stem completion task? Evidence from behavioural and event-related-potential data. Psychol. Res. 73, 871–82.

Kronbichler, M., Hutzler, F., Wimmer, H., Mair, A., Staffen, W., Ladurner, G., 2004. The visual word form area and the frequency with which words are encountered: evidence from a parametric fMRI study. NeuroImage 21, 946–53.

Kuchinke, L., 2007. Implicit and explicit recognition of emotionally valenced words. Thesis, Free University Berlin.

Kuchinke, L., Brockhaus, W.-R., Hofmann, M., Jacobs, A.M., 2010. Sequential Dependencies in the Lexical Decision Task: A Role of the Basal Ganglia. In: Frontiers in Neuroscience Conference Abstract: Decision Neuroscience From Neurons to Societies. Berlin, pp. 1–1.

Kuchinke, L., Fritzemeier, S., Hofmann, M.J., Jacobs, A.M., 2013. Neural Correlates of Episodic Memory: Associative Memory and Confidence Drive Hippocampus Activations. Behav. Brain Res.

Kuchinke, L., Hofmann, M.J., Jacobs, A.M., Frühholz, S., Tamm, S., Herrmann, M., 2011. Human striatal activation during adjustment of the response criterion in visual word recognition. NeuroImage 54, 2412–2417.

Kuchinke, L., Jacobs, A.M., Grubich, C., Võ, M.L.-H., Conrad, M., Herrmann, M., 2005. Incidental effects of emotional valence in single word processing: an fMRI study. NeuroImage 28, 1022–32.

Kuchinke, L., Jacobs, A.M., Võ, M.L.-H., Conrad, M., Grubich, C., Herrmann, M., 2006. Modulation of prefrontal cortex activation by emotional words in recognition memory. Neuroreport 17, 1037–41.

Kumaran, D., McClelland, J.L., 2012. Generalization through the recurrent interaction of episodic memories: a model of the hippocampal system. Psychol. Rev. 119, 573–616.

Kuperberg, G.R., Holcomb, P.J., Sitnikova, T., Greve, D., Dale, A.M., Caplan, D., 2003. Distinct

patterns of neural modulation during the processing of conceptual and syntactic anomalies. J. Cogn. Neurosc. 15, 272–93.

Kutas, M., Federmeier, K.D., 2011. Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). Ann. Rev. Psychol. 62, 621–47.

Kutas, M., Hillyard, S.A., 1984. Brain potentials during reading reflect word expectancy and semantic association. Nat. 307, 161–3.

Lachaux, J. P., Rodriguez, E., Martinerie, J., Varela, F. J., 1999. Measuring phase synchrony in brain signals. Hum. Brain Mapp., 8(4), 194–208.

Landauer, T.K., Dumais, S.T., 1997. A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. Psychol. Rev. 104, 211–240.

Lau, E.F., Phillips, C., Poeppel, D., 2008. A cortical network for semantics: (de)constructing the N400. Nat. Rev. Neurosci. 9, 920–33.

Lucas, M. (2000). Semantic priming without association: a meta-analytic review. Psychon. Bull. Rev., 7, 618–30.

Malmberg, K.J., 2008. Recognition memory : A review of the critical findings and an integrated theory for relating them. J. Mem. Lang..

Mandler, G., 1980. Recognizing: The judgment of previous occurrence. Psychol. Rev. 87, 252–271.

Maratos, E. J., Allan, K., Rugg, M. D. 2000. Recognition memory for emotionally negative and neutral words: an ERP study. Neuropsychologia, 3*8*(11), 1452–65.

Martin, R.C., Byrne, M.D., 2006. Why opening a door is as easy as eating an apple: A reply to Thompson-Schill and Botvinick (2006). Psychon. Bull. Review 13, 409–411.

Massaro, D.W., 1988. Some criticisms of connectionist models of human performance. J. Mem. Lang. 27, 213–234.

Massaro, D. W. 1989. Testing between the TRACE model and the fuzzy logical model of speech perception. Cogn. Psychol. 21, 398-421.

Massaro, D., Cohen, M., 1991. Integration Influence versus Interactive Activation : The Joint of Stimulus and Context in Perception. Psychol. Rev. 614, 558–614.

McCarthy, G., Nobre, A.C., Bentin, S., Spencer, D.D., 1995. Language-related field potentials in the anterior-medial temporal lobe: I. Intracranial distribution and neural generators. J. Neurosci. 15, 1080–9.

McClelland, J.L., 1991. Stochastic Interactive Context Processes and the Effect of on Perception. Cogn. Psychol. 23, 1–44.

McClelland, J.L., 1993. Toward a Theory of Information Processing in Graded, Random, and Interactive Networks, in Meyer, D.E., Kornblum, S., Attention and Performance XIV: Synergies in Experimental Psychology, Artificial Intelligence, and Cognitive Neuroscience, pp. 655–688.

McClelland, J.L., Chappell, M., 1998. Familiarity breeds differentiation: a subjective-likelihood approach to the effects of experience in recognition memory. Psychol. Rev. 105, 724–60.

McClelland, J.L., McNaughton, B.L., O'Reilly, R.C., 1995. Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. Psychol. Rev. 102, 419–57.

McClelland, J.L., Rumelhart, D.E., 1981. An Interactive Activation Model of Context Effects in Letter Perception: Part 1. An Account of Basic Findings. Psychol. Rev. 5, 375–407.

McKoon, G., Ratcliff, R., 1992. Spreading activation versus compound cue accounts of priming:

mediated priming revisited. J. Exp. Psychol. Learn. Mem. Cogn. 18, 1155–72.

McNeely, H.E., Dywan, J., Segalowitz, S.J., 2004. ERP indices of emotionality and semantic cohesiveness during recognition judgments. Psychophysiol. 41, 117–29.

Mechelli, A., Humphreys, G.W., Mayall, K., Olson, A., Price, C.J., 2000. Differential effects of word length and visual contrast in the fusiform and lingual gyri during reading. Proc. R. Soc. Lond. B 267, 1909–13.

Miller, G.A., 1956. The magical number seven, plus or minus two: some limits on our capacity for processing information. 1956. Psychol. Rev. 101, 343–52.

Miller, G.A., Galanter, E., Pribram, K.H., 1960. Plans and the structure of behavior. Holt, New York.

Mitchell, T.M., Shinkareva, S. V, Carlson, A., Chang, K.-M., Malave, V.L., Mason, R.A., Just, M.A., 2008. Predicting human brain activity associated with the meanings of nouns. Sci. 320, 1191–5.

Morton, J., 1969. Interaction of information in word recognition. Psychol. Rev. 76, 165–178.

Newell, A. (1990). Unified theories of cognition. Harward University Press, England.

Nobre, A.C., McCarthy, G., 1995. Language-related field potentials in the anterior-medial temporal lobe: II. Effects of word type and semantic priming. J. Neurosci. 15, 1090–8.

Norris, D., Kinoshita, S., 2012. Reading through a noisy channel: why there's nothing special about the perception of orthography. Psychol. Rev. 119, 517–45.

Paap, K.R., Chun, E., Vonnahme, P., 1999. Discrete threshold versus continuous strength models of perceptual recognition. Can. J. Exp. Psychol. 53, 277–93.

Paap, K.R., Newsome, S.L., McDonald, J.E., Schvaneveldt, R.W., 1982. An activation-verification model for letter and word recognition: the word-superiority effect. Psychol. Rev. 89, 573–94.

Page, M., 2000. Connectionist modelling in psychology: a localist manifesto. Behav. Brain Sci. 23, 443–512.

Pascual-Marqui, R.D., 2002. Standardized low resolution brain electromagnetic tomography (sLORETA): technical details. Methods and Findings in Experimental and Clinical Pharmacology 24, 5-12.

Patterson, K., Nestor, P.J., Rogers, T.T., 2007. Where do you know what you know? The representation of semantic knowledge in the human brain. Nat. Rev. Neurosci. 8, 976–87.

Penolazzi, B., Hauk, O., Pulvermüller, F., 2007. Early semantic context integration and lexical access as revealed by event-related brain potentials. Biol. Psychol. 74, 374–88.

Perry, C., Ziegler, J.C., Zorzi, M., 2007. Nested incremental modeling in the development of computational theories: the CDP+ model of reading aloud. Psychol. Rev. 114, 273–315.

Perry, C., Ziegler, J. C., Zorzi, M. 2010. Beyond single syllables: large-scale modeling of reading aloud with the Connectionist Dual Process (CDP++) model. Cogn. Psychol. 61, 106–151.

Petersen, S.E., Fox, P.T., Posner, M.I., Mintun, M., Raichle, M.E., 1988. Positron emission tomographic studies of the cortical anatomy of single-word processing. Nat. 331, 585–589.

Phelphs, E.A., LaBar, K.S., Anderson, A.K., O'Connor, K.J., Fullbright, R.K., Spencer, D.D., 1998. Specifying the Contributions of the Human Emotional Memory : A Case Study Amygdala to. Neurocase 4, 527–540.

Phelps, E A., LaBar K. S., 1997. The role of organization in recall for affective words. Abstracts of the Psychonomic Society, 2, 4.

Pitt, M. A., Myung, I.J., Zhang, S., 2002. Toward a method of selecting among computational models of cognition. Psychol. Rev. 109, 472–491.

Plaut, D.C., and Mcclelland, J.L. 2010. Locating Object Knowledge in the Brain: Comment on Bowers ' s (2009) Attempt to Revive the Grandmother Cell Hypothesis. Psychol. Rev., 117, 284–290.

Polich, J., 1985. Semantic categorization and event-related potentials. Brain Lang. 26, 304–21.

Popper, K., 1935. Logik der Forschung. Julius Springer Verlag, Wien.

Price, C.J., Devlin, J.T., 2003. The myth of the visual word form area. NeuroImage 19, 473–481.

Price, C.J., Devlin, J.T., 2011. The interactive account of ventral occipitotemporal contributions to reading. Trends Cogn. Sci. 15, 246–53.

Pulvermüller, F., Fadiga, L., 2010. Active perception: sensorimotor circuits as a cortical basis for language. Nat. Rev. Neurosci. 11, 351–360.

Quasthoff, U., Richter, M., Biemann, C., 2006. Corpus Portal for Search in Monolingual Corpora. In: Proceedings of LREC-06. Genova, pp. 10–13.

Quiroga, R. Q., and Kreiman, G. (2010). Measuring Sparseness in the Brain : Comment on Bowers (2009). Psychol. Rev., 117, 291–297.

Ranganath, C., 2010. A unified framework for the functional organization of the medial temporal lobes and the phenomenology of episodic memory. Hippocampus 20, 1263–90.

Rapp, R., 2002. The computation of word associations: comparing syntagmatic and paradigmatic approaches, in Association for Computational Linguistics, Proceedings of the 19th international conference on Computational linguistics-Volume 1, pp. 1-7. http://acl.ldc.upenn.edu/coling2002/proceedings/data/area-21/co-024.pdf

Rapp, R., Wettler, M. 1991. Prediction of Free Word Associations Based on Hebbian Learning, in Proceedings of the Joint Conference on Neural Networks, Singapore, pp. 25–29.

Rauss, K., Schwartz, S., Pourtois, G., 2011. Top-down effects on early visual processing in humans: a predictive coding framework. Neurosci. Biobehav. Rev. 35, 1237–53.

Recio, G., Conrad, M., Hansen, L.B., Jacobs, A.M. in press. On Pleasure and Thrill: The Interplay between Arousal and Valence during Visual Word Recognition. Brain Lang.

Reilly, R. G., Radach, R., 2006. Some empirical tests of an interactive activation model of eye movement control in reading. Cognitive Systems Research, 7, 34–55.

Rey, A., Dufau, S., Massol, S., Grainger, J., 2009. Testing computational models of letter perception with item-level event-related potentials. Cogn. Neuropsychol. 26, 7 – 22.

Richardson, F.M., Seghier, M. L., Leff, A.P., Thomas, M. S.C., Price, C. J. 2011. Multiple Routes from Occipital to Temporal Cortices during Reading. J. Neurosc., 31, 8239–8247.

Roediger, H.L., Balota, D., Watson, J.M., 2001. Spreading Activaiton and Arousal of False Memories. In: Roediger, Henry L., Nairne, J.S., Neath, I., Surprenant, A.M. (Eds.), The Nature of Remembering: Essays in Honor of Robert G. Crowder. American Psychological Association, Washington DC.

Roediger, H.L.I., McDermott, K.B., 1995. Creating False Memories : Remembering Words Not Presented in Lists. J. Exp. Psychol. Learn. Mem. Cogn. 21, 803–814.

Rossell, S.L., Price, C.J., Nobre, A.C., 2003. The anatomy and time course of semantic priming investigated by fMRI and ERPs. Neuropsychol. 41, 550–64.

Rumelhart, D.E., Mcclelland, J.L., 1982. An Interactive Activation Model of Context Effects in Letter Perception : Part 2 . The Contextual Enhancement Effect and Some Tests and Extensions of the Model. Psychol. Rev. 89, 60–94.

Rumelhart, D.E., Siple, P., 1974. Process of recognizing tachistoscopically presented words. Psychol.

Rev. 81, 99–118.

Saussure, F. de 1959. Course in general linguistics. Philosophical Library, New York. http://books.google.de/books?id=FSpZAAAAMAAJ.

Schrott, R., Jacobs, A.M. 2011. Gehirn und Gedicht: Wie wir unsere Wirklichkeiten konstruieren [Brain and poem: How we construct our realities]. Hanser, München.

Schurz, M., Sturm, D., Richlan, F., Kronbichler, M., Ladurner, G., Wimmer, H., 2010. A dual-route perspective on brain activation in response to visual words: evidence for a length by lexicality interaction in the visual word form area (VWFA). NeuroImage 49, 2649–61.

Schurz, M., Kronbichler, M., Crone, J., Richlan, F., Klackl, J., Wimmer, H., 2014. Top-down and bottom-up influences on the left ventral occipito-temporal cortex during visual word recognition: An analysis of effective connectivity. Hum. Brain Mapp., 35(4), 1668–80.

Seidenberg, M.S., McClelland, J.L., 1989. A distributed, developmental model of word recognition and naming. Psychol. Rev. 96, 523–68.

Sereno, S.C., Rayner, K., 2003. Measuring word recognition in reading : eye movements and event-related potentials. Trends Cogn. Sci. 7, 489–493.

Sereno, S.C., Rayner, K., Posner, M.I., 1998. Establishing a time line of word recognition : evidence from eye event-related potentials. Neuroreport 9, 2195–2200.

Shaoul, C., Westbury, C.F., 2006. Word frequency effects in high-dimensional co-occurrence models : a new approach. Behavior Res. Meth. 38, 190–195.

Shenhav, A., Botvinick, M.M., Cohen, J.D., 2013. The expected value of control: an integrative theory of anterior cingulate cortex function. Neuron 79, 217–40.

Shiffrin, R.M., Steyvers, M., 1997. A model for recognition memory: REM-retrieving effectively from memory. Psychon. Bull. Rev. 4, 145–66.

Skrandies, W., 1998. Evoked potential correlates of semantic meaning--A brain mapping study. Cogn. Brain Res. 6, 173–83.

Smith, S.M. 2012. The future of FMRI connectivity. NeuroImage 62, 1257–1266.

Spieler, D.H., Balota, D. A., 1997. Bringing Computational Models of Word Naming Down to the Item Level. Psychol. Sci. 8, 411–416.

Squire, L.R., Wixted, J.T., Clark, R.E., 2007. Recognition memory and the medial temporal lobe: a new perspective. Nat. Rev. Neurosci. 8, 872–83.

Sternberg, S., 1969. The discovery of processing stages: extensions of Donders' Method. Acta Psychol. 30, 276–315.

Steyvers, M., Griffiths, T.L., Dennis, S., 2006. Probabilistic inference in human semantic memory. Trends Cogn. Sci. 10, 327–34.

Sun, R., Coward, L. A., and Zenzen, M. J. 2005. On levels of cognitive modeling. Philos. Psychol., 18, 613–637.

Talmi, D., Moscovitch, M., 2004. Can semantic relatedness explain the enhancement of memory for emotional words? Mem. Cogn. 32, 742–51.

Taylor, J.S.H., Rastle, K., Davis, M.H., 2012. Can Cognitive Models Explain Brain Activation During Word and Pseudoword Reading? A Meta-Analysis of 36 Neuroimaging Studies. Psychol. Bull. 45, 0–26.

Teodorescu, A.R., Usher, M., 2013. Disentangling decision models: from independence to competition.

Psychol. Rev. 120, 1–38.

Thompson-Schill, S.L., Botvinick, M.M., 2006. Resolving conflict: A response to Martin and Cheng. Psychon. Bull. Rev. 13, 402–408.

Thompson-Schill, S.L., D'Esposito, M., Aguirre, G.K., Farah, M.J., 1997. Role of left inferior prefrontal cortex in retrieval of semantic knowledge: a reevaluation. Proc. Natl. Acad. Sci. U.S.A. 94, 14792–7.

Treisman, M., 1978. A theory of the identification of complex stimuli with an application to word recognition. Psychol. Rev. 85, 525–570.

Ungerleider, L.G., Mishkin, M., 1982. Two cortical visual systems. In: Ingle, D.J., Goodale, M.A., Mansfield, R.J.W. (Eds.), Analysis of Visual Behavior. pp. 549–586.

Vinckier, F., Dehaene, S., Jobert, A., Dubus, J.P., Sigman, M., Cohen, L., 2007. Hierarchical coding of letter strings in the ventral stream: dissecting the inner organization of the visual word-form system. Neuron 55, 143–56.

Võ, M.L.-H., Conrad, M., Kuchinke, L., Urton, K., Hofmann, M.J., Jacobs, A.M., 2009. The Berlin Affective Word List Reloaded (BAWL-R). Behav. Res. Meth. 41, 534–8.

Wagenmakers, E.-J., Van Der Maas, H.L.J., Grasman. R.P.P.P., 2007. An EZ-diffusion Model for Response Time and Accuracy. Psychon. Bull. Rev. 14, 3–22.

Wagner, A.D., Paré-Blagoev, E.J., Clark, J., Poldrack, R.A., 2001. Recovering meaning: left prefrontal cortex guides controlled semantic retrieval. Neuron 31, 329–38.

Wible, C.G., Han, S.D., Spencer, M.H., Kubicki, M., Niznikiewicz, M.H., Jolesz, F. A., McCarley, R.W., Nestor, P., 2006. Connectivity among semantic associates: an fMRI study of semantic priming. Brain Lang. 97, 294–305.

Windmann, S., Kutas, M., 2001. Electrophysiological correlates of emotion-induced recognition bias. J. Cogn. Neurosc. 13, 577–92.

Wixted, J.T., 2007. Dual-Process Theory and Signal-Detection Theory of Recognition Memory. Psychol. Rev. 114, 152–176.

Wixted, J.T., Mickes, L., 2013. On the Relationship Between fMRI and Theories of Cognition: The Arrow Points in Both Directions. Perspect. Psychol. Sci. 8, 104–107.

Woolrich, M.W., Stephan, K.E., 2013. Biophysical network models and the human connectome. NeuroImage 80, 330–8.

Yeung, N., Botvinick, M.M., Cohen, J.D., 2004. The Neural Basis of Error Detection: Conflict Monitoring and the Error-Related Negativity. Psychol. Rev. 111, 931–959.

Yonelinas, A.P., 1994. Receiver-Operating Characteristics in Recognition Memory: Evidence for a Dual-Process Model. J. Exp. Psychol. Learn. Mem. Cogn. 20, 1341–1354.

Yonelinas, A.P., 2002. The Nature of Recollection and Familiarity: A Review of 30 Years of Research. J. Mem. Lang. 46, 441–517.

Yonelinas, A.P., Otten, L.J., Shaw, K.N., Rugg, M.D., 2005. Separating the Brain Regions Involved in Recollection and Familiarity in Recognition Memory. J. Neurosc. 25, 3002–3008.

Ziegler, J.C., Goswami, U. 2005. Reading Acquisition , Developmental Dyslexia , and Skilled Reading Across Languages : A Psycholinguistic Grain Size Theory. Psychol. Bull. 131, 3–29.

Ziegler, J.C., Jacobs, A.M., Klüppel, D., 2001. Pseudohomophone effects in lexical decision: Still a challenge for current word recognition models. J. Exp. Psychol. Hum. Percept. Perform. 27, 547–559.

Figures

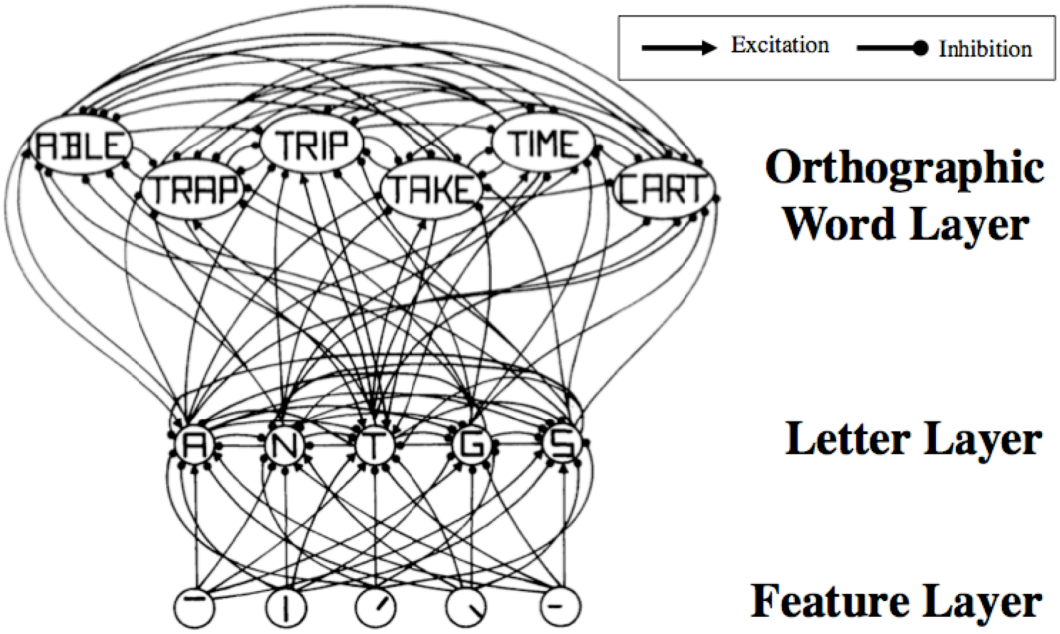
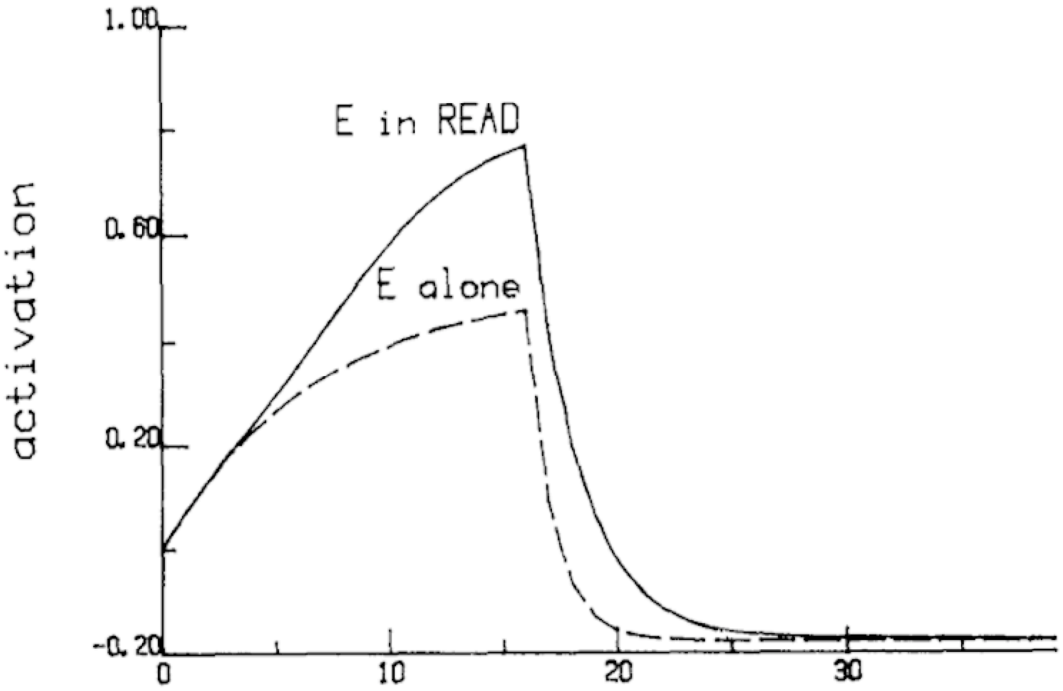


Figure 1. Architecture of the classic IAM. For each letter position, there are visual feature units in a feature layer. For instance, if a „T“ is presented to the model at the first position, the visual features „|“ and „¯“ activate the unit „T“ at the letter layer, which in turn activates all units at the orthographic word layer starting with a T, e.g. trip or take (Figure taken from McClelland and Rumelhart, 1981).

Figure 2. Example simulations that account for the word superiority effect. Perceptual identification of



a letter is faster, if it is contained in a word. The classic IAM can account for this by the letter level activations shown at the y-axis. The x-axis represents model cycles. When the identified target letter obtains excitation from the orthographic word unit ‚READ', its activation becomes greater than when the letter is presented in isolation. While greater activations indicate greater evidence that the letter has been presented, the IAM accounts for a faster and less error-prone identification of letters in words (Figure taken from McClelland and Rumelhart, 1981).

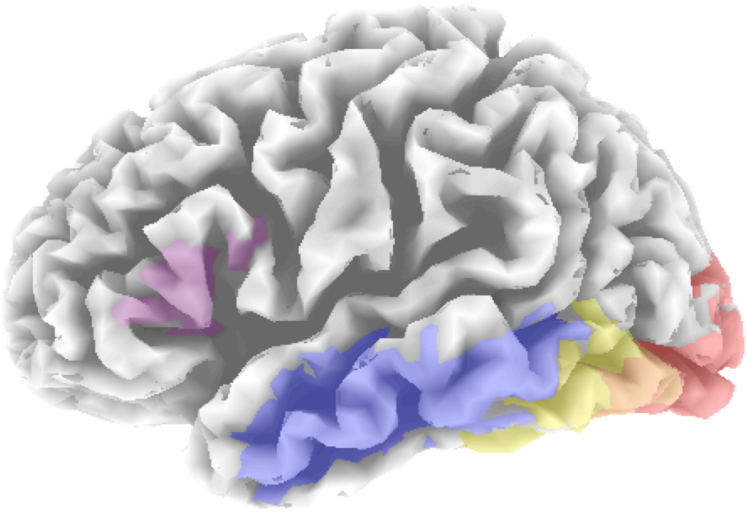


Figure 3. Target regions of IAMs. Visual features are represented in the occipital cortex (red). Bottom-up driven neural activation then propagates to more anterior regions, which represent greater levels of abstraction (e.g. Vinckier et al., 2007). The posterior fusiform gyrus represents orthographic word forms (yellow). Finally, semantic units are represented in the temporal lobe (blue), and semantic competition is represented in the left inferior frontal gyrus (cf. sections 3.4 and 3.5 below; Figures generated by sLORETA, Pascual-Marqui, 2002).

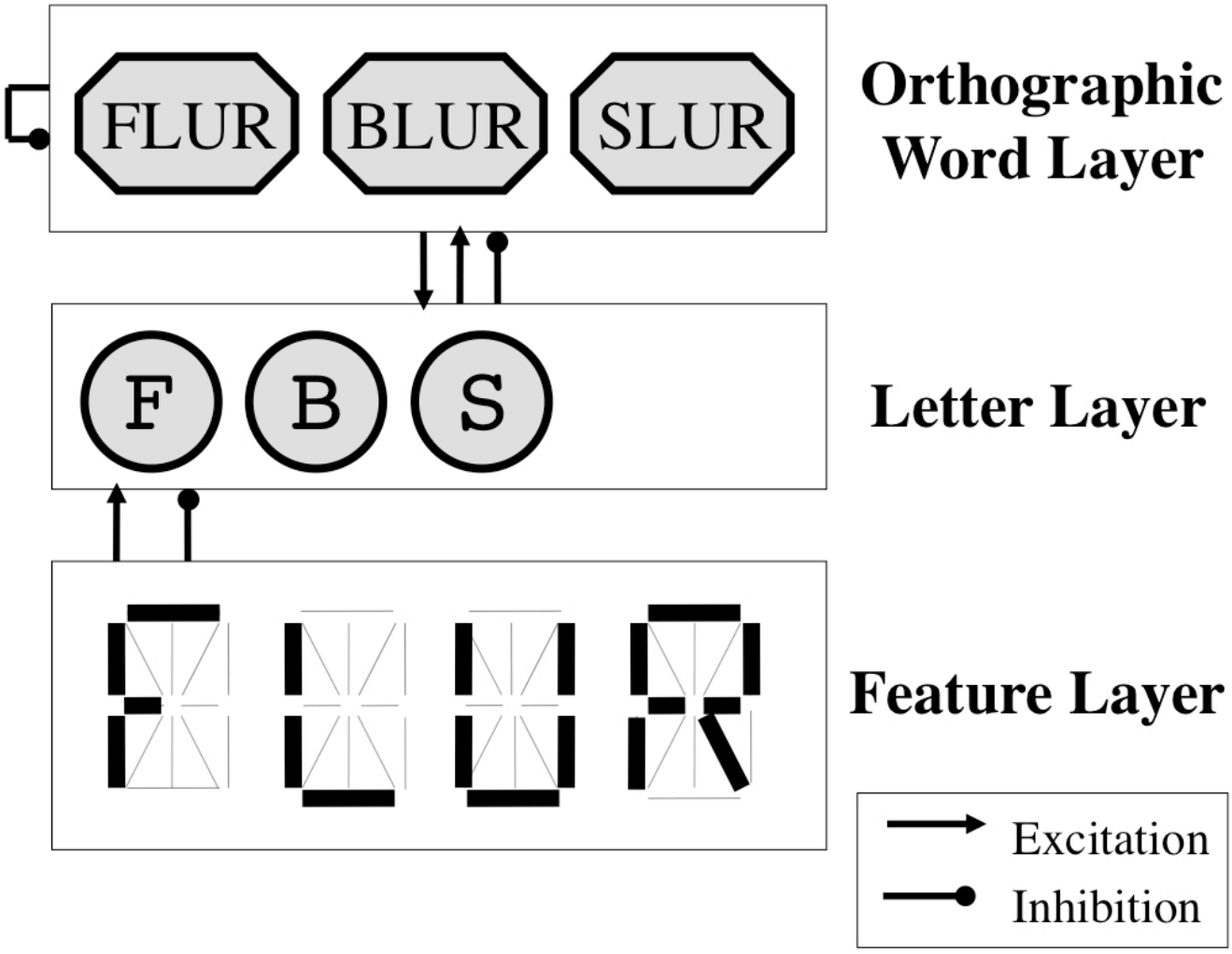


Figure 4. Sketch of an IAM. Stimuli are presented to the feature layer. For each letter, there is a letter unit. The first letter of the English nonword FLUR activates the units 'F', 'B', and 'S', because they share several features. Therefore, the orthographic word units 'FLUR', 'BLUR' and 'SLUR' all become activated.

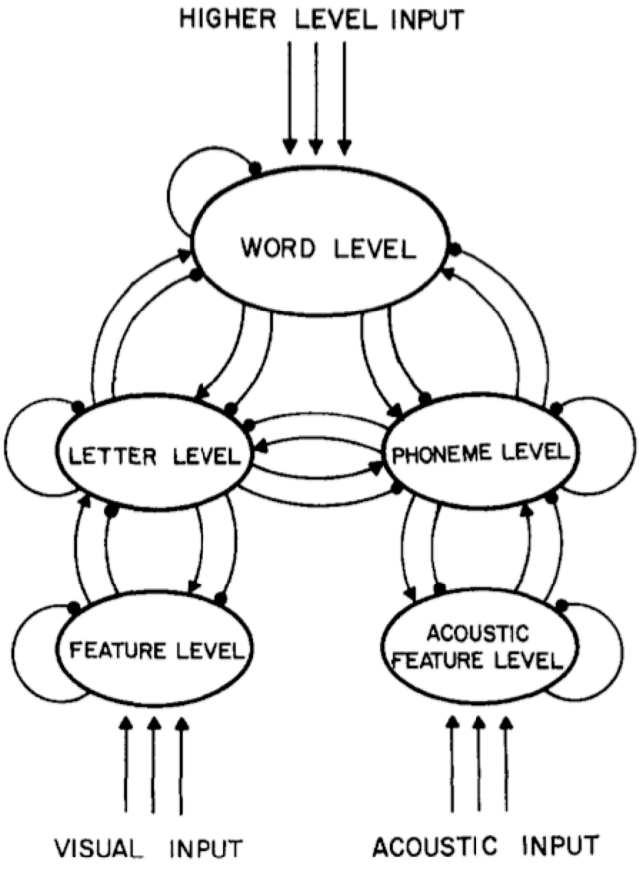


Figure 5. The broader conceptual framework of IAMs. While the original simulation model only implemented orthographic and visual representations (cf. Figure 1), it also sketched the role of phonology, as well as “higher level input” (taken from McClelland and Rumelhart, 1981).

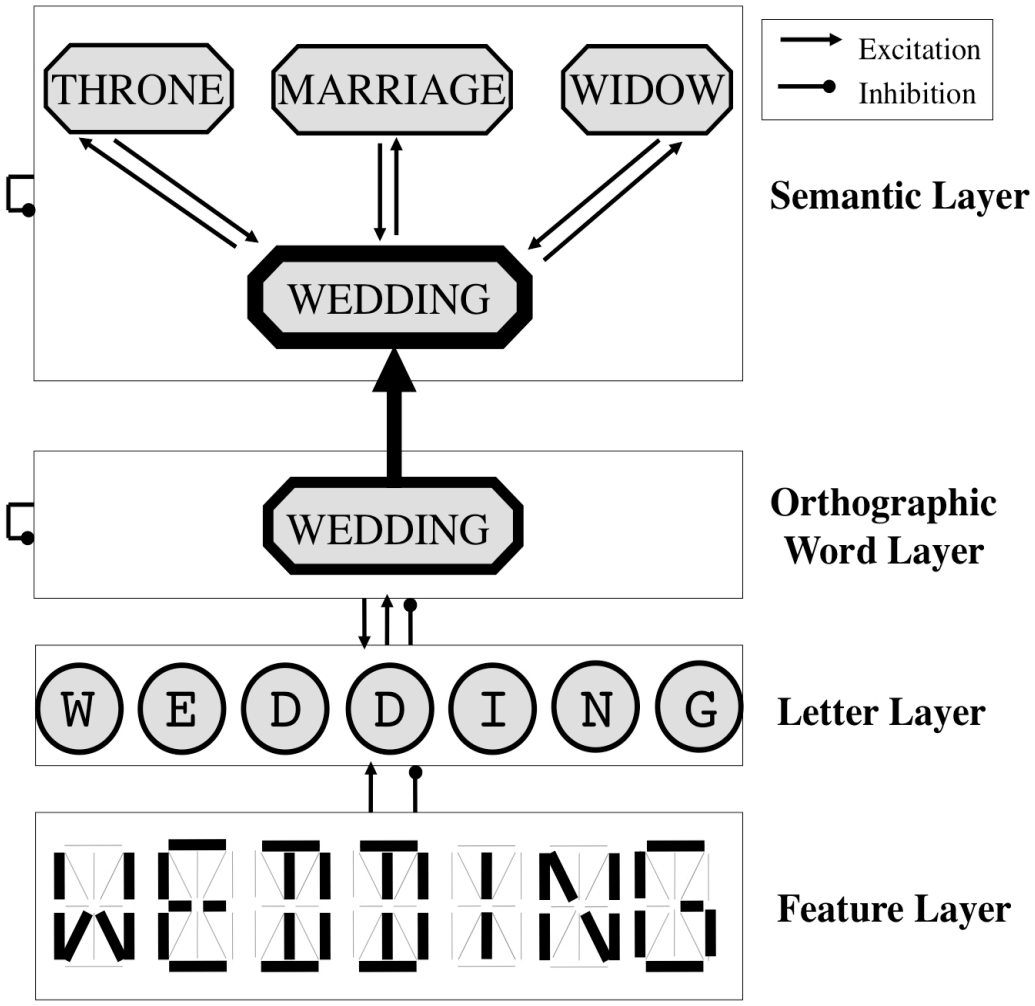


Figure 6. Sketch of the Associative Read-Out Model (AROM). The lower three layers correspond to an IAM. The semantic layer receives input from the orthographic word layer. The semantic unit of the presented word activates associated units, which feed activation back. Therefore, semantic associates in the context increase the activation of a target word unit (adopted from Hofmann et al., 2011).

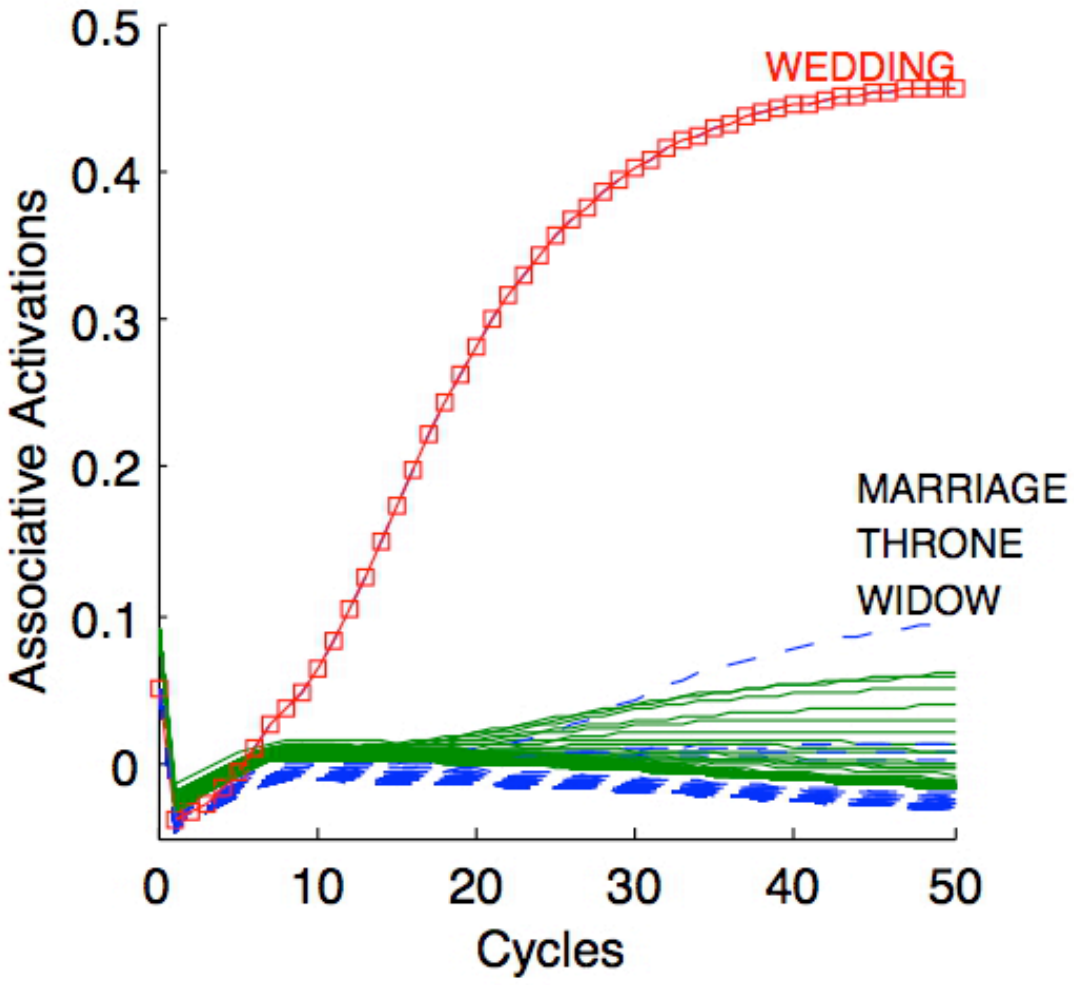


Figure 7. An example simulation of a stimulus with a high number of associates. The x-axis shows simulation cycles, and the y-axis displays the relative activations of semantic word units. When 'wedding' (red) is presented to the model, the most strongly co-activated items in the stimulus set are 'marriage' – a non-studied item (blue dashed) – and the studied items 'throne' and 'widow' (green). These associates drive the activation of the target word unit, and thus account for false memory effects (adopted from Hofmann et al., 2011).

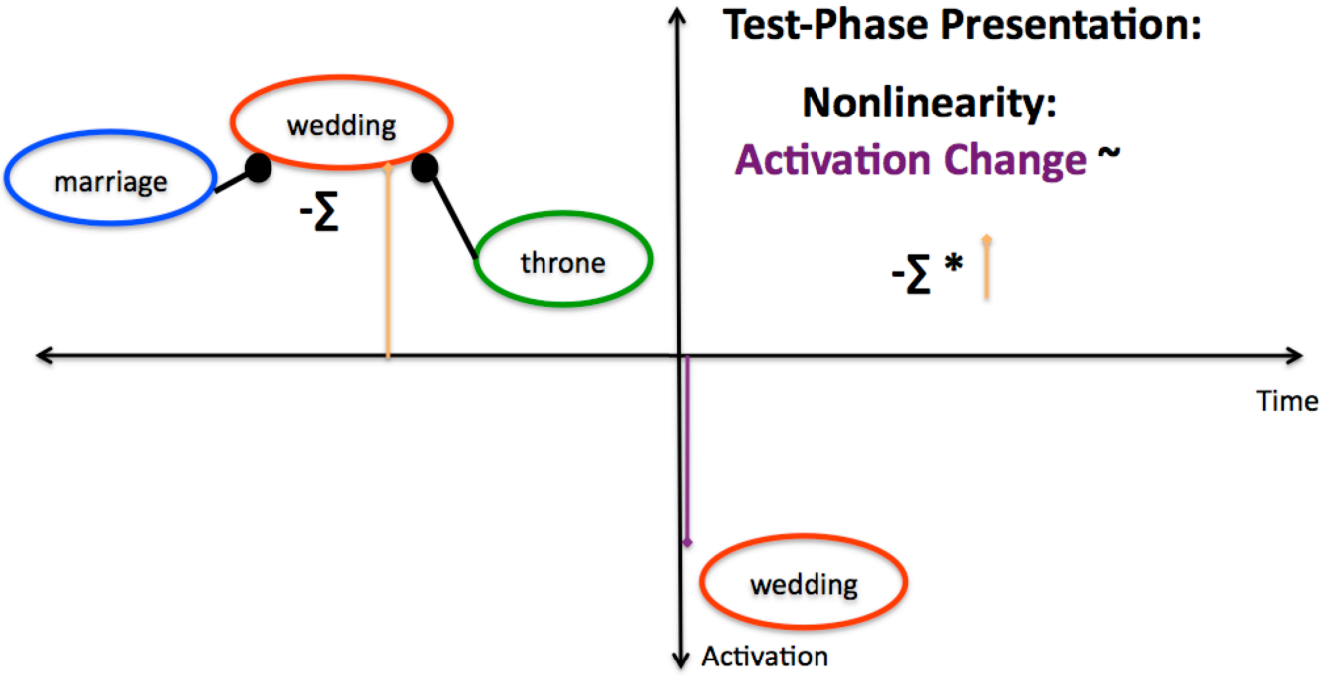


Figure 8. Illustration of the IAM's functional mechanism that accounts for a greater memory signal variance for studied than for non-studied items in the test phase of a recognition memory task. The x-axis represents time and the y-axis shows relative activation of the memory traces, i.e. the semantic layer units. To the left of the y-axis is the time period before the test-phase presentation of the studied example item 'wedding' (red, cf. Figures 6 and 7). The AROM shares the first ad-hoc assumption with the unequal variance model of recognition memory. Therefore, it represents learning in a study-phase episode by setting a greater resting level (orange line) to studied than to non-studied items. A critical process that causes between-item variability is inhibition from studied (green) and previously presented non-studied items (blue). Such inhibition is obtained for studied and non-studied items alike. The inhibition is summed (-∑) before the well-established assumption of nonlinearity causes the actual activation change of the semantic unit of the presented word. This nonlinearity assumption predicts that the activation change of one item is

proportional to the product of the summed net inhibition (-∑) and the resting level (orange). Because the resting level is greater in studied items, the first activation change in cycle 1 (violet) is greater for a high-resting-level studied item than for a low-resting level non-studied item. Thus, if the distribution of net inhibition across all items remains constant, the greater variability of studied-item variances can be predicted by the resting-level difference: Every item distribution is scaled by the resting level. In consequence, the AROM predicts greater memory trace variability from studying an item. That is, the semantic layer's resting level logically results in greater memory signal variances for studied items. The second ad-hoc assumption of the unequal variance model of recognition memory can be predicted – and thus can be saved (cf. Appendix A for a formal demonstration).

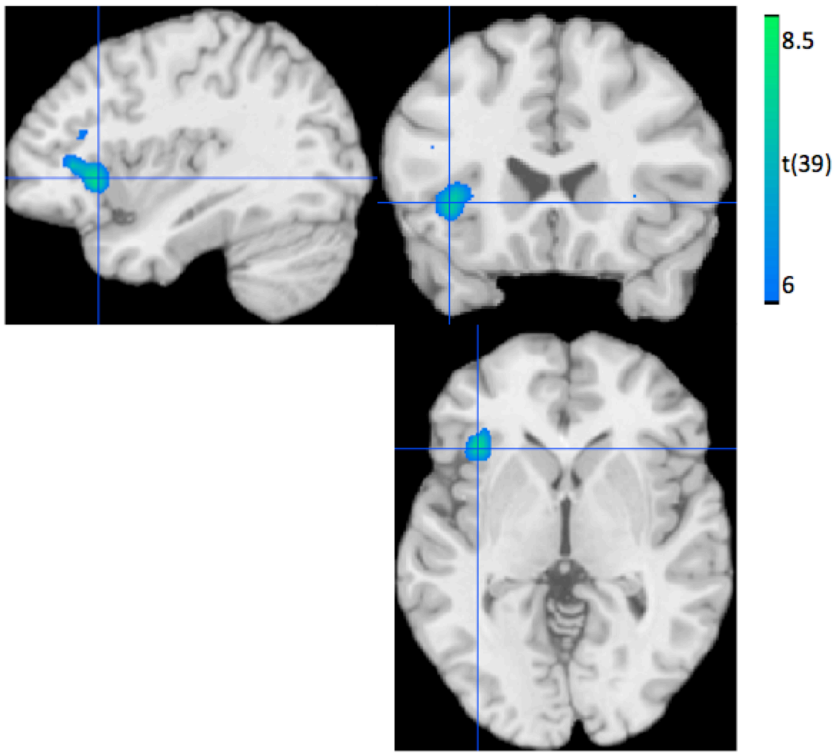


Figure 9. Association strength and left IFG activation. In this region, the association strength of the AROM was a Bonferroni-corrected significant predictor at the item-level.

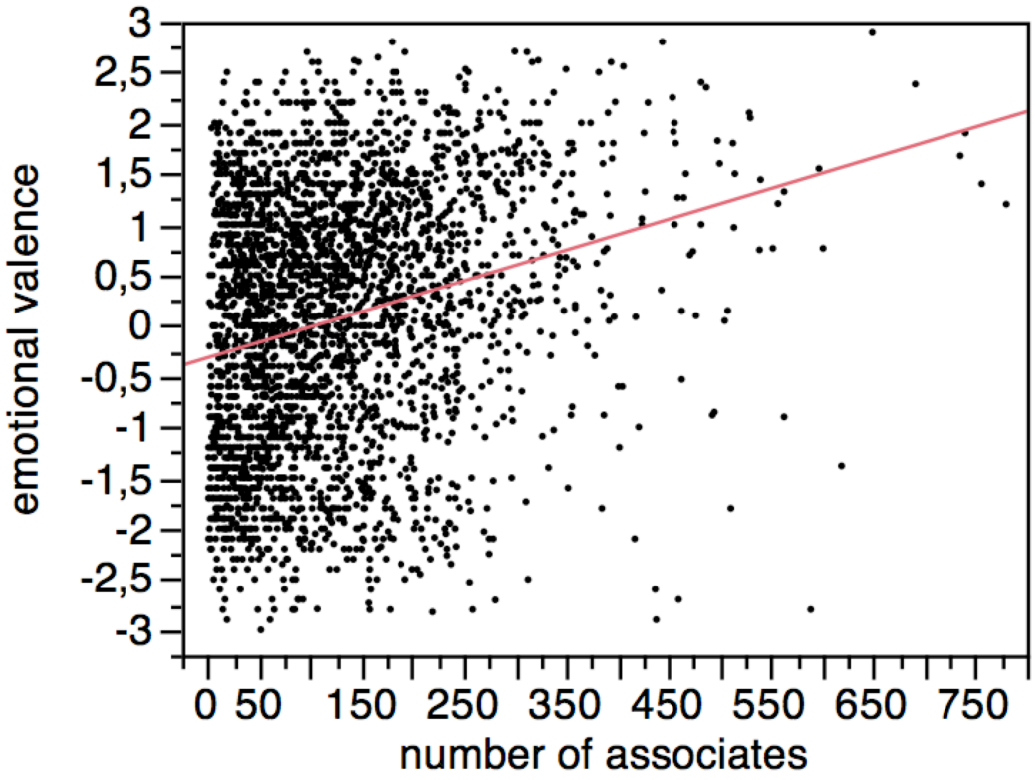


Figure 10. Relationship between the number of associates of a word and emotional valence in the BAWL-R. Though this only accounts for a minor portion of variance (R2 = 0.063), there is a highly significant effect (P < 0.001). The more positive a word is the larger is its amount of associates.

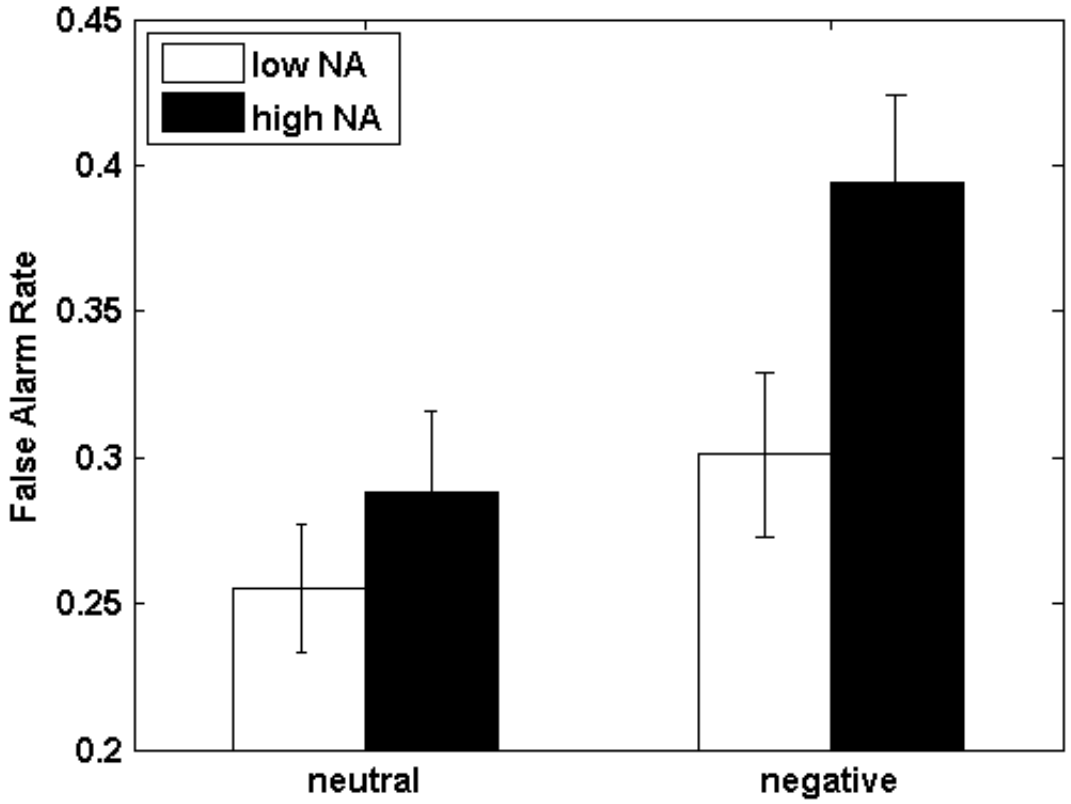


Figure 11. False alarm rates (SE) as a function of negative valence and semantic cohesiveness. The left bars correspond to neutral words, the right bars to negative words. Words with a low number of inter-item associations (NOA) are given in white bars and high-NOA words in black ones. There were significant main effects of negative valence and NOA, as well as a significant interaction.

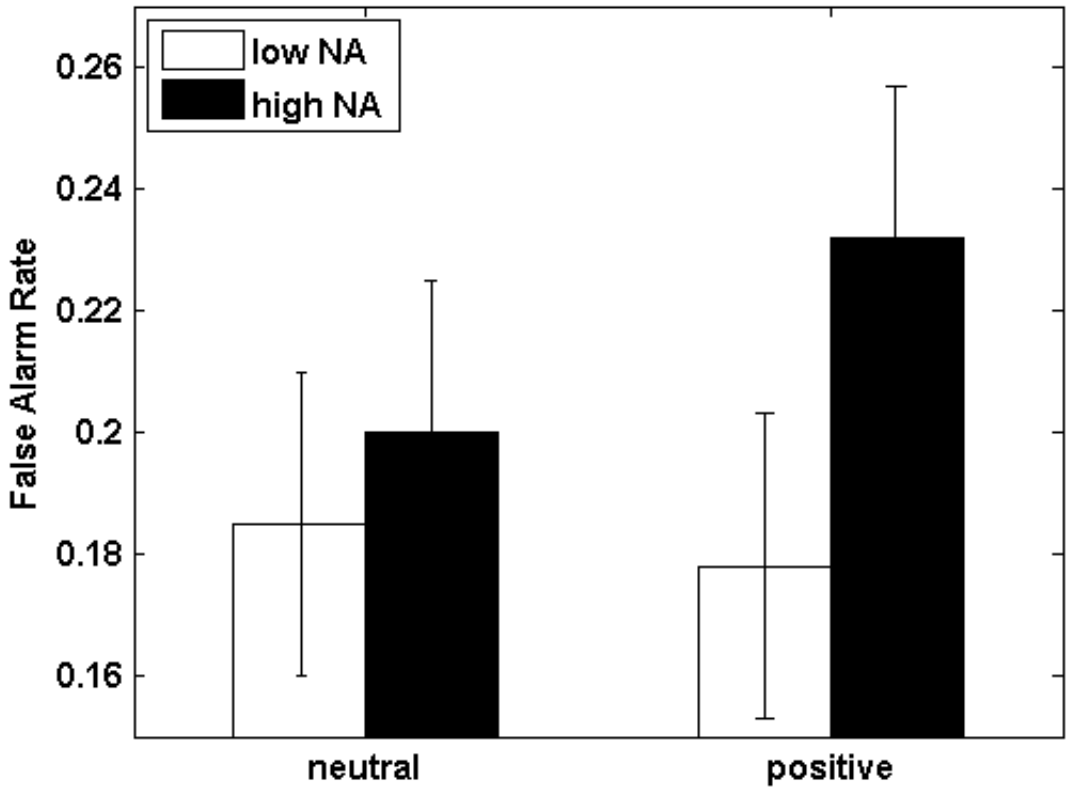


Figure 12. False alarm rates (SE) as a function of positive valence and semantic cohesiveness. The left bars correspond to neutral words, the right bars to positive words. Words with a low number of associations in the stimulus set (low-NOA) are shown in white bars and high-NOA words are in black. There was a significant main effects of NOA, but neither a main effect of positive valence, nor an interaction.